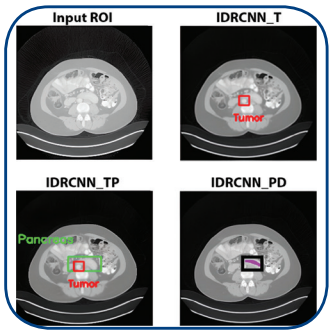
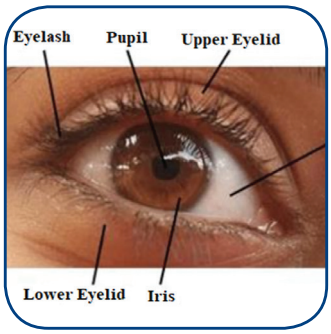
Volume 15, Number 1, 2024 ISSN 1847-6996

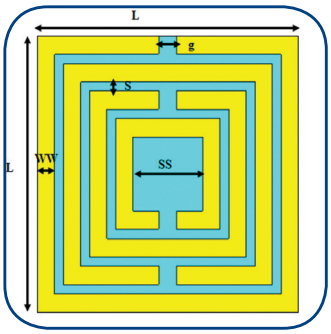
FERIT
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY **OSIJEK**

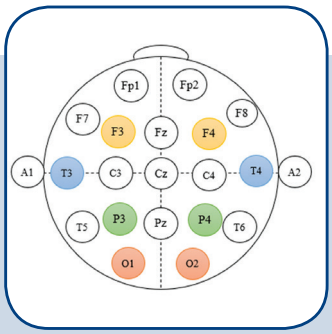
IJECES International Journal of Electrical and Computer Engineering Systems

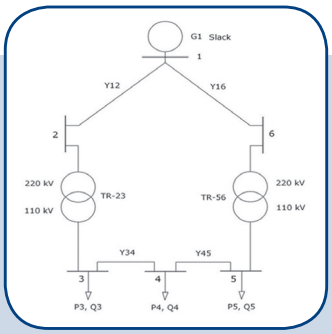
International Journal of Electrical and Computer Engineering Systems

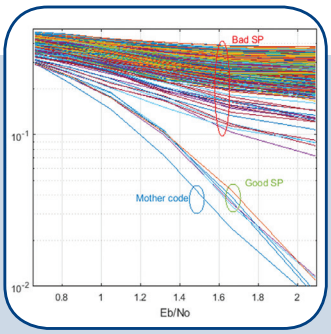












ISSN: 1847-6996

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia

Osijek, Croatia | Volume 15, Number 1, 2024 | Pages 1 - 120

The International Journal of Electrical and Computer Engineering Systems is published with the financial support of the Ministry of Science and Education of the Republic of Croatia

CONTACT

International Journal of Electrical and Computer Engineering Systems (IJECES)

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b, 31000 Osijek, Croatia
Phone: +38531224600, Fax: +38531224605
e-mail: ijeces@ferit.hr

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.
Giro account: 2390001 - 1100016777,
Croatian Postal Bank

Journal is referred in:

- Scopus
- Web of Science Core Collection (Emerging Sources Citation Index - ESCI)
- Google Scholar
- CiteFactor
- Genamics
- Hrčak
- Ulrichweb
- Reaxys
- Embase
- Engineering Village

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003
Published: quarterly
Circulation: 300

IJECES online

https://ijeces.ferit.hr

TABLE OF CONTENTS

Design and simulation of a building-based off-grid photovoltaic microgrid using PVsyst: A case study 1
Case Study
Anjana Saikia | Santanu Sharma

Machine Learning Approaches to Advanced Outlier Detection in Psychological Datasets 13
Original Scientific Paper
Khoula Al. Abri | Manjit Singh Sidhu

Empirical Forecasting Analysis of Bitcoin Prices: A Comparison of Machine Learning, Deep Learning, and Ensemble Learning Models 21
Original Scientific Paper
Nrusingha Tripathy | Sarbeswara Hota | Debahuti Mishra | Pranati Satapathy | Subrat Kumar Nayak

Intelligent and secure real-time auto-stop car system using deep-learning models 31
Original Scientific Paper
Hiba Ali Ahmed | Mohammed A. Noaman Al-hayanni | Muayad Sadik Croock

Multimodal emotion recognition based on the fusion of vision, EEG, ECG, and EMG signals 41
Original Scientific Paper
Shripad Bhatlawande | Swati Shilaskar | Sourjadip Pramanik | Swarali Sole

Black hole algorithm along edge detector and circular hough transform based iris projection with biometric identification systems 59
Original Scientific Paper
Tara Othman Qadir | N.S.A.M Taujuddin

A Survey of Sentiment Analysis and Sarcasm Detection: Challenges, Techniques, and Trends 69
Original Scientific Paper
Ahmed Derbala Yacoub | Salwa O. Slim | Amal Elsayed Aboutabl

An improved algorithm of generating shortening patterns for polar codes 79
Original Scientific Paper
Issame El Kaime | Abdessalam Ait Madi | Reda Benkhouya | Soufiane Hajbi

Wideband circularly polarized fractal antenna with SSRR metasurface for 5G applications 89
Original Scientific Paper
Maalim Al-Abbasi | Tarik Abdul Latef

Recyclability and ecological-economic analysis of a simple photovoltaic panel 99
Original Scientific Paper
Goran Rozing | Dina Jukić | Hrvoje Glavaš | Matej Žnidarec

Comprehensive Analysis of Power System: Exploring Load Factor, Power Balance, Active Load Variation, and Increment Factors with Iterative Implications 105
Original Scientific Paper
Agron Bislimi

Design and simulation of a building-based off-grid photovoltaic microgrid using PVsyst: A case study 113
Case Study
Anjana Saikia | Santanu Sharma

About this Journal
IJECES Copyright Transfer Form

IDRCNN: A Novel Deep Learning Network Model for Pancreatic Ductal Adenocarcinoma Detection on Computed Tomography

Original Scientific Paper

K. Laxminarayanamma
Department of Information Technology, Institute of Aeronautical Engineering,
Hyderabad 500043, India
k.laxminarayanamma@iare.ac.in

R.V. Krishnaiah
Electronics and Communication Engineering, Chebrolu Engineering College,
Chebrolu (Vi & Md)-522212, Guntur, Andhra Pradesh, India
principal@rvit.edu.in

P. Sammulal
Computer Science and Engineering, JNTUH, Kondagattu,
Karimnagar, Jagtial, Telangana 505501, India.
sam@jntuh.ac.in

Abstract *– Early identification of pancreatic ductal adenocarcinoma (PDAC) improves prognosis. Still, it is difficult since lesions are generally smaller and difficult to define on contrast-enhanced computed tomography images (CE-CT). Ineffective PDAC diagnosis has recently been achieved using deep learning models, but the output localized and identified images are of poor quality. This research focuses on small lesions and presents a new, efficient automatic deep-learning network model for PDAC detection. The Improved Deep Residual Convolutional Neural Network (IDRCNN) detects PDAC. The hyperparameter is optimized using the Tunicate Swarm Optimization Algorithm (TSOA) algorithm. A better diagnosis is made due to segmenting the surrounding anatomy structure effects, such as PD. We train a proposed IDRCNN model for segmenting and detecting lesions automatically using CE-CT images. Two more IDRCNN models are trained with the aim of investigating the effects of anatomy integration: (i) segmentation of tumor and pancreas (IDRCNN_TP), and (ii) segmentation of pancreatic Duct (IDRCNN_PD). The three networks' performance was assessed using an external, publicly available test set. Due to its effective classification results, the proposed method produces improved identification results for automated preliminary diagnosis of PDAC in cervical cancer clinics and hospitals. The performance of the proposed method is evaluated using a publicly assessable CT image dataset. It outperforms the existing state-of-the-art methods and achieved 98.67% accuracy, 97.26% recall, 98.52% precision, 97.65% sensitivity, and 98.45% specificity for pancreatic tumor detection.*

Keywords*: computed tomography, deep learning, surrounding anatomy, and PDAC detection*

1. INTRODUCTION

The most prevalent kind of pancreatic cancer is pancreatic ductal adenocarcinoma (PDAC). It has the poorest prognosis among all cancers globally [1, 2]. Pancreatic cancer is becoming more common, and by 2030, it is anticipated to surpass lung carcinoma as the second-most common reason for cancer-related deaths in Western societies. [3, 4]. Patients with early disease stages have a significantly higher 3-year survival rate (82%) than other patient groups [5]. At the time of diagnosis, about 80–85% of patients had either metastatic or resectable disease since early detection of tumors is extremely rare [6, 7]. These facts clearly show that improving patient outcomes requires the early identification of PDAC [8, 9].

Early PDAC detection is difficult because most patients exhibit particular symptoms in the late stages of the illness, and a general population screening is prohibitively expensive with currently available technology [10]. Additionally, PDAC tumors are challenging to see on CT scans, the preferred method for early diagnosis, as the lesions have asymmetrical shapes and ill-defined margins [11]. This presents a considerably greater issue in the disease's early stages because lesions are

frequently iso-attenuating and tiny (2 cm), making them readily missed even by seasoned radiologists [12, 13]. According to a new study that tracked the course of pre-diagnostic PDAs' CT alterations, suspicious abnormalities could have been noticed 18 to 12 months earlier [14]. However, just 44% of patients were referred for additional study by radiologists because they were ineffective at spotting those alterations [15].

By utilizing large volumes of imaging data, artificial intelligence (AI) can help radiologists discover PDAC early [16]. CNNs are a subgroup especially effective at analyzing images among deep learning models. Using images has proven to be accurate in detecting numerous types of cancer [17]. The significant features of the images are extracted by using CNN after completing the process of a series of convolution and pooling for diagnosis [18].

In the input image, clinically useful computer-aided diagnostic tools should be able to locate the lesion and, if cancer is present, determine its existence with relatively little user involvement [19]. Deep learning methods are now being researched in diagnosing PDAC automatically [20]. However, several investigations only accomplish the binary classification without associated lesion location. We developed a more effective deep-learning model for accurate PDAC lesion detection and localization in this research. We provide a completely automated deep-learning method; it produces highly accurate segmented tumor images in the result using abdominal CE-CT images. Additionally, we investigate the effects of integrating the surrounding anatomy.

The main contribution of the research is

- A segmentation-oriented approach is proposed in the present research on the process of PDAC diagnosis. The diagnosis and segmentation of PDAC are made using a new, effective deep learning-based IDRCNN model.
- For the detection and localization of PDAC, three distinct IDRCNN models were trained in this research: (1) Tumor segmentation (IDRCNN_T), (2) Tumor and Pancreas segmentation (IDRCNN_TP), and (3) Pancreatic Duct segmentation (IDRCNN _PD).
- The hyper-parameters of the networks are optimized by using TSOA. The hyper-parameters of the proposed network model, such as the number of iterations, position angle, wide angle, population size, and end condition, are effectively updated by this optimization algorithm. This proposed network produced results with greater accuracy in the diagnosis of PDAC.
- The proposed method is evaluated on publicly accessible datasets. The experiments are performed on the Python platform. According to the experimental data, the proposed approach outperforms the state efficiency concerning all other approaches.

The structure of the presented research is as follows: related prior research is shown and discussed in Section 2. Section 3 provides a detailed explanation of the proposed methodology. The experimental findings and discussions are presented in Section 4. Finally, Section 5 provides conclusions of the present research.

2. RELATED PRIOR WORKS

For PDAC detection, we review some recent deep-learning models in this section. For the segmentation of pancreatic histopathological images, the first deep convolutional neural network architecture was developed by Fu et al. [21]. The two stages, including the WSI level and patch level, are used for the training process, and it is built on a two-step framework based on various recognizing objects. The hidden features for the patch-level classification are extracted from the training set using the CNN model. The cancer regions are predicted and located using the U-Net for patch-level segmentation.

A new model was created by Qureshi et al. [22] to categorize people at high risk for PDAC by automatically recognizing pre-diagnostic CT images. Using radiomic analysis of images from the internal dataset, numerous characteristics that could be used as PDAC predictors were discovered. The identified predictors were then used to train the Nave Bayes classifier, automatically classifying CT scans into groups of healthy controls or pre-diagnostic individuals.

For PDAC mass and surrounding vessel segmentation in CT images, an approach based on CNN was presented by Mahmoudi et al. [23] while also integrating potent traditional features. Using the original image's 3D Local Binary Pattern (LBP) map, the pancreas region is first localized from the entire CT volume. Then, Texture Attention U-Net (TAU-Net) is used to segment PDAC mass. The benefits of both networks are then added together using a 3D-CNN ensemble model.

Transfer learning was utilized by Zhang et al. [24] to develop a CNN model for feature extraction from CT images. According to this study, most radiomics and transfer learning features show only marginally positive linear associations, indicating that these two feature sets may be complementary. Additionally, we evaluated feature fusion techniques to examine their predictive ability for overall survival.

A PDAC resection specimen's discrete cellularity regions were examined by Jungmann et al. [25] using standard histopathology procedures. Iodine concentrations and regional tumor cellularity were matched, as were Hounsfield Units obtained from CT scans. The association between various levels of cellularity in traditional iodine map reconstructions and virtual monoenergetic is assessed using one-way ANOVA and pairwise t-tests.

For high-accuracy automatic pancreatic segmentation, a different two-phase method was developed by Dogan et al. [26] using CT images. The developed ap-

proach involves two steps: (1) Pancreas Localization, where The Mask R-CNN model is used to identify the approximate pancreatic location on the 2D CT slice, and (2) Pancreas Segmentation, where the candidate pancreatic area is further refined with 3D U-Net on the 2D sub-CT slices created in the first stage to produce the segmented pancreas region.

To detect key characteristics and malignant growths in the pancreas in the CT images, Dinesh et al. [27] developed a deep learning-based approach for predicting pancreatic cancer at an early stage. CNN and YOLO model-based CNN (YCNN) are developed in this paper. Images of pancreatic cancer are predicted using the CNN model. Additionally, to help with classification, we employ the CNN (YCNN) built on the YOLO concept. The threshold parameters booked as markers are used for predicting the proportion of pancreatic cancer feasts and key CT scan characteristics.

Most of the existing research on PDAC detection and segmentation produces better results. However, it has some difficulties such as (i) Due to the enormous number of images needed for model training, a lot of computer memory is needed to prevent model crashes when the memory load is exceeded, (ii) additional time for training and highly complex process, and (iii) The segmented image quality is poor. To overcome these difficulties, we propose a new, effective hybrid deep learning network model in this research.

3. PROPOSED METHODOLOGY

3.1. DATASET DESCRIPTION

Model testing was conducted using two accessible abdominal CE-CT datasets that combined images from the portal venous phase: (1) From Memorial Sloan Kettering Cancer Centre in Manhattan, New York, the training set is created for "The Medical Segmentation Decathlon" pancreatic dataset. It has pancreas and lesion voxel-level annotations as well as 281 patients with pancreatic malignancies [32], and (2) From the Clinical Centre at the National Institutes of Health in the US, 80 people with a healthy pancreas are included in the "Cancer Imaging Archive" dataset, along with corresponding voxel-level annotations [33].

3.2. IMAGE ACQUISITION AND LABELING

Five scanners were used to get the CE-CT scans. The available image sizes were 1024 x 1024 pixels (10 images) or 512 x 512 pixels (232 images), and the slice thickness ranged from 1.0 to 5.0 mm. Images with a resolution of 1024 x 1024 were downscaled to 512 x 512 before being used in the model creation. With 17 years of experience, a pancreatic radiologist in abdominal radiology evaluated and corrected all segmentations. The pancreatic duct, pancreas parenchyma, and tumor were all segmented in the annotations.

3.3. AUTOMATIC PDAC DETECTION FRAMEWORK

For automatic PDAC diagnosis and localization, this work employs a segmentation-oriented method. In the proposed pipeline, the cutting-edge, effective automatic medical image segmentation network known as IDRCNN is used to develop the proposed model. The inference pipeline is depicted schematically in Fig. 1 from the initial picture input to the tumor.

For the non-PDAC and PDAC cohorts, pancreas-related manually extracted regions of interest (ROIs) are identified from the other anatomical features, training the anatomy segmentation network for separating the pancreas with the help of the PDAC cohort's reconstructed ROIs from images. From the non-PDAC cohort, ROIs are automatically annotated by this network. With the manually annotated PDAC cohort, these non-PDAC cohorts are subsequently combined for training the three different IDRCNN models in the process of PDAC detection. The three different networks are (1) segmentation of tumor only (IDRCNN_T), (2) segmentation of tumor and pancreas (IDRCNN_TP), and (3) segmentation of Pancreatic Duct (IDRCNN _PD).

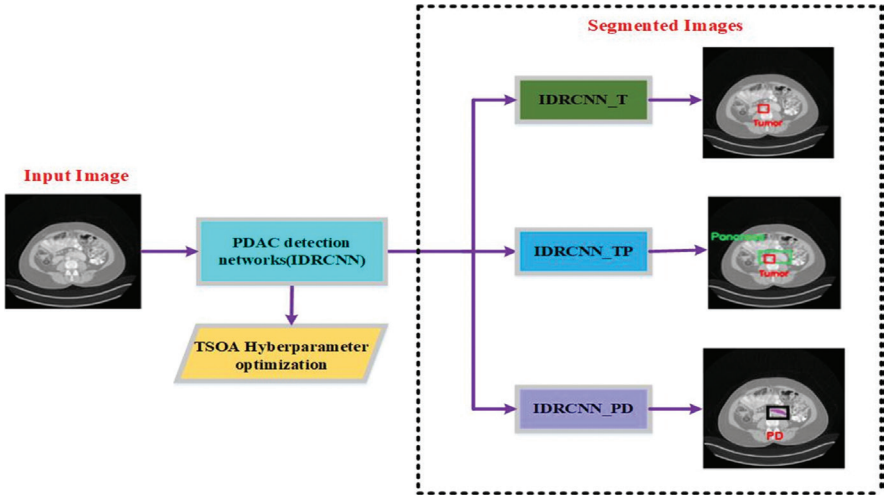


Fig. 1. An illustration of the proposed architecture for automatic PDAC detection

The models were produced for each configuration using identical splits using five-fold cross-validation from these networks. The cross-entropy loss function is better suited for segmentation-oriented detection tasks. It is chosen by default in the IDRCNN framework; it was employed for the PDAC detection networks. Additionally, From the PDAC cohort, a low-resolution pancreatic segmentation network was trained using the complete CE-CT images that had been down-sampled to 256 × 256. This was subsequently used during inference for extracting the pancreatic ROI automatically from hidden images. The segmented tumor images produced using models like IDRCNN_T, IDRCNN_TP, and IDRCNN_PD for PDAC detection display the image regions.

3.4. IDRCNN-TSOA BASED ON TRAINING AND CLASSIFICATION

In this research, the IDRCNN is applied for training and classifying data. The IDRCNN is often very effective at enhancing the classification process's reliability and accuracy. Several researchers have used CNN to recognize PDAC from the provided images. This shows how CNN has made considerable progress in image processing. One deep learning method that permits weights and biases to various elements of images is CNN. CNN performs better at classification and can extract the features automatically. In the area of computer vision, CNN offers more accurate classification. An advanced level of depth occurs due to the residual connection being established as the neural network gets deeper.

Deep residual networks address the difficulties of network degradation, vanishing gradients, and exploring gradients. To detect and localize PDAC lesions on CE-CT scans, IDRCNN is preferred. A loss function is produced during the IDRCNN process, and it can potentially reduce the detection's performance. The hyperparameters should be modified for loss function reduction. The TSO algorithm is combined with the IDRCNN approach to adjust the hyperparameters.

A residual network's usual design consists of various techniques and arranges the fundamental components, such as normalization of pooling, non-linear mapping, and convolution sequentially. As a result, the IDRCNN's mathematical observation is considered as $H(x)$. For fitting the residual mapping, the residual function learned is known as $F(x)$ and is shown in the equation below.

$$F(x) = H(x) - x \tag{1}$$

The output of a conventional CNN has been used to represent the final mapping's residual learning. In Eq. (2), the convolutional method of calculation can be expressed;

$$V_{ij} = f\left(\sum_m \sum_{h=0}^{H_{1-1}} \sum_{w=0}^{W_{1-1}} k_{jm}^{hw} V_{(1-1)m}^{(x+h)(y+w)} + b_{ij}\right) + b_{ij} \tag{2}$$

The convolutional kernel size is represented as $H_{1\text{-}1}$ & $W_{1\text{-}1}$, the number of keyframes is represented as m, the maximum number of iterations is described as (*i&j*), and the linear coefficient is represented by k and b.

Each image must train a convolution kernel before performing convolution processing. In general, accuracy and over-fitting issues affect the entire network due to the network's calculation efficiency and training speed. Different feature extraction algorithms are used for obtaining high-level features, which are more accurate due to the deep network structure. Batch normalization (BN), which successfully eliminates gradient explosion and vanishing gradient, is added after each convolution layer. After a network has been modified to include a residual structure more suitable for solving these problems, the gradient is propagated more readily and effectively due to shortcuts. The IDRCNN network's structure is also shown in Fig. 2.

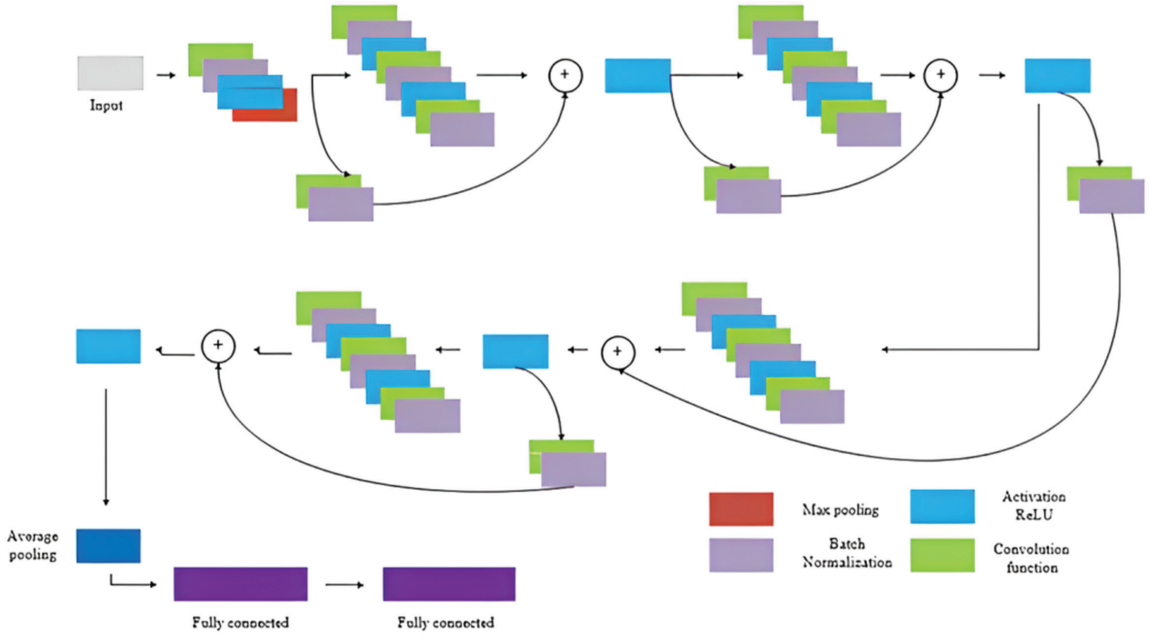


Fig. 2. The structure of the IDRCNN model

Internal parameters of the IDRCNN are optimized, and each layer's weight values are assigned to it. Another challenge in the network is the selection of hyper-parameters.

The selection domains $\Lambda_1, \Lambda_2, ..\Lambda_n$ of each hyperparameter are used to choose the assumed hyperparameters $\lambda_1, \lambda_2, \lambda_3, ...\lambda_n$. The model's domain space for selecting its hyper-parameters is then defined as $\Lambda=\Lambda_1*\Lambda_2*..*\Lambda_n$. Eq. (3) describes the hyperparameter optimization process;

$$f(\lambda) = \frac{1}{k}\sum L(\lambda, {D_{train}}^i, {D_{valid}}^i) \quad (3)$$

Where a loss of function is represented $L(.)$, the training set is depicted as D_{train}, and the validation set is depicted as D_{valid}. Manual and grid search (violent search) are the two most used hyperparameter optimization techniques. However, optimizing for hyper-parameters has always been laborious since they could be more efficient. Because of how they are structured in a search area intended for automatic modification, hyper-parameters must be manually initialized. The images are finally divided into various classes using the Softmax classifier. This classifier works well for multi-class classification issues in general. The expression presented below defines the function of the Softmax layer.

$$S_i = \frac{e^{V_i}}{\sum_i^c e^{V_i}} \quad (4)$$

Where V_i represents the classifier's output value for class I, the number of classes is represented by c, and relative probability is defined by S_i. The method determines the relative likelihood of each type's output result, and the class's most significant relative probability group determines the classification outcomes.

3.5. HYPERPARAMETER OPTIMIZATION USING TSOA

TSO completed adjusting the IDRCNN's hyperparameter following the training and classification of data. The performance of the training approach is significantly impacted by the hyper-parameters, which are crucial for influencing the training approach's behavior. Due to the hyperparameter tuning, it was simpler to maintain a more significant number of experiments, and a better selection of hyperparameters resulted in a more efficient algorithm. In this suggested work, TSO is hybridized with IDRCNN to adjust the hyper-parameters of the classification algorithm. The TSO algorithm is one of the metaheuristic optimization algorithms with a bio-inspired design. During foraging and navigating, this algorithm simulates the actions of the swarm and jet populations of tunicates. The optimization strategy determines the ideal values by choosing the parameters for optimizing and describing the ranges to investigate. The ability of the tunicate to locate food sources has been described in the TSOA overview.

In this research, the food source has been considered as an optimization technique. Knowing the search term in the given search area is necessary for the suggested optimization strategy to solve all optimization issues. For hyperparameter optimization, the IDRCNN uses the TSO algorithm as a result. When compared to selecting hyper-parameters manually, this optimization technique selects them quickly. This contributes to a better training process as well.

The number of iterations, position angle, wide angle, population size, and end condition have all been initialized by TSOA. Here, $\vec{A}$ it is employed to determine the optimal position for the PDAC region while preventing conflict between the search agents.

$$\vec{A} = \frac{\vec{G}}{\vec{M}} \quad (5)$$

$$\vec{G} = C_2 + C_3 - \vec{F} \quad (6)$$

$$\vec{F} + 2.C_1 \quad (7)$$

Where G represents the target function in Eq. (6), and the search agent's flow of advection is considered by F in Eq. (7). Then $C_1, C_2, \& C_3$ have been regarded as search agent variables; therefore, a force between the searches agents are represented by $\vec{M}$, and Eq. (8) explains the calculation;

$$\vec{M} = \left[P1min_{max_{min}}\right] \quad (8)$$

Where P_{maxmin} represents the minimum and maximum intensity iterations. Thus, the neighbor search agent's position is determined by using Eq. (9);

$$\vec{P}D = \left|\vec{F}S - r_{and}.\vec{P}p(x)\right| \quad (9)$$

The tumor region's position is represented as $\vec{F}S$, the distance of the location is depicted as $\vec{P}D$, the position of the tunicate is expressed as $\vec{P}p(x)$, the current iteration is indicated as x, and the random number is considered as r_{and} in a wide range [0, 1]. The following equation in Eq. (10) has been calculated to keep the search agent in its position toward the best search;

$$\vec{P}p(x) = \begin{cases} \vec{F}S + \vec{A}.\vec{P}D, if \\ \vec{F}S - \vec{A}.\vec{P}D, if \end{cases} r_{and} \geq 0.5 r_{and} < 0.5 \quad (10)$$

Here, the search agent's updated position is $\vec{P}p(x)$. As well as being utilized for updating the tumor region's position and location, the behavior of the TSOA has been proposed as a means of fine-tuning the hyperparameter of the IDRCNN. As a result, the following equation in Eq. (11) has been used to determine the behavior of the TSOA as the best global search;

$$\vec{P}p(x+1) = \frac{\vec{P}p(x) + \vec{P}p(x+1)}{2+C_1} \quad (11)$$

Additionally, the global position value has been regarded as the IDRCNN's weight updating where the hyperparameters' wide range of values is [0, 1]. The TSOA may also cover temporal and space complexity. The proposed pseudo-coding algorithm's TSOA has been calculated as follows;

Pseudo Code of IDRCNN-TSOA

Input: initialize the population size $\vec{P}p$

Output: optimal fitness value $\vec{F}S$

Initialize the hyperparameters{(λ_1, λ_2,...,λ_n), $\vec{A}, \vec{G}, \vec{F}, \vec{M}$}, and choose the Max number of iterations.

While ($x<m_{axiteration}$)***do***

for all hyper-parameters

Compute fitness function ($\vec{P}p$)

Generate r_{and}( ) with the range [0,1]

Update the search agent's position using equation (11)

If ($r_{and} \leq 0.5$)then

$\vec{F}S + \vec{A}.\vec{P}D$

else

$\vec{F}S - \vec{A}.\vec{P}D$

end if

Update $\vec{P}p$ for the optimal solution until the best solution

end for

Update the hyper-parameters

$x \leftarrow x+1$

end while

Return optimized hyperparameter value

The proposed IDRCNN-TSOA model assists in the best possible detection and localization of PDAC lesions. At this stage, the IDRCNN technique detects PDAC and the TSO algorithm is used to hyper-tune the parameters to minimize the obtained loss function. Thus, using the provided input pictures, the proposed IDRCNN-TSOA technique successfully segments the tumor, the tumor and pancreas, and the pancreatic duct.

4. RESULT AND DISCUSSION

This section covers the many performance metrics utilized to assess the suggested strategy and contrast it with the recent related previous approaches. We display the results of our experimental study on how cost-sensitive learning affects categorization ability.

This research offers a thorough and systematic application of deep learning methods (CNN) for precisely identifying and segmenting PDAC. Hybrid deep-learning models were used to classify and segment the CT image collection. The segmentation of PDAC images is developed using the hybrid IDRCNN-TSOA model, where the network training is done by using 70% of the dataset, and network validation is performed by using 30% of the dataset. The i5 processor and 8 GB of RAM were the computer resources used to obtain the results on the Python platform.

4.1. PERFORMANCE METRICS

Recall, precision, sensitivity, specificity, and accuracy are performance metrics used to assess segmentation performance, given below.

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Sensitivity = \frac{TP}{FN+TP} \quad (14)$$

$$Specificity = \frac{TN}{TN+FP} \quad (15)$$

$$Accuracy = \frac{TP+TN}{TN+FN+TP+FP} \quad (16)$$

For the PDAC diagnosis, the number of true positives is known as the TP. Similarly, false positives are represented as FP, the true negatives are represented as TN, and the false negatives are indicated as FN.

4.2. ANALYSIS

The receiver operating characteristic (ROC) curve is employed for assessing patient performance. The free-response receiver operating characteristic (FROC) curve was used to assess lesion performance. The false positive rate (1-specificity) and actual positive rate (sensitivity) are contrasted within the ROC analysis at various model output thresholds to determine the model's confidence in the presence or absence of a tumor, referred to as the tumor likelihood map's highest value. For every specific lesion prediction at different thresholds, the false positive rate versus the average number of true positives per image is plotted in the FROC analysis to determine whether the model correctly detected the lesion.

The trained models were individually applied to testing the three PDAC-detection settings to compare them. The partial area under the FROC curve (pAUC-FROC) and the area under the ROC curve (AUC-ROC) are then compared to determine whether there were any statistically significant changes. The statistical significance was determined at a level of 97.5% confidence. Each configuration's final performance was identified by combining the models' forecasts.

4.3. EXPERIMENTAL RESULTS

Table 1 displays the results of the internal five-fold cross-validation sets using the three different PDAC detection network setups. With an AUC-ROC of 0.993, IDRCNN_PD performs best at the patient level. All networks produce a similar pAUC-FROC for lesion localization, with IDRCNN_PD and IDRCNN_TP performing somewhat better than IDRCNN_T.

Table 1. Internal five-fold cross-validation for each setup using the AUC-ROC and pAUC-FROC metrics

Configuration	Mean AUC-ROC (95%CI)	Mean pAUC-FROC (95%CI)
IDRCNN_T	0.974	3.795
IDRCNN_TP	0.987	4.245
IDRCNN_PD	0.993	4.660

Fig. 3 displays the mean FROC and ROC curves produced for each PDAC detection network design using the external test set, which correspond to the 95% confidence intervals. These curves were produced for each configuration using the proposed IDRCNN model (with five-fold cross-validation). Compared to training with just tumor segmentation, the findings on the external test set demonstrate that incorporating the pancreas parenchyma has a clear advantage at the patient level. Both IDRCNN_TP and IDRCNN_PD produced a significantly higher AUC-ROC than IDRCNN_T. The IDRCNN_TP and IDRCNN_PD networks performed identically, nevertheless. In contrast, the three FROC curves clearly distinguished themselves at the lesion level across the whole test set (Figure 3), with pAUC-FROC being significantly greater for IDRCNN_PD than for the other two configurations. For localizing PDAC lesions, this displays how the incorporation of surrounding anatomy improves the model's capabilities.

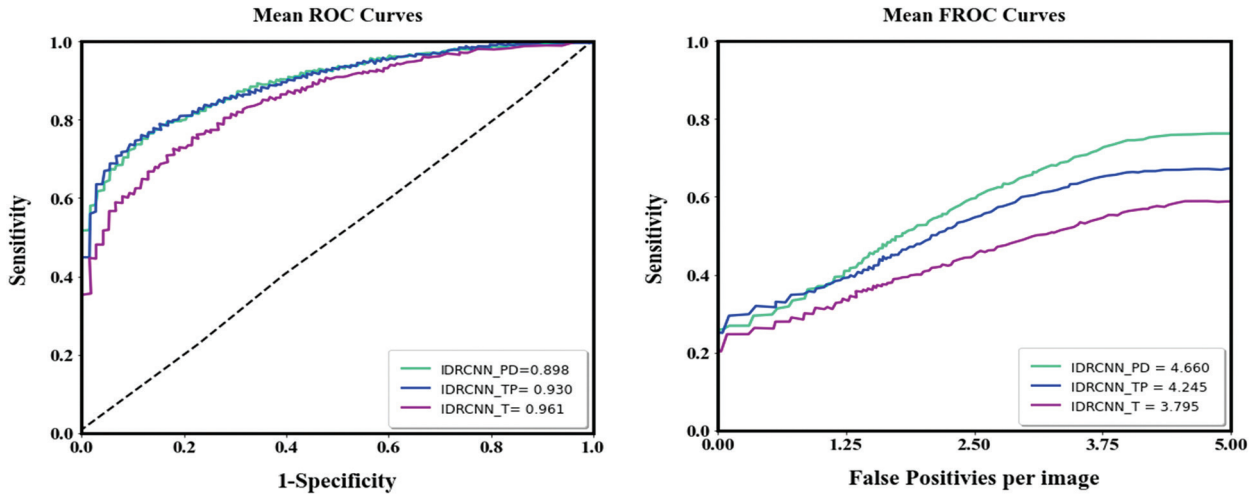


Fig. 3. Curves for the mean ROC and FROC for the external test set.

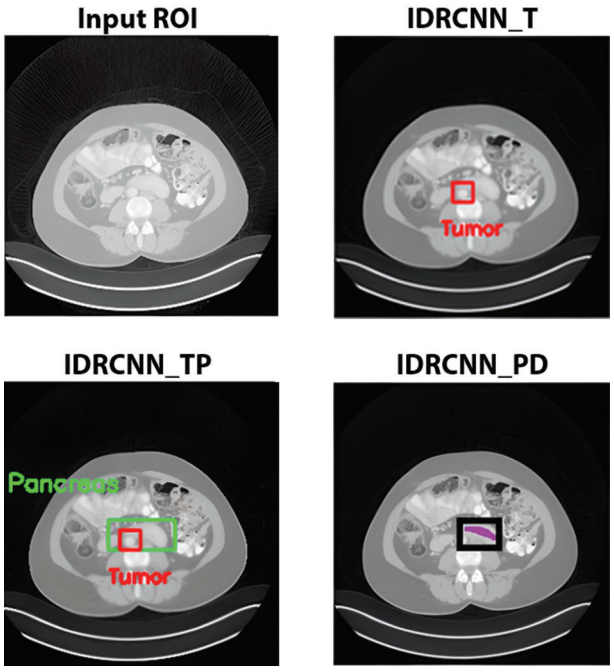


Fig. 4. Results of experiments using the proposed approach

In the instance of a pancreatic duct-obstructing isodense lesion, the benefit of anatomical integration is demonstrated in Fig. 4 and causes it to enlarge. Given that the tumor and healthy pancreatic parenchyma cannot be distinguished from one another, the IDRCNN_TP and IDRCNN_T models are unable to detect this lesion. The dilated duct termination, however, allows IDRCNN_PD to locate its position in the pancreas precisely. In the pancreas parenchyma, the neural model can concentrate more on the remaining regions by the duct's segmentation through supervised training; this could be responsible for its capacity to identify small tumors. Additionally, the radiologist can interpret the network output regarding the tumor with the help of the multi-structure segmentation offered by IDRCNN_PD.

The proposed CNN architecture produces A high segmentation rate, making it the most efficient system for providing segmentation results based on CT images with the largest dataset of images used ever. The suggested model successfully extrapolates traits that describe the inter-scale heterogeneity of the illnesses, enhancing classification performance.

In the proposed segmentation network model, the residual connections make it easier to train deeper networks, which is crucial for capturing complex spatial relationships in images. The skip connections aid in the flow of gradients during backpropagation, making it feasible to train networks with many layers without suffering from vanishing gradients. Information from earlier layers can directly contribute to later layers, allowing the network to reuse essential features, which is beneficial for tasks like image segmentation, where different levels of features are necessary. Based on these advantages, our proposed IDRCNN segmentation network effectively segments the tumor region.

4.4. COMPARISON RESULTS

Our proposed model's PDAC identification findings are compared with those of earlier studies over CT images in this section. The performance comparison of the proposed and existing approaches are given in Table 2 regarding recall, precision, specificity, sensitivity, and accuracy.

This research created a completely automated system for finding and locating PDAC tumors on CE-CT images using the cutting-edge, self-configuring medical segmentation framework IDRCNN. The effect of incorporating surrounding anatomy was examined as well. Abdominal CE-CT images have only a small area for the pancreas. And within that area, lesions are a much smaller target. These are used to make deep learning for PDAC detection extremely difficult. The entire CE-CT images would be exceedingly resource-intensive for training and testing the networks and offer practical information about the other organs. In this method, it is required to pick a low level of interest surrounding the pancreas. Still, it would take more time and resources to physically annotate the pancreatic before processing every image via the network, significantly reducing the model's clinical value. A smaller volume of interest is automatically extracted in the entire CE-CT images using our PDAC detection framework in the first step for solving this difficulty.

Table 2. Comparative analysis of proposed approach and state-of-the-art approaches

Reference	Method	Accuracy (%)	Recall (%)	Precision (%)	Sensitivity (%)	Specificity (%)
Mahmoudi et al. [23]	TAU-Net	87.36	88	82	-	-
Tureckova et al [28]	CNN	72.53	68	70	-	-
Si et al. [29]	FEE-DL	82.7	81.27	84.61	-	-
Qiu et al. [30]	MSTA	86.02	-	-	87.39	91.12
Ma et al. [31]	CNN	95.47	-	-	91.58	98.27
Althobaiti et al. [37]	CNN-CTPCD	95.49	-	-	91.50	98.64
Khdhir et al. [38]	ALO-CNN-GRU	95	-	-	-	-
Proposed Approach	IDRCNN	98.67	97.26	98.52	97.65	98.45

For PDAC segmentation, TAU-Net and 2D attention U-Net were developed by Mahmoudi *et al*. [23]. From the whole CT volume, the pancreas region is localized by using a 3D-CNN architecture with the help of the original images 3D Local Binary Pattern (LBP) map in this paper. TAU-Net achieves 87.36% accuracy in detecting the PDAC. A CNN model with deep supervision and attention gates was studied by Tureckova *et al*. [26], and in this paper, PDAC is detected with an accuracy of 72.53%. End-to-end deep learning (FEE-DL) developed by Si *et al*. [27] achieves 82.7% accuracy in detecting PDAC. Qiu *et al*. [28] studied a new MSTA (multi-resolution-statistical texture analysis) system for PDAC, which achieves an accuracy of 86.02%. For PDAC classification, a powerful CNN model was developed by Ma *et al*. [29]. It achieves a specificity of 98.27%, a sensitivity of 91.58%, and an overall accuracy of 95.47% for plain scans. Using CT images, an optimal deep learning-based pancreatic tumor and non-tumor classification (ODL-PTNTC) model is developed by Althobaiti *et al*. [31], and it achieves an average sensitivity of 91.50%, specificity of 98.64%, and accuracy of 95.49 % for pancreatic disease classification. For CT image-based pancreatic tumor segmentation and classification, the Antlion Optimization-Convolutional Neural Network-Gated Recurrent Unit (ALO-CNN-GRU) model was developed by Khdhir *et al*. [33], and it achieves an overall accuracy of 95% for pancreatic tumor classification. Our proposed IDRCNN model achieves a high performance with specificity of 98.45%, sensitivity of 97.65%, accuracy of 98.67%, recall of 97.26%, and precision of 98.52%.

Compared to the prior models, the proposed IDRCNN model achieves better results, and it effectively identifies the PDAC and segments the region of the tumor, pancreas, and pancreatic duct. The essential advantage of the proposed IDRCNN model is that it avoids overfitting and has no detrimental effects on network performance due to the classification and segmentation process.

The proposed DRCNN model optimizes the flow of information through both convolutional and residual connections; it could lead to improved gradient flow during training. This could result in faster convergence and better generalization performance than CNN models. In the TAU-Net model, the texture attention mechanisms can introduce additional computational complexity and memory requirements. This might result in slower training and inference times, making the architecture less suitable for real-time applications or resource-constrained environments. Still, our proposed approach utilizes minimum memory requirements and low com computational complexity. The quality of the input images influences the effectiveness of multi-resolution analysis. Poorly captured images, noise, or artifacts can affect the accuracy of texture analysis at different scales. Hence, our proposed approach produces defective lesion regions with high-quality images during segmentation.

The end-to-end automatic diagnosis is performed by the proposed system, which is a significant advantage of this research, and the time it takes for each patient to

obtain a diagnosis after entering the initial abdomen CT image is 18.6 seconds. It is appropriate for clinical usage with significant promise for evaluating and recommending treatments since it can manage and usefully interpret vast volumes of data quickly, correctly, and affordably. The model, for instance, may be used to help diagnose patients in low-level hospitals with limited resources or for widespread pre-diagnosis during physical examinations. The model's capacity to generate saliency maps, highlighting the region's most crucial to its diagnostic decision-making, is a last feature that can increase its dependability. While clinicians have access to additional data, such as patient health records and testimony, our method only uses evidence from CT images. As a result, definitive diagnoses and treatment plans should still be based on the clinical evaluation of specialists, not just the output of a deep-learning system.

5. CONCLUSIONS

In CE-CT scans for tumor localizing, this research develops an effective and automatic deep-learning-based network model to localize the tumor in CE-CT scans and determine if a patient has PDAC. To precisely detect and localize PDAC tumors, the novel, effective IDRCNN model is employed. The maximum AUC of 0.961 is acquired by the proposed model, demonstrating that our proposed approach is an effective tool for recognizing tiny PDAC lesions and may help support radiologists in the early diagnosis of PDAC. In addition, we demonstrate that adding local anatomical information considerably improves model performance regarding lesion localization. The performance comparison demonstrates that while the proposed IDRCNN model adds a small amount of computing overhead, the performance is significantly improved. Finally, our automatic system has proven effective at detecting PDAC, making it more versatile than the current state of the art while still performing on a level with more specialized techniques.

In our future research, we will use the combination of two deep learning models for PDAC detection and also perform segmentation of several surrounding anatomical structures of the tumor.

Conflict of interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. ACKNOWLEDGMENTS

We declare that this manuscript is original, has not been published before, and is not currently being considered for publication elsewhere.

Funding: No funding was received to assist with the preparation of this manuscript.

Availability of data and material: Data will be available when requested

Authors' contributions: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Ethics approval: This material is the author's original work, which has not been previously published elsewhere. The paper reflects the author's research and analysis in a truthful and complete manner.

7. REFERENCE

[1] Y. Fukuda, D. Yamada, H. Eguchi, T. Hata, Y. Iwagami, T. Noda, T. Asaoka, K. Kawamoto, K. Gotoh, S. Kobayashi, Y. Takeda, "CT density in the pancreas is a promising imaging predictor for pancreatic ductal adenocarcinoma", Annals of surgical oncology, Vol. 24, 2017, pp. 2762-2769.

[2] H. Hayashi, N. Uemura, K. Matsumura, L. Zhao, H. Sato, Y. Shiraishi, Y. I. Yamashita, H. Baba, "Recent advances in artificial intelligence for pancreatic ductal adenocarcinoma", World Journal of Gastroenterology, Vol. 27, No. 43, 2021, p. 7480.

[3] J. D. Kang, S. E. Clarke, A. F. Costa, "Factors associated with missed and misinterpreted cases of pancreatic ductal adenocarcinoma", European Radiology, Vol. 31, 2021, pp. 2422-2432.

[4] R. P. Sah, A. Sharma, S. Nagpal, S. H. Patlolla, A. Sharma, H. Kandlakunta, V. Anani, R. S. Angom, A. K. Kamboj, N. Ahmed, S. Mohapatra, "Phases of metabolic and soft tissue changes in months preceding a diagnosis of pancreatic ductal adenocarcinoma", Gastroenterology, Vol. 156, No. 6, 2019, pp. 1742-1752.

[5] W. Qiu, N. Duan, X. Chen, S. Ren, Y. Zhang, Z. Wang, R. Chen, "Pancreatic ductal adenocarcinoma: machine learning–based quantitative computed tomography texture analysis for prediction of histopathological grade", Cancer Management and Research, Vol. 11, 2019, p. 9253.

[6] J. Chang, Y. Liu, S. A. Saey, K. C. Chang, H. R. Shrader, K. L. Steckly, M. Rajput, M. Sonka, C. H. Chan, "Machine-learning based investigation of prognostic indicators for oncological outcome of pancreatic ductal adenocarcinoma", Frontiers in oncology, Vol. 12, 2022, p. 6660.

[7] T. Tong, J. Gu, D. Xu, L. Song, Q. Zhao, F. Cheng, Z. Yuan, S. Tian, X. Yang, J. Tian, K. Wang, "Deep learning radiomics based on contrast-enhanced ultra-

sound images for assisted diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis", BMC Medicine, Vol. 20, No. 1, 2022, pp. 74.

[8] C. An, D. Li, S. Li, W. Li, T. Tong, L. Liu, D. Jiang, L. Jiang, G. Ruan, N. Hai, Y. Fu, "Deep learning radiomics of dual-energy computed tomography for predicting lymph node metastases of pancreatic ductal adenocarcinoma", European Journal of Nuclear Medicine and Molecular Imaging, Vol. 49, 2022, pp. 1187-1199.

[9] Z. M. Zhang, J. S. Wang, H. Zulfiqar, H. Lv, F. Y. Dao, H. Lin, "Early diagnosis of pancreatic ductal adenocarcinoma by combining relative expression orderings with machine-learning method", Frontiers in Cell and Developmental Biology, Vol. 8, 2020, p. 582864.

[10] L. C. Chu, S. Park, S. Kawamoto, Y. Wang, Y. Zhou, W. Shen, Z. Zhu, Y. Xia, L. Xie, F. Liu, Q. Yu, Application of deep learning to pancreatic cancer detection: lessons learned from our initial experience. Journal of the American College of Radiology, Vol. 16, No. 9, 2019, pp. 1338-1342.

[11] S. Ziegelmayer, G. Kaissis, F. Harder, F. Jungmann, T. Müller, M. Makowski, R. Braren "Deep convolutional neural network-assisted feature extraction for diagnostic discrimination and feature visualization in pancreatic ductal adenocarcinoma (PDAC) versus autoimmune pancreatitis (AIP)", Journal of clinical medicine, Vol. 9, No. 12, 2020, p. 4013.

[12] A. Patra, K. Panagiotis, G. Suman, A. Panda, S. K. Garg, A. Goenka, "Abstract PO-084: Automated detection of pancreatic ductal adenocarcinoma (PDAC) on CT scans using artificial intelligence (AI): Impact of inclusion of automated pancreas segmentation on the accuracy of 3D-convolutional neural network (CNN)", Clinical Cancer Research, Vol. 27, 2021.

[13] G. A. Kaissis, S. Ziegelmayer, F. K. Lohöfer, F. N. Harder, F. Jungmann, D. Sasse, A. Muckenhuber, H. Y. Yen, K. Steiger, J. Siveke, H. Friess, "Image-based molecular phenotyping of pancreatic ductal adenocarcinoma", Journal of Clinical Medicine, Vol. 9, No. 3, 2020, p. 724.

[14] W. Qiu, N. Duan, X. Chen, S. Ren, Y. Zhang, Z. Wang, R. Chen, "Pancreatic ductal adenocarcinoma: machine learning–based quantitative computed tomography texture analysis for prediction of histopathological grade", Cancer Management and Research, Vol. 11, 2019, p. 9253.

[15] T. G. W. Boers, Y. Hu, E. Gibson, D. C. Barratt, E. Bonmati, J. Krdzalic, F. van der Heijden, J.J. Hermans, H.J. Huisman, "Interactive 3D U-net for the segmentation of the pancreas in computed tomography scans", Physics in Medicine & Biology, Vol. 65, No. 6, 2020, p. 065002.

[16] T. A. Qureshi, C. Lynch, L. Azab, Y. Xie, S. Gaddam, S. J. Pandol, D. Li, "Morphology-guided deep learning framework for segmentation of pancreas in computed tomography images", Journal of Medical Imaging, Vol. 9, No. 2, 2022, p. 24002.

[17] M. Kriegsmann, K. Kriegsmann, G. Steinbuss, C. Zgorzelski, A. Kraft, M. M. Gaida, "Deep learning in pancreatic tissue: Identification of anatomical structures, pancreatic intraepithelial neoplasia, and ductal adenocarcinoma", International Journal of Molecular Sciences, Vol. 22, No. 10, 2021, p. 5385.

[18] C. Cen, L. Liu, X. Li, A. Wu, H. Liu, X. Wang, H. Wu, C. Wang, P. Han, S. Wang, "Pancreatic ductal adenocarcinoma at CT: a combined nomogram model to preoperatively predict cancer stage and survival outcome", Frontiers in Oncology, 2021, p. 1980.

[19] M. A. Attiyeh *et al*. "Survival prediction in pancreatic ductal adenocarcinoma by quantitative computed tomography image analysis", Annals of surgical oncology, Vol. 25, 2018, pp. 1034-1042.

[20] E. J. Koay *et al*. "A visually apparent and quantifiable CT imaging feature identifies biophysical subtypes of pancreatic ductal adenocarcinoma," Clinical Cancer Research, Vol. 24, No. 23, 2018, pp. 5883-5894.

[21] H. Fu *et al*. "Automatic pancreatic ductal adenocarcinoma detection in whole slide images using deep convolutional neural networks", Frontiers in oncology", Vol. 11, No. 1, 2021, p. 665929.

[22] T. A. Qureshi, S. Gaddam, A. M. Wachsman, L. Wang, L. Azab, V. Asadpour, W. Chen, Y. Xie, B. Wu, S. J. Pandol, D. Li, "Predicting pancreatic ductal adenocarcinoma using artificial intelligence analysis of prediagnostic computed tomography images", Cancer Biomarkers, Vol. 33, No. 2, 2022, pp. 211-217.

[23] T. Mahmoudi, Z. M. Kouzahkanan, A. R. Radmard, R. Kafieh, A. Salehnia, A. H. Davarpanah, H. Arabalibeik, A. Ahmadian, "Segmentation of pancreatic ductal adenocarcinoma (PDAC) and surrounding vessels in CT images using deep convolutional neural networks and texture descriptors", Scientific Reports, Vol. 12, No. 1, 2022, p. 3092.

[24] Y. Zhang, E. M. Lobo-Mueller, P. Karanicolas, S. Gallinger, M. A. Haider, F. Khalvati, "Improving prognostic performance in resectable pancreatic ductal adenocarcinoma using radiomics and deep learning features fusion in CT images", Scientific Reports, Vol. 11, No. 1, 2021, pp. 1-11.

[25] F. Jungmann, G. A. Kaissis, S. Ziegelmayer, F. Harder, C. Schilling, H. Y. Yen, K. Steiger, W. Weichert, R. Schirren, I. E. Demir, H. Friess, "Prediction of tumor cellularity in resectable PDAC from preoperative computed tomography imaging", Cancers, Vol. 13, No. 9, 2021, p. 2069.

[26] R. O. Dogan, H. Dogan, C. Bayrak, T. Kayikcioglu, "A two-phase approach using mask R-CNN and 3D U-Net for high-accuracy automatic segmentation of pancreas in CT imaging", Computer Methods and Programs in Biomedicine, Vol. 207, 2021, p. 106141.

[27] M. G. Dinesh, N. Bacanin, S. S. Askar, M. Abouhawwash, "Diagnostic ability of deep learning in detection of pancreatic tumour", Scientific Reports, Vol. 13, No. 1, 2023, p. 9725.

[28] A. Turečková, T. Tureček, Z. Komínková Oplatková, A. Rodríguez-Sánchez, "Improving CT image tumor segmentation through deep supervision and attentional gates", Frontiers in Robotics and AI, Vol. 7, No. 1, 2020, p. 106.

[29] K. Si, Y. Xue, X. Yu, X. Zhu, Q. Li, W. Gong, T. Liang, S. Duan, "Fully end-to-end deep-learning-based diagnosis of pancreatic tumors", Theranostics, Vol.11, No.4, 2021, p. 1982.

[30] J. J. Qiu, J. Yin, W. Qian, J. H. Liu, Z. X. Huang, H. P. Yu, L. Ji, X. X. Zeng, "A novel multiresolution-statistical texture analysis architecture: radionics-aided diagnosis of PDAC based on plain CT images", IEEE Transactions on Medical Imaging, Vol. 40, No. 1, 2020, pp. 12-25.

[31] H. Ma, Z. X. Liu, J. J. Zhang, F. T. Wu, C. F. Xu, Z. Shen, C. H. Yu, Y. M. Li, "Construction of a convolutional neural network classifier developed by computed tomography images for pancreatic cancer diagnosis", World Journal of Gastroenterology, Vol. 26, No. 34, 2020, p. 5156.

[32] A.L. Simpson *et al*. "A large annotated medical image dataset for the development and evaluation of segmentation algorithms", arXiv:1902.09063v1, 2019.

[33] Pancreas-CT, "The Cancer Imaging Archive (TCIA)", Public Access-Cancer Imaging Archive Wiki, Vol. 1, 2022.

[34] Y. Bian, Z. Zheng, X. Fang, H. Jiang, M. Zhu, J. Yu, H. Zhao, L. Zhang, J. Yao, L. Lu, J. Lu, "Artificial intelligence to predict lymph node metastasis at CT in pancreatic ductal adenocarcinoma", Radiology, Vol. 306, No. 1, 2023, pp. 160-169.

[35] K. Sekaran, P. Chandana, N. M. Krishna, S. Kadry, "Deep learning convolutional neural network (CNN) With Gaussian mixture model for predicting pancreatic cancer, Multimedia Tools and Applications", Vol. 79, No. 15, 2020, pp. 10233-10247.

[36] T. Vaiyapuri, A. K. Dutta, I. H. Punithavathi, P. Duraipandy, S. S. Alotaibi, H. Alsolai, A. Mohamed, H. Mahgoub, "Intelligent deep-learning-enabled decision-making medical system for pancreatic tumor classification on CT images", Healthcare, Vol. 10, No. 4, 2022, p. 677.

[37] M. M. Althobaiti, A. Almulihi, A. A. Ashour, R. F. Mansour, D. Gupta, "Design of optimal deep learning-based pancreatic tumor and nontumor classification model using computed tomography scans", Journal of Healthcare Engineering, Vol. 2022, 2022, pp. 1-15.

[38] R. Khdhir, A. Belghith, S. Othmen, "Pancreatic Cancer Segmentation and Classification in CT Imaging using Antlion Optimization and Deep Learning Mechanism", International Journal of Advanced Computer Science and Applications, Vol. 14, No. 3, 2023.

Machine Learning Approaches to Advanced Outlier Detection in Psychological Datasets

Original Scientific Paper

Khoula Al. Abri
Universiti Tenaga Nasional
Department of College of Computing and Informatics
Kajang, Selangor Malaysia
Khoula.Alabri@hotmail.com

Manjit Singh Sidhu
Universiti Tenaga Nasional
Department of College of Computing and Informatics
Kajang, Selangor Malaysia
manjit@uniten.edu.my

Abstract *– The core aim of this study is to determine the most effective outlier detection methodologies for multivariate psychological datasets, particularly those derived from Omani students. Due to their complex nature, such datasets demand robust analytical methods. To this end, we employed three sophisticated algorithms: local outlier factor (LOF), one-class support vector machine (OCSVM), and isolation forest (IF). Our initial findings showed 155 outliers by both LOF and IF and 147 by OCSVM. A deeper analysis revealed that LOF detected 55 unique outliers based on differences in local density, OCSVM isolated 44 unique outliers utilizing its transformed feature space, and IF identified 76 unique outliers leveraging its tree-based mechanics. Despite these varying results, all methods had a consensus for just 44 outliers. Employing ensemble techniques, both averaging and voting methods identified 155 outliers, whereas the weighted method highlighted 151, with a consensus of 150 outliers across the board. In conclusion, while individual algorithms provide distinct perspectives, ensemble techniques enhance the accuracy and consistency of outlier detection. This underscores the necessity of using multiple algorithms with ensemble techniques in analyzing psychological datasets, facilitating a richer comprehension of inherent data structures.*

Keywords*: Outlier Detection, psychological dataset, machine learning techniques, ensemble methods*

1. INTRODUCTION

Outliers are data points that differ substantially from the overall pattern of a dataset [1] and have long been a topic of interest in various scientific fields, including psychology. In questionnaire-based psychological research, outliers pose unique challenges [2]. Considering the inherent variability in human responses and psychological traits, the presence of outliers can severely affect the accuracy and consistency of the findings. This is mainly because outliers can skew statistical measures and lead to biased or misleading conclusions. Consequently, detecting and managing outliers is critical in ensuring the integrity and credibility of research in psychology [3].

Identifying outliers in datasets presents significant challenges, especially in the context of multivariate datasets [4]. Our research is specifically designed for detecting outliers in observational data. The driving force behind our approach is to enhance its robustness, preparing our dataset for deeper, more precise subsequent analyses [5 , 6].

However, traditional techniques encounter substantial limitations when applied to multivariate datasets, especially those that lack predetermined labels for outlier identification. These constraints expose a significant gap in our methodological framework, emphasizing the immediate need for the design and implementation of advanced techniques [7]. Ideally, such techniques would be capable of effectively addressing the complex characteristics inherent in multivariate psychological datasets.

With advancements in machine learning, several cutting-edge algorithms have emerged as promising alternatives for outlier detection. In this study, we examine three algorithms: LOF, OCSVM, and IF, which have demonstrated effectiveness in various models for outlier detection [8, 9, 10]. The LOF algorithm measures how isolated a data point is from its neighbors, allowing

us to identify outliers that have a significantly different density from their neighbors [8]. The OCSVM algorithm effectively determines the decision boundary that distinguishes outliers from the remaining data points [10]. The IF algorithm employs a unique approach based on randomly partitioning the dataset and isolating outliers based on the number of partitions [9]. We use an integrative approach by combining these algorithms using ensemble methods such as averaging, voting, and weighted combining approaches. This ensures the results are more reliable and stable. Ensemble methods have the advantage of mitigating the weaknesses of individual algorithms while harnessing their strengths [11].

Outlier detection in psychological datasets has attracted significant attention due to its crucial impact on the accuracy of findings. However, there is a noticeable gap in research on unlabeled multivariate datasets [4]. Most of the existing methods focus on outlier detection in numerical data, predominantly using proximity-based techniques [7]. In their research, "Ahmad A. A. Alkhatib" and "Qusai Abed-Al" explored the use of ensemble methods for outlier detection in multivariate a forest fire environment dataset. Their findings emphasized the advantages of leveraging multiple algorithms to enhance accuracy and robustness.

The main contribution of our research is the detailed analysis of a unique dataset derived from Omani students, consequently bridging a significant geographical and cultural gap in the literature. By utilizing three advanced outlier detection algorithms—specifically LOF, OCSVM, and IF—our study highlights the data-driven aspects of identifying fabricated responses in psychological datasets. To further enhance our methodological robustness, we have integrated ensemble techniques, achieving heightened accuracy in outlier detection. This multi-layered approach, when combined with a comparative analysis of algorithmic outputs, demonstrates our work's position at the intersection of psychology and machine learning. It effectively addresses critical methodological challenges and provides a robust framework to detect anomalies in multivariate psychological datasets.

This paper is structured as follows: Section II introduces the materials and methodology, explaining the algorithms and ensemble strategies employed for outlier detection. Section III details the results, presenting a comparative analysis of the algorithms and their performance alongside ensemble techniques. Section IV offers a discussion, focusing on the integration between individual algorithmic strengths and their combined capabilities. Finally, Section V concludes the study by highlighting insights from the examination of outlier detection algorithms and emphasizing the complex nature of anomalies in datasets.

2. MATERIALS AND METHOD

This research utilized a dataset obtained from tenth-grade students in Oman over four months, with the assistance of the Ministry of Education and career experts. The dataset comprises 1004 observations, each corresponding to a unique student. Within the dataset, there are 142 features, encompassing both psychological measures and demographic attributes. This research primarily focused on detecting outliers in the observations and did not delve into feature extraction or selection. Fig.1 displays a selected subset of the unlabeled multivariate psychological dataset.

Gender	Region	Parents Marital Status	Result in grade 9	Weight	Height
Female	Al Batinah North	Married	80% - 90%	50 kg - 60 kg	150 cm - 155 cm
Male	Musandam	Married	80% - 90%	80 kg - 90 kg	175 cm - 180 cm
Male	Musandam	Married	80% - 90%	70 kg - 80 kg	175 cm - 180 cm
Male	Musandam	Married	80% - 90%	70 kg - 80 kg	170 cm - 175 cm
Female	Al Batinah North	Married	70% - 80%	50 kg - 60 kg	160 cm - 165 cm
Female	Al Batinah North	Married	90% - 100%	40 kg - 50 kg	160 cm - 165 cm
Female	Al Batinah North	Married	80% - 90%	40 kg - 50 kg	150 cm - 155 cm
Female	Al Batinah North	Married	60% - 70%	50 kg - 60 kg	150 cm - 155 cm
Male	Ash Sharqiyah North	Married	50% - 60%	40 kg - 50 kg	170 cm - 175 cm
Female	Al Batinah North	Married	50% - 60%	30 kg - 40 kg	145 cm - 150 cm
Female	Al Batinah North	Married	80% - 90%	40 kg - 50 kg	145 cm - 150 cm
Female	Ad Dhahirah	Married	90% - 100%	30 kg - 40 kg	145 cm - 150 cm
Female	Ad Dhahirah	Married	60% - 70%	50 kg - 60 kg	150 cm - 155 cm
Female	Al Batinah North	Married	70% - 80%	30 kg - 40 kg	160 cm - 165 cm
Female	Ad Dhahirah	Married	80% - 90%	70 kg - 80 kg	150 cm - 155 cm
Female	Al Batinah North	Married	80% - 90%	50 kg - 60 kg	150 cm - 155 cm
Female	Al Batinah North	Married	80% - 90%	70 kg - 80 kg	160 cm - 165 cm

Fig. 1. A subset of unlabeled psychological

2.1. MULTIFACETED APPROACHES: THREE ALGORITHMS IN OUTLIER DETECTION

This research aims to identify outliers using LOF, OCSVM, and IF. These different algorithms were chosen because they are known to be flexible and successful in detecting outliers in high-dimensional datasets. These algorithms also exhibit excellent outlier sensitivity, indicating a strong capacity to distinguish between outliers and inliers.

Parameters play a pivotal role in determining the efficacy of an algorithm's outlier detection capability. For this specific dataset and outlier detection task, the techniques were optimized by adjusting these parameters [12]. Specific adjustments include the number of neighbors and contamination level for the LOF algorithm, the “Nu” value and kernel type for the OCSVM, and the number of estimators and contamination for the IF algorithm. Such parameters can significantly influence the sensitivity of outlier detection. Table 1 provides a detailed breakdown of these algorithm parameters. In this section, we detail the methodologies and experimental setups for both the selected algorithms and ensemble techniques.

A. Local outlier factor

Within the scope of this study, the LOF algorithm is configured with two primary parameters. The first parameter, "n_neighbors," is adjusted to 20. This defines the number of neighboring data points that the algorithm considers when computing local density, thereby allowing each data point to be understood about its 20 closest neighbors in the dataset. In our visual analyses, n_neighbors=20 showed an optimal balance, offering a clear distinction between outliers and inliers, and ensuring consistent outlier identification across the dataset [13].

As illustrated in Fig. 2, the choice of the number of neighbors, n_neighbors, significantly impacts the LOF scores. Observing the visual patterns, it becomes evident that setting n_neighbors to 20 provides a clear boundary between inliers and outliers. This decision was largely driven by visual analyses, ensuring a balance between detection accuracy and specificity while minimizing overlaps and potential over-sensitivity. The visuals effectively validate our selection, emphasizing its capability to accurately identify genuine outliers.

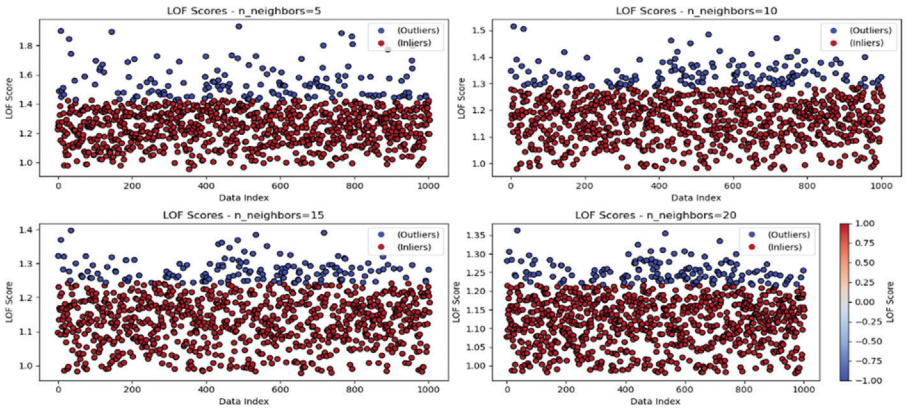


Fig. 2. Comparison of LOF Scores for Various Numbers of Neighbors

The second parameter, "contamination," is utilized to determine the threshold for outlier identification. By setting the "contamination" parameter to 0.15%, we define the acceptable ratio of observations that can be classified as outliers without raising false positives.

In the LOF algorithm, a key computation is the "reachability distance" (RD) between two data points, represented as y_1 and y_2 [14]. The RD incorporates both the local density around y_2 and the Euclidean distance separating y_1 and y_2. Mathematically, RD is the greater of two distances: the distance from y_2 to its k-th closest data point and the direct distance between y_1 and y_2.This is formally captured in equation (1) as:

$$\mathrm{RD}(\mathrm{y}_1, \mathrm{y}_2) = \max\{\mathrm{k} - \mathrm{dist}(\mathrm{y}_2), \mathrm{dist}(\mathrm{y}_1, \mathrm{y}_2)\} \quad (1)$$

Here, " *k-dist* (y_2)" represents the distance from y_2 to its k-th nearest neighbor, and " dist (y_1,y_2)" denotes the Euclidean distance between y_1 and y_2 [8].

Subsequently, we calculate the Local Reachability Density (LRD) for each data point, specifically y_1 [15]. LRD is essentially the inverse of the average RD of y_1 based on its k nearest neighbors. It involves summing the RD to each of y_1's k neighbors and computing the inverse of the average. LRD provides a quantitative measure of the density of data points in the vicinity of y_1. LRD is determined using the equation (2):

$$\mathrm{LRD}(\mathrm{y}_1) = \frac{1}{\frac{\Sigma \mathrm{RD}(\mathrm{y}_1, \mathrm{y}_2)}{\mathrm{K}}} \quad (2)$$

LOF evaluates the local density of y_1 in comparison to that of its neighbors. Data points with densities comparable to their neighbors possess LOF values around 1, while outliers with significantly lower local densities have LOF values much higher [8]. The equation (3) represents the LOF:

$$LOF(y_1) = \Sigma \frac{\left(\frac{\mathrm{LRD}(y_2)}{\mathrm{LRD}(y_1)}\right)}{\mathrm{k}} \quad (3)$$

Utilizing the LOF algorithm in this manner isolates outliers by analyzing variations in local density relative to neighboring data points. This ensures a robust and accurate distinction between outliers and inliers based on their surrounding data context.

B. One class support vector machine

The OCSVM algorithm was utilized with a specific configuration comprising a Radial Basis Function (RBF) kernel, a 'nu' parameter set at 15%, and a 'gamma' parameter defined as 'auto', which corresponds to 1 divided by the number of observations. The core functionality of OCSVM involves the identification of a hyperplane within a transformed feature space, which effectively segregates the dataset from the origin [16]. The fundamental objective is to ascertain the optimal hyperplane by minimizing the function using equation (4):

$$\frac{1}{2} * ||\mathrm{w}||^2 + \left(\frac{1}{\nu * \mathrm{n}}\right) * \Sigma\xi_\mathrm{i} - \rho \quad (4)$$

subject to the constraints using equation 5:

$$\mathrm{w}^\mathrm{T}\varphi(\mathrm{y}_\mathrm{i}) \geq \rho - \xi_\mathrm{i}, \text{and } \xi_\mathrm{i} \geq 0 \quad (5)$$

The weight vector, represented by 'w', is fundamental in defining the decision hyperplane. This vector directly impacts the alignment of the hyperplane in the feature space. In contrast, φ represents the feature mapping function responsible for transforming input data into a different space. This transformation is driven by the selected kernel, which in our setup is the Radial Basis Function (RBF). The parameter 'ρ' is vital since it sets the decision boundary and is determined based on the dataset. Meanwhile, 'ξ' denotes the slack variables which grant the model some tolerance by allowing certain data points to reside on the undesirable side of the decision boundary. This flexibility is essential to strike a balance between optimizing the margin and managing outliers in the training set. The decision function is formulated in equation 6 as follows:

$$F(y) = \text{sgn}(w^T \varphi(y) - \rho) \quad (6)$$

Data points adhering to the criterion $w^T \varphi(y) - \rho > 0$ are located within the decision boundary and are consequently classified as inliers. Alternatively, instances for which $w^T \varphi(y) - \rho < 0$ are designated as outliers.

Moreover, the decision hyperplane is typified by the collection of points x in the feature space that satisfies the equation (7):

$$w^T \varphi(y) - \rho = 0 \quad (7)$$

This hyperplane acts as the decision boundary in the transformed feature space, with the points lying on one side classified as inliers and those on the opposite side classified as outliers [16].

C. Isolation forest

In this study, the IF algorithm is employed with a specific configuration that includes 100 trees (estimators) and a contamination level of 0.15%. The algorithm calculates an anomaly score, $S(y, x)$, for each data point y. This score is calculated using the equation (8):

$$S 2^{-\frac{E(h(y))}{c(x)}} \quad (8)$$

where $E(h(y))$ represents the average path length from the root to the terminal node for instance x across all trees, and $c(x)$ is the anticipated average path distance of a failed search in a Binary Search Tree [17]. To separate outliers from inliers, a threshold T is calculated based on the contamination level. This is done using the equation (9):

$$T = quantile\ (S,\ contamination\_Level\ ) \quad (9)$$

Data points with an anomaly score less than 0 are marked as outliers in equation (10), while those with an anomaly score equal to or greater than 0 are marked as inliers in equation (11). In mathematical terms:

$$x \text{ is an outlier if } S(x, n) < 0 \quad (10)$$

$$x \text{ is an inlier if } S(x, n) >= 0 \quad (11)$$

This method enables the algorithm to distinguish between inliers and outliers' data points effectively by isolating the points that have shorter paths in the trees, which usually signifies that they are less like the other data [18].

2.2. SYNERGY IN DETECTION: AN EXPLORATION OF ENSEMBLE METHODS

The objective of ensemble methods in outlier detection is to harness the strengths of multiple algorithms and enhance the overall performance and reliability of the outlier detection process. Rather than depending on a singular or single algorithm, ensemble methods combine predictions or outlier scores from multiple algorithms to make more robust decisions about the outlier status of data points. Ensemble methods in outlier detection employ different techniques to integrate the predictions or outlier scores from multiple algorithms [19].

A. Averaging Approach

This method calculates the average outlier scores provided by different algorithms for each data point. By taking the means of these scores, this approach seeks to provide a more consistent and balanced outlier score that reduces the individual biases inherent in any single algorithm [19]. Let S_{LOF}, S_{OCSVM}, and S_{IF} represent the outlier scores from LOF, OCSVM, and IF algorithms respectively for a particular data point [20]. The combined score $S_{Average}$ for that data point using the averaging approach is given by Equation 12.

$$S_{Average}=(S\_LOF+S\_OCSVM+S\_IF)/3 \quad (12)$$

B. Voting Approach

The voting approach in outlier detection determines the status of a data point based on a majority vote from multiple algorithms. Each of these algorithms produces a binary vote, classifying the data point as either an outlier (1) or an inlier (0) using a predetermined threshold [21]. S_{LOF}, S_{OCSVM}, and S_{IF} represent the binary classifications from the LOF, OCSVM, and IF algorithms, respectively, for a particular data point.

A data point is classified as an outlier if the sum of these classifications is 2 or 3, yielding an overall score of vote is 1. Conversely, if the sum is 0 or 1, the data point is deemed an inlier, giving the vote's score a value of 0. Formally, the voting can be represented in equation 13 as:

$$\begin{cases} 1 \ if\ (S_{LOF} + S_{OCSVM} + S_{IF}) \geq 2 \\ 0 \quad Othewise \end{cases} \quad (13)$$

C. Weighted Sum Approach

The weighted sum approach assigns variable weights to the outlier scores based on criteria such as the performance or reliability of each algorithm. By assigning different weights, it allows for giving more importance to the outlier scores from algorithms that have demonstrated better performance [22]. They $S_{weighted}$ are computed by multiplying the outlier scores S_{LOF}, S_{OCSVM}, and S_{IF} from LOF, OCSVM, and IF by their respective weights (W_{lof}, W_{OVSVM}, W_{IF}), and then summing them as shown in equation 14.

$$Sweighted = (SLOF \text{ x } Wlof + SOCSVM \text{ x } WOVSVM + SIF \text{ x } WIF) \quad (14)$$

3. RESULTS

In the results framework of our research, we investigate a methodical assessment of algorithmic benchmarks in outlier detection, comparing singular methodologies with the robustness of ensemble techniques. Our objective is to provide an analytical perspective on the effectiveness of individual algorithms and the synergistic capabilities of aggregated ensemble systems in detecting outliers (fabricated responses).

3.1. ALGORITHM PERFORMANCE IN OUTLIER DETECTION

Upon examining the outcomes from the LOF, OCSVM, and IF algorithms, we identified several critical

insights related to their effectiveness in detecting outliers. Firstly, the LOF algorithm demonstrated its capability by successfully identifying a total of 151 outliers from the dataset. Significantly, these outliers displayed an average outlier score of 0.6992, emphasizing their distinct characteristics when compared to the remaining data points. Furthermore, the standard deviation of the outlier scores was determined to be 0.7149, indicating a moderate level of variability in the identified outliers' scores.

In the provided scatter plot (Fig. 3), the Local Outlier Factor (LOF) scores depict data point anomalies based on local density. Most data points, represented as yellow dots, fall within the LOF score range for inliers: (0.9797, 1.2154), signifying they are inliers in compact regions. Conversely, the purple dots, scattered higher on the chart with scores ranging from 1.2161 to 1.3629, denote outliers in sparser areas. This distinction highlights the core concept of LOF: measuring data point deviation from a typical distribution. Thus, the plot provides a clear visual contrast between the inliers and the outliers.

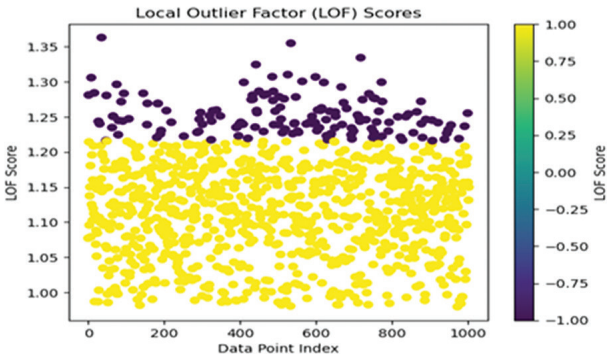


Fig. 3. Outlier detection for LOF algorithm

Subsequently, the OCSVM algorithm identified 147 outliers, which is slightly lower than anticipated. These outliers are depicted as red points in Fig. 4 and have an impressive average score of 1.840. The significant standard deviation of 2.271 for these scores indicates a broad range, bearing witness to OCSVM's sensitivity. Specifically, the OCSVM score range for inliers was (0.0000, 22.8469), and for outliers was (-9.8256, -0.0000).

Fig. 4 visually presents the difference between inliers and outliers. The dense cluster of blue points represents inliers, but they don't congregate close to zero. Some blue inliers, especially those around the 15 mark, illustrate the unique nature of the data. Despite their distance from the main group, they're still categorized as normal. On the other hand, the scattered red points show the outliers' positions, demonstrating OCSVM's ability to detect points that deviate significantly from standard patterns.

Finally, the Isolation Forest (IF) algorithm identified an equivalent number of outliers as the LOF algorithm. The IF algorithm, however, displayed a lower average outlier score of 0.0263, signifying a notable deviation from the mean score. Furthermore, the standard deviation for these scores was significantly reduced, indicating a more concentrated distribution of anomaly scores among the detected outliers, suggesting that the IF algorithm can identify outliers that closely conform to the standard deviation, differentiating it from the LOF and OCSVM algorithms.

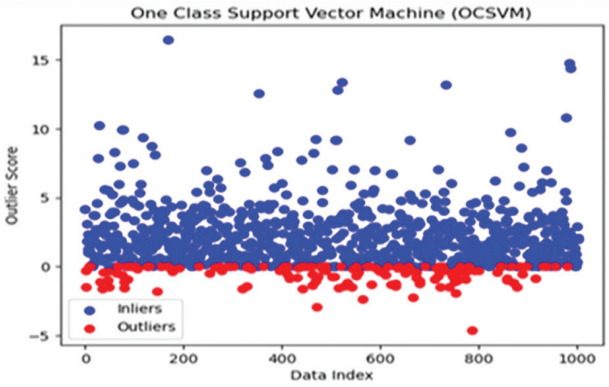


Fig. 4. Outlier detection of OCSVM algorithm

Fig. 5 presents the results obtained from the IF algorithm. The blue points represent inliers, clustering mostly around the higher regions of the graph. Specifically, the score range for inliers was between (0.0000, and 0.0946). Conversely, the red points, dispersed mostly in the lower part, highlight the outliers. The score range for these outliers was between (-0.0883, -0.0000).

This distinction emphasizes the efficiency of the IF algorithm in distinguishing between regular data points and those that deviate from the norm. The concentration of red points near the -0.000 to -0.040 range further highlights the close alignment of detected outliers with the mentioned standard deviation. IF algorithm demonstrates a precise and efficient approach to outlier detection.

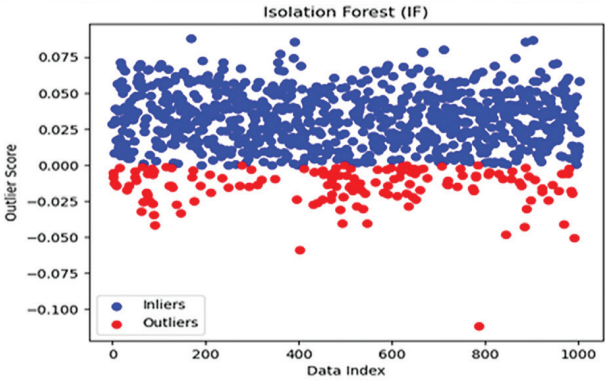


Fig. 5. Outlier detection for IF

Table. 1. Outlier Detection Overview

	LOF	OCSVM	IF	Common
Total Outliers	155	147	155	46
Unique Outliers	55	44	76	

Table 1 provides a comparative analysis of outliers detected using the three algorithms: LOF, OCSVM, and IF. Both the LOF and IF algorithms pinpointed the same number of outliers, amounting to 155, while the OCSVM identified a slightly fewer number, 147. On initial observation, this similarity in count might indicate consistent and standardized results across the three models. However, a deeper investigation of the data reveals nuanced variations.

When examining the unique outliers detected by each algorithm, we observe variations that suggest different underlying algorithms and sensitivities in anomaly detection. LOF, which utilizes the method of measuring local deviation of density relative to its neighbors, identified 55 unique outliers. OCSVM, using the SVM technique to separate normal and outlier data points in a transformed feature space, detected 44 unique outliers. The IF algorithm, employing tree-based partitioning strategies, identified 76 unique outliers. This difference in the unique anomalies found by each algorithm highlights the variations in their detection mechanisms.

Furthermore, even with the proximity in the overall outlier counts, the actual number common across all three algorithms is only limited to 41 outliers. This signifies that while the models have some overlapping observations, they also bring their unique perspectives in anomaly detection.

In essence, while the three algorithms display similar overall counts of outliers, their observations diverge considerably. Such variations underline the importance of employing multiple algorithms in anomaly detection tasks to capture a broader and more varied understanding of the underlying data.

3.2. OUTLIER DETECTION THROUGH ENSEMBLE TECHNIQUE

The Averaging algorithm consistently detected 155 outliers across various iterations. This consistency in identifying outliers illustrates both the robustness of the method and the effective coordination among the integrated algorithms. Fig. 6 provides an in-depth visualization of the Averaging approach. On the x-axis, data points range from 0 to 1000, while the y-axis presents combined scores, which extend from -1 to 1. The continuous blue line signifies the averaged combined scores for each data point. The distribution of these scores appears quite consistent, indicating a steady data trend.

The red dots concentrated mainly in the lower section of the graph represent the detected outliers. Their position suggests that these points deviate from the average scores by a significant margin. The clustering of the outliers around specific score values emphasizes the effectiveness of the Averaging approach in detecting anomalous data points that deviate from the standard.

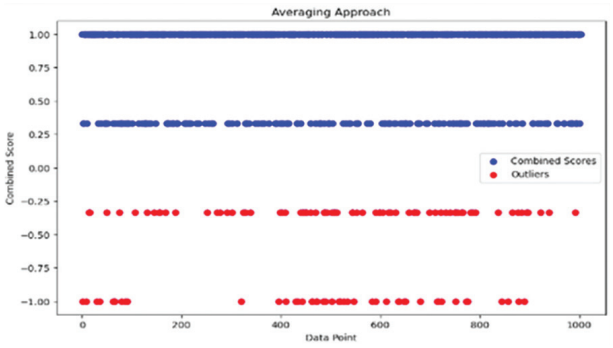


Fig. 6. Averaging approach

Similarly, the Voting method consistently detected 155 outliers across multiple iterations. The coherence in output illustrates the compatibility among the involved algorithms and confirms the efficiency of the Voting technique. In Fig. 7, we observe the Voting method in detail. As in the previous graph, the horizontal axis plots the data points, while the vertical axis illustrates voting scores, which range between -3 and 3.

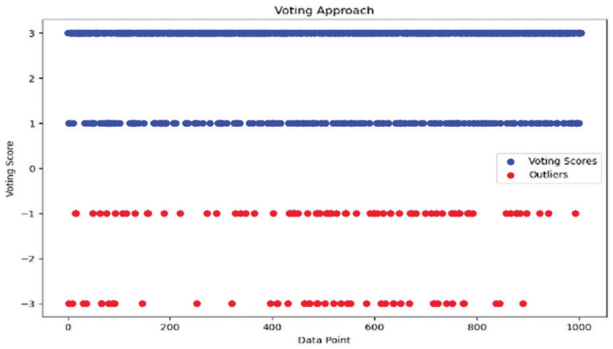


Fig. 7. Voting approach

Blue dots indicate the inliers, providing a visualization of how each data point was interpreted across the ensemble. Most of these scores are clustered within a narrow range, demonstrating a consensus among the involved algorithms. In contrast, the red dots, which represent the outliers, differ significantly from most of the blue dots. This points out the consistent identification of these points as anomalies across multiple iterations.

Subsequently, the Weighted approach, demonstrating performance metrics comparable to the Averaging and Voting methods, detected 151 outliers across multiple iterations. This uniformity strengthens the reliability and precision of the Weighted system. Fig. 8 presents the results of the Weighted approach. The graph's design aligns with Fig. 6 and 7, featuring data points on the horizontal axis and weighted combined scores on the vertical, varying from -1 to 5.

Blue dots represent the weighted scores, indicating the variations in weights allocated to various data points. A large portion of these points have scores centered around the middle, implying a consistent distribution of weights from different algorithms. Conversely, the red dots, which denote outliers, are more spread

out compared to the Averaging and Voting methods, suggesting that the Weighted method provides a more detailed identification of outliers based on the ensemble algorithms' weighted scores.

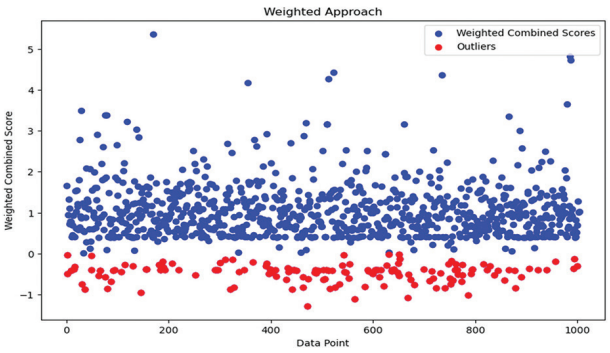


Fig. 8. Weighted approach

After a comprehensive evaluation of the three algorithms, it became evident that 150 outliers, interpreted as fabricated responses or imprecisions, were consistently identified, highlighting a robust consistency across the combined approaches.

4. DISCUSSION

In the domain of current algorithmic research, the synergy between the capabilities of individual algorithms and their combined implementations in outlier detection is crucial and remains a focal point in contemporary technology research. The findings presented in this study highlight the deep complexities embedded in these integrations.

From an individual algorithm perspective, LOF, OC-SVM, and IF each demonstrated unique strengths in detecting outliers, with their underlying principles and methodologies varying significantly. While there was a similarity in the overall counts, the number of outliers identified consistently by all three algorithms was relatively few, pointing to the distinct observation patterns intrinsic to each method. Such diversity indicates the notion that relying on a single detection technique might not capture all facets of outliers present in a dataset.

The table on outlier detection provides an insightful overview of the strengths and weaknesses of various algorithms. While the overall data indicates consistency in their results, a detailed analysis shows that each algorithm identifies distinct data points that might be missed by others. It's crucial to observe that even an algorithm with a high success rate might sometimes fail to detect certain outliers that other algorithms capture. This element emphasizes the significance of comprehending the functionality of each algorithm, especially when there's a requirement to optimize or adjust them for specific datasets.

A core finding from our research highlights the advantages of using integrated methods for outlier detection. The consistent results from techniques such as Averaging, Voting, and Weighted methods demonstrate the benefits of combining the capabilities of different algorithms. Each of these combined approaches identified outliers in the range of 151-155, showcasing their stable performance. Moreover, the identification of 150 common outliers among these methods shows a significant overlap, strengthening confidence in the accuracy of these aggregated results.

It's crucial to emphasize that the effectiveness of using combined methods is not only evident in their results but also in their ability to detect outliers potentially overlooked by individual algorithms. Striking a balance between comprehensive analysis and detailed outlier identification becomes essential for developers and researchers in determining the best detection method.

5. CONCLUSION

From the comprehensive exploration of outlier detection algorithms, it's evident that individual algorithms offer specific strengths in the field of data analysis. Their diverse observational patterns and the distinct outliers they detect illustrate the multifaceted nature of anomalies in datasets. In this study, the use of a combined approach proves invaluable. Techniques like averaging, weighted sum, and voting have been more consistent and trustworthy. When compared, single algorithms offer insights into some aspects of the data, but when combined, they present a more comprehensive view.

Thus, while individual algorithms provide unique perspectives on outlier detection, combined methods offer a holistic view. However, as with all computational methods, tailoring the approach to the specific needs of the dataset and research question is paramount. The results of this study pave the way for more in-depth discussions and research in the realm of outlier detection in computer science.

Given the complexity of contemporary datasets, it's essential for data analysts to continually refine and explore these combined methods. By doing so, they can more effectively interpret intricate data, thereby facilitating the identification of anomalous patterns. This, in turn, assists various industries in making informed decisions. Future studies should concentrate on optimizing the amalgamation of these techniques to ensure that data analysis remains robust and precise.

6. REFERENCES

[1] V. N. Tran, J. Kim, "Robust and efficient uncertainty quantification for extreme events that deviate significantly from the training dataset using polynomial chaos-kriging", Hydrology, Vol. 609, No. 6, 2022, p. 127716.

[2] R. Cao, X. Liu, J. Zhou, D. Chen, D. Peng, T. Chen, "Outlier Detection for Spotting Micro-expressions", Proceedings of the IEEE International Conference on

Bioinformatics and Biomedicine, Houston, TX, USA, 9-12 December 2021, pp. 3006-3011.

[3] X. Zhang, H. Ren, L. Gao, B. Shia, M. Chen, L. Ye, R. Wang, L. Qin, "Identifying the predictors of severe psychological distress by auto-machine learning methods", Informatics in Medicine Unlocked, Vol. 39, No. 4, 2023, p. 101258.

[4] K. Matsue, M. Sugiyama, "Unsupervised feature extraction from multivariate time series for outlier detection", Intelligent Data Analysis, Vol. 26, No. 6, 2022, pp. 1451-1467.

[5] S. Li, Y. Wang, X. Xie, "Prediction of uniaxial compression strength of limestone based on the point load strength and SVM model", Minerals, Vol. 11, No. 12, 2021, p. 1387.

[6] B. Dastjerdy, A. Saeidi, S. Heidarzadeh, "Review of Applicable Outlier Detection Methods to Treat Geomechanical Data", Geotechnics, Vol. 3, No. 2, 2023, pp. 375-396.

[7] H. M. Campbell, A. E. Murata, J. T. Mao, B. McMahon, G. H. Murata, "A novel method for handling pre-existing conditions in multivariate prediction model development for COVID-19 death in the Department of Veterans Affairs", Biological Methods and Protocols, Vol. 7, No. 1, 2022, pp. 1-18.

[8] O. Alghushairy, R. Alsini, T. Soule, X. Ma, "A Review of Local Outlier Factor Algorithms for Outlier Detection in Big Data Streams", Big Data and Cognitive Computing, Vol. 5, No. 1, 2020, pp. 1-24.

[9] M. Heigl, K. A. Anand, A. Urmann, D. Fiala, M. Schramm, R. Hable, "On the Improvement of the Isolation Forest Algorithm for Outlier Detection with Streaming Data", Electronics, Vol. 10, No. 13, 2021, pp. 15-34.

[10] E. H. Budiarto, A. E. Permanasari, S. Fauziati, "Unsupervised Anomaly Detection Using K-Means, Local Outlier Factor and One-Class SVM", Proceedings of the 5th International Conference on Science and Technology, Yogyakarta, Indonesia, 30-31 July 2019, pp. 1-5.

[11] A. M. Jabbar, "Local and Global Outlier Detection Algorithms in Unsupervised Approach: A Review", Iraqi Journal for Electrical and Electronic Engineering, Vol. 17, No. 1, 2021, pp. 76-87.

[12] H. Alazzam, A. Sharieh, K. E. Sabri, "A lightweight intelligent network intrusion detection system using OCSVM and pigeon inspired optimizer", Applied Intelligence, Vol. 52, No. 4, 2022, pp. 3527-3544.

[13] S. Zhong, S. Fu, L. Lin, X. Fu, Z. Cui, R. Wang, "A novel unsupervised anomaly detection for gas turbine using Isolation Forest", Proceedings of the IEEE International Conference on Prognostics and Health Management, San Francisco, CA, USA, 17-20 June 2019, pp. 1-6.

[14] Z. Chen, K. Xu, J. Wei, G. Dong, "Voltage fault detection for lithium-ion battery pack using local outlier factor", Measurement, Vol. 146, No. 16, 2019, pp. 544-556.

[15] P. R. M. Júnior, T. Boult, J. Wainer, A. Rocha, "Open-Set Support Vector Machines", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 52, No. 6, 2022, pp. 3785-3798.

[16] W. B. Richardson, J. Meyer, S. V. Solms, "Towards Machine Learning and Low Data Rate IoT for Fault Detection in Data-Driven Predictive Maintenance", Proceedings of the IEEE World AI IoT Congress, Seattle, WA, USA, 10-13 May 2021, pp. 202-208.

[17] W. Zhang, Z. Lin, X. Liu, "Short-term offshore wind power forecasting - A hybrid model based on Discrete Wavelet Transform (DWT), Seasonal Autoregressive Integrated Moving Average (SARIMA), and deep-learning-based Long Short-Term Memory (LSTM)", Renewable Energy, Vol. 185, No. 5, 2022, pp. 611-628.

[18] S. Borrohou, R. Fissoune, H. Badir, "Data cleaning survey and challenges – improving outlier detection algorithm in machine learning", Journal of Smart Cities and Society, Vol. 2, No. 3, 2023, pp. 125-140.

[19] S. Ardabili, A. Mosavi, A. R. Várkonyi-Kóczy, "Advances in machine learning modeling: Reviewing hybrid and ensemble methods", Proceedings of the International Conference on Global Research and Education: Engineering for Sustainable Future, Balatonfüred, Hungary, 4-7 September 2019, pp. 215-227.

[20] C. Wang, X. Li, J. Sun, X. Zhang, Y. Li, X. Peng, J. Liu, Z. Jiao, "Ensemble Learning Model of Power System Transient Stability Assessment Based on Bayesian Model Averaging Method", Proceedings of 6th IEEE Conference on Energy Internet and Energy System Integration, Changsha, China, 8-10 November 2022, pp. 1467-1471.

[21] R. Atallah, A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method", Proceedings of the 2nd International Conference on New Trends in Computing Sciences, Amman, Jordan, 9-11 October 2019, pp. 1-6.

[22] S. Cai, R. Sun, S. Hao, S. Li, G. Yuan, "An efficient outlier detection approach on weighted data stream based on minimal rare pattern mining", China Communications, Vol. 16, No. 10, 2019, pp. 83-99.

Empirical Forecasting Analysis of Bitcoin Prices: A Comparison of Machine Learning, Deep Learning, and Ensemble Learning Models

Original Scientific Paper

Nrusingha Tripathy
Department of Computer Science and Engineering,
Siksha 'O' Anusandhan Deemed to be University,
Bhubaneswar, India
nrusinghatripathy654@gmail.com

Pranati Satapathy
Department of IMCA,
Utkal University,
Bhubaneswar, Odisha, India
satapathy.pranati@gmail.com

Sarbeswara Hota
Department of Computer Application,
Siksha 'O' Anusandhan Deemed to be University,
Bhubaneswar, India
sarbeswarahota@soa.ac.in

Subrat Kumar Nayak
Department of Computer Science and Engineering,
Siksha 'O' Anusandhan Deemed to be University,
Bhubaneswar, India
subratsilicon28@gmail.com

Debahuti Mishra
Department of Computer Science and Engineering,
Siksha 'O' Anusandhan Deemed to be University,
Bhubaneswar, India
debahutimishra@soa.ac.in

Abstract *– Bitcoin has drawn a lot of interest recently as a possible high-earning investment. There are significant financial risks associated with its erratic price volatility. Therefore, investors and decision-makers place great significance on being able to precisely foresee and capture shifting patterns in the Bitcoin market. However, empirical studies on the systems that support Bitcoin trading and forecasting are still in their infancy. The suggested method will predict the prices of all key cryptocurrencies with accuracy. A number of factors are going to be taken into account in order to precisely predict the pricing. By leveraging encryption technology, cryptocurrencies may serve as an online accounting framework and a medium of exchange. The main goal of this work is to predict Bitcoin price. To address the drawbacks of traditional forecasting techniques, we use a variety of machine learning, deep learning, and ensemble learning algorithms. We conduct a performance analysis of Auto-Regressive Integrated Moving Averages (ARIMA), Long-Short-Term Memory (LSTM), FB-Prophet, XGBoost, and a pair of hybrid formulations, LSTM-GRU and LSTM-1D_CNN. Utilizing historical Bitcoin data from 2012 to 2020, we compared the models with their Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The hybrid LSTM-GRU model outperforms the rest with a Mean Absolute Error (MAE) of 0.464 and a Root Mean Squared Error (RMSE) of 0.323. The finding has significant ramifications for market analysts and investors in digital currencies.*

Keywords*: Bitcoin, Cryptocurrency, Arima, LSTM, Prophet, XGBOOST, Hybrid model, Prediction*

1. INTRODUCTION

Using freely available software and peer-to-peer networks, the digital currency known as Bitcoin was developed to serve as a private, untraceable payment system because Bitcoin is not backed by any institution or government and has no physical existence. The concept of cryptocurrencies, or digital currency, served as the inspiration for Bitcoin. Cryptocurrencies are digital, computer-generated money that only exists online. Cryptocurrency uses cryptographic methods to secure transactions and is designed to be used as a means of exchange [1]. Today's financial systems employ cryptocurrencies, which are virtual or digital money. All cryptocurrencies are built on very complicated blockchain technology, which promises to stock data in a manner that is dreadful to hack and modify. It is difficult to create phony cryptocurrencies because of the additional security provided by cryptography for these currencies. El Salvador became the first country to recognize Bitcoin under the US dollar as legal money

in November 2021 [2], even though cryptocurrencies are still difficult to forecast if they will ever be extensively used in international markets. The most appealing market for financial speculation at the moment is the one for cryptocurrency. By making use of digital market speculators and marketing investment schemes that incorporate Bitcoins, Ethereum, and other cryptocurrencies yet are replete with unforeseen traps, many people have gained significant sums of money.

However, cryptocurrencies are extremely volatile. The element that raises the volatility of cryptocurrencies is their smaller market cap than that of equities. Due to the still-small size of the cryptocurrency market, even a tiny investment can have a significant impact on a cryptocurrency's price. Most trading algorithms include various parameters, largely rely on historical data, and fail to handle high volatility [3, 4]. There are several methods for them such as ARIMA and LSTM-based on Recurrent Neural Networks (RNNs). Nevertheless, understanding their behaviour is difficult because each of them requires a lot of parameters.

These algorithms work well for patterns that persist, but in the world of cryptocurrencies, trends and patterns may change quickly. To make more accurate forecasts, an algorithm can handle all of these problems including holidays that are identified in advance, lacking observations, substantial outliers, and seasonal impacts carried on by humans. The main contributions of the paper are summarised as follows:

- We take the historical Bitcoin data of the past eight years, from January 2012 to September 2020. Initially, the data is pre-processed and we make the data balanced.
- The hybrid LSTM-GRU model gives the best result by taking RMSE and MAE as parameters, after that with the help of our best model we go for future prediction.
- Various machine learning and deep learning methods were investigated, but the hybrid LSTM-GRU model performs remarkably well.

2. LITERATURE SURVEY

Given the fact that Bitcoin is still a relatively new technology, there aren't many models available for predicting its price. Time series models interact with information gathered from regular time sequences, 10-minute intervals, and intervals of ten seconds. These models produced three sets of time series data for periods of 30, 60, and 120 minutes, and then three linear models utilizing GLM/Random Forest were produced from the datasets. For predicting the price of Bitcoin, all three of these models are linearly constructed from information obtained from daily time sequences, 10-minute periods, and 10-second intervals [5].

Forecasting the stock's trend instead of predicting its price in the years, a pattern may be deduced from the trend. These might offer both quick forecasts (for a day or a week) and extended ones (for months). The longer projections were shown to be more accurate (79%) and to have better outcomes. The network's achievement of the evaluation standard, which is based on expected output, is also taken into consideration in the research. For Bitcoin prediction, the authors employ machine learning strategies that include gradient descent, linear search, and both deep learning and regression approaches. They are not founded on the same data-related assumptions as ARIMA and Prophet, including time series stationary motion or the availability of a Date field. The disadvantage of LSTM-based RNNs is that it is difficult to comprehend and anticipate how they will behave. To acquire good results, meticulous hyperparameter adjustment is also required. The accuracy of the LSTM is significantly impacted by substantial simultaneous seasonality in the context of cryptocurrency [6].

The cryptocurrency trading market changes rapidly over only a brief amount of time even though little is known about its characteristics, periodicity can occur on a daily, weekly, or even every-hour basis on average, data contains large outliers, and it doesn't just depend on historical data. Each of them could have a detrimental influence on how well the ARIMA model performs. For techniques like ARIMA to provide decent results, further tinkering may be necessary, which is typically out of scope for many people who lack the necessary understanding. In this study, a special model based on Facebook Prophet is created to alleviate the limitations of the ARIMA model in terms of Bitcoin trading [7].

Prophet's hyperparameters don't need much tweaking because their main goal is to find correlations in commercial time series. Outliers and trend shifts based on new goods and market events are often manageable by the Prophet. Prophet resists missing data and changes in trend. In opposition to Auto-Arima, Prophet shows a seasonal pattern that is more accurate, despite the fact the absolute values are considerably off from the real information. Prophet is exceptional in that it doesn't necessitate any prior expertise or expertise in forecasting time series data due to its capacity to instinctively recognize the fundamental seasonal patterns within the data and give a set of "simple to understand" criteria.

Prophet was created to handle planned holidays, missing data, and significant outliers [8]. It may therefore be utilized by non-statisticians to produce conclusions that are generally on par with, if not superior to, those produced by experts. The most efficient method for learning from sequential data is probably LSTM-based neural network, of which time series are a special example. Using 4857377 the digital currency Bitcoin data source.

We undertake a fair study on the Auto-Regressive Integrated Moving Average (ARIMA), Long-Short-Term Memory (LSTM), FB-Prophet, XGBoost, and two hybrid formulations, LSTM-GRU and LSTM-1D_CNN.

3. MATERIAL AND METHODS

3.1. DATASET

The major focus of this study is the time-series forecast of BTC prices using different machine learning and deep learning models. An accumulation of data values for a sequence of time points is referred to as a time series. We collected the "Bitcoin Historical Data" from Kaggle competitions. Here, historical market data for a few exchanges that accept Bitcoin is made available at 1-minute intervals. The OHLC (Open, High, Low, Close) minute-by-minute updates, volume in BTC and the designated currency, and adjusted Bitcoin price, covers the period from January 2012 to September 2020. The Open and Close columns display the daily opening and closing prices. Utilizing historical Bitcoin data from 2012 to 2020, we contrasted the models.

3.2. METHODOLOGY

We used the Bitcoin data to apply current machine learning and front-line deep learning models. We used well-known models including the ARIMA, LSTM, Prophet, XGBoost, LSTM-GRU and LSTM-1D CNN model. Because the data were spatial-temporal, models that utilized LSTM and GRU approaches produced superior results. We provide a framework for better analysis, as shown in Fig. 1 The following phases are used to explain the framework in detail.

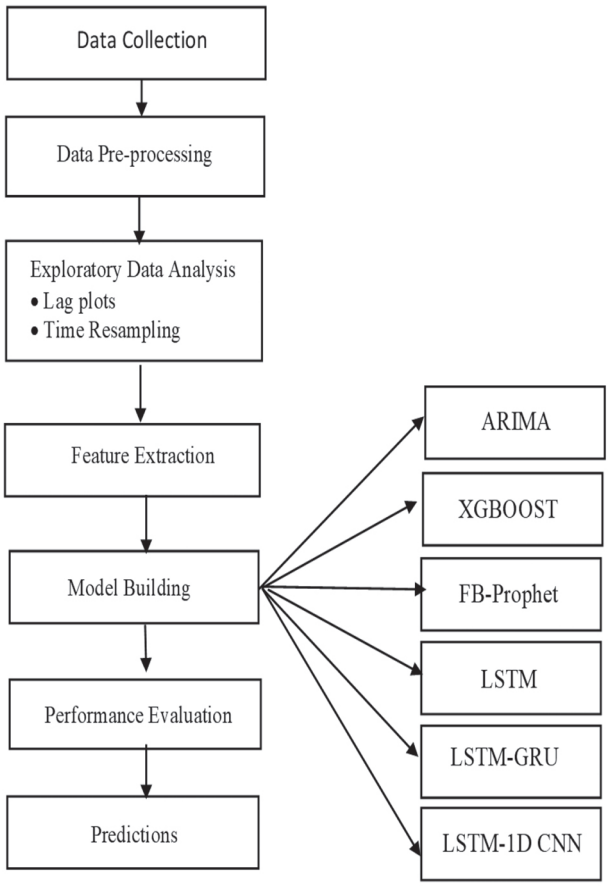


Fig. 1. Proposed Methodology

3.2.1. HANDLING MISSING VALUES

Missing data occurs when no information is provided for a couple of elements. In the actual world, inadequate data is a big problem. NA (Not Available) values may also be referred to as missing values in pandas. Several datasets in data frames sometimes contain missing values, either as a result of the data never being collected or as a result of the data being available but not being recorded.

Table 1. Attribute names with missing values

Attribute Name	Total Missing Values	Missing %
Timestamp	0	0.00000
Open	1243608	25.60246
High	1243608	25.60246
Low	1243608	25.60246
Close	1243608	25.60246
Volume_(BTC)	1243608	25.60246
Volume_(BTC)	1243608	25.60246
Weighted Price	1243608	25.60246

To overcome this situation, we use Fill values forward shortly called 'ffill' or 'pad', and Fill values backward similarly called 'bfill' or 'backfill', and the linear interpolations method was used as three imputation strategies to replace the missing data. The observed value is used to replace NaNs in the 'ffill' technique. The next observed value is used to replace NaNs in the 'bfill' procedure. These two techniques allow for the filling of a sizable amount of the missing information. The linear interpolation technique is used to fill in the missing data. It is a restoration strategy that, on the premise that there is an immediate relationship between the data points, substitutes a missing data point with non-missing values from nearby data points [9]. No null values are discovered when we use these three methods on our dataset. Table 1 describes different attribute names and their missing values.

3.2.2. Exploratory Data Analysis

a. Visualizing the weighted price

Visualizing time-series data may reveal a lot when dealing with it. To assist in highlighting certain observations or specific occurrences in the time series, markers can be added to the plot. Since the cryptocurrency market is open around the clock and cryptocurrencies are prone to wild price fluctuations, traders are always seeking methods to take advantage of openings [10]. Visualizing the weighted price enables traders to decide when to purchase or sell. Fig. 2 shows the weighted price vs. years plotting.

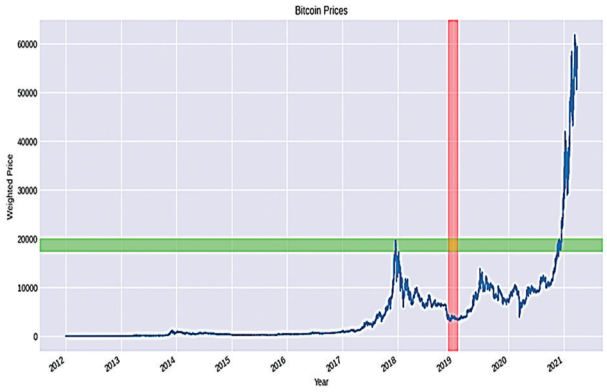


Fig. 2. Weighted price vs. Years

b. Visualizing using Lag Plots

To establish whether the outcomes in a collection of data or a series of values are random or not, lag graphs are utilized. If the information being analyzed is random, the lag plot fails to show any discernible patterns.

Is there going to be more unwanted mail because of user risky behavior (e.g. questionable registrations around the World Wide Web) [10]?

- Is there going to be more unwanted mail because users leave their addresses around on the Internet [11]?
- Does it hold that even a careful e-mail user is obligated to eventually start receiving unwanted mail?

If the facts are not randomly distributed, the lag plot will clearly show a pattern. The type of pattern can assist the user in identifying the irregular structure in the data. Outliers can also be found using lag graphs. The minutely, hour, making it, and day lag graphs all exhibit a positive correlation, as can be shown [11, 12]. We discovered no relationship for month lag plots. We replicated all of our information to the everyday level maximum to preserve autocorrelation. Fig. 3 shows the lag plots with different time intervals.

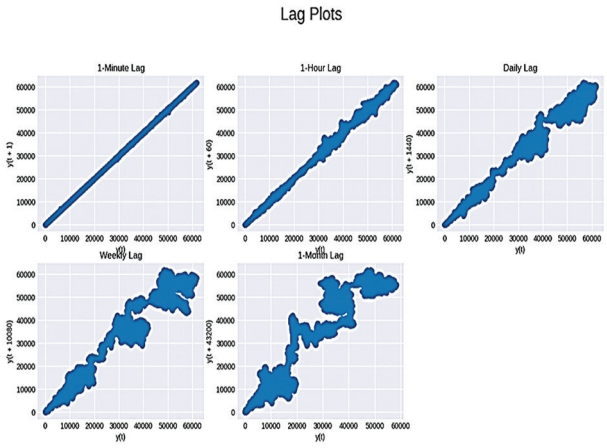


Fig. 3. Lag plots

c. Time Series Decomposition

Trend, seasonal, and other components can be separated into a time series. The series may be created by multiplying or adding the starting point, trend, periodic index, and residual [13]. Then, using statistical analyses such as the KPSS and enhanced Dickey-Fuller tests, we verified stationarity. After decomposing the time data, we discover no indication of seasonality. Because the variance, covariance, and mean are not constant, the series is also non-stationary. Fig. 4 shows the time series decompositions plotting.

Fig. 4. Time series Decompositions

d. Auto Correlation Function (ACF)

Autocorrelation is the relationship between a given time series and an extended version of itself. The time series' relationship with itself is the initial lag of the automatic correlation function (ACF), which results in a correlation of 1.

Fig. 5. Auto Correlation Function (ACF) for Weighted Price

A computed number called Auto Correlation Function (ACF) is used to show how closely a value in a time series resembles a prior value. Python's Statsmoldels package greatly simplifies the process of computing autocorrelation [14, 15]. Fig. 5 shows the Auto Correlation Function (ACF) for Weighted Price.

e. Partial Auto Correlation Function (PACF)

In time series analysis, the partly autocorrelation function (PACF) regresses the time series' components at all shorter lags to determine the complete relation-

ship between a stationary time series and its individual independently delayed values. Unlike the autocorrelation function, it ignores additional delays. The partial autocorrelation at lag k is the synchronization between X_t_t and X_(t-k) that is ignored by delays 1 through k1. By setting the method option to "ols" (regression of time series dependent on it and on constant lags), we will use the plot_pacf function of the statsmodels.graphics.tsaplots package [16]. Fig. 6 shows the Partial Auto Correlation Function (PACF).

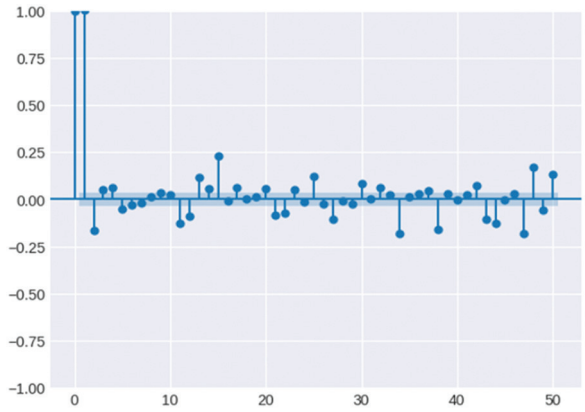


Fig. 6. Partial Auto Correlation Function (PACF)

3.2.3. Feature Extraction

Time series data could be noisy as a result of the substantial market movements. It might be difficult to spot a pattern or trend in the data. There is a tonne of noise when evaluating everyday data. Our data's noise level has significantly decreased, and the upward trend is now more obvious than the real data. There is an excessive amount of noise in the daily figures that we are looking at. We use a rolling mean to average this out by one week, which is handy. The rolling mean, sometimes called the moving average, is a data processing method that helps remove noise from data by averaging [17]. Simple division and agglomeration of the data into windows based on factors like mean, median, count, etc. The trend is now more obvious than the actual data since we use a rolling mean for 3, 7, and 30 days.

4. MODEL BUILDING

4.1. AUTO-REGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODEL

To better comprehend the data set or forecast future changes, time series data are employed in a statistical analysis model called ARIMA. ARIMA stands for Auto-Regressive Integrating Moving Average. It's a collection of models that may be applied to time series data to depict a range of frequent temporal patterns [18, 19]. This acronym is descriptive and encapsulates the key model characteristics. The model forecasts upcoming securities or changes in financial markets by focusing on discrepancies between values compared to actual values across the series.

4.2. FACEBOOK PROPHET

Facebook's Prophet is a freely available approach for creating time-series models that include both classic ideas and cutting-edge advances. It succeeds at simulating time series with various seasonality and does not share the drawbacks of other algorithms. It depends on an element of the model in which seasonality (weekly, yearly, plus vacation) and non-linear patterns are fitted together. Even in the context of incomplete data, Prophet can identify major outliers and variations in the trend. Furthermore, it precisely calculates the mixed data without requiring manual labor. Prophet was created to benefit from Facebook business predictions. Examples include trends with non-linear growth curves, significant outliers, trend changes, and daily, monthly, and yearly historical observations [20, 21].

4.3. XGBOOST

A universal gradient boosting library called XGBoost designed to be rapid and scalable for training machine learning models. XGBoost, which stands for Extreme Gradient Boosting, is a popular and powerful machine learning algorithm used for both regression and classification tasks. It belongs to the ensemble learning family of algorithms and is known for its exceptional predictive performance. The ensemble learning approach is used to aggregate results from a variety of weak models to get a more accurate forecast. The "ds" column among them is used to store date-time series. The values for the relevant time series in the data frame are kept in the "y" column, which is the other column. As a result, the outline works fairly well with seasonal time series and offers a few options for handling the dataset's seasonality.

To employ libraries like xgboost in Python, which provide user-friendly APIs for building and training XGBoost models. This library offers extensive documentation and various tuning options to help and get the most out of the algorithm.

4.4. LONG SHORT-TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) were introduced by Sepp Hochreiter and Jurgen Schmidhuber in 1997 to address the vanishing gradient problem that can occur when training traditional Recurrent Neural networks (RNNs). LSTMs are designed to capture long-range dependencies in sequential data by maintaining a memory cell that can store information over long sequences. They achieve this by using a more complex structure than simple RNNs, which allows them to learn and remember information over longer time intervals [22]. RNNs come in a variety of types, with LSTM networks being the most popular. The input and forget gates modulate the inner contents of the memory cell. Since both segue are closed, the information in the memory cell will not change from one time step to the next. This allows groups of information to flow across many time steps. As a result, the LSTM model can correct-

ly overcome the vanishing gradient issue that most RNN models have [23, 24].

4.5. LSTM-GRU MODEL

We originally used a composite LSTM-GRU model. LSTM is used to overcome the problem of vanishing gradients in backpropagation. The three gates that make up an LSTM are the input gate (IG), forget gate (fg), and output gate (og). Gates are used in memory to store information. Information is kept there in analog format. Sigmoid function ranges between 0 and 1 are multiplied element-by-element on these gates. If the gate's value is zero, the data is ignored or deleted.

Tan(h) is a popular non-linear activation function that has a range of 1 to +1. The information does not fade when a second derivative is used. A sigmoid function has values between 0 and 1. In simple terms, the method is used to suggest to memory components called gates whether data should be destroyed or maintained. Equations discuss input gate, forget gate, and output gate (og) mathematical equations that were taken from the literature and changed. The update gate (ug) and reset gate (rg) are GRU's two gates, in contrast. This approach required feeding the LSTM the output of the GRU. At a specific time and location, the input set of characteristics exit contains hybrid feature space (start-node, end-node, way-id, week, hour, agg-minutes, quarter, holiday, peak hour, etc).

4.6. LSTM-1D_CNN MODEL

Nine layers make up the suggested algorithm for predicting bitcoin prices. In Fig. 7, the model architecture is shown.

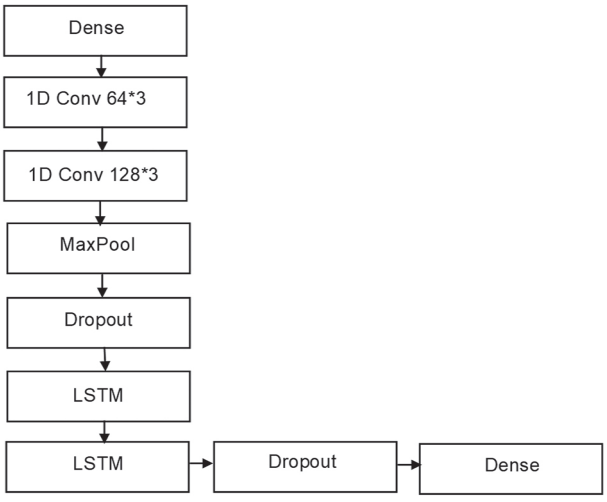


Fig. 7. Architecture of LSTM-1D_CNN

There are 2 LSTM layers and 2 1D CNN layers. Relu serves as the activation function for each convolutional layer. To prevent overfitting, two dropout layers have also been included. Both 0.5 and 0.4 are dropout rates. The output unit has the number 5. There are 200 in the batch.

The most dependable performance is determined by examining a range of input circumstances. Two hybrid deep learning models work incredibly well for forecasting time series. It has two key benefits, namely: (1) In order to make accurate predictions, the model does away with the time series decomposition procedure and (2) strengthens its high-level temporal representation.

5. PERFORMANCE MEASURE OF HYBRID DEEP LEARNING MODELS

The approach's performance and pattern are evaluated using well-known and current performance evaluation measures. Additionally, it suggests the ideal model for achieving output label efficiency. We have utilized established performance assessment measures like RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) to assess the solution.

The information and formulae for the current performance evaluation measures are obtained. The abbreviation RMSE stands for root-mean-square deviation. The mean squared variance between intended output and projected output is what this term represents.

$$\text{MSE} = \sqrt{(\sum_i (Y_{\text{dsi}} - Y_{\text{psi}})^2 / n} \text{ (number of observations)}$$

Where,

Y_{dsi} = desired, Y_{psi} = predicted speed, n= number of observations.

The MAE (Mean-Absolute-Error) in our example measures how far away from the required output speed our anticipated output speed is. We can get a mathematical explanation for this.

$$MAE = 1/n \sum_i |(Y_{dsi} - Y_{psi})|$$

6. RESULTS AND ANALYSIS

The graphical depiction of the results and analysis on the test dataset is demonstrated in this part. Fig. 8 displays the ARIMA projected Bitcoin price, where the x and y axes represent the months and the sample price rate. Fig. 9 and Fig. 10 illustrate the XG Boost and FB-Prophet predicted Bitcoin price, similarly, Fig. 11, Fig. 12 and Fig.13 illustrate the LSTM, hybrid LSTM-1D_CNN and LSTM-GRU predicted Bitcoin price respectively. In this collection, the test data is taken from January 2020 through September 2020.

Table 2, shows a comparison analysis of proposed method and previous method in Bitcoin price prediction. we quickly contrast the outcomes of earlier research in the field with our newly proposed method. Table 3 shows that the RMSE value for the ARIMA is 18100.14, whereas the values for the FB-Prophet, XG Boost, and LSTM are 1309.49, 13571.91, and 421.29 respectively but when we hybridize the combinations of LSTM-GRU and LSTM-1D_CNN the hybrid model performs better than others. The RMSE value of LSTM-1D_CNN is 83.408 and LSTM-GRU is 0.323.

Overall, the suggested hybrid deep learning based prediction model LSTM-GRU result is significant as compared to others.

Both individual investors and asset managers must be able to predict the price of bitcoin. Despite the fact that Bitcoin is a currency, it cannot be analysed similarly to other traditional currencies where economic theories about future cash flow models, purchasing power parity, and uncovered interest rate parity are relevant. This is due to the fact that the market for digital currencies like Bitcoin does not allow for the application of certain traditional criteria relating to the connection between supply and demand. Contrarily, Bitcoin has a number of characteristics that make it useful for investors, including transaction speed, dispersion, decentralisation, and the sizable online community of people interested in talking about and sharing important information concerning digital currencies, especially Bitcoin.

Table 2. Comparison of proposed method and previous method

Reference	Dataset	Method	Results
Yan Li et al. [18]	Bitcoin	BP, CNN, LSTM, CNN-LSTM	RMSE of BP = 515.35 RMSE of CNN = 261.90 RMSE of LSTM = 297.97 RMSE of CNN-LSTM= 258.31
Kervanci et al. [19]	Bitcoin	4-layer LSTM, LSTM and BO, 4-layer GRU	RMSE of 4-LSTM = 32.98 RMSE of LSTM & BO= 4260 RMSE of 4-GRU= 251.60
Hashish et al. [20]	Bitcoin	ARIMA, LSTM, HMM-LSTM	RMSE of ARIMA = 141.96 RMSE of LSTM = 7.006 RMSE of HMM-LSTM= 5.82
Proposed work (ensemble deep model)	Bitcoin	ARIMA, LSTM, FB Prophet, XGBoost, LSTM-1D_CNN LSTM-GRU	RMSE of ARIMA = 18100 RMSE of LSTM = 421 RMSE of Prophet= 1309 RMSE of XGBoost= 13571 RMSE of LSTM-1D_CNN = 83.408 RMSE of LSTM-GRU= 0.323

Fig. 8. ARIMA Predicted BTC Price

Fig. 9. XGBoost Predicted BTC Price

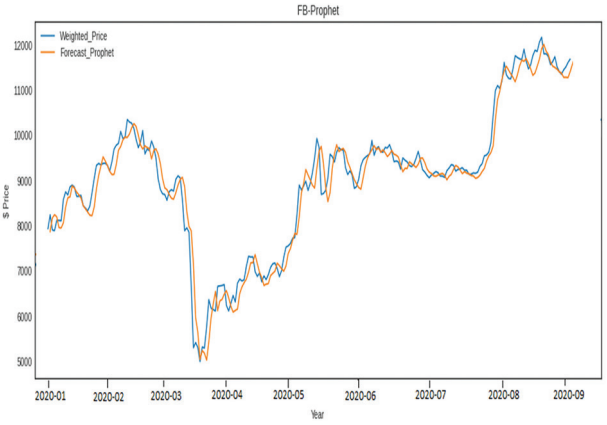


Fig. 10. FB Prophet Predicted BTC Price

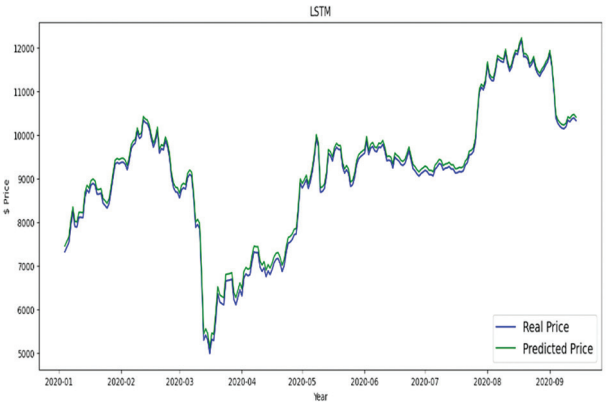


Fig. 11. LSTM Predicted BTC Price

Fig. 12. LSTM-1D_CNN Predicted BTC Price

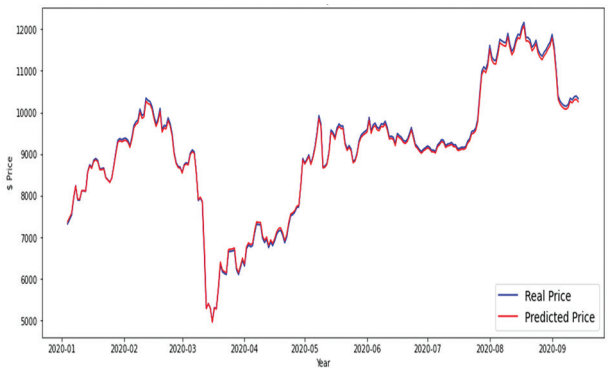


Fig. 13. LSTM-GRU Predicted BTC Price

Table 3. RMSE, MAE values of models

Model Name	RMSE Value	MAE Value
ARIMA	18100.14	10733.36
XG_Boost	13571.91	6500.36
FB-Prophet	1309.49	690.08
LSTM	421.29	20.518
LSTM-1D_CNN	83.408	9.140
LSTM-GRU	0.323	0.464

7. CONCLUSION

With the swings of the financial trading market and the precision of forecasting models, the market for Bitcoin mechanics alters dynamically. A comparison analysis must be done to properly comprehend the parallels and contrasts between various financial enterprises using cryptocurrency. Several deep-learning prediction methods are introduced in this work for forecasting Bitcoin's market behaviour. we use two ensemble models, both of which give an eye-catching result. The hybrid LSTM-1D_CNN gives an RMSE value of 83.408 and an MAE value of 9.140, while the LSTM-GRU gives an RMSE score of 0.323 and an MAE score of 0.464. The hybrid LSTM-GRU model performed well as compared to others.

We employed time series forecasting since output in the past was unpredictable. Deep learning techniques can recognize and benefit from the links and patterns contained in a data set through a self-learning process. Contrary to conventional approaches, algorithms based on deep learning can analyse the linkages and undetected trends within the data to effectively represent this type of data and offer a reliable forecast. Multivariate time series research employs several deep learning methods. The models might potentially be trained using small datasets and short-term predictions in the future. This study might be enhanced by using other hybrid deep learning techniques to forecast Bitcoin performance.

8. REFERENCE

[1] R. Schmidt, M. Möhring, D. Glück, R. Haerting, B. Keller, C. Reichstein, "Benefits from Using Bitcoin: Empirical Evidence from a European Country", International Journal of Service Science, Management, Engineering, and Technology, Vol. 7, 2016, pp. 48-62.

[2] Z. Shahbazi, Y. C. Byun, "Knowledge Discovery on Cryptocurrency Exchange Rate Prediction Using Machine Learning Pipelines", Sensors, Vol. 22, 2022, p. 1740.

[3] G. Prashanth, G. V. S. S. K. R. Naganjaneyulu, M. Revanth, A. V. Narasimhadhan, "Multi Indicator based Hierarchical Strategies for Technical Analysis of Crypto market Paradigm", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 7, 2023, pp. 765-780.

[4] W. Chen, Z. Zheng, M. Ma, J. Wu, Y. Zhou, J. Yao, "Dependence structure between Bitcoin price and its influence factors", International Journal of Computational Science and Engineering, Vol. 21, 2020, pp. 334-345.

[5] S. F. Wamba, J. R. K. Kamdjoug, R. E. Bawack, J. G. Keogh, "Bitcoin Blockchain and Fintech: A systematic review and case studies in the supply chain", Production Planning & Control, Vol. 31, 2019, pp. 115-142.

[6] B. Manujakshi, M. G. Kabadi, N. Naik, "A Hybrid Stock Price Prediction Model Based on PRE and Deep Neural Network", Data, Vol. 7, 2022, p. 51.

[7] N. Tripathy, S. Hota, D. Mishra, "Performance analysis of bitcoin forecasting using deep learning techniques", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 31, No. 3, 2023, pp. 1515-1522.

[8] E. Loh, S. Ismail, "Emerging Trend of Transaction and Investment: Bitcoin Price Prediction using Machine Learning", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 9, 2020, pp. 100-104.

[9] M. J. Hamayel, A. Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRULSTM and bi-LSTM Machine Learning Algorithms", AI, Vol. 2, No. 4, 2021, pp. 477-496.

[10] T. Shintate, L. Pichl, "Trend Prediction Classification for High Frequency Bitcoin Time Series with Deep Learning", Journal of Risk and Financial Management, Vol. 12, No. 1, 2019, p. 17.

[11] V. Derbentsev, V. Babenko, K. Khrustalev, H. Obruch, S. Khrustalova, "Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices", International Journal of Engineering, Transactions A: Basics, Vol. 34, 2021, pp. 140-148.

[12] S. Yogeshwaran, M. J. Kaur, P. Maheshwari, "Project Based Learning: Predicting Bitcoin prices using deep learning", Proceedings of the IEEE Global Engineering Education Conference, Dubai, United Arab Emirates, 8-11 April 2019, pp. 1449-1454.

[13] N. Tripathy, S. Hota, S. Prusty, S. K. Nayak, "Performance Analysis of Deep Learning Techniques for Time Series Forecasting", Proceedings of the International Conference in Advances in Power, Signal, and Information Technology, Bhubaneswar, India, 9-11 June 2023, pp. 639-644.

[14] L. G. N. De Leon, R. C. Gomez, M. L. G. Tacal, J. V. Taylar, V. V. Nojor, A. R. Villanueva, "Bitcoin Price Forecasting using Time-series Architectures," Proceedings of the International Conference on ICT for Smart Society, Bandung, Indonesia, 10-11 August2022, pp. 1-6.

[15] M. A. Fauzi, N. Paiman, Z. Othman, "Bitcoin and cryptocurrency: Challenges opportunities and future works", Journal of Asian Finance Economics and Business, Vol. 7, No. 8, 2020, pp. 695-704.

[16] S. Lahiri, S. Bekiros, "Deep Learning Forecasting in Cryptocurrency High-Frequency Trading", Cognitive Computation, Vol. 13, No. 2, 2021, pp. 485-487.

[17] P. D. DeVries, "An analysis of cryptocurrency Bitcoin and the future", International Journal of Business Management and Commerce, Vol. 1, No. 2, 2016, pp. 1-9.

[18] Y. Li, W. Dai, "Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model", The Journal of Engineering, Vol. 2020, No. 13, 2020, pp. 344-347.

[19] I. S. Kervanci, M. F. Akay, E. Özceylan, "Bitcoin price prediction using LSTM, GRU and hybrid LSTM-GRU with bayesian optimization, random search, and grid search for the next days", Journal of Industrial and Management Optimization, 2023.

[20] I. A. Hashish, F. Forni, G. Andreotti, T. Facchinetti, S. Darjani, "A hybrid model for bitcoin prices prediction using hidden Markov models and optimized LSTM networks", IEEE International Conference on Emerging Technologies and Factory Automation, Zaragoza, Spain, 10-13 September 2019, pp. 721-728.

[21] N. Tripathy, S. K. Nayak, J. F. Godslove, I. K. Friday, S. S. Dalai, "Credit Card Fraud Detection Using Logistic Regression and Synthetic Minority Oversampling Technique (SMOTE) Approach", International Journal of Computer and Communication Technology, 2022, 8(4), 4.

[22] J. Liu, A. Serletis, (2019), "Volatility in the cryptocurrency market", Open Economies Review, 2019, 30, 779-811.

[23] Z. Ftiti, W. Louhichi, H. Ben Ameur, "Cryptocurrency volatility forecasting: What can we learn from the first wave of the COVID-19 outbreak?", Annals of Operations Research, 2021, 1-26.

[24] N. Tripathy, I. K. Friday, D. Rath, D. S. K. Nayak, S. Nayak, "Deep Learning-based Gated Recurrent Unit Approach to Stock Market Forecasting: An Analysis of Intel's Stock Data", International Journal of Smart Sensor and Adhoc Network, 2023, 4 (1).

Intelligent and secure real-time auto-stop car system using deep-learning models

Original Scientific Paper

Hiba Ali Ahmed
Department of Control and System Engineering
University of Technology-Iraq,
Baghdad, Iraq
cse.21.15@grad.uotechnology.edu.iq

Mohammed A. Noaman Al-hayanni
Department of Electrical Engineering
University of Technology-Iraq,
Baghdad, Iraq
mohammed.a.noaman@uotechnology.edu.iq

Muayad Sadik Croock
Department of Control and System Engineering
University of Technology-Iraq,
Baghdad, Iraq
muayad.s.croock@uotechnology.edu.iq

Abstract *– In this study, we introduce an innovative auto-stop car system empowered by deep learning technology, specifically employing two Convolutional Neural Networks (CNNs) for face recognition and travel drowsiness detection. Implemented on a Raspberry Pi 4, our system is designed to cater exclusively to certified drivers, ensuring enhanced safety through intelligent features. The face recognition CNN model accurately identifies authorized drivers, employing deep learning techniques to verify their identity before granting access to vehicle functions. This first model demonstrates a remarkable accuracy rate of 99.1%, surpassing existing solutions in secure driver authentication. Simultaneously, our second CNN focuses on real-time detecting+ of driver drowsiness, monitoring eye movements, and utilizing a touch sensor on the steering wheel. Upon detecting signs of drowsiness, the system issues an immediate alert through a speaker, initiating an emergency park and sending a distress message via Global Positioning System (GPS). The successful implementation of our proposed system on the Raspberry Pi 4, integrated with a real-time monitoring camera, attains an impressive accuracy of 99.1% for both deep learning models. This performance surpasses current industry benchmarks, showcasing the efficacy and reliability of our solution. Our auto-stop car system advances user convenience and establishes unparalleled safety standards, marking a significant stride in autonomous vehicle technology.*

Keywords*: auto-stop car system, CNN, deep learning, drowsiness recognition, face recognition*

1. INTRODUCTION

Many incidents, such as robberies and unexpectedly unwanted entrances, happen now. Security is therefore important in this day and age [1]. People are constantly preoccupied with daily activities in a safe, including private properties, houses, and cars [2, 3]. To access traditional systems, a user needs to identify himself using different ways, such as passwords, and intelligent cards. These security measures have certain flaws; for instance, they are susceptible to forgetfulness and theft by unauthorized parties. As a result, software that ensures a better security level is needed to be developed. Our brain's capacity to solely think in images rather than words is one of its distinctive qualities [4, 5]. You might once forget where you put your car key, but you have never forgotten to bring a face with you. Each individual's face is different and the most essential portion of the body, Thus, it can represent various feelings, including identification and weariness. It also reflects the emotional effects and the sleeping or drowsiness of people. The facial expression can explain the fatigue

problem in people as well as the other indicators, including vital parameters. In the driving mode of people, different vital parameters can be measured, such as heart rate, fever, eye activities, and visual activities, such as listening and speaking. Moreover, biometrics are also can be considered. This is done using sensors to measure movement, touch, availability, and so on. These biometric approaches have high detection accuracy, but they need more installation efforts to obtain the necessary measures. As a result, they are more challenging to implement in the real world than contactless methods [6, 7]. Because of this, most research in the field concentrates on non-intrusive detection techniques based on the driver's behavior to identify driver weariness or distraction. These non-intrusive techniques employ embedded sensors in various locations throughout the car to track steering wheel movement, lane deviation, and steering wheel angle. Due to the road, driving conditions are variable, the real-time systems with regular updating are necessary [8].

This paper presents a novel and intelligent car auto-stop system designed to enhance driver safety during vehicle operations. Leveraging state-of-the-art CNN-deep learning algorithms, our system integrates face and drowsiness-based eye recognition models to identify and monitor the driver's identity and alertness in real time. A key innovation of our approach is the seamless combination of these models within the proposed car auto-stop system, which is designed to activate a stop mode when the system detects that the driver is unconscious.

The implementation is carried out on a Raspberry Pi 4, utilizing a camera for efficient face and eye activity detection. Additionally, the system incorporates a touch sensor and a speaker for enhanced interactivity. Notably, the face and drowsiness models are trained on meticulously collected datasets, resulting in impressive testing accuracy rates of 99.1% for both. This high accuracy underscores the robustness and reliability of our proposed system.

By addressing the critical aspects of driver identity verification and real-time drowsiness detection, our work contributes significantly to intelligent transportation systems. The seamless integration of our models into a practical auto-stop system offers a tangible solution to enhance road safety by preventing potential accidents caused by unconscious driving. This paper serves as a valuable resource for researchers, practitioners, and policymakers interested in the intersection of deep learning, computer vision, and automotive safety.

2. RELATED WORK

The studied field was considered by several researchers around the world. In this section, we divided the related studies into two parts for simplicity. The recognition works for people's faces as well as emotional face gestures.

2.1. FACE RECOGNITION

The facial recognition system is a computer vision model designed to match a human face with a digital image or video frame. Facial detection or face verification poses a challenge in computer vision, and various real-time face detection applications exist. Subsequently, research efforts shifted towards developing a model that automatically detects human faces [9, 10].

In [11], the authors propose the Fisher face method, which stands out as one of the most widely used face recognition algorithms. Fisher Face utilizes Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) methods to extract image characteristics and identify faces based on the reduction in face space dimension. Fisher's face also demonstrates resistance to issues such as blurring and noise-induced image problems. The proposed system accuracy is 93%. In [12], the authors suggest that the memory constraints of embedded devices, owing to the complex structure of CNN, can be addressed by employing a Field Programmable Gate Array (FPGA) based accelerator for face feature extraction. This approach facilitates the acceleration of the entire CNN. Neural networks find applications in various fields, including healthcare, aerial picture categorization, and facial recognition. According to [13], the Raspberry Pi, a smart embedded device with a built-in camera capable of capturing images and videos, can detect motion using a motion detector sensor. The information detected can be transmitted to the administrator via the Wireless Fidelity (Wi-Fi) module of this advanced embedded device. The proposed system can detect any security threat with an accuracy of 95.5% and a precision of 91%. In [14], present research frameworks rely on extracting low-level features and representing mid-level features, but recent studies have prioritized utilizing deep learning models. Over the past decade to fifteen years, numerous distinctive algorithms for detecting human faces have been developed, and these algorithms find application in platforms like Facebook, WhatsApp, biometric verification, and autonomous vehicles. The highest accuracy achieved for Virtual Makeup (VMU), face recognition, and 14 celebrity datasets is 98%, 98.24%, 89.39%, and 95.71%, respectively.

2.2. DROWSINESS RECOGNITION

Using feature extraction-based algorithms, efforts in literature have been made to address the challenge of detecting weariness and sleepiness (drowsiness) [15]. The primary focus of these studies has been monitoring the driver's attention deficit. Techniques based on feature extraction, as indicated in [16] and [17], provided faster identification and lower computing complexity. However, they were more dependent on the quality of imaging and lighting. Recently, researchers have shifted their focus to driver face monitoring using deep learning, eliminating explicit feature extraction from raw images. When appropriately trained, machine

learning algorithms exhibit high precision and reliability in identifying drowsiness.

In [18], the proposed work establishes a drowsy detection and accident avoidance system based on eye blink duration. Initially, the open and closed states of the eyes are detected using the Eye-Aspect Ratio (EAR). Subsequently, the blink duration or count during transitions from open to closed states is analyzed. The system identifies drowsiness when the blink duration exceeds a certain limit, sending an alert message to the driver through an alarm. The developed system demonstrated an accuracy of approximately 92.5%. In [19], the authors introduced volunteer eye-blinking as a Human-Computer Interaction (HCI), employing an advanced computer vision detector for real-time processing with a generic camera. Eye-blink detection was performed in addition to eye-state identification, involving face recognition, modeling, Region of Interest (ROI) extraction, and eye-state classification. The system included a rotation compensator, an ROI evaluator, and a moving average filter. Two additional datasets, the Youtube Eye-state Classification (YEC) dataset and the Analysis of Biological Data (ABD), were created, achieving an accuracy of 97.44% for the CNN and Support Vector Machine (SVM) on various datasets. In [20], eye-blinking-based drowsiness detection was implemented, analyzing eye-blink patterns. Custom data were used for model training, and real-time fluctuation representations of eye landmarks were measured using deep learning methods. Experimental analysis demonstrated a correlation between yawning and closed eyes, categorized as drowsy, with an overall performance of 95.8% accuracy for drowsy-eye detection and 0.84% for yawning detection. In [21], the authors initiated their work on drowsiness detection by detecting yawning. They proposed a three-step method using the "Face Boxes" face detector to locate the driver's face and a DNN model to classify the face and detect yawning. Deep Neural Networks (DNN) model was trained using a dataset obtained from the developed Drive Safely system, achieving an accuracy of 95.2%. In [22], a deep learning-based approach was presented for detecting driver drowsiness using CNNs. eye and yawning were used for detecting drowsiness, achieving an average accuracy of 96% on the Yawing Detection Dataset (YawDD).

Finally, in [23], a model for evaluating driver fatigue based on eye state and yawning. Utilizing CNN and Visual Geometry Group (Vgg16), the model detected facial sleepiness expressions classified into four categories (open, closed, yawning, and no yawning). Testing on a dataset of 2900 images resulted in high accountability, with the CNN model achieving an accuracy rate of 97%.

3. PROPOSED AUTO-STOP CAR SYSTEM

The proposed auto-stop car system performs two phases of jobs. The first phase identifies the driver, while the second phase monitors the driver's status if entering the drowsiness mode. In this case, the system performs the auto-stop for a car. This section is divided into subsections to ease the structure of the proposed system.

3.1. GENERAL STRUCTURE

This work utilizes two deep-learning models to recognize four individuals and monitor them while driving for the cause of drowsiness. The face images of the four individuals are employed as the input dataset for the first deep learning model in the training phase. This particular model categorizes the images into four distinct groups corresponding to the previously mentioned, individuals, enabling us to identify a specific person. The second deep learning model is designed to identify drowsiness by analyzing the status of the person's eyes, whether open or closed. To accomplish this, we employ facial images of the four individuals, including closed and open-eye instances, as the input dataset for the second deep-learning model. This approach allows us to make a definitive determination regarding their drowsiness status. In the notification of the driver's closed eyes mode, the touch sensor ensures that the driver is catching the driving wheel, while the sound messages are sent to the driver through the speakers to wake up. If there is no response from the driver, the auto-stop car mode is activated. A general block diagram of the proposed system is presented in Fig. 1.

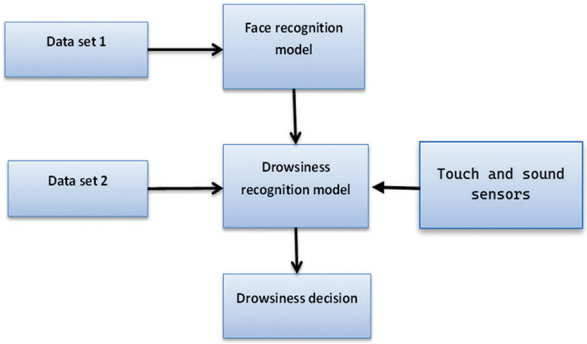
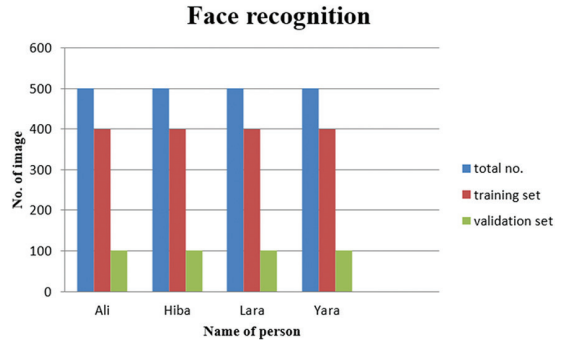


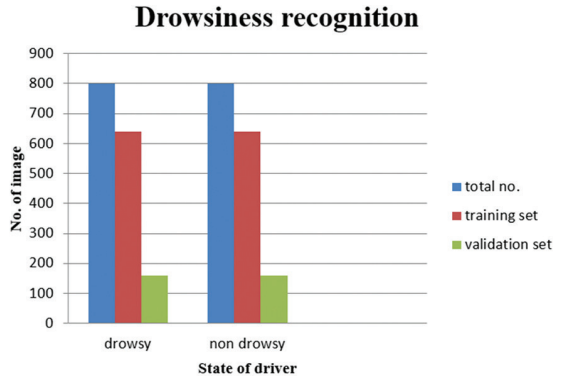
Fig. 1. General structure model

3.2. DATASET

In the proposed system, numerous images are collected to support training the presented deep learning models. In the driver identification model, 2000 images are prepared for the investigated four individuals: Ali, Hiba, Lara, and Yaraa. Each individual, class, is represented by 500 images (400 photos for training, constituting 80%, and 100 photos for validation, constituting 20%), as visually depicted in Fig. 2(a). Subsequently, we conducted an additional set of 1600 photos focusing on the same group of individuals to enable classification for drowsiness(conscious) detection. As illustrated in Fig. 2(b), each individual, or class, is depicted by 800 photos (640 photos allocated for training, which constitutes 80%, and 160 photos designated for validation, which constitutes 20%). This extensive dataset enhances the accuracy of our drowsiness classification process.



(a) Dataset for person face recognition



(b) Dataset for drowsiness recognition

Fig. 2. Dataset allocation for each stage of the proposed deep learning model

Typically, image preprocessing techniques are applied as a preliminary step before the training process of the models including reducing computational complexity and adjusting dimensions to enhance overall performance. In the case of this dataset, the original images are standardized to a resolution of 224 × 224 pixels. The image's pixel values were normalized to manipulate the intensity pixel values within a more manageable range. This normalization process entails transforming the pixel values, which originally fell within the range of (0-255), to a normalized range of (0-1) simplifying subsequent calculations.

3.3. PROPOSED DEEP LEARNING MODELS

The proposed backbone deep learning model for face and drowsiness recognition uses 15 CNN layers. The designed deep learning layers are shown as follows:

- Starting from the first layer, performed as the input layer. It carries out images from the pre-processing stage described in the pre-processing stage section.
- There are three stages of convolution layers; each stage consists of convolution and rectified linear units (Relu), which are the activation function, max-pooling, and dropout ranges (from 25% to 50%) layers.
- One fully connected layer is implemented.
- The dropout layer is adopted with a probability ratio of 50% before the final layer (sigmoid layer). It uses four classes of face images in the CNN model for face recognition. In comparison, the softmax layer uses two classes of face images in the CNN model for drowsiness recognition.

It is important to note that the differences between both proposed deep learning models in terms of sophisticated descriptions of each layer, including output shape and parameters, are illustrated in Tables 1 and 2.

Table 1. CNN's layer of face recognition

Layer (type)	Output Shape	Param #
image_input (InputLayer)	[(None, 224, 224, 3)]	0
layer_1 (Conv2D)	(None, 224, 224, 32)	896
layer_2 (Conv2D)	(None, 224, 224, 64)	18496
layer_3 (MaxPooling2D)	(None, 112, 112, 64)	0
layer_4 (Conv2D)	(None, 112, 112, 64)	36928
layer_5 (MaxPooling2D)	(None, 56, 56, 64)	0
dropout_1 (Dropout)	(None, 56, 56, 64)	0
layer_6 (Conv2D)	(None, 56, 56, 128)	73856
layer_7 (MaxPooling2D)	(None, 28, 28, 128)	0
dropout_2 (Dropout)	(None, 28, 28, 128)	0
fc_1 (Flatten)	(None, 100352)	0
layer_8 (Dense)	(None, 64)	0
dropout_3 (Dropout)	(None, 64)	0
pridictions (Dense)	(None, 2)	130

Table 1. CNN's layer of face recognition

Layer (type)	Output Shape	Param #
image_input (InputLayer)	[(None, 224, 224, 3)]	0
layer_1 (Conv2D)	(None, 224, 224, 32)	896
layer_2 (Conv2D)	(None, 224, 224, 64)	18496
layer_3 (MaxPooling2D)	(None, 112, 112, 64)	0
layer_4 (Conv2D)	(None, 112, 112, 64)	36928
layer_5 (MaxPooling2D)	(None, 56, 56, 64)	0
dropout_1 (Dropout)	(None, 56, 56, 64)	0
layer_6 (Conv2D)	(None, 56, 56, 128)	73856
layer_7 (MaxPooling2D)	(None, 28, 28, 128)	0
dropout_2 (Dropout)	(None, 28, 28, 128)	0
fc_1 (Flatten)	(None, 100352)	0
layer_8 (Dense)	(None, 64)	6422592
dropout_3 (Dropout)	(None, 64)	0
pridictions (Dense)	(None, 4)	260

3.4. Proposed algorithm

Fig. 3 illustrates the proposed algorithm for the auto-stop system as a flowchart. When an image is acquired through the Raspberry Pi camera, the face recognition model determines whether the person is authorized. The authorized drivers can now drive the car, meanwhile; the drowsiness recognition model monitors them focusing on eyes for closure exceeding 3 seconds. If the touch sensor is not activated during this period, the system responds by sending a sound message through the speaker. If the person's eyes remain closed and unresponsive, the system autonomously stops the car and sends a GPS message to assist the driver.

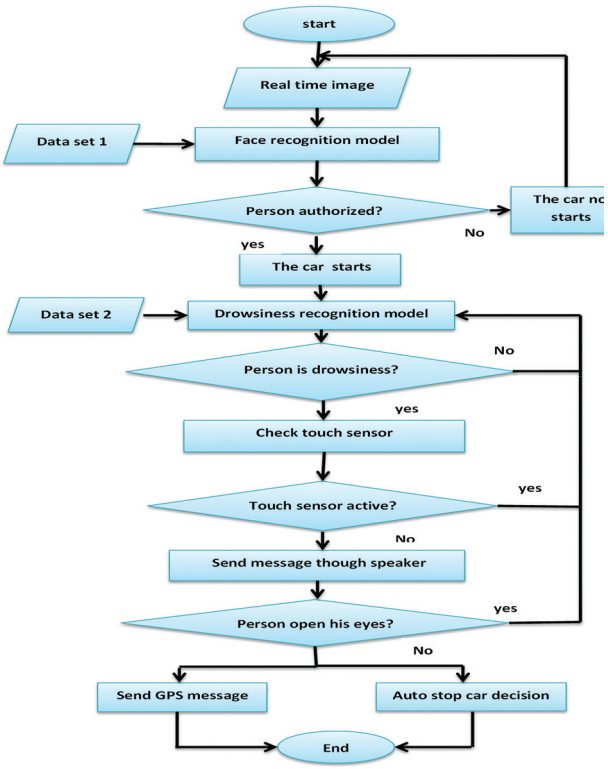


Fig. 3. Proposed algorithm flowchart

3.5. PROPOSED SYSTEM IMPLEMENTATION

The CNN models, proposed in the presented auto-stop car system, are trained using Python within the Anaconda Python 3.7 environment, employing Keras as the backend framework (utilizing Tensor Flow). The training takes place on a Windows 10 laptop equipped with an Intel R Core (TM) i7-6820 HQ CPU, featuring 4 physical and 8 logical cores, 8 GB of RAM.

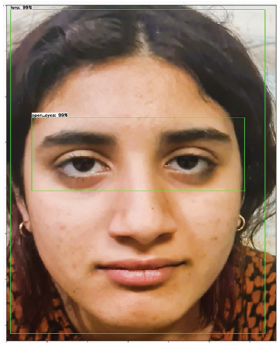


Fig. 4. Test image after converting the model to Tf_light

The weight updates are based on training data, while validation data provides insights into the network's progressive improvements. The hyperparameters that yielded the highest accuracies are optimizer: Adam, learning rate: 0.0001, batch size: 4, and epochs:100, a balance between accuracy and training time. Subsequently, the model is converted to Tensor Flow Lite format, and its performance is evaluated on previously unseen images from the "test" folder. These images are not part of the training data, providing a genuine measure of the model's accuracy. As indicated in Fig. 4, the model demonstrated a remarkable accuracy of 99% in detecting open eyes and recognizing the authorized person, "Lara".

Following the conversion to Tensor Flow Lite and creating a trained .tf _lite model, the next step involves deploying it on a Raspberry Pi. To execute the model, the Tensor Flow Lite Runtime is installed on the Raspberry Pi 4, and the Python environment and directory structure are configured for running the application. Four Python scripts are developed to accommodate various input sources, including images, videos, web streams, or webcam feeds. The proposed model specifically utilized TFLite_detection_webcam.py. The hardware components used in this research are a Raspberry Pi 4 model B (8 GB), a memory card, a Pi-Camera 5MP, a display screen, a speaker, a touch sensor, and LEDs as shown in Fig. 5. The LEDs are used to indicate the actuation performance of the system in different cases. For example, if one of the authorized individuals is recognized for the driver identification phase, a green LED is activated; otherwise, a yellow LED is illuminated. While in the drowsiness phase, An emergency is declared if the driver neglects to grip the sensor while their eyes are closed for more than three seconds, the system responds by activating a speaker to deliver a voice message, confirming the driver's alertness and alerting them to the situation. If the driver's eyes remain closed and unresponsive. The car initiates safety protocols, including activating warning lights, gradually reducing speed, and executing a controlled right turn to bring the vehicle to a stop. All these instructions are represented by a red LED illumination.

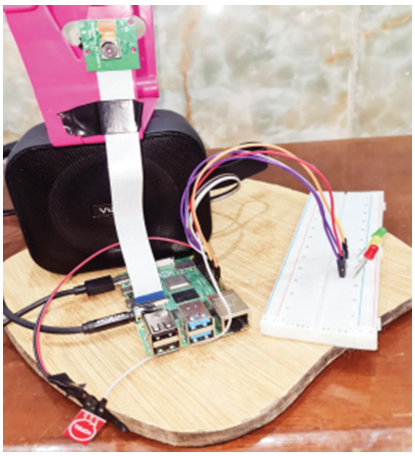
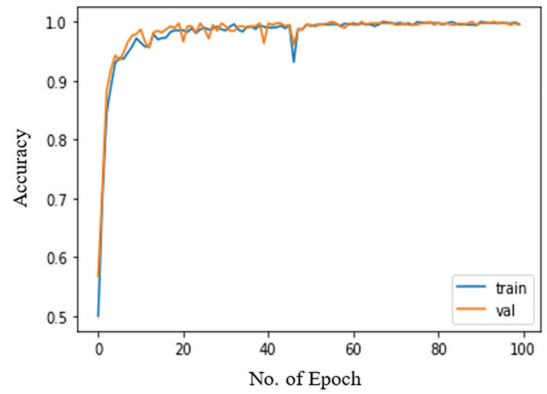


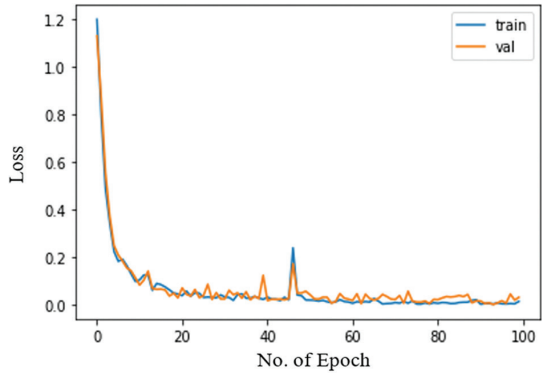
Fig. 5. The proposed system implementation

4. RESULTS

First, the proposed deep learning models are tested by adopting the training and testing accuracy and loss factors. The accuracy and loss for training and validation sets of the proposed face recognition model (driver recognition) are shown in Fig. 6. The accuracy ratio reaches saturation over 99% after 60 epochs in both the training and validation processes. Meanwhile, the losses occurred to be minimal immediately after 60 epochs.

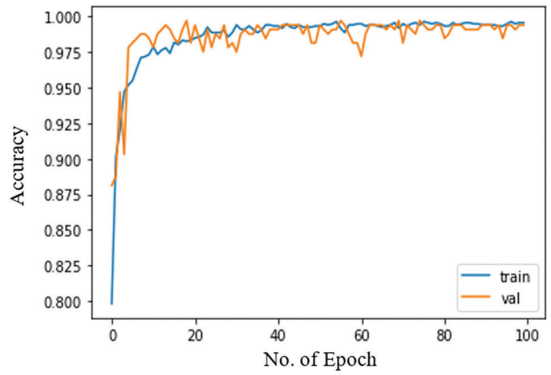


(a) Training and validation accuracy

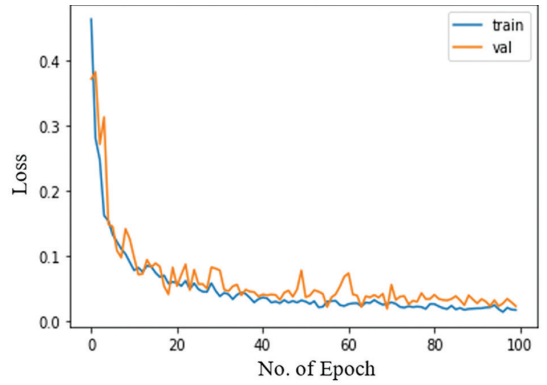


(b) Training and validation loss

Fig. 6. Accuracy and loss in training and validation of Face Recognition Model



(a) Indicating the accuracy



(b) Indicating the loss

Fig. 7. Accuracy and loss in training and validation for drowsiness model

In Fig. 7(a), the accuracy of the proposed model for detecting drowsiness during both training and validation phases is depicted. Meanwhile, Fig. 7(b) illustrates a different aspect: the variation in loss over time. The loss curve exhibits a decline beginning at the 67th epoch, followed by a period of stabilization, and then a gradual deceleration for the subsequent 100 epochs. Remarkably, despite the dataset containing only 800 image samples for each class, the suggested model performs remarkably well, achieving impressive accuracy rates of 99.54% for the training set and 99.37% for the validation set.

To test the performance of the proposed system, different case studies are adopted to simulate the possible situations that might face the proposed auto-stop car system.

4.1. CASE STUDY 1: UNAUTHORIZED DRIVER

If the driver is not an authorized driver (not one of the four authorized individuals), the system does not allow the car to start up. In the proposed system implementation prototype, a yellow LED is turned on to indicate that the driver is not authorized as shown in Fig. 8.

Fig. 8. Unauthorized driver case study

4.2. CASE STUDY 2: AUTHORIZED DRIVER

If the person is an authorized driver, a green LED is turned on to indicate that the proposed system recognizes the driver and the car starts up as shown in Fig. 9. Then, the system performs the drowsiness recognition monitoring mode.

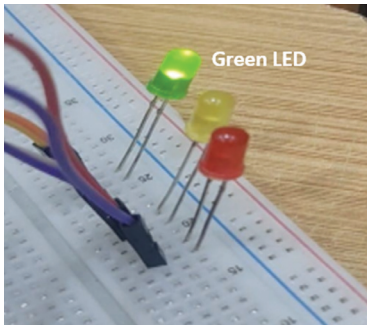


Fig. 9. Authorized driver case study

4.3. CASE STUDY 3: DRIVER TOUCHES SENSOR

While the authorized driver performs the driving, the touch sensor indicates that he/she is in driving mode, yet the system continues to monitor the driver's eyes.

4.4. CASE STUDY 4: DRIVER DOESN'T TOUCH SENSOR (3-SECOND TIMER)

If the authorized driver does not touch the touch sensor on the steering wheel, the system sends a sound message to alert the driver and check the level of drowsiness. If the driver opens his/her eyes within 3 seconds, the system returns to monitoring the driver's eyes.

4.5. CASE STUDY 5: DRIVER DOESN'T OPEN EYES (3-SECONDS TIMER EXPIRES)

If the authorized driver does not open his/her eyes within 3 seconds, the system sends another sound message to alert the driver and check their level of drowsiness. Then if the driver is still in closed eyes mode, a red LED is turned on as an emergency signal in the prototype to represent the auto-stop car mode, where the system sends a message through GPS to request help for the driver. Fig. 10 demonstrates how the detecting system can distinguish between drowsy drivers. A driver-alerting warning is promptly delivered once sleepiness signs are recognized during the real-time detection procedure, which has a 99% accuracy rate.

Fig. 10. Real-time drowsiness detection system

One essential key to evaluating the proposed system's accuracy is finding the accuracy ratio. The accuracy ratio is calculated as follows [24]:

$$A_{cc} = \frac{N_{cc}}{T_S} \times 100\%. \quad (1)$$

Where A_{cc} is the accuracy of classification, N_{cc} is the number of correctly classified images, and T_s is the total number of samples.

The proposed system is tested in real-time mode by considering numerous cases of drivers for both driver recognition and drowsiness recognition models.

This test is performed on the presented prototype to ensure the proposed models work properly. This assessment consisted of two stages: first recognizing the drivers and then determining their emotional states. The face recognition model has been tested with 120 images (30 images per person), yielding an average accuracy of 99.1%, as shown in Table 3. The drowsiness model is tested with 120 images, 30 for each driver, and an accuracy of 99.1% is achieved as shown in Table 4. This proves that the proposed models in the presented auto-stop car system work highly and efficiently.

On the other hand, the proposed drowsiness recognition model is compared with different models in the literature to validate the proposed work as shown in Table 5. This table, it is shows the superiority of the proposed deep learning model amongst the literature works in terms of accuracy to have 99.1% in the hardware implementation prototype not in the simulations as in the literature works.

Table 3. The effectiveness of the proposed CNN model for face recognition

Person name	Total test Images	No. of the correct image in face detection	No. of the incorrect image in face detection	Image percentage for face detection
Ali	30	30	0	100%
Hiba	30	30	0	100%
Lara	30	30	0	100%
Yara	30	29	1	96.6%
Total	120	119	1	99.1

Table 4. The effectiveness of the proposed CNN model for drowsiness recognition

State of recognition image	Total test Images	No. of the correct recognition image	No. of the incorrect recognition image	Image recognition percentage
Drowsy	60	59	1	98.3%
Non-Drowsy	60	60	0	100%
Total	120	118	1	99.1%

Table 5. Accuracy Comparison of drowsiness detection model with literature works.

Reference	Year	Method	Accuracy
[18]	2021	Eye-aspect ratio (EAR).	92.5%
[21]	2021	DNN model	90%
[19]	2022	CNN and SVM	97.44%
[22]	2022	CNN model	96%
[20]	2023	Deep learning methods	95.8%
[23]	2023	CNN and VGG16	97%.
		Our proposed CNN model	99.1%

5. CONCLUSIONS

An intelligent auto-stop car system was presented using deep learning models. The systems adopted two deep learning models with the same backbone design, one for driver recognition and the other for drowsiness recognition while driving. It also considered touch and camera sensors as well as a speaker and microphone for ensuring the driver's status in driving mode. The proposed system was implemented as a prototype in a Raspberry Pi device connected with the utilized sensors and group of LEDs. These LEDs indicate the actuation modes performed after the system makes decisions. The proposed deep learning models are tested in simulation and real-time phases and the obtained results proved the claim of this research in terms of accuracy and efficiency. At the same time, the auto-stop car system was tested in real-time mode with different case studies and the results were promising. The real-time accuracy, achieved using the prototype, was 99.1% for both deep learning models.

6. REFERENCES

[1] M. Zamir et al. "Face Detection & Recognition from Images & Videos Based on CNN & Raspberry Pi", Computation, Vol. 10, No. 9, 2022, p. 148.

[2] A. A. Mohammed, M. A. N. Al-hayanni, H. M. Azzawi, "Detection and segmentation the affected brain using ThingSpeak platform based on IoT cloud analysis", Periodicals of Engineering and Natural Sciences, Vol. 9, No. 2, 2021, pp. 858-867.

[3] M. A. Yaseen, A. K. Abass, "Performance evaluation of fiber bragg grating strain sensor integrating in WDM communication system", The Iraqi Journal of Computers, Communications, Control, and Systems Engineering, Vol. 20, No. 2, 2020, pp. 74-78.

[4] M. A. Uthaib, M. S. Croock, "Multiclassification of license plate based on deep convolution neural networks", International Journal of Electrical and Computer Engineering, Vol. 11, No. 6, 2021, p. 5266.

[5] A. M. Mohammad, S. A. AL-Samarraie, "Robust controller design for flexible joint based on backstepping approach", The Iraqi Journal of Computers, Communications, Control, and Systems Engineering, Vol. 20, No. 2, 2020, pp. 58-73.

[6] A. H. Hadi, W. F. Shareef, "In-situ pipeline monitoring system design for identifying and locating damaging activities based on wireless sensor network", The Iraqi Journal of Computers, Communications, Control, and Systems Engineering, Vol. 20, No. 2, 2020, pp. 33-46.

[7] S. T. Ahmed, R. Al-Hamdani, M. S. Croock, "Developed third iterative dichotomizer based on feature decisive values for educational data mining", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 18, No. 1, 2020, pp. 209-217.

[8] Z. Li, S. E. Li, R. Li, B. Cheng, J. Shi, "Online detection of driver fatigue using steering wheel angles for real driving conditions", Sensors, Vol. 17, No. 3, 2017, p. 495.

[9] M. Sajid, N. Ali, S. H. Dar, B. Zafar, M. K. Iqbal, "Short search space and synthesized-reference re-ranking for face image retrieval", Applied Soft Computing, Vol. 99, 2021, p. 106871.

[10] X. Mou, Y. Song, R. Wang, Y. Tang, Y. Xin, "Lightweight Facial Expression Recognition Based on Class-Rebalancing Fusion Cumulative Learning", Applied Sciences, Vol. 13, No. 15, 2023, p. 9029.

[11] M. Anggo and L. Arapu, "Face recognition using fisherface method", Journal of Physics: Conference Series, Vol. 1028, No. 1, 2018, p. 12119.

[12] R. Ding, G. Su, G. Bai, W. Xu, N. Su, X. Wu, "A FPGA-based accelerator of convolutional neural network for face feature extraction", Proceedings of the IEEE International Conference on Electron Devices and Solid-State Circuits, Xi'an, China, 12-14 June 2019, pp. 1-3.

[13] A. K. Majumder, J. A. Izaguirre, "A smart IoT security system for smart-home using motion detection and facial recognition", Proceedings of the 44th Annual Computers, Software, and Applications Conference, Madrid, Spain, 13-17 July 2020, pp. 1065-1071.

[14] H. Wang, L. Guo, "Research on face recognition based on deep learning", Proceedings of the 3rd International Conference on Artificial Intelligence and Advanced Manufacture, Manchester, United Kingdom, 23-25 October 2021, pp. 540-546.

[15] A. Dasgupta, A. George, S. L. Happy, A. Routray, "A vision-based system for monitoring the loss of attention in automotive drivers", IEEE Transactions on Intelligent Transportation Systems, Vol. 14, No. 4, 2013, pp. 1825-1838.

[16] H. Hajjdiab, I. Al Maskari, "Plant species recognition using leaf contours", Proceedings of the International Conference on Imaging Systems and Techniques, Batu Ferringhi, Malaysia, 17-18 May 2011, pp. 306-309.

[17] R. Laganière, H. Hajjdiab, A. Mitiche, "Visual reconstruction of ground plane obstacles in a sparse view robot environment", Graphical models, Vol. 68, No. 3, 2006, pp. 282-293.

[18] N. N. Pandey, N. B. Muppalaneni, "Real-time drowsiness identification based on eye state analysis", Proceedings of the International Conference on Artificial Intelligence and Smart Systems, Coimbatore, India, 25-27 March 2021, pp. 1182-1187.

[19] P. A. de Lima Medeiros et al. "Efficient machine learning approach for volunteer eye-blink detection in real-time using webcam", Expert Systems with Applications, Vol. 188, 2022, p. 116073.

[20] F. Safarov, F. Akhmedov, A. B. Abdusalomov, R. Nasimov, Y. I. Cho, "Real-time deep learning-based drowsiness detection: leveraging computer-vision and eye-blink analyses for enhanced road safety", Sensors, Vol. 23, No. 14, 2023, p. 6459.

[21] F. Hasan, A. Kashevnik, "State-of-the-art analysis of modern drowsiness detection algorithms based on computer vision", Proceedings of the 29th Conference of Open Innovations Association, Tampere, Finland, 12-14 May 2021, pp. 141-149.

[22] A. Rajkar, N. Kulkarni, A. Raut, "Driver Drowsiness Detection Using Deep Learning", Applied Information Processing Systems, Springer Singapore, 2022, pp. 73-82.

[23] M. I. B. Ahmed et al. "A deep-learning approach to driver drowsiness detection", Safety, Vol. 9, No. 3, 2023, p. 65.

[24] M. R. Hidayah, I. Akhlis, E. Sugiharti, "Recognition number of the vehicle plate using Otsu method and K-nearest neighbour classification", Scientific Journal of Informatics, Vol. 4, No. 1, 2017, pp. 66-75.

Multimodal emotion recognition based on the fusion of vision, EEG, ECG, and EMG signals

Original Scientific Paper

Shripad Bhatlawande
Dept. of E&TC, VIT,
Pune, India
shripad.bhatlawande@vit.edu

Sourjadip Pramanik
Dept. of E&TC, VIT,
Pune, India
sourjadip.pramanik20@vit.edu

Swati Shilaskar
Dept. of E&TC, VIT,
Pune, India
swati.shilaskar@vit.edu

Swarali Sole
Dept. of E&TC, VIT,
Pune, India
swarali.sole20@vit.edu

Abstract *– This paper presents a novel approach for emotion recognition (ER) based on Electroencephalogram (EEG), Electromyogram (EMG), Electrocardiogram (ECG), and computer vision. The proposed system includes two different models for physiological signals and facial expressions deployed in a real-time embedded system. A custom dataset for EEG, ECG, EMG, and facial expression was collected from 10 participants using an Affective Video Response System. Time, frequency, and wavelet domain-specific features were extracted and optimized, based on their Visualizations from Exploratory Data Analysis (EDA) and Principal Component Analysis (PCA). Local Binary Patterns (LBP), Local Ternary Patterns (LTP), Histogram of Oriented Gradients (HOG), and Gabor descriptors were used for differentiating facial emotions. Classification models, namely decision tree, random forest, and optimized variants thereof, were trained using these features. The optimized Random Forest model achieved an accuracy of 84%, while the optimized Decision Tree achieved 76% for the physiological signal-based model. The facial emotion recognition (FER) model attained an accuracy of 84.6%, 74.3%, 67%, and 64.5% using K-Nearest Neighbors (KNN), Random Forest, Decision Tree, and XGBoost, respectively. Performance metrics, including Area Under Curve (AUC), F1 score, and Receiver Operating Characteristic Curve (ROC), were computed to evaluate the models. The outcome of both results, i.e., the fusion of bio-signals and facial emotion analysis, is given to a voting classifier to get the final emotion. A comprehensive report is generated using the Generative Pretrained Transformer (GPT) language model based on the resultant emotion, achieving an accuracy of 87.5%. The model was implemented and deployed on a Jetson Nano. The results show its relevance to ER. It has applications in enhancing prosthetic systems and other medical fields such as psychological therapy, rehabilitation, assisting individuals with neurological disorders, mental health monitoring, and biometric security.*

Keywords: *Emotion Recognition (ER), Analysis of Mental Health, Feature Fusion, Machine Learning (ML), Computer Vision, Physiological Signals*

1. INTRODUCTION

Emotion recognition (ER) is a fascinating field that aims to identify and understand human emotions through different modalities, such as facial expressions, speech, physiological signals, and behavioral patterns. It has become a research topic in areas such as medicine, machine learning (ML), and psychology [1]. ER technology can potentially be used in prosthetic arms to improve their usability and functionality, including adjusting their sensitivity and responsiveness for prosthetic arm wearers. Facial emotion recognition (FER) finds relevance in numerous applications, including identification processes for citizenship, identification cards, social security cards, and even intrusion detection [2]. Two prominent approaches in this domain are ER using physiological signals and FER using the ML model. Many FER systems employ ML techniques to recognize accurate emotions. One such ML-based FER system is proposed in [3], which constructs a multi-layer classifier based on a carefully curated dataset of 7 individuals, employing Haar-cascade features and histogram of oriented gradients (HOG) for feature extraction while employing Support Vector Machine (SVM) as the classifier.

Emotions can be recognized using different physiological signals, namely Electroencephalogram (EEG), skin temperature, Electrocardiogram (ECG), Electromyography (EMG), blood pressure, respiration rate, heart rate, Blood Volume Pressure, and Galvanic Skin Response, but the collection and processing of these signals become hard due to some added noise. For our system to work as intended, filtering these signals is necessary. Physiological signals, namely ECG, can reflect the relationship

between changes in emotions and heartbeat. Different emotions can be recognized by extracting heart rate variability (HRV). In the field of emotion identification, EEG signals are getting more attention day by day as they accurately reflect the feelings of any person. Signals collected from peripheral nervous systems, including ECG and EMG, can also be used for the same [4]. The studies in brain-computer interface (BCI) with different techniques that recognize emotions mostly made use of EEG, as EEG responds in time and is sensitive to changes in affective states [5]. EMG-based systems for ER have the capability to identify a person's genuine emotions.

ER systems can prove useful for understanding someone's emotional state and for achieving better communication. Harnessing the power of computer vision and ML advancements, researchers and developers are now utilizing facial expression analysis to automate ER processes with remarkable precision. Multimodal emotional datasets can be used, or the fusion of two or more signals improves accuracy and provides the relationship between different bio-signals. These signals can be fused together for the unique identification of emotions. Fusion of EEG, ECG, and EMG signals can be done to achieve better performance. Among the diverse approaches in this domain, FER holds particular significance.

In this paper, a multi-model system is proposed that considers the features from both the fusion of EMG, ECG, and EEG and facial expressions. The subsequent sections delve into the process of implementing it, aiming to contribute to the development of a more robust and reliable system.

2. RELATED WORK

Researchers have done notable research in the domain of ER using bio-medical devices due to their potential applications in various areas such as mental health and human-computer interaction. Hwang et al. made use of the SEED dataset, a publicly available dataset, recorded from 15 participants. The sample rate for the EEG signal was chosen as 1000 Hz [6]. Ferdinando Hany et al. made use of the ECG signals, which were taken from the Mahnob-HCI database, containing records of 27 participants, while participants were shown images and videos [7]. A 32-channel EEG was recorded, and a sampling frequency of 256 Hz was used in [8]. A diagram and positioning of electrodes over the face for EMG and experimental setup have been proposed in [9], which helps in getting insight into montage placement. According to research, EEG signals can be acquired using silver chloride electrodes. The EEG signal has a relatively low amplitude (5–500 uV), making it difficult to capture and evaluate. As a result, an amplifier was used to amplify the signals to a desirable level, which helped in achieving better accuracy [10].

Kumar Nitin et al. collected data from participants using music and videos as external stimuli, with almost 40 trials for each participant. 32 electrodes were placed according to a 10–20 worldwide system for EEG with a sampling frequency of 512 Hz. Down-sampling was initially carried out from 512 to 128 Hz. The Noise was removed using a bandpass filter in the frequency range of 5–45 Hz. A Butterworth filter was used for the filtering signal [11]. Signals, which were collected using 16 channels PowerLab with a sampling rate of 400 Hz, were filtered by a digital notch filter at a frequency of 50 Hz [12]. HoSeung Cha et al. used a Riemann manifold for feature extraction, and a pattern recognition-based myoelectric interface was built based on Linear Discriminant Analysis (LDA) implementation. Recall, F1 score, and precision were calculated for concluding LDA adaptation conditions. The various results were successfully reflected in the user's current state between virtual and real [13]. Time domain characteristics such as First and Second differences, Root mean square, Line length, Signal power, and Total Wavelet Energy, Frequency domain characteristics such as Dominant frequency, Total Wavelet Energy, and entropy-based characteristics such as spectral entropy, Shannon entropy, and sample entropy were employed, as well as various classifier techniques including SVM, ANN, and Naïve Bayes were used for classification in [14]. Silvio Barra performed feature extraction using a* peak detection method. The ECG signals were characterized by detecting repeating peaks consisting of Q, R, and S waves [15]. An analysis was performed by the authors using Normalization, mean, and standard deviation of the original signal for all 11 channels [16].

Many researchers have utilized a variety of validation and searching techniques to optimize the classification process. Different classification techniques were used and reviewed in [17]. In [18], the Bio Vid Emo DB dataset was used to validate the proposed method. The experimental conditions included a classifier, namely SVM using the RBF kernel, which obtained a 79.51% maximum accuracy in differentiating positive and negative emotions. P. Sarkar et al. fed emotion identification weights into a neural network with fully connected layers that were trained to categorize emotions. Databases like Dreamer, Seed, and Swell were used to test the results, with the maximum being an improvement over the 96.9% accuracy [19]. Min Chen et al. [20] used a minimal quantity of multimodal labeled data; the proposed LLEC first trains the neural network model. The unlabeled data is then automatically labeled and added to the training set, utilizing improved hybrid label-less learning to boost model detection accuracy even more. Wu et al. [21] developed a prototype of a wearable emotion-detection headband using EEG. The temporal window of 0.5 to 4 sec is suggested, with a short delay of <1 sec in between the 5 bands of the signal. The prototype included an EEG- measuring front end in the form of a headband for acquiring and pre-amplifying EEG data. Ante Topic et al. [22] built a model showing that the holographic feature map technique clearly outperforms topographic feature maps.

A successful expression recognition system has the potential to have robust features to effectively recognize the face's appearance [23]. Alghamdi et al. gave a comparison between various technologies that can be used for facial recognition. Types of detection algorithms like SVM, the Viola-Jones (VJ) algorithm, the Kanade Lucas Tomasi algorithm, the AdaBoost algorithm, the hybrid face detection algorithm, and the Elman Neural Network have been introduced [24]. Li et al. discussed the future development direction and potential application prospects of FER. Techniques like Principal Component Analysis (PCA) and LDA have been discussed. For classification, SVM, Ada-boost, small samples, and neural network techniques were compared. Some deep learning-based techniques were also reviewed [25]. He et al. proposed Laplacianfaces, which are based on the Laplacian Eigenmaps approach. Laplacianfaces is an algorithm utilizing the Laplacian Eigenmaps approach for face recognition by constructing a graph using the similarity of the face images, which is then used to compute the eigenfaces [26]. The paper [27] focuses on conducting a comparative analysis of various convolutional neural network (CNN) architectures for face recognition. The authors emphasize the significance of face recognition in numerous applications and highlight the growing use of CNN in this field. It provides an overview of popular CNN architectures, including AlexNet, VGGNet, GoogLeNet, and ResNet, explaining their design principles and features relevant to face recognition. The authors conducted experiments by training and evaluating the selected CNN architectures on the dataset, employing evaluation metrics like accuracy, precision, recall, and F1 score to assess performance.

Chowdary et al. discussed various deep-learning approaches used for FER. This includes CNN, Recurrent Neural Networks (RNN), or a combination of both. The authors described the architecture and configuration of the neural network used in their study. The authors also explained preprocessing techniques applied to the images, such as normalization or data augmentation, to enhance the performance of the model. The training process of the deep learning model, including details on the optimization algorithm, loss function, and hyperparameter tuning has been presented in [28]. In [29], two databases named Bosphorus Face Database (consisting of 4666 face images of 105 subjects) and the University of Milano Bicocca (UMB) Face Database (consisting of 1473 face images of 143 subjects) were used. Multiple ML models were tested on them. Enhancement attacks like blurring, sharpening, histogram equalization, and median filtering; geometric attacks like rotation and cropping resizing; and noise attacks like Gaussian attacks, speckle attacks, and poison attacks were implemented. Koonsanit et al. created a framework for ER by capturing the facial images of users. Facial expressions were categorized into seven categories, i.e., Angry, Disgust, Fear, Happy, Sad, Surprise, and Neutral. Classifiers like SVM, Logistics Regression (LR), K-nearest neighbor (KNN), and Multi-Layer Perceptron (MLP) were used [30]. Liu et al. collected a custom dataset for EEG signals from 16 participants. Higuchi Fractal Dimension (FD) Spectrum was used for analyzing non-linear properties of EEG signal [31]. Ergin et al. used the Empirical mode decomposition (EMD) method for EEG signals, to obtain Intrinsic Mode Functions (IMFs). Previously used methods have been described in Table 1.

Numerous solutions have been put forth for ER through the utilization of physiological signals. However, these methods have limitations, which call for further research to improve their effectiveness. Through using a multi-modular system, a more robust and reliable system can be obtained. The contemporary ER system includes the use of music, games, and videos as external stimuli, with the assembly including a complex module for analysis where the limitations include poor performance in participant-independent ER. This results in reduced robustness and increases the chance of misdiagnosis or biased decisions. Also, the use of deep learning techniques results in complex models requiring more resources to deploy in real-time.

Table 1. Review of previous technologies used for ER

Ref No.	Year	Signal	Dataset	Algorithm	Performance
[6]	2020	EEG	SEED	Input signal -> Band-pass filtering -> Short-Time Fourier Transform (STFT) -> Applying single-task DNN, multi-task DNN and adversarial DNN -> Model performance evaluation	Accuracy- 75.31%
[9]	2014	EMG, ECG, GSR	Custom	Data acquisition of EMG, EEG and GSR signals using audio-visual stimuli -> Filtering using notch filter and average filter -> Applying Higher order statistics (HOS) for feature extraction -> Classification using KNN	Accuracy- 69%
[31]	2014	EEG	Custom, DEAP	Data acquisition for EEG signal from 16 subjects -> Higuchi Fractal Dimension Spectrum -> Classification using SVM -> Validation on DEAP dataset	Accuracy- 85.38%
[32]	2019	EEG	Custom	Data acquisition for EEG signal from 25 subjects -> Filtering of data -> Empirical Mode Decomposition -> Selection of intrinsic mode functions -> Classification using SVM	Accuracy- 84.3%

Thus, a need for a lightweight system using ML arises. Even though the result of emotional identification is significant, further study is required to uncover the elements that may simply and efficiently recognize emotional patterns. The importance of a model that considers the various physiological signals together and makes decisions accordingly is taken into consideration. As a result, to increase the accuracy and other performance parameters of the final result, a fusion of EEG, ECG, EMG, and facial expressions was undertaken. An Affective Response System (ARS), which is based on stimuli from sudden news, audio, video, speech, and sudden actions, was used to provide better results. The physiological system proposes a better variable for choosing the Butterworth filter, window slicing techniques, Exploratory Data Analysis (EDA), and PCA. The model based on facial expressions leverages several feature extraction techniques to provide a comprehensive analysis of emotional states. Decision Tree, Random Forest, and Optimized Random Forest were selected as ML classifiers. The model was tested for various performance parameters using the f1 score, precision, accuracy, sensitivity, and Receiver Operating Characteristic Curve (ROC) curve. By incorporating additional modalities, the proposed system potentially enhances the accuracy and reliability of ER systems, leading to more robust and reliable results.

3. METHODOLOGY

This paper presents an approach for ER by fusing physiological signals, including EEG, ECG, EMG, and facial expressions. The proposed system is organized into two distinct sections, with one dedicated to physiological signals and the other to facial expressions. In the physiological domain, the system integrates signals from the EMG, EEG, and ECG, extracting relevant features. Simultaneously, facial emotions were captured by analyzing facial cues from images. Emotions can be broadly categorized as positive, namely surprise and happiness, and negative, namely sadness, anger, and fear. This paper focuses on recognizing three emotions, fear, neutrality, and surprise. The data utilized for this study was acquired from a sample of 10 healthy participants. The collected data was subjected to preprocessing to remove any artifacts, followed by feature extraction. These features were carefully fused and optimized to form a comprehensive feature fusion vector, which would enable a more robust ER model.

In the context of FER, a custom dataset was curated for the experiment. A model was trained using this dataset to classify the above-mentioned emotions. The final decision-making process involved a voting mechanism between the predicted facial emotion and the emotion inferred from physiological signals. The final result is the generation of a comprehensive medical report, facilitated by GPT-3, leveraging the combined insights from both physiological and facial expression data sources. This approach takes advantage of the strengths of both modalities, leading to more accurate and reliable ER outcomes. An overview of the proposed system can be seen in Fig. 1.

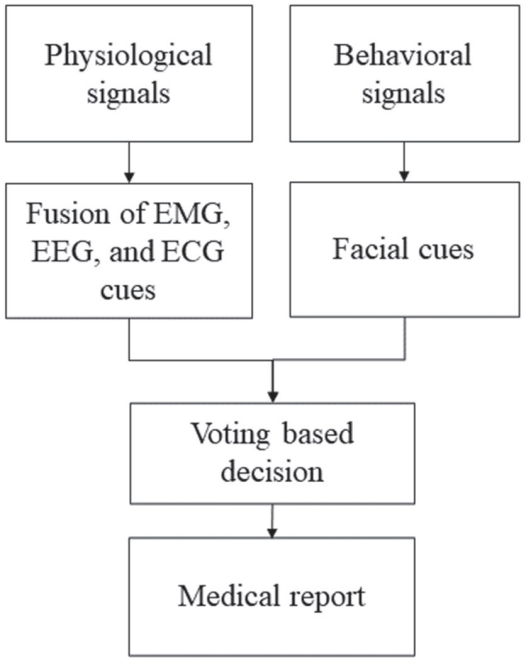


Fig. 1. Overview of the proposed multimodal ER system

3.1. IMPORTANCE OF ECG, EEG, EMG, AND FACIAL SIGNALS

There is a need to understand why vision, EEG, ECG, and EMG were selected for this experiment before delving deep into the process of multimodal ER. Fusion of vision, EEG, ECG, and EMG signals is an interdisciplinary approach that combines physiological and behavioral data to gain a comprehensive understanding of human emotions. Together, these signals provide valuable insights into how fear, surprise, and neutral emotions are represented and expressed.

3.1.1. VISION

Visual information is one of the most prominent and informative source for understanding emotions. Facial expressions, body language, and eye movements are crucial indicators of emotional states. Fear and surprise caused widened eyes, raised eyebrows, and open mouths. Neutral emotions typically result in neutral facial expressions. In neutral, there wasn't any noticeable change in the facial expressions. The analysis of these visual cues provides insights into the intensity and type of emotion being experienced.

3.1.2. EEG

EEG measures electrical activity in the brain and is particularly useful for studying cognitive and emotional processes. Different brain regions and frequency bands are associated with various emotions and cogni-

tive functions. Fear and surprise lead to increased activity in the amygdala and other brain regions associated with emotional processing, while neutral emotions have a more balanced brain activity pattern. EEG can capture these differences in neural activation during different emotional states.

3.1.3. ECG

The heart rate and heart rate variability are closely linked to emotional responses. ECG measures the electrical activity of the heart and reflects changes in autonomic nervous system activity during emotions. Fear and surprise typically lead to an increase in heart rate and reduced heart rate variability due to activation of the sympathetic nervous system ("fight or flight" response). In contrast, neutral emotions often result in relatively stable heart rate patterns. ECG can quantify these physiological changes.

3.1.4. EMG

EMG measures electrical activity in skeletal muscles, which provides information about facial expressions and emotional responses involving muscle contractions. Fear and surprise lead to increased muscle tension in the facial muscles, such as in the brows and around the mouth. Neutral emotions result in a more relaxed facial muscle state. EMG data can capture these subtle muscle activity changes.

3.2. MODEL BASED ON THE FUSION OF PHYSIOLOGICAL SIGNALS

The model architecture was divided into six main parts: data acquisition, pre-processing, feature extraction, feature selection, feature fusion, and classification. Each part plays a crucial role in the overall process of physiological ER.

3.2.1. DATA ACQUISITION

A well-designed setup was implemented to collect the physiological signals of the participants accurately. The setup for the physiological system differs depending on the type of biomedical devices used to detect the physiological signals, including EEG, EMG, and ECG.

A custom dataset for all three ECG, EMG, and EEG signals for ER was collected. The emotions that were selected for the proposed model were fear, neutrality, and surprise. The channels were selected considering the effect of emotion on the lobes of the brain, the electrical activity of muscles, and the heart. In the case of 4-channel EMG, the frontalis and masseter muscles were used, which are described in Fig. 2(a) with the ground placed at the forehead. The ECG setup included a 2-channel electrode placed on the arm, as shown in Fig. 2(b). AgCl electrodes were placed at positions namely F3-F4, T3-T4, P3-P4, and O1-O2 for a 4-channel EEG, which can be seen in Fig. 2(c). Setup for each of the devices was carried out separately.

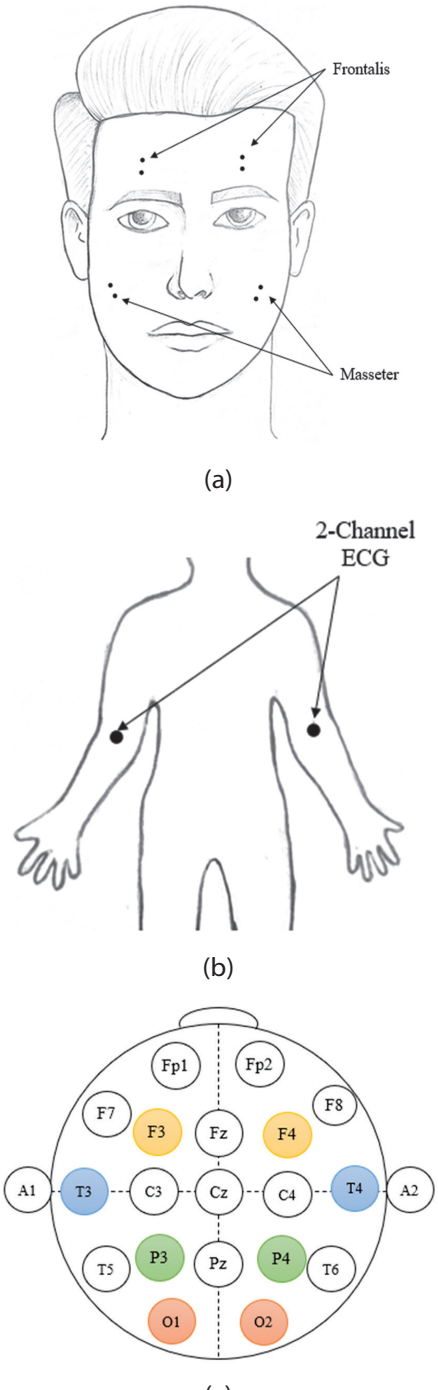


Fig. 2. Montages for (a) EMG, (b) ECG, and (c) EEG used in the System

The data acquisition setup included subject preparation and environmental setup. The chosen subjects were between the ages of 19 and 21, were healthy, and had no prior medical history. They were prepared in a manner to minimize any human artifacts. By ethical guidelines, all participants willingly signed consent forms before their involvement in the experiment.

Participants were given experimental guidance at the start of the experiment and were asked to sit in a laboratory environment. An affective response system (ARS) was used for detecting and analyzing emotion in response to audio and video stimuli. Simultaneously, a video of each

participant was recorded, which was used for testing the facial model based on the physiological emotions. Different video clips corresponding to different emotions were shown to 50 people. Based on their voting, video clips with more than 90% votes were finalized for the experiment. For each of the emotions, 3 clips were chosen, bringing the total to 9 clips with a length of 2 minutes. The clips chosen were carefully selected to induce the desired emotion. The data were recorded from 10 participants. While recording the data, each participant was asked to relax for one minute at the start of the experiment, then a signal was collected for two minutes while a video clip was shown to the participant. In this period of 2 minutes, the signal was recorded for 1 minute and 30 seconds, depending on the time for which emotion was induced. The subject was provided to rest in between each session for one minute, as described in Fig. 3. This process was repeated for each participant. For each emotion, every participant underwent three separate test sessions for signal acquisition, following an identical procedure that was repeated for ECG, EMG, and EEG measurements. Data were collected separately for the different signals. After collecting the signal readings, the data was exported as Excel sheets and formatted accordingly.

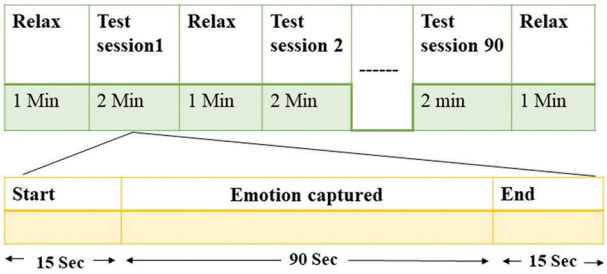


Fig. 3. Experimental setup for physiological signal acquisition of EEG, ECG, and EMG

3.2.2. DATA PRE-PROCESSING

In the process of bioelectric data collection, noise may be added due to the participant's body movements, eye blinking, or other sources. Notch and Butterworth filters were used to remove these artifacts. A default-frequency notch filter was maintained during the collection of data to remove artifacts from the signals. A notch filter was used to eliminate 50Hz supply frequency noise. The range of this band-stop filter was 49–51 Hz. As shown in Fig. 3, a window of 90 seconds was selected to capture the signals generated by the subject when subjected to a particular emotion. The window size was based on noticing the frequency and changes in the signal during the emotion. It was chosen as 5 seconds for EEG, EMG, and ECG, respectively. The culmination of all the windows was added to the final Excel file, along with the other necessary steps such as imputing missing values and clearing out outliers. All the collected data was formatted according to their specific requirements, making it organized for feature extraction. The process flow for signal acquisition, preprocessing, feature extraction, and classification is described in Fig. 4, with the end result being an accurate classifier that recognizes emotions.

3.2.3. FEATURE EXTRACTION

Feature extraction extracts the different properties of signals. Different feature extraction techniques, domains, and methods exist for bio-signal processing. For this experiment, time, frequency, and wavelet domain features were considered, as shown in Table 2.

3.2.3.1. TIME DOMAIN FEATURES

Signal properties that pertain to the temporal dimension are designated as time-domain features. These features are helpful in analyzing the temporal characteristics of signals. Root Mean Square, Integral, Variance, Mean Absolute Value, Amplitude Change, Difference absolute standard deviation value, Average Wavelength, Willison amplitude, Zero-Crossing, and Myopulse percentage rate were all taken into account.

3.2.3.2. FREQUENCY DOMAIN FEATURES

In this domain, the signals are analyzed with respect to frequency rather than time. They provide insights into the spectral content of a signal. For the frequency domain, features, namely Mean Power, Mean Frequency, Frequency ratio, and total power were selected.

3.2.3.3. WAVELET DOMAIN FEATURES

The analysis incorporates techniques that examine the signal simultaneously in both time and frequency domains, offering a comprehensive and nuanced understanding of the signal characteristics. Wavelet decomposition up to four levels was utilized to extract features from the signal. The chosen wavelet features include Mean, Standard Deviation, Energy, and Entropy of the coefficients.

In addition, Kurtosis, Max-Min, and H2-H1 were computed for each bio-electrical signal. The combination of these features with all the domain features resulted in a total of 35 features. These features were extracted from each channel of ECG, EMG, and EEG, whose details can be found in Table 2.

Table 2. Selection of features according to different domains on a single channel for EEG, EMG, and ECG

Domain	Features Extracted
Time domain	Variance, Difference absolute standard deviation value, Willison amplitude, Zero Crossing and Myopulse percentage rate, Root Mean Square, Integral, Mean Absolute Value, Wavelength, Average Amplitude Change
Frequency domain	Mean Power, Mean Frequency, Frequency ratio, Total Power
Wavelet domain	Mean, Kurtosis, Standard Deviation, Energy and Entropy of the coefficients

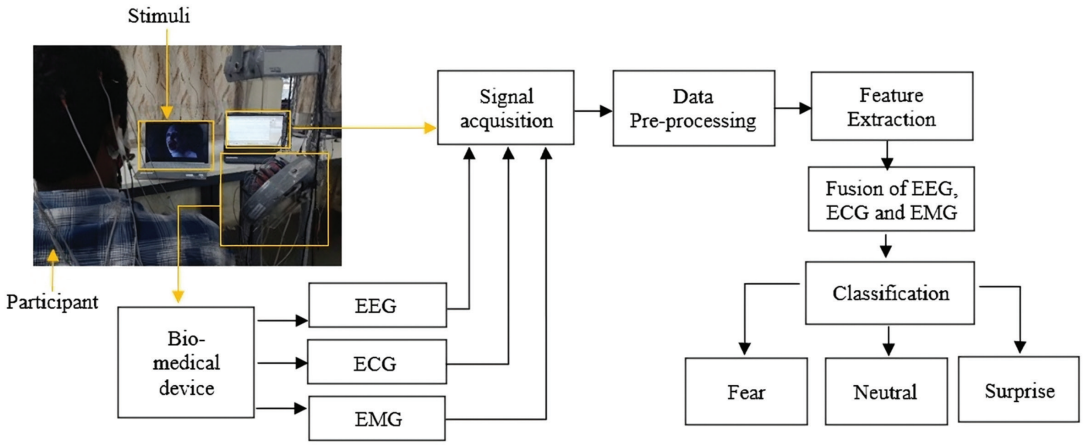


Fig. 4. Subject placement and laboratory setup for implemented workflow

3.2.4. FEATURE SELECTION AND FUSION

After extracting 35 features, for EMG (4 channel), the size of the feature vector was (280, 140) for a single participant, and for EEG (4 channel EEG + 1 channel ECG), it was (1070, 175). EDA was performed on these extracted features to categorize them according to their importance in classifying emotions. 11 less important features based on visual inspection of mean and variance were removed. The final size of the feature vector becomes (280, 129) for EMG and (1070, 164) for EEG.

As the sampling rates for EEG and EMG were different, the number of data points generated was also different. The number of data points generated for EEG and ECG was higher than for EMG. As a result, up-sampling (ups) and down-sampling (ds) techniques were used to prevent a biased model, as shown in Fig. 5. One such synthetic sampling technique is the Synthetic Minority Oversampling Technique (SMOTE), whose formula is given in (1). The SMOTE process involves identifying the S closest neighbors to every sample in the minority class and then using those neighbors to generate synthetic samples. This is achieved by computing the dissimilarity between a sample from the minority class and one of its S nearest neighbors and multiplying it by a random number between 0 and 1 to generate a new synthetic sample with a slightly different feature set. This preserves important information about the minority class and helps to improve model performance.

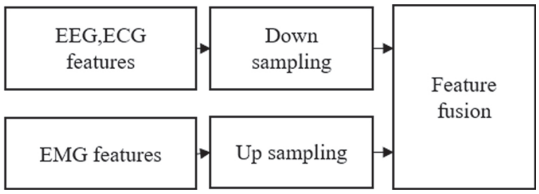


Fig. 5. EEG, ECG, EMG feature data with synthetic sampling

For a given sample from the minority class, denoted as w, SMOTE generates a synthetic y as:

$$y = w + \lambda^* (neighbor\_i\text{-}x) \quad (1)$$

where λ is a random number between 0 and 1, and $neighbor\_i$ is one of the nearest neighbors to w, where s is the number of neighbors to consider. The resultant matrix was formed after using the smote function on the EMG dataset.

Algorithm 1: Fusion of EEG, ECG and EMG features

Input: Subject-wise dataset on EMG, EEG and ECG
Output: Fusion Dataset for Model Training
Initialization
1: fv(EMG,EEG,ECG)←No. of feature vector
2: Nn←Nearest Neighbours of a Sample
3: kN(class samples)←Minority Samples
4: N←No. of Samples
5: Compute $N1$ ←Number of Classes
6: if fv(EMG)<fv(EEG,ECG) then
7: fvo ← Split fv(EMG) into $N1$ frame
8: end if
9: for i==class do:
10: if kN≥Nn then
11: pre set ←$kN(i)$:N(Classes-i)
12: smote (y) ← Ups(pre set)
13: emg ($y1$) ←y+fvo
14: if $y1$<fv(EEG,ECG) then
15: $N2$ ←$fv(y1)$
16: $y2$←ds[N(EEG,ECG)]
17: end if
18: end if
19: end for
20: Final model= $y1$+$y2$
21: return Final model
end

The EMG dataset was first divided into training and testing sets before applying SMOTE. The subsequent technique was only applied to the training set. This training set was further split into subject-wise sets, and each test was further divided using labels assigned to the emotions.

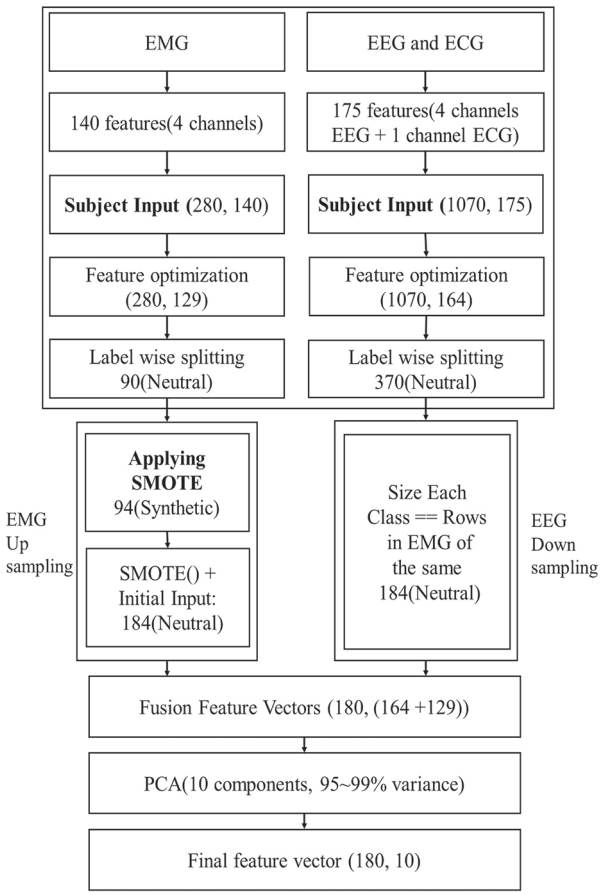


Fig. 6. Description of feature vector size at each step of the methodology

Considering a single participant, the feature vector of 280 rows was split label-wise, resulting in the ratio for neutral, fear, and surprise being 90:96:94. This helped to separate out the data for each subject's emotions. These were then randomly selected with a ratio of S: Majority class samples among the chosen emotion label vs. the rest of the labels, resulting in class unbalancing among the EMG labels for every subject.

The value of neighbor_i was observed to be 7. Hence, the initial minority input for each individual label S was chosen to be greater or equal to 7, i.e., 10. The resultant ratio was 10:96, which was stored in a separate set. The Smote technique was subsequently applied to this dataset. The process occurred recursively for each subject's label, resulting in a unique ratio of 96:96.

The original feature vector was then appended to these synthetically generated sets, which resulted in a ratio of 184: 192: 190. This provided the final resultant training set for each subject as (566, 129) for the EMG.

As the feature vector of the final resultant training set of EMG (566, 129) is less than the total size of the feature vector of EEG and ECG (1070, 164), a down-sampling technique was used. Random down-sampling was applied to the EEG dataset for each subject on separate labels to equalize the rows. This method randomly chooses a subset of samples from the majority class to match the size of the minority class. This created separate datasets with the same rows and ratios as in the EMG set. These separate data frames were then appended together to form the final resultant training set for each subject as (566, 164) for the EEG. The resultant training set for EEG, ECG, and EMG was merged to produce a fusion model for the recognition of emotions. This fusion model was further optimized for classification. The complete process of change in the feature vector size can be visualized in Fig. 6.

3.2.5. FEATURE OPTIMIZATION AND CLASSIFICATION

Feature optimization is crucial in ER as it aims to enhance the accuracy and efficiency of classification models by selecting the most relevant and significant features from the dataset. Feature optimization streamlines the classification process, resulting in improved performance and a better understanding of emotional patterns by reducing dimensionality and focusing on essential attributes.

3.2.5.1. FUSION DATASET

The resulting training sets for EEG, ECG, and EMG are amalgamated to form a fusion model capable of ER. The fusion set consisted of (3898, 292) samples. Feature vector tuning techniques were employed to further optimize this fusion set.

A dimensionality reduction technique, PCA, was applied to the fused multi-dimensional dataset to reduce computational complexity and enhance classification accuracy. It was used to identify and extract the most significant features from the resultant fusion dataset into a new set of variables. Here, a set of 10 principal components was chosen, and the final optimized feature matrix was obtained, which was of size (3898,10).

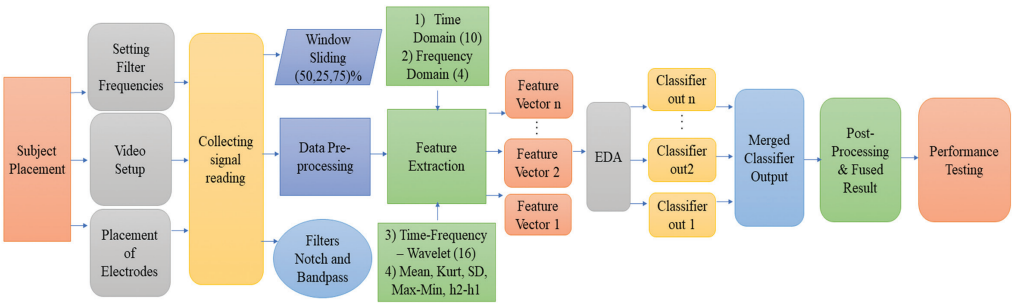


Fig. 7. Schematic of the model representing systematic steps from subject placement to model testing

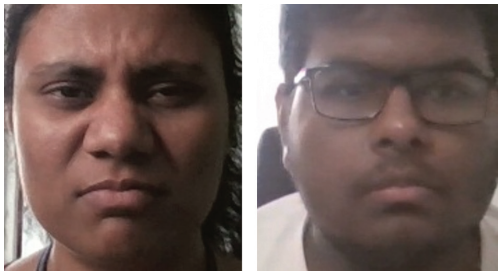
The model testing was done in two stages. First, the individual datasets of EEG, EMG, and ECG were used and optimized. Secondly, the fusion model was used. The classifiers used were decision tree classifier, random forest classifier, and Optimized random forest. After performing classification, the outputs of the classifiers were merged, followed by post-processing optimization. Different performance parameters were calculated, including accuracy, F1-score, precision, recall, ROC, and Area Under Curve (AUC). The complete process for physiological signal-based ER is illustrated in Fig. 7.

3.3. MODEL BASED ON FACIAL EXPRESSIONS

The facial model architecture was implemented in four main steps: facial data acquisition, image pre-processing, feature optimization, and classification. Each part plays a crucial role in the overall process of physiological ER.

3.3.1. FACIAL DATA ACQUISITION

A custom dataset was created for the face detection process. Sample images are illustrated in Fig. 8. The data acquisition includes participants exposed to a variety of stimuli and scenarios carefully designed to elicit specific emotions. A total of 10 participants were chosen, and video clips for various emotions were shown to them. As participants responded to these stimuli, their facial expressions were captured by the camera, resulting in a diverse dataset of facial images depicting different emotional states. This dataset consisted of face images captured against a uniform background. Firstly, normal images capturing various facial expressions were obtained. These images were then cropped manually to isolate the face region for more precise facial feature analysis. A total of 4,000 facial images were collected. These facial images were then categorized into four emotions, namely fear, surprise, neutral, and "other emotion". To enhance the visibility of facial cues during data acquisition, appropriate lighting conditions, facial angles, distances, and orientations were maintained. This ensured the proper capture of facial cues during each emotion.



(a) (b)

Fig. 8. Images for (a) fear and (b) neutral in the dataset

3.3.2. DATA PRE-PROCESSING

The collected images were resized to 64x64 and converted from RGB scale to grayscale for the facial dataset.

To enhance the system's generalization capability, further preprocessing was performed on the images, including normalization, standardization, and histogram equalization. Fig. 9 depicts the output of the preprocessing stage. This pre-processed dataset was subsequently partitioned into training, testing, and validation sets, which were then forwarded for feature extraction.

(a) (b)

Fig. 9. Pre-processed images

3.3.3. FACIAL FEATURES

Facial features represent the different properties and characteristics of facial images. Different techniques were used for extracting the required facial key points from the dataset. To select the features from the face, different techniques were studied that focused on extracting texture and shape information. Gabor filters, Histograms of Oriented Gradients (HOG), Local Binary Patterns (LBP), and Local Ternary Patterns (LTP) filters were selected and applied to the pre-processed images to extract facial landmarks. The selection of these features was based on their effectiveness in capturing multiple aspects of facial expressions.

The pre-processed dataset images were passed through the feature extraction filters, and the resulting features were used to train a classification model. In the presented method, techniques for feature extraction, namely Gabor, LBP, LTP, and HOG, were used. These techniques were designed to capture different aspects of facial expressions, such as texture, shape, and spatial relationships between facial features. The method of feature extraction is illustrated in Fig. 10.

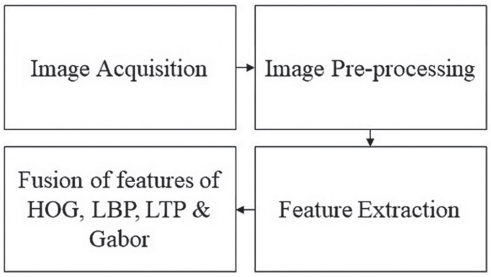


Fig. 10. Facial feature extraction

HOG was used to detect edges and oriented features in images. It detected the presence of facial expressions such as smiles, frowns, and raised eyebrows. This was achieved by detecting the changes in orientation

and gradient magnitude of the facial features. The applied HOG filter with a cell size of 16*16 pixels on the pre-processed image obtained a vector size of (1,700). LBP captures the texture information of the image and detects facial expressions such as wrinkles, dimples, and other small facial features. These features were extracted by comparing the neighboring pixels' values and their intensities. LTP, a variant of LBP, captures the texture information of an image in a more robust way. It helped to detect facial expressions that involve subtle changes in texture, such as those associated with emotions such as surprise and fear. Gabor features, derived from Gabor filters, were used to detect features at different scales and orientations. It detected the presence of facial expressions such as wrinkles, creases, and other fine details that are associated with emotions such as fear and surprise. The features mapped from the GABOR came around to (1,200).

The resulting feature vectors from all four feature extraction methods were concatenated to form the final feature vector that represents the facial expression. The final feature vector had a size of (21445, 964) and was then optimized for better performance using PCA and K-Means.

3.3.4. FEATURE OPTIMIZATION AND CLASSIFICATION

The optimization of the facial features involved two steps, i.e., clustering and feature selection. K-Means was used for clustering the extracted data features. The number of clusters K is a hyperparameter, which was chosen as 18, based on the elbow plot. This technique helped to group the similar facial emotion feature set together, which was then split into six histogram bins for creating feature columns. These feature columns were further optimized using PCA to select important features and reduce their dimensionality, making them easy for further processing. Features representing 95% of the total variance contributed by 13 principal components were chosen to create a lower-dimensional representation of the data that still retains most of the important information. The results were then used as input for a classifier for better classification performance. The implementation stages for the facial model are illustrated in Fig. 11.

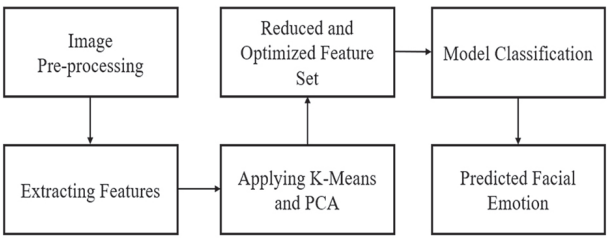


Fig. 11. FER process

Classification techniques like KNN, XGBoost, Random Forest, and decision trees are employed on the optimized feature vector. Decision trees recursively split the input space to form a tree-like structure for class prediction. KNN, which used k-nearest neighbors was initialized with a value of 59. For Random Forest, multiple decision tree predictions were combined to yield a more robust outcome, and it was initialized with 100 features. Similarly, XGBoost, an advanced gradient boosting algorithm, utilized 100 decision trees in each iteration, with the best one being 73. These techniques were then applied to the facial dataset for classification based on the provided labels. The models were trained on the extracted features, and their performance was subsequently evaluated.

3.4. VOTING BASED DECISION AND IMPLEMENTATION ON JETSON NANO

The facial and bio-electric fusion models were carefully selected with the consideration that the system needed to be deployed in real-time, aiming for a balance between performance and time requirements. In the video capture process, where subjects displayed specific emotions, the video stream was segmented into individual frames.

Algorithm 2: Voting based Decision on the Candidate Emotions

Input: Video input from camera & Raw input from the Datasheet
Output: Voting based Decision on the final output
Initialization
$f(x) \leftarrow$ features for face emotion, $f(y) \leftarrow$ features for bio-emotion
$r \leftarrow$ ROI (Face)
$fe(m) \leftarrow$ Facial Emotion Trained Model
$be(t) \leftarrow$ Fused Bio-Emotion Trained Model
1: for frame $f(x,y)$ in video input
2: if frame then
3: $f(y) \leftarrow$ Feature Extraction (Signal Input)
4. Pred_bio_emotion $\leftarrow be(f(y))$
5: ROI $\leftarrow$ Haar Cascade(Frame)
6: if $face=True$ then
7: $Op \leftarrow Feature\ Extraction$ (ROI)
8: $Pred\_f \leftarrow fe$(Op)
9: end if
10: end if
11: end for
12: Count Number of $Pred\_f$ & $Pred\_b$
13: final decision = $Max\ vote(emotion)$
14: return final decision
end

Accurate real-time ER relies on the identification and interpretation of human facial expressions, making the localization of the face within each frame crucial to identifying the Region of Interest (ROI). To accomplish this, the Haar Cascade Localization technique was employed for face detection. Once the face was successfully recognized through the HAAR cascade, the captured image underwent pre-processing, includ-

ing feature extraction and filtering. This pre-processed data was then fed into a classification model, allowing the real-time prediction of emotions based on this comprehensive facial analysis. A review of facial model implementation is given in Fig. 12.

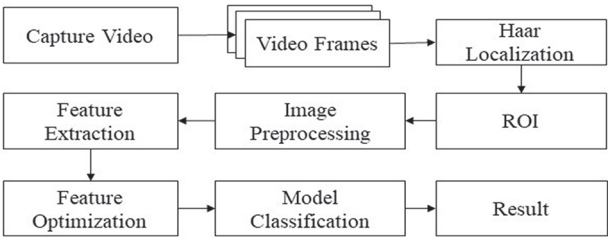


Fig. 12. Complete implementation of facial model from data acquisition to model classification

Jetson Nano was selected as the embedded development board as it is suitable for handling real-time, scalable ML models. Here, a Logitech C270 camera was interfaced with the Jetson Nano Board to capture the real-time video. The Jetson Nano was configured accordingly, and then a bio-datasheet was stored in it. The developed system was a stand-alone system in which the video captured by the camera was processed to predict facial emotion, which was fed to the final voting mechanism. Simultaneously, the bio-emotion predicted by the bio-electric model was also integrated into the system. Depending on the frequency of emotions, for every 2-second duration of video input, the final output emotion was given as described in Algorithm 2. This resulted in five voting classifier decisions for a duration of 10 seconds of video input.

This array of outputs was further given to the GPT language model to generate a medical report.

3.5. CONCLUSIVE REPORT GENERATION USING GPT

GPT is a state-of-the-art natural language processing technology developed by OpenAI. It is a type of artificial neural network that has gained significant attention in the fields of AI and ML due to its remarkable ability to generate human-like text. GPT models are pre-trained on massive amounts of text data from the internet and then fine-tuned for specific tasks, making them highly versatile for various natural language understanding and generation tasks. The core of GPT technology is the Transformer architecture, which is designed to handle sequential data efficiently. Transformers use a mechanism called attention to process and generate sequences of text. GPT models are accessed and utilized through an Application Programming Interface (API) provided by OpenAI, allowing for a wide range of natural language processing tasks.

The final output from the voting classifier based on the predictions from the physiological and vision classifiers is given to the GPT using this API. The use of GPT here is to synthesize these diverse emotional cues into a comprehensive conclusion. For a 10-second video, the voting classifier produces a decision every 2 seconds, resulting in 5 sequential emotions. Using this array of emotions, GPT generates a comprehensive medical report. This report succinctly encapsulates the array of emotions exhibited by the subject throughout the 10-second timeframe. This approach allows for a holistic understanding of the subject's emotional state, bridging the gap between physiological and visual indicators to provide valuable insights.

4. RESULTS AND DISCUSSION

The major part of the proposed study focuses on offering a comprehensive and systematic approach to ER, effectively combining physiological signals and facial expressions to achieve more accurate and nuanced emotion estimation. Utilizing GPT, a medical conclusion derived from these emotions yields promising implications for medical applications.

4.1. CHANGES IN PHYSIOLOGICAL SIGNALS AND FACIAL EXPRESSIONS

In this experiment, several alterations were observed while participants were expressing a particular emotion. These variations in traits provided valuable insights into the unique patterns exhibited in physiological and facial signals across different subjects.

4.1.1. VISION

During emotions like fear and surprise, there is an observable increase in electrical activity in the muscles responsible for controlling eye and brow movements, reflected in the EMG signal. The amplitude and voltage of the EMG signal are elevated as the electrical impulses to these muscles intensify. This heightened electrical activity corresponds to widened eyes, raised eyebrows, and a more open mouth, all of which are indicative of the emotional response.

4.1.2. EEG

Fear and surprise lead to increased electrical activity in specific brain regions, particularly the amygdala. This heightened electrical activity is reflected as an increase in the amplitude of the EEG signal. The amygdala's role in emotional processing and arousal results in these elevated electrical signals. Conversely, during neutral emotions, the EEG signal shows a more balanced amplitude across various brain regions, reflecting a state of relative emotional calm.

4.1.3. ECG

During fear and surprise, there is an increase in the electrical signal's amplitude, which directly corresponds to a rise in heart rate. The sympathetic nervous system's activation during these emotions leads to the generation of stronger electrical impulses in the heart,

causing an elevated heart rate. Simultaneously, heart rate variability may decrease as the electrical signals governing heartbeat rhythms become more regular.

4.1.4. EMG

During fear and surprise, there is an increase in the amplitude and voltage of the EMG signal in facial muscles, such as those controlling eyebrow and mouth movements. These changes reflect heightened muscle tension in response to emotional arousal. In contrast, during neutral emotions, the EMG signal exhibits a more relaxed state with lower amplitude and voltage, indicating a lack of pronounced muscle contractions. In the case of EMG, during fear and surprise, there is an increase in the amplitude and voltage of the electrical signals detected in facial muscles. These changes are a direct result of heightened muscle tension and increased electrical impulses in the muscles controlling facial expressions, such as those around the eyebrows and mouth. These electrical signals are indicative of the muscle contractions associated with the emotional response.

4.2. EVALUATION OF CLASSIFIERS

EDA was performed on both the facial and physiological features to visualize and identify the important features. These features were further optimized and then provided to the classifier. The performance of the classifier was estimated using different parameters, including accuracy, recall, precision, F1 score, and ROC.

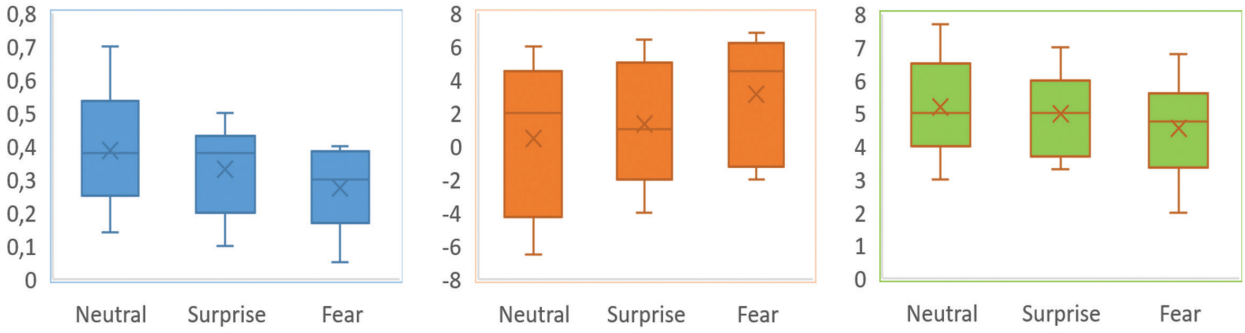


Fig. 13. Data analysis of some of the important features (a) Myopulse percentage rate (channel – 2), (b) Kurtosis (channel – 3), and (c) Mean frequency (channel – 1) considered in the model for different channels of device

4.2.1. PERFORMANCE OF BIO-ELECTRIC CLASSIFIER

EDA performed on the bioelectric features produced different variations depending on the emotions chosen. This helped to separate out the important features among the set of features and remove the redundant features. Features from the wavelet domain, namely mean and kurtosis showed distinct categorization in the minimum and maximum values of their boxplots. The myopulse percentage rate is an important feature in the time domain, showing variations in its EDA visualization aiding in the categorization of the different emotions captured by the EMG biomedical device. These important features were depicted through their boxplots in Fig. 13. Changes in performance parameters were observed using low frequencies (LF) and high frequencies (HF).

The frequency range was first chosen depending on the highest performance parameters and then changed according to the distribution of mean band amplitude over the EEG frequency bands. For EMG, when 2 and 10 were selected for LF and HF, respectively, the accuracy was around 63%. In the case of EEG and ECG, when 5 and 20 were selected for LF and HF, respectively, the accuracy was around 59%. Among all of these above classifiers, different classifiers for different scenarios have achieved better accuracy. The performance of different classifiers for each EEG, EMG, and ECG has been visualized. Accuracy was boosted up to 71% and 62% in the cases of EMG and EEG-ECG, respectively, by changing the notch filter frequency. These changes in the performance parameters have been visualized in Fig. 14 and Fig. 15.

Fig. 14. Parameter comparison of classifiers on the EMG dataset

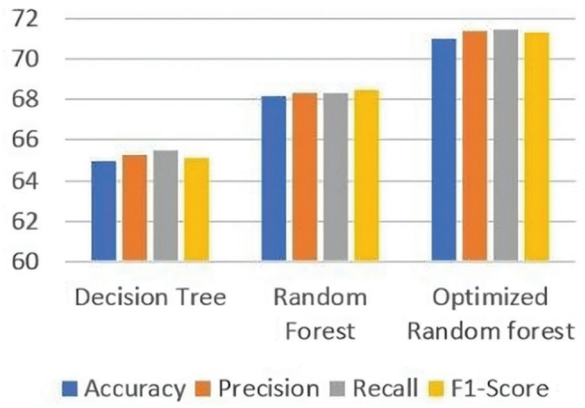


Fig. 15. Comparison of parameters of classifiers on EEG and ECG dataset

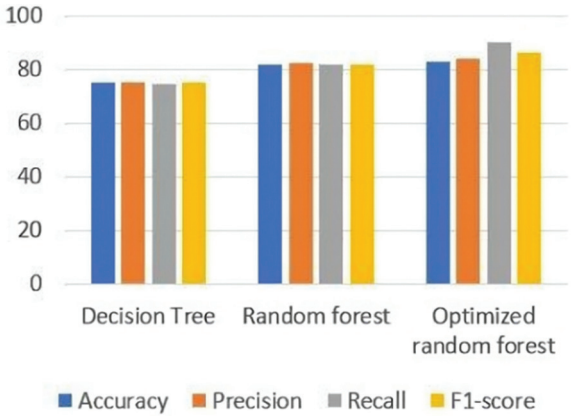


Fig. 16. Parameters comparison for models on class balanced fusion dataset of EEG, EMG, ECG with smote technique

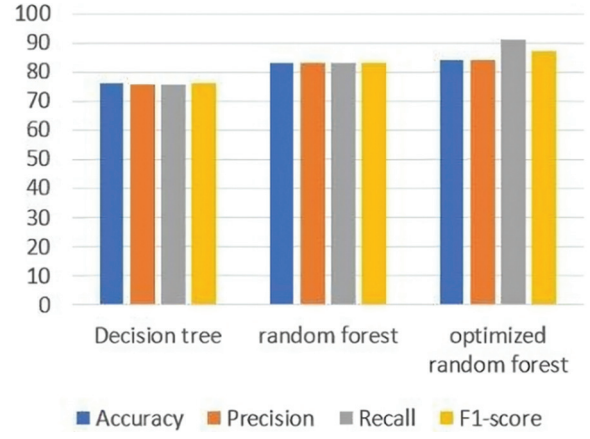


Fig. 17. Comparison of parameters for models on optimized fusion dataset after PCA

The resultant data set was obtained using fusion techniques. Further optimization was carried out by choosing 10 linearly uncorrelated variables and ordering them in terms of their importance in explaining 99% of the variance in the original data. These chosen principal components helped in identifying the key features or variables that have the most significant impact on the outcome variable, which aids in the final classification. After optimization, the accuracy was increased up to 84% in the case of EMG and 76% for EEG and ECG, as can be seen in Fig. 16 and Fig. 17.

The results highlight the importance of considering multiple performance metrics, as different metrics provide different insights into the performance of the model. Here, the parameters were calculated based on the Weighted-average which considers the relative importance of different performance parameters. The Weighted average precision was calculated based on precision per class and takes into account the number of samples of each class in the data, providing the highest result compared to other weights.

The fusion model showed a boost in classifier performance compared to individual datasets. The 84% accuracy of the optimized random forest model suggests that it was able to detect positive emotions effectively, which is important for clinical applications. The decision tree model achieved an accuracy of 76%. The precision of the models was 85% and 76.4%, respectively, providing the ability to correctly identify a particular emotion (e.g., neutral) among all the emotions it predicts. The recall of the models was compared, and the optimized random forest achieved the highest recall of 91%, showing its ability to identify particular emotions (e.g., fear) among all the samples that actually represent that emotion. This ensures that almost no positive cases are missed, which can improve patient outcomes. Both recall and precision are important in clinical settings and for prosthetic patients, as they provide insights into the model's ability to accurately detect positive and negative cases of a particular medical condition. It is important to strike a balance between precision and recall. Accounting for this F1 score, a harmonic mean of precision and recall, which provides a balance between the two metrics, was calculated. The models achieved an F1 score of 88% and 79%, respectively.

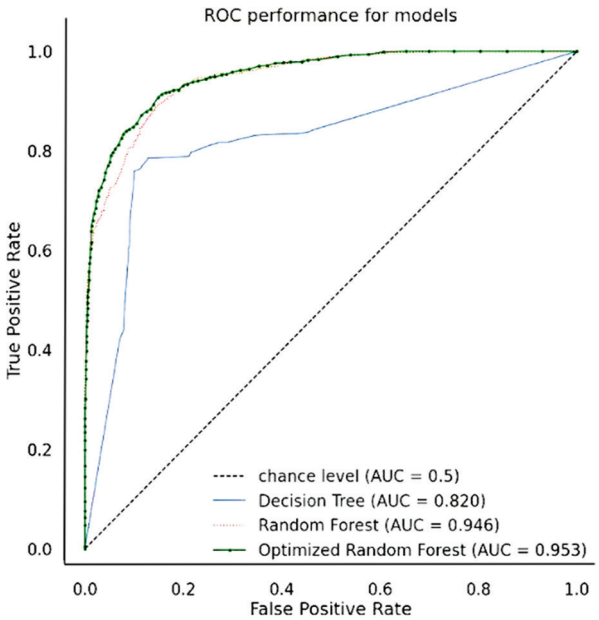


Fig. 18. ROC curve and AUC curve for neutral vs. all for the highest-achieving models

A multiclass ROC curve is visualized in Fig. 18. The multiclass ROC curve extends the binary ROC curve to multiple classes by considering each class as a positive class (neutral) and all other classes as the negative class (surprise & fear), plotting the ROC curve for each class, and then averaging them. The curve illustrates the trade-off between the true positive rate and false positive rate for each class, providing a comparison of model performance across different classes.

The optimized random forest classifier achieved the highest Area Under Curve of 0.95, demonstrating the model's capability to differentiate between the positive and negative classes. This model was then saved for implementation on the embedded system.

4.2.2. PERFORMANCE OF FACIAL FEATURE CLASSIFIER

Facial features were extracted from each facial image, with a specific focus on facial key points. Here, HOG, LBP, LTP, and Gabor feature techniques were chosen according to their ability to capture important facial cues. Depending on the emotion, distinct facial cues were identified. The visualization of these different facial features is illustrated in Fig. 19.

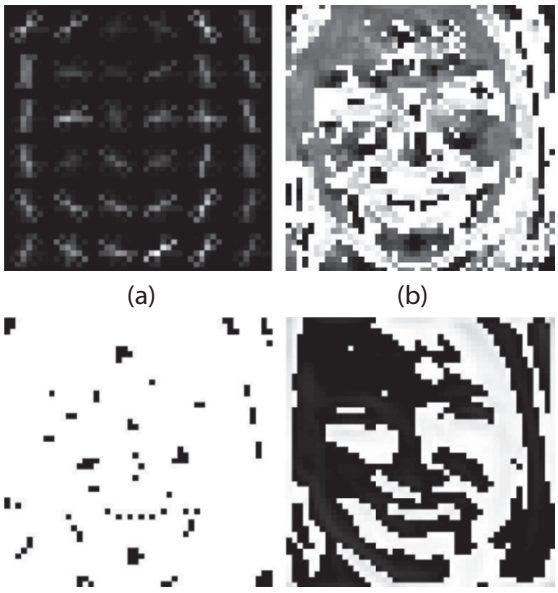


Fig. 19. Facial feature visualization on applying different feature extraction techniques (a) Hog, (b) Gabor, (c) LTP, and (d) LBP

This helped to visualize facial features and remove redundant features. Further optimization was carried out using clustering and PCA. This feature vector was then provided to the ML algorithms. Their performance metrics were calculated to provide useful insights for FER. Similar to the bio-electric model, here the parameters were calculated based on the Weighted average which considers the relative importance of different performance parameters. Fig. 20 summarizes the classification performance of each classifier for ER. KNN achieved the highest accuracy of 84.6%, followed by Random Forest with 74.3%, Decision Tree with 67%, and Xg-Boost with 64.5%. KNN also achieved the highest precision and F1-score for each emotion category, indicating its superior performance compared to the other classifiers. This model was saved and further used for embedded deployment.

Fig. 20. Classification performance of each classifier for emotion recognition

4.3. FUNCTIONAL/USABILITY EVALUATION

The models to be deployed on the embedded board were chosen based on the highest ratio of accuracy to time taken. A comparison of model accuracy and the time taken to predict the emotion after the input is given is shown in Table 3.

Model	Classifier	Accuracy	Time Taken
FER	KNN	84.6%	0.6 s
	Random forest	74.3%	0.8 s
	Decision Tree	67%	0.5 s
	Xg-Boost	64.5%	1.4 s
Bio-signal Fusion based ER	Optimized Random Forest	84%	0.3 s
	Optimized decision tree	76%	0.2 s

Table 3. Comparison of accuracy and time taken to generate prediction for bio-electric and facial models

Accordingly, for physiological signals and facial expression-based emotion prediction, optimized random forest and KNN were selected, respectively, to be used in the implemented system, which is illustrated in Fig. 21. Video recorded during participant under the ARS system was used as a basis to test the implemented system, and the facial expression video was mapped according to the physiological emotions. This helped to compare the ground-truth emotions with the physiologically and facially predicted emotions.

The EEG, ECG, and EMG fusion data was directly provided to the bio-electric classifier on the embedded system. The real-time video of the subject was captured at the same time. This helped process the video to predict the facial emotion and compare it with the predicted physiological emotion. The predicted emotions

from both sources were fed into the voting mechanism, which determined the final output based on the highest frequency of the recurring predicted emotion. Out of a total of 17 predictions made for the fusion of facial and bio-electric models, the average of 15 output predictions matched the ground truth emotion. The average accuracy for the fusion model was found to be 87.5%.

The predicted emotions were further subjected to a voting mechanism where, depending on the highest frequency of emotions in a 2-second window, the final output emotion is provided. This results in a total of five final emotion predictions during a 10-second video input.

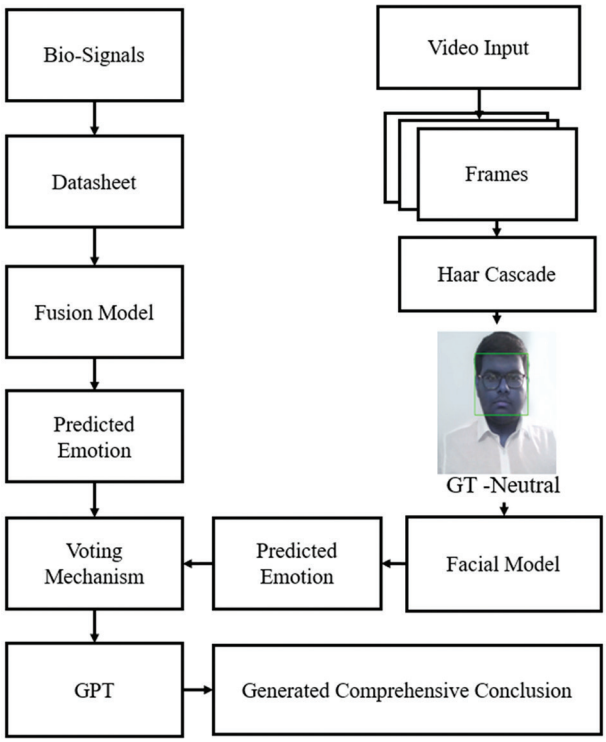


Fig. 21. Implementation of the system on a standalone embedded board

These emotions were provided to the GPT-3 model to generate a suitable, comprehensive conclusion. Its performance was compared by observing the similarity between the ground truth for the patient over the period of video input and the generated conclusion. The GPT-generated conclusion for an array of emotions, namely fear, is: "Based on the predicted outcome of the patient's emotions, it can be concluded that they experience fear, and it is important for healthcare professionals to provide emotional support and resources to help patients cope with it". This helps to provide a brief conclusion on the subject's emotions that are conveyed during the test.

4.4. SIGNIFICANCE AND COMPARISON WITH EXISTING WORK

The fusion of EEG, ECG, and EMG signals provides a holistic understanding of emotional states by capturing neural, cardiac, and facial muscle insights.

These signals collectively represent the diverse dimensions of human emotions. An integrated decision, combined with facial insights, is essential due to the intricate nature of emotions, accommodating individual variations, and enhancing accuracy. This integration adds robustness, overcoming challenges like noise, and cross-validation across signals further boosts credibility, and minimizes the risk of misclassification.

Table 4. Comparison of proposed system with existing approaches

Name	Dataset	Multi-Modularity	Fusion	Accuracy	Real-Time Implementation
Proposed System	Custom	Yes	Yes	87.5%	Yes
[6]	Available	No	No	75.31%	No
[9]	Custom	Yes	Yes	69.5%	No
[31]	Custom & Available	No	No	85.38%	No
[32]	Custom	No	No	84.3%	No

This proposed system has been compared with existing approaches in two ways, i.e., by comparing complete systems and by comparing only datasets. The effectiveness of the proposed facial model was assessed through testing on established benchmark datasets, including FER2013 [33] and CK+ [34], resulting in accuracy rates of 75% on FER and 90% on CK+. Due to the innovative nature of the fusion model incorporating EEG, EMG, and ECG signals, a dedicated testing dataset tailored to this unique approach was not available for assessment. A complete system comparison has been shown in Table 4 by using different parameters like multi-modularity, real-time nature, and system performance.

The proposed system model shows high accuracy in real-time systems compared to other existing ones. It uses a custom dataset and focuses on multi-modularity. The development of this system represents a significant advancement in the understanding and application of emotions within prosthetic systems and as general medical solutions. The integration of vision and physiological signals plays an integral part in user-machine interactions, allowing prosthetic arm users to experience an emotionally intuitive control interface.

5. CONCLUSION

The recognition of emotions can be achieved using signals such as EEG, ECG, and EMG, as well as facial expressions. The fusion of these signals and facial emotions has demonstrated a notable improvement in system performance. This study explored an ER system that uses EEG, ECG, and EMG data to identify the three emotional

states of fear, neutral, and surprise. Concurrently, facial expressions were also incorporated into the analysis. In this work, a dataset of fear, neutral, and surprise emotions for EEG, ECG, and EMG signals was meticulously collected. The subsequent steps involved the extraction of diverse domain-specific features, followed by the visualization of feature vectors. Important features were discerned and subsequently refined through the application of EDA and PCA The resultant fusion matrix was formed using fusion and class-balancing techniques on a combined dataset. These matrices were then subjected to classification employing specific classifiers. The accuracy of 84% and 76% were obtained using the optimized random forest model and the optimized decision tree, respectively, for the bio-signal-based model. Comparatively, when EEG-ECG and EMG signals were chosen individually, accuracy was lower due to the consideration of an unimodular system.

A boost in performance parameters was observed for the fusion dataset. Optimization also helped in increasing this accuracy. In the case of facial emotion, features, namely LBP, LTP, HOG, and Gabor were extracted and optimized by using K-means clustering and PCA. An accuracy of 84.6%, 74.3%, 67%, and 64.5% was achieved using KNN, Random Forest, Decision tree, and Xg-Boost Classifiers, respectively. Both modalities were used to determine the optimal emotion. To consolidate the results of both physiological and facial emotion analyses, a voting classifier was employed to determine the final emotional classification. This model was deployed in a real-time ER system by interfacing a Logitech C270 camera with the Jetson Nano board.

In this context, the system was beset with several issues, such as subject independence, which provides less accuracy owing to limited data. Expanding sample sizes is crucial to enhancing the generalizability of findings. By using more participants' data, higher accuracy can be achieved. Additional model tuning can be done using a greater number of channels and features. Other techniques and classifiers can be explored to improve the current performance estimators. Other preprocessing steps could be explored for recognizing facial key points.

The emotions recognized by this model can be used to improve the functionality and usability of prosthetic devices. By enabling users to control their prosthetics using their emotions, it can provide a more intuitive and natural user experience. Its inception followed a comprehensive survey of existing prosthetic system users, addressing the need for more personalized and responsive assistive technologies. The system not only enhances the functionality and comfort of the prosthetic system but also opens doors to a wide range of medical applications. It allows for emotionally aware control, enhances the user experience, and blurs the lines between human and machine. This novel integration not only improves the functionality of prosthetic arms but also profoundly impacts users' comfort and confidence. Moreover, while its primary application is within prosthetic arms, its adaptability and versatility make it usable for general medical systems. Its potential impact spans across the spectrum of medical and assistive technologies, making it a pioneering and transformative development.

Future research could focus on improving ER accuracy by incorporating other modalities, such as voice and body language, and using larger and more diverse datasets. This system holds promise for further refinements and broader applications. Additional parameters, such as Galvanic Skin Response (GSR) sensors, voice analysis, and temperature sensors, can be integrated to enhance its accuracy and versatility. This expansion of sensor types will enable even more precise recognition and response to emotional states, further improving the quality of life for prosthetic system users. The incorporation of these sensors may lead to applications in mental health support, user experience enhancement, and medical diagnosis, marking an exciting path for future research and development. The proposed work highlights its significance in the domain of ER and suggests the potential for further exploration in related fields.

6. REFERENCES:

[1] J. Chen, T. Ro, Z. Zhu, "Emotion Recognition With Audio, Video, EEG, and EMG: A Dataset and Baseline Approaches", IEEE Access, Vol. 10, 2022, pp. 13229-13242.

[2] S. Sharma, M. Bhatt, P. Sharma, "Face recognition system using machine learning algorithm", Proceedings of the 5th International Conference on Communication and Electronics Systems, Coimbatore, India, 10-12 Jun 2020, pp. 1162-1168.

[3] S. Kim, G. H. An, S. Kang, "Fusion expression recognition system using machine learning", Proceedings of the 14th International SoC Design Conference, Seoul, Korea, 5-8 Nov 2017, pp. 266-267.

[4] A. Kumar, N. Garg, G. Kaur, "An emotion recognition based on physiological signals", International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No. 9, 2019, pp. 335-341.

[5] H. E. Houssein, H. Asmaa, A. A. Abdelmgeid, "Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review", Neural Computing and Applications, Vol. 34, No. 15, 2022, pp. 12527-12557.

[6] S. Hwang, M. Ki, K. Hong, H. Byun, "Subject-independent EEG-based emotion recognition using

adversarial learning", Proceedings of the 8th BCI International Winter Conference on Brain-Computer Interface, Gangwon, Korea, 26-28 February 2020, pp. 1-4.

[7] H. Ferdinando, T. Seppänen, E. Alasaarela, "Comparing features from ECG pattern and HRV analysis for emotion recognition system", Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, Chiang Mai, Thailand, 5-7 October 2016, pp. 1-6.

[8] S. Katsigiannis, N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices", IEEE Journal of Biomedical and Health Informatics, Vol. 22, No. 1, 2017, pp. 98-107.

[9] S. Jerritta, M. Murugappan, W. Khairunizam, S. Yaacob, "Emotion recognition from facial EMG signals using higher order statistics and principal component analysis", Journal of the Chinese Institute of Engineers, Vol. 37, No. 3, 2014, pp. 385-394.

[10] M. Hafsa, B. B. Hajira, B. L. Meghana, Y. A. Arpitha, H. R. Niveditha, "Human Basic Emotion Recognition from EEG Signals using IOT", International Journal of Engineering Research & Technology IETE, Vol. 8, No. 11, 2020, pp. 47-50.

[11] N. Kumar, K. Kaushikee, M. S. Hazarika, "Bispectral analysis of EEG for emotion recognition", Procedia Computer Science, Vol. 84, 2016, pp. 31-35.

[12] A. Goshvarpour, A. Abbasi, A. Goshvarpour, "An accurate emotion recognition system using ECG and GSR signals and matching pursuit method", Biomedical Journal, Vol. 40, No. 6, 2017, pp. 355-368.

[13] H. S. Cha, C. H. Im, "Performance enhancement of facial electromyogram-based facial-expression recognition for social virtual reality applications using linear discriminant analysis adaptation", Virtual Reality, Vol. 26, No. 1, 2022, pp. 385-398.

[14] M. R. Kose, M. K. Ahirwal, A. Kumar, "A new approach for emotions recognition through EOG and EMG signals", Signal, Image and Video Processing, Vol. 15, No. 8, 2021, pp. 1863-1871.

[15] B. Silvio, A. Casanova, M. Fraschini, M. Nappi, "EEG/ECG signal fusion aimed at biometric recognition", Proceedings of the 18th International Conference on Image Analysis and Processing, Genoa, Italy, 7-8 September 2015, pp. 35-42.

[16] S. Ganguly, R. Singla, "Electrode channel selection for emotion recognition based on EEG signal", Proceedings of the 5th International Conference for Convergence in Technology, Bombay, India, 29-31 March 2019, pp. 1-4.

[17] K. P. Wagh, K. Vasanth, "Electroencephalograph (EEG) based emotion recognition system: A review", Proceedings of the 6th Innovations in Electronics and Communication Engineering, Hyderabad, India, 21-22 July 2017, pp. 37-59.

[18] Z. Cheng, L. Shu, J. Xie, C. P. Chen, "A novel ECG-based real-time detection method of negative emotions in wearable applications", Proceedings of the International Conference on Security, Pattern Analysis, and Cybernetics, Shenzhen, China, 15-17 December 2017, pp. 296-301.

[19] P. Sarkar, A. Etemad, "Self-supervised learning for ECG-based emotion recognition", Proceedings of the 45th International Conference on Acoustics, Speech and Signal Processing, Barcelona, Spain, 4-8 May 2020, pp. 3217-3221.

[20] M. Chen, Y. Hao, "Label-less learning for emotion cognition", IEEE Transactions on Neural Networks and Learning Systems, Vol. 31, No. 7, 2019, pp. 2430-2440.

[21] Y. Wu, Y. Wei, J. Tudor, "A real-time wearable emotion detection headband based on EEG measurement", Sensors and Actuators A: Physical, Vol. 263, 2017, pp. 614-621.

[22] A. Topic, M. Russo, "Emotion recognition based on EEG feature maps through deep learning network", Engineering Science and Technology, an International Journal, Vol. 24, No. 6, 2021, pp. 1442-1454.

[23] A. Majumder, L. Behera, V. Subramanian, "Automatic Facial Expression Recognition System Using Deep Network-Based Data Fusion", IEEE Transactions on Cybernetics, Vol. 48, No. 1, 2016, pp. 103–114.

[24] J. Alghamdi, R. Alharthi, R. Alghamdi, W. Alsubaie, R. Alsubaie, D. Alqahtani, R. Ramadam, L. Alqarni, R. Alshammari, "A survey on face recognition algo-

rithms", Proceedings of the 3rd International Conference on Computer Applications & Information Security, Riyadh, Saudi Arabia, 19-21 March 2020, pp. 1-5.

[25] L. Li, X. Mu, S. Li, H. Peng, "A review of face recognition technology", IEEE Access, Vol. 8, 2020, pp. 139110-139120.

[26] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using Laplacian faces", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, 2005, pp. 328-340.

[27] T. Ahmed, P. Das, F. Ali, and F. Mahmud, "A comparative study on convolutional neural network-based face recognition", Proceedings of the 11th International Conference on Computing, Communication and Networking Technologies, Kharagpur, India, 1-3 July 2020, pp. 1-5.

[28] M. K. Chowdary, T. N. Nguyen, D. J. Hemanth, "Deep learning-based facial emotion recognition for human–computer interaction applications", Neural Computing and Applications, Vol. 35, No. 32, 2021, pp. 1-18.

[29] S. Sharma, V. Kumar, "Performance evaluation of machine learning based face recognition techniques", Wireless Personal Communications, Vol. 118, No. 4, 2021, pp. 3403-3433.

[30] K. Koonsanit, N. Nishiuchi, "Classification of user satisfaction using facial expression recognition and machine learning", Proceedings of the IEEE Region 10 Conference, Osaka, Japan, 16-19 November 2020, pp. 561-566.

[31] Y. Liu, O. Sourina, "EEG-based subject-dependent emotion recognition algorithm using fractal dimension", Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, USA, 5-8 October 2014, pp. 3166-3171.

[32] T. Ergin, M. Ozdemir, A. Akan, "Emotion recognition with multi-channel EEG signals using visual stimulus", Proceedings of the Medical Technologies Congress, Izmir, Turkey, 3-5 October 2019, pp. 1-4.

[33] L. Goodfellow et al. "Challenges in representation learning: A report on three machine learning contests", Proceedings of the 20th International Conference on Neural Information Processing, Daegu, Korea, 3-7 November 2013, pp. 117-124.

[34] L. Patrick, J. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews, "The extended Cohn-Kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition – Workshops, San Francisco, CA, USA, 13-18 June 2010, pp. 94-101.

Black hole algorithm along edge detector and circular hough transform based iris projection with biometric identification systems

Original Scientific Paper

Tara Othman Qadir
Universiti Tun Hussein Onn Malaysia (UTHM)
Faculty of Computer Science and Information Technology
Parit Raja 86400, Johor, Malaysia.
alsaraftara@gmail.com

N.S.A.M Taujuddin
Universiti Tun Hussein Onn Malaysia (UTHM)
Faculty of Electrical and Electronic Engineering
Parit Raja 86400, Johor, Malaysia.
shahidah@uthm.edu.my

Abstract *– The circular parameters between the pupil and the iris are found using current iris identification techniques but the accuracy creates an issue for the detection process during image processing. The procedure of extracting the iris region from an eye image using circular parameters can be improved via approximately too many approaches in literature but remain some portions under slightly unconstrained circumstances. In this study, we presented a Black Hole Algorithm (BHA) along the Canny edge detector and circular Hough transform-based optimization technique for circular parameter identification of iris segmentation. The iris boundary is discovered using the suggested segmentation approach and a computational model of the pixel value. The BHA looks for the central radius of the iris and pupil. The system uses MATLAB to test the CASIA-V3 database. The segmented images exhibit 98.71% accuracy. For all future access control applications, the segmentation-based BHA is effective at identifying the iris. The integration of the BHA with the Hough transforms and Canny edge detector is the main method by which the iris segmentation is accomplished. This novel technique improves the accuracy and effectiveness of iris segmentation, with potential uses in image analysis and biometric identification.*

Keywords: *identification, segmentation, biometric, image-processing, iris detection, accuracy, edge detector, circular hough transformation*

1. INTRODUCTION

Iris recognition has quickly emerged as a popular study issue due to its potential use in personal identification considering the rising security demands in our daily lives [1-5]. The iris of the human eye is the annular portion positioned between the white sclera and the black pupil. It has a rich texture created by numerous unique minutes (such as circles, inscriptions, and furrows), among others. Iris is especially helpful for personal identification because it is regarded to be highly discriminative between eyes and stable during a person's lifespan. The human eye is considered a complicated and extremely distinctive biological feature in the field of iris recognition. In this technology, the iris is a tiny, round structure that surrounds the pupil which is also an essential part of the eye. Each person's iris is unique because of the complex patterns of crypts, furrows, and other distinguishing characteristics that make it unique. Iris patterns are generated during a person's development and are astonishingly stable over the course of a person's lifetime, making them a permanent and trustworthy type of biometric identification. Tucked away inside the iris, the pupil controls how much light enters the eye and adjusts to the surrounding lighting. Iris recognition systems use this interaction between the iris and pupil to offer extremely precise and safe authentication and identification techniques, utilizing the biological wonder of the human eye for cutting-edge technical applications [6, 7]. Fig. 1 depicts the entire human eye, with all of its distinguishing features.

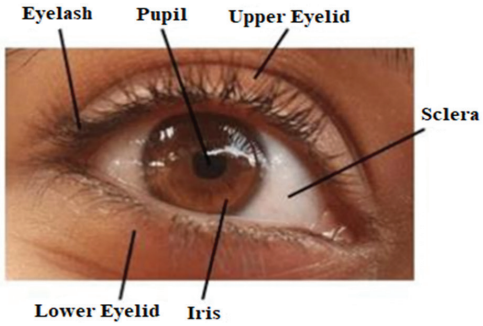


Fig. 1. A generic view of the human eye [6]

For many security applications, rapid development of recognition systems has occurred. This research is driven by the rising market need for automated access control systems with a biometric recognition platform [8]. For human recognition systems, a variety of identification traits have been investigated, including the iris, fingerprint, face, mouth, and ear [9]. The most reliable and consistent recognition biometric, however, is iris-based personal identity systems [10]. In 1993, Daugman [11] wrote about the first successful application of an iris recognition system. A few processes are involved in iris recognition systems, including edge localization, segmentation, normalization, feature extraction, and matching. Accurate segmentation can considerably enhance the performance of the recognition system. The characteristics of the iris' texture vary between any two individuals and between the left and right eyes of the same individual. A sample of the tested database of iris photographs is shown in [12] (Fig. 2). It should be emphasized that the white area is the sclera, and the small black circle represents the pupil of the eye.

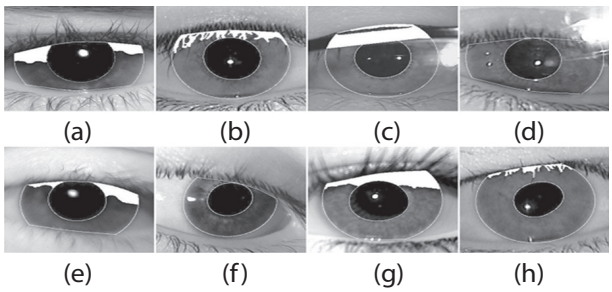


Fig. 2. Examples of iris segmentation by the proposed methods on various challenging iris images. Iris images with (a) severe eyelid occlusion, (b) eyelash occlusion, (c) glass frame occlusion, (d) specular reflections, € defocusing, (f) off-axis view angle, (g) motion ghost, and (h) pupil deformation

The iris of the eye is located between the pupil and sclera, though. The key challenge to improving identification accuracy is precisely identifying both the inner and exterior edges of the iris region. It is typical to examine an incomplete iris image, nonetheless. Whereas the area of interest for an iris is obscured by both the eyelid and the eyelash. The outside boundary area cannot be localized using traditional localization techniques. Additionally, the iris's inner area limits, radius, center, and shape are constantly altered by the amount of light falling on the eye pupil.

Due to conflicting boundaries and centers, both the inner and outer circular parameters must be obtained separately [13]. Therefore, a key challenge for biometric iris identification systems is the precise segmentation of incomplete iris boundaries.

High-accuracy iris systems require an accurate segmentation technique. The system's computing complexity may grow, and the time required for iris segmentation may be cut in half. As a result, the iris segmentation problem has been the subject of numerous studies in literature. To determine the iris' boundary, Daugman presented an integro-differential method-based iris identification system [11]. The Hough transform approach was then presented by Wildes to determine the boundaries of the iris [14, 15]. The parameters of the iris boundaries must be searched for using these two traditional methods across the entire parameter space. In [16], the authors used the watershed transform and circle fitting to locate the outer circle, while the inner circle was found using the Canny edge and fitting circle approach [16]. An iris localization approach for imperfect iris data was put forth in [17]. The iris boundaries were located using visible-wavelength light. Edge detection and the identification of the inner and outer circles were performed using the computation of image intensity and the integro-differential operator. To acquire the pupil and iris edges, the researchers also developed an adaptive intensity threshold approach [18]. The boundaries of the inner and outer circles were localized using the circular Hough transformation. Sardar [19] presented hole filling and rough entropy-based iris segmentation. The sclera edge is used to roughly compute the outer circle. However, the next phase of the normalizing process requires more than just utilizing the binarization algorithm. According to [20], the Daugman-rubber sheet normalization requires the parameters of both the inner and outer circles.

The BHA for data clustering was suggested in [21]. The BHA is straightforward to use and has an uncomplicated structure. It features a quick search method for an objective function's ideal parameters. In this research, the BHA is used to determine the iris' inner and outer circular characteristics. This study is divided into the following sections: First, describe some previous studies, and after this, the methodology in which the mechanism of BHA is described. The iris segmentation process is examined through BHA, Canny edge detector, and circular Hough transform. The results and a discussion of the suggested model are covered in the following section. Finally, the conclusion brings this study to a close end.

The BHA's goals for iris border detection and segmentation are useful for vetting accurate and trustworthy iris recognition systems. The purpose is to successfully isolate and extract the iris region from an image.

The end effect of this is to detect and segment the iris boundary with great precision. This guarantees that the extracted iris region closely resembles the iris's true limits.

Here, Section 1 covers the introduction to iris research, and Section 2 presents the research that is pertinent to this matter. The algorithm and its primary parts, as well as the associated methods in the form of methodology, are the foundation of Section 3. Section 4 displays the results of the experiments, and Section 5 concludes this investigation.

2. RELATED WORK

In unregulated situations, iris segmentation approaches can be categorized in a variety of ways. To efficiently investigate the development of the study on this topic, this can be provided three distinct categorizations for these methodologies in this section.

Table 1. Various methods used/developed under some unconstrained factors

Reference	Method	Image scale	Datasets	Limitations
[22]	Adaptive mean shift procedure	Gray-scale	CASIA-V3	Overall performance and reliability of the system
[23]	Robust post-processing Algorithm	Color and Gray-scale	CASIA-V4	Can improve recognition performance with other approaches and improve the unstable bits in the iris code, further investigation is needed for the robust human verification, can search the best possible configuration for feature extraction
[24]	Deep multi-task learning framework, named as IrisParseNet	Color and Gray-scale	CASIA-V1	Can improve the efficiency of the post-processing step or integrate it into the iris segmentation and localization system to form an end-to-end model
[25]	Deep neural network with augmentation method	Gray-scale	CASIA-V1	Advanced augmentation process along with numerical analysis for segmentation
[26]	Dense-Fully Convolutional Network (DFCN), Batch Normalization (BN)	Color and Gray-scale	CASIA-V4	Dense connections are time-consuming, require more training parameters, design for more robust iris segmentation algorithms under non-ideal conditions, and more effective ways of labeling
[27]	Deep-learning classification models	Gray-scale	CASIA-V4	The matching process of computational complexity
[28]	A deep learning framework called SIP-SegNet	Gray-scale	CASIA-V1	This can be expanded upon to create a powerful multi-modal biometric identification system.

Table 1 lists a few well-known techniques for iris segmentation in unrestricted environments along with their limitations. Some methods introduce a novel preprocessing phase for iris segmentation to shorten the search time and lessen the sources of errors. The skin and sclera regions were separated from the iris picture by Sahmoud and Abuhaiba [29] using the K-means algorithm as a preprocessing step. Some other references [23] also used the boundary- and pixel-based approaches.

Among the first suggested and often used techniques that exploit iris boundaries in iris segmentation are Daugman's integro-differential operator [30] and Wildes' method [3]. As a result, research regarding iris region searching is [23] in process. For the reduction of the time searching and minimized the errors are [22]. In [23], the researchers suggested an approach based on a deep multi-task learning architecture named IrisParseNet. To improve the iris identification procedure, it takes advantage of the relationships among the iris, sclera, as well as the pupil.

The researchers developed a deep neural network approach to precisely segment severely damaged iris areas selected from such devices to improve the authentication of the wearable glasses [25]. By combining convolutional neural networks with dense blocks, Chen, Ying, et al. [26] offer an innovative design for segmenting the iris. Their design, DFCN, is where their name originates. Additionally, several well-known optimization techniques like BN and dropout are used to improve the performance.

Table 2. Comparison accuracy of the iris segmentation with other approaches

Reference	Method	Accuracy (%)
[19]	Rough entropy with circular sector analysis	97.12
[27]	Deep-learning	96.90
[28]	SIP-SegNet	95.11
Proposed algorithm	BHA with Canny edge detector and hough transform	98.71

As can be seen from Table 2, different approaches consistently perform operations with good values across all metrices on different datasets for iris segmentation along with its accuracy comparison. Using a technique dubbed segmentation-less polar representation for iris division, reference [27] combined the noise identification and equalization steps alongside the iris identification stage. This approach can prevent the pupil dilatation issue which may happen under unrestricted contexts because it is built on a deep-learning classification system. In recent research [28], the authors have determined to construct algorithms that concurrently segment more than one visual feature.

The scholars have developed an automated system for iris recognition based on 2D iris images to reduce the dimensionality of iris characteristics without losing pertinent information, feature extraction and feature selection techniques have been applied, including principal component analysis and genetic algorithms. Levenberg-Marquardt's learning rule is used to create the back propagation neural network for iris recognition [31]. Utilizing a voting mechanism, the researchers identify and merge the architectures with the Softmax classifier, concentrating on consensus orientation, using convolutional neural networks with consensus between the architectures. Applying optimization strategies will also prevent overfitting and quicken the learning process [32]. Using geometrical information, the Viola-Jones algorithm divides the human eye into parts and intensifies contrast, all the while designating a circular region that holds the iris. Working on an efficient approach based on the Lagrange interpolating polynomial, it obtains non-circular iris outlines. This system performed better and achieved a higher accuracy rate [33]. The attention mechanism, also known as iris segmentation, is intended to be carried out using an end-to-end encoder-decoder model built on enhanced UNet++. To decrease the number of network parameters and increase training time, efficientNetV2 is chosen as a convolutional block of UNet++. UNet++ incorporates an attention module during the down-sampling phase to reduce irrelevant noise interference and improve the network's capacity to recognize the iris region's discriminability. The technique uses a pruning scheme to generate four distinct performance networks that can be used to recognize iris in a variety of conditions. The method's strong iris segmentation and generalization capability is demonstrated by the experimental findings on iris datasets [34]. An end-to-end, unified deep learning system without normalization to increase iris segmentation and recognition accuracy. Iris segmentation and recognition are handled by a dense spatial attention network (DSANet) and a multi-attention dense connection network (MADNet) in the system. Therefore, several ablation tests are carried out to show how successful MADNet and DSANet are. The best segmentation and recognition performance on low-quality iris images without associated GT data is achieved, according to experiments conducted on three databases in use [35].

3. MATERIALS AND METHODS

The following sections explain the theoretical and experimental aspects of the selected algorithm called BHA which is applied to the iris segmentation along the Canny edge detector and circular Hough transform.

3.1. BLACK HOLE ALGORITHM

The concept of the BHA in space was identified in the last two centuries by Michell and Laplace [36].

Since starlight theoretically cannot pass, it was drawn to the BH. Due to the large mass of matter being compressed into a compact area, a strong gravitational field is detected. According to the phenomenon, anything traveling towards the BH is consumed by the BH before dissipating. The event horizon is a spherical shape that forms the BH boundary. The event horizon's radius is determined by the following equation (1), as shown below [36, 37].

$$R = \frac{2GM}{c^2} \tag{1}$$

Where G, M, and c are the gravitational constant, mass, and the speed of light.

The candidates for the BHA are produced and spread at random throughout the search space [21]. The best candidate with the lowest fitness cost from the previous iteration is the one the candidates turn toward next. Equation (2) shows how the stars are moving in the direction of the black hole. Each star's Euclidean distance from the BHA is computed. Any star that is in the event horizon of the BHA recovery is removed from the optimization. New stars are generated at random within the BHA region's perimeter in the following iteration. The fitness values of each new candidate are then assessed once more. The least-suited candidate is selected as the new BHA. The event horizon radius in the optimization is referred to as the Schwarzschild radius. Schwarzschild radius is calculated by (1) in real space, while it is calculated by equation (3) in BHA [21]:

$$Xi\,(t+1) = Xi(t) + rand \times (XBH - Xi(t)),\ i = 1,2,\ldots,N \tag{2}$$

$$R = \frac{f_{BH}}{\sum_{i=1}^{N} f_i} \tag{3}$$

Where $Xi(t)$ and $Xi(t+1)$ signify the locations of the ith star at iterations t and t+1, respectively. rand also denotes a uniform distribution with a 0–1 range. N stands for the number of stars. In the exploration space, X_{BH} identifies the black hole's location. R stands for the event horizon's radius, f_i for the i_{th} star's fitness value, and f_{BH} for the black hole's fitness value.

3.2. CANNY EDGE DETECTOR

By using edge detection, the search space is reduced while maintaining an image's basic structural integrity. One of the common edge detection methods is the Canny edge detection method which has some steps. To reduce noise, the image is smoothed using a Gaussian filter. By using the Pythagorean theorem, the gradient magnitudes of the image are determined with the help of the following equation (4):

$$G = \sqrt{G_x^2 + G_y^2} \tag{4}$$

where G_x and G_y, respectively, represent the corresponding horizontal as well as vertical gradients. The boundaries are then designated in accordance with the gradients' big magnitudes. Edges can be designated as those pixels that are thought to form a portion of an edge through non-maximum suppression. While examination of the appropriate edges is performed via

a thresholding procedure. Besides this, suppressing all edges that are not related to an effective edge to determine the final edges.

3.3. CIRCULAR HOUGH TRANSFORM

This method is used for the searching process of the circles shapes in images this can be found through the equation (5) as stated below:

$$r^2 = (x-a)^2 + (y-b)^2 \quad (5)$$

Where r is considered as the radius of the circle and two points of the circle are x and y whereas the coordinates of the center are a and b. The geometric formulas used for the points of the circle are as follows in equation (6):

$$x = a + r\cos\theta; \text{ and } y = b + r\sin\theta \quad (6)$$

Each point (x, y) on a circle in the original image with a known radius may identify a round centered at (x, y) in the parametric space while the algorithm runs. If all the circles in the parameter space come together at just one location, this location is the exact center of the initial circle (Fig. 3).

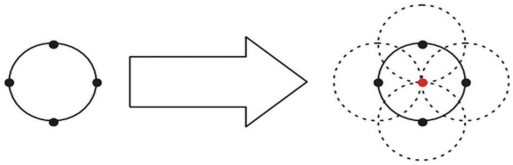


Fig. 3. The principle rule of circular Hough transformation

3.4. IRIS SEGMENTATION

Some sub-level steps are required for this process which is necessary for the iris segmentation.

Step-1: Submission of the paper: Fig. 4 presents the suggested framework of the given conceptual structure. The segmentation model-based BHA optimization is presented using a few procedures along with its activities. As a result, the best fitness value is calculated through the procedure and jumps to the next point which is evaluation.

Step-2: Database: The suggested iris segmentation model is tested using the CASIA-V3 interval database [12]. All the iris photographs were captured using a near-infrared light source and are 8-bit grayscale JPEG files. 2655 photos in all, having a 320 by 280-pixel resolution, were captured from 249 different subjects.

Step-3: Pre-processing: Before testing an image, the following procedures are put into place: The tested image is initially put into the MATLAB workspace. Second, the name of the image that was tested is noted. One of the tested images is called image-1 from the CASIA-V3 database (Fig. 5). The size of the image is examined in the third. The image will become grayscale if it has three layers. The picture matrix is then imported into the MATLAB workspace along with the values of its pixels. The imparted image matrix is now prepared to determine the iris image border.

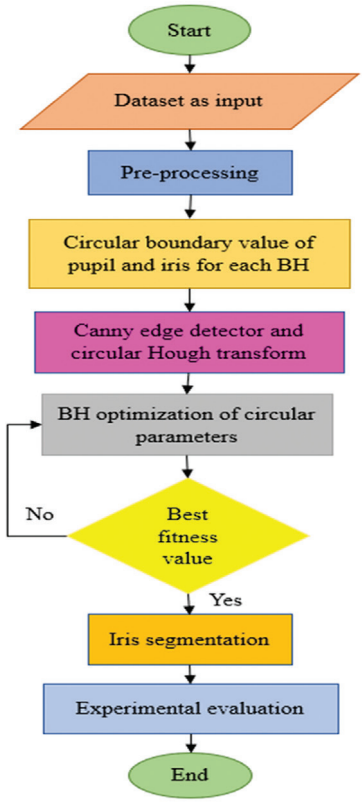


Fig. 4. The framework

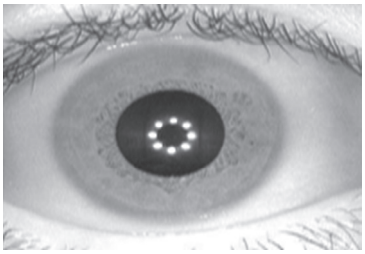


Fig. 5. The input image-1 from the CASIA-V3 for iris segmentation empirical testing

Step-4: Circular boundary detection: The initial problem in this paper is the border identification of the iris image based on pixel values.

As in the vector representations of the iris, which are one-dimensional. In this case, it would seem logical to assume that a boundary pixel should exist between the fifth and sixth pixels.

53	58	55	62	68	123	151	145	141

It would be harder to say that there should be a border in the corresponding region if the intensity difference between the fifth and sixth pixels were larger and the intensity differences between the adjacent neighboring pixels (1st to 4th and 6th to 9th) were smaller. Furthermore, it may be argued that there are numerous edges to this instance. Considered to be a boundary, the fifth pixel can be determined using equation (7).

$$f(x) = \frac{\sum_{i=x+1}^{i=x+n} I(i) - \sum_{i=x-n}^{i=x-n} I(i)}{n} \quad (7)$$

Where $n = 4$ is the fourth-order detection, $I(i)$ is the value of a specific pixel, and $f(x)$ is the boundary value. MATLAB is used to analyze the population of pixels before doing the histogram analysis for the tested image. Fig. 6 shows three populations in three significant regions (pupil, iris, and sclera). Therefore, it is necessary to determine a precise threshold for the size of the intensity difference between two adjacent pixels. Therefore, to determine the boundary of the pupil and iris, two threshold values are required. The multi-level Otsu's technique [38] in MATLAB, which analyses the tested images using the Otsu method, can be used to generate these thresholds. The following results are derived using the multilevel thresholds: the pixel value of the iris region (124 to 165), the pupil (20 to 61), and the sclera area (185 to 219). As a result, considering Equation (7) the pupil boundary's minimal value, $f(x)$ =63 pixels, and the iris's minimum value, $f(x)$, is 20 pixels along a boundary. That establishes the $f(x)$ of BHA. Using limit optimization in MATLAB between 20 and 70. During optimization, any pixel's $f(x)$ = (20 to 70) is thought of as a boundary pixel.

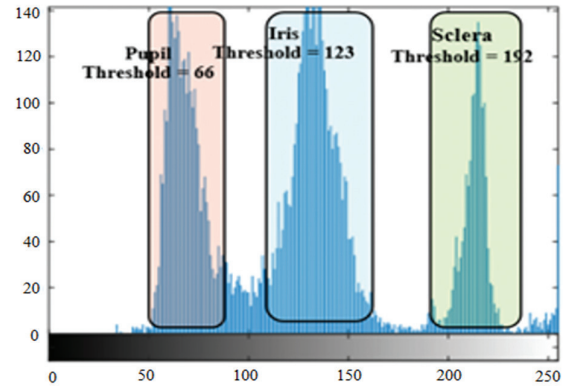


Fig. 6. The numerical values of empirical testing regarding significant regions of the iris

Four times are used to detect border pixels at various angles: 0° and 180°, 45° and 225°, 90° and 270°, and 135° and 315°. That might get worse. The precision of the calculation for various boundary directions (Fig. 7).

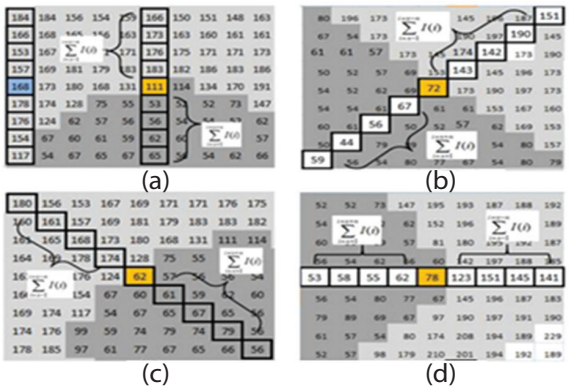


Fig. 7. Recognition of boundary pixels with the help of different angles

Step-5: The BH optimization for pupil boundary: This work uses the BHA search technique to segment the iris. The BHA can quickly and with fewer computations arrive at the ideal parameters. First, the optimization problem's objective function is set to the circle equation (8). The program looks for the pupil circle's three primary parameters, Rp, Xp, and Yp, where Rp is the pupil circle's radius in pixels. Additionally, Xp and Yp show where the circle's center is. Second, the data are analyzed, and the fitness function is determined using the border pixel value equation (7). Thirdly, the BHA generates 100 stars (100 circles) at random for each iteration. Each candidate receives random values for Rp, Xp, and Yp (Fig. 8). During optimization, the image's size is considered to ensure that all 100 circles are contained within the image.

$$x = r\,cos(\theta) + cx; \quad and\ y = r\,sin(\theta) + cy \quad (8)$$

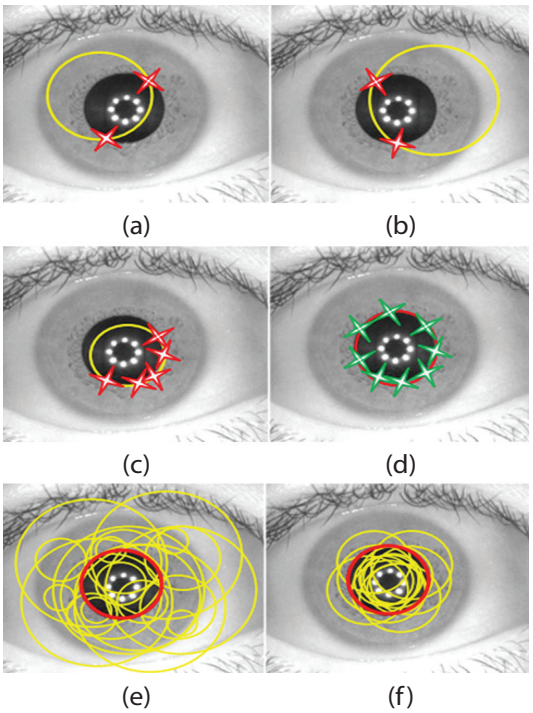


Fig. 8. The BHA operation, (a) star number 7, (b) star number 31, (c) star number 57, (d) star number 73 which is the BH of the first iteration, (e) first iteration the BHA generates randomly 100 stars (100 circles), and (f) second iteration generates 100 circles close to the pupil

To demonstrate (8), the star parameters use a 360-degree step-size circle. Fourth, a black mask zeros image (Istar) with the same dimensions as the input image (320 x 280) is used to plot this circle. In the original image, an identical circle is formed, but it doesn't touch the image borders. The equation is used to calculate the pixels at the circle's edge (7). When the pixel is located at the border between the iris and the sclera or between the iris and the pupil, the equation (7)'s maximum value might be reached. Fifth, the pixel's computed value that is higher than the thresholds is tallied and recoded using equation (7). For various

angles surrounding this pixel, this calculation is used. The same procedure is applied to the additional pixels that are situated at the circle's edge. For this star, the total number of recoded pixels (the circular border is true) is calculated. Sixth, this candidate's outcome just displays the recoded pixels. The fitness value of the first star (first circle) for unrecorded pixels is calculated using equation (9).

$$f_{star} = 1 - \frac{P_{recorded}}{P_{star}} \quad (9)$$

Where $r = Rp$, θ is the angle, $cx = Xp$, $cy = Yp$, f_{star} is the fitness value of the star, $P_{recoded}$ is the total number of recoded pixels, and $P_{star} = 360$ is the number of star's pixels. For instance, Fig. 8(a) shows the star number 7 produced by the first BHA iteration. On the tested iris image, the seventh star in Fig. 8(a) is shown as a yellow circle. Only at two points does the pupil's border cross that of the yellow circle. Considering this, equation (9)'s calculation yields the high fitness value $f_{7\ star}$ = 1-(47/360) = 0.87. Furthermore, as shown in Fig. 8, the computation of star number 31 yields a fitness value of $f_{31\ star}$ = 1- (61/360) = 0.83 Fig. 8(b). The star number 57's fitness value is calculated as $f_{57\ star}$ = 1- (143/360) = 0.60. Fig. 8(c). Additionally, the star number 73's computation yields a fitness value of $f_{73\ star}$ = 1-(257/360) = 0.28. Fig. 8(d). According to the framework proposed in Fig. 4, the created circle of star 73 observes the lowest cost fitness. As a result, during the second iteration, star No. 73 is regarded as the pupil's black hole. A further 100 stars are produced in the second iteration using equation (3) BHA. These new stars, however, have circular characteristics that are comparable to those of star 73 from the initial iteration. According to the Equation, the stars' parameters have horizon radius conditions (3). All the newly suggested circles will be placed near the student boundary as a result. Consequently, in just 2 seconds, the best circle representing the pupil boundary is found.

The fitness values of all the stars as well as the remaining 99 stars are subject to this process. Seventh, the BHA updates the location and the black hole candidate for the following iteration by evaluating the fitness values for all the stars. Eighth, in the following iteration, the star whose fitness value is lower than the existing black hole's fitness becomes the new black hole. The fresh random 100 stars migrate in the following iteration toward the new black hole. The BHA generates 100 more circles with parameters that are very similar to those of the new BH and they have the best fitness from the previous iteration. The eight procedures that were used in the initial iteration are then carried out once more. The second iteration's top fitness star advances to the third iteration's BH position, and so forth. The third iteration updates both the BH's location and the stars' proximity to the ideal iris boundary. Finally, this technique can quickly determine the ideal parameters without calculating all the Rp, Xp, and Yp probabilities. The pupil boundary's recognized circle (Fig. 9).

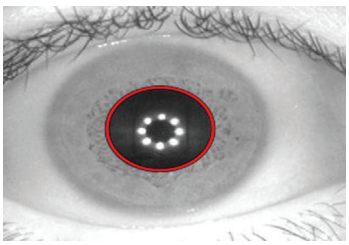


Fig. 9. The boundary of the pupil along its identification

Step-6: The BH optimization for iris boundary: For each star, the BHA can recommend a set of random circle parameters. The top fitness performer from the initial iteration devolves into a black hole for the subsequent iteration. The fitness value of the goal function is gradually increased. The ideal circle that shows the iris border is given once the stop criteria have been met (Fig. 10). The BHA accurately determines the iris boundary's center and radius.

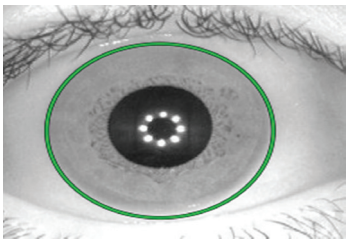


Fig. 10. The boundary of the iris along its identification

Step-7: Iris mask: There are two ways to obtain a logical array mask. First, it is used solely to extract the segmented iris region. Second, it is employed to assess the segmented system's correctness. A 320 x 280 zeros array is created to produce the mask image (I-mask). The iris region is then given one logical value by plotting a filled circle on the I-mask after that. The pupil region is then given 0 logical values by plotting a filled circle on the I-mask. By applying element-wise multiplication, the proposed mask of the segmented iris image (I-mask) is multiplied with the input image (image-1) (Fig. 11), where only the segmented portion of the iris is visible (segmented).

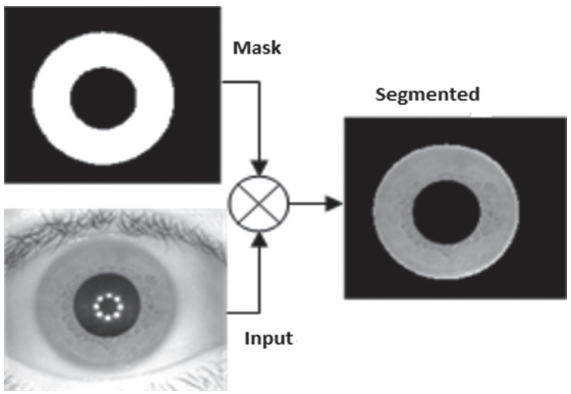


Fig. 11. The process of masking as image-1 and segmenting as results

4. EXPERIMENTAL RESULTS

In this experiment, the accessible iris database is utilized. The chosen segmentation method is tested using the CASIA-V3 database. This database was chosen because it has 2639 images, whereas most of the images in CASIA-V3 had less-than-ideal iris conditions (partial occlusion). Some of the images that were segmented using the BHA are shown in Fig. 12. The outcome shows that the limits of the pupil and the iris are accurately identified. Even though Fig. 12 depicts images with various pupil or iris occlusions and radiuses, the suggested technique is successful in segmenting the iris region.

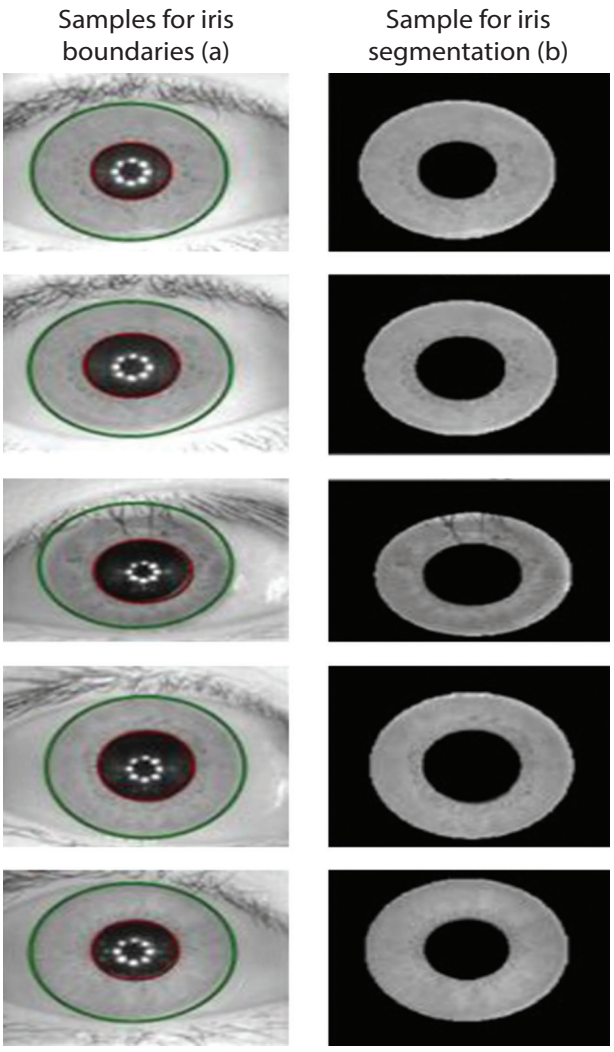


Fig. 12. CASIA-V3 Dataset samples for the detection of the iris edges and segmentation.

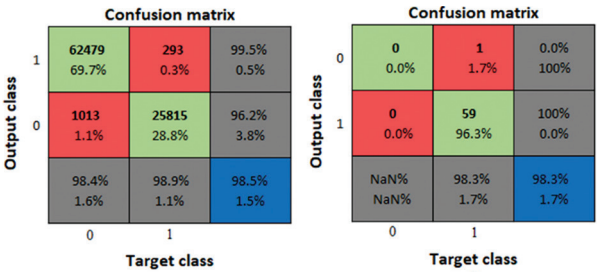
Additionally, the suggested algorithm's average accuracy is assessed using the model in [19]. The correctness of each extracted mask (I-mask) is tested using the Images Ground Truth (IGT) of the CASIA-V3 in [39]. The I-mask XOR IGT operation is used to generate mismatched localized pixels (mismatch). Utilizing the I-mask and IGT's confusion matrix function, the performance matrix is computed. According to equation (10), the segmented image's accuracy is expressed as the proportion of correctly predicted pixels, where TN, TP, FN, and FP stand for true-negative, true-positive, false-negative, and false-positive, respectively. The accuracy of the segmented image-1 using BHA is estimated to be 98.71% (Fig. 13 (a)).

$$Accuracy=(TP+TN)/(TP+TN+FN+FP) \quad (10)$$

The average accuracy of all the images is calculated using mismatched images. If image segmentation has a segmented image that, when utilizing (10), has an accuracy lower than 90%. (FN). However, any image segmentation utilizing (10) that achieves an accuracy of more than 90% is correct (TP). The average accuracy of the BHA is tested in this experiment by randomly choosing many images of various subjects. Fig. 13 (b) depicts the confusion matrix of the tested images. The computed average accuracy of all examined images is 8.

$$Average\ accuracy= (Number\ of\ correct\ segmented\ images)/ (Total\ number\ of\ images)\ x\ 100\% \quad (11)$$

The average accuracy of the suggested BHA segmentation is 98.71 % according to Equation 11. On average, the suggested BH method has higher segmentation accuracy than [19].



(a) (b)

Fig. 13. The confusion matrix along all its parameters percentage

5. CONCLUSION

There are many issues that are frequently formulated as optimization problems, and low-performance optimization techniques suffer from weak or slow convergence to the best solutions for high-performance optimization methods that produce high-quality solutions while consuming the fewest steps. Due to its dependability and practically flawless detection rates, iris recognition has gained popularity. The three primary phases of an iris recognition system are image preprocessing, feature extraction, and template matching. The success of the next feature extraction and template matching stages depends on the iris segmentation stage of the pre-processing procedure. In this study, the CASIA-V3 database has been segmented accurately by using the BHA along the Canny edge detector and circular Hough transform which provides the most appropriate improvement in accuracy. The pupil and iris are segmented using optimization edge boundaries. When optimizing, the complete boundary of the image was deemed satisfactory because of the high segmentation

accuracy of 98.71% obtained by utilizing the combination of the algorithm with the selected methods and the value of pixels. The suggested technique removed the noise from both the eyelid and eyelash consideration in our future study and only split the circular borders. Any iris biometric security system can be made more effective by using the suggested algorithm.

A second way to enhance the algorithm's convergence is to change the randomization parameter so that it gradually drops as the optimum approach. For future work, to improve performance evaluation according to the iris recognition system in non-ideal conditions. It will look for ways to increase accuracy as well as speed so that it can operate in real-time. Future research should focus on expanding this algorithm to color images and evaluating their accuracy for human recognition. Future work should also involve utilizing a minimum of two iris databases to ensure comprehensive experimentation and robust validation of the results.

6. REFERENCES:

[1] A. K. Jain, A. Ross, S. Prabhaker, "An Introduction to Biometric Recognition", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 1, 2004, pp. 4-20.

[2] J. Daugman, "How Iris Recognition Works", The essential guide to image processing, Academic Press, 2009, pp. 715-739.

[3] R. P. Wildes, "Iris Recognition: An Emerging Biometric Technology", Proceedings of the IEEE, Vol. 85, No. 9, 1997, pp. 1348-1365.

[4] K. W. Bowyer, K. Hollingsworth, P. J. Flynn, "Image Understanding for Iris Biometrics: A Survey", Computer Vision and Image Understanding, Vol. 110, No. 2, 2008, pp. 281-307.

[5] C. Tisse, L. Martin, L. Torres, M. Robert, "Person Identification Technique Using Human Iris Recognition", Proceedings of the Vision Interface, 27 May 2002, pp. 294-299.

[6] M. Pathak, N. Srinivasu, V. Bairagi, "Effective segmentation of sclera, iris and pupil in noisy eye images", Telkomnika (Telecommunication Computing Electronics and Control), Vol. 17, No. 5, 2019, pp. 2346-2354.

[7] H. D. Rafik, S. A. Mahmoudi, A. Reda, M. Boubaker, "A Model of A Biometric Recognition System Based On The Hough Transform Of Libor Masek and 1-D Log-Gabor Filter", Proceedings of the 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications, Marrakesh, Morocco, 24-26 November 2020, pp. 1-9.

[8] M. Dua, R. Gupta, M. Khari, R. G. Crespo, "Biometric iris recognition using radial basis function neural network", Soft Computing, Vol. 23, No. 22, 2019, pp. 11801-11815.

[9] H. Purohit, P. K. Ajmera, "Optimal feature level fusion for secured human authentication in multimodal biometric system", Machine Vision and Applications, Vol. 32, No. 24, 2021, pp. 1-12.

[10] N. Ahmadi, G. Akbarizadeh, "Hybrid robust iris recognition approach using iris image pre-processing, two-dimensional gabor features and multi-layer perceptron neural network/PSO", IET Biometrics, Vol. 7, No. 2, 2018, pp. 153-162.

[11] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, No. 11, 1993, pp.1148-1161.

[12] T. Tan, "Note on CASIA-IrisV3", Chinese Academy of Sciences, 2011.

[13] M. Choudhary, V. Tiwari, U. Venkanna, "Enhancing human iris recognition performance in unconstrained environment using ensemble of convolutional and residual deep neural network models", Soft Computing, Vol. 24, No. 15, 2020, pp. 11477-11491.

[14] H. G. Daway, H. H. Kareem, A. R. Hashim, "Pupil detection based on color difference and circular hough transform", International Journal of Electrical and Computer Engineering, Vol. 8, No. 5, 2018, pp. 3278-3284.

[15] B. Zhang, J. Fei, "An Iris Location Algorithm Based on Gray Projection and Hough Transform", Proceedings of the International Conference in Communications, Signal Processing, and Systems, Springer, Singapore, 14 August 2019, pp. 1323-1330.

[16] B. Deshpande, D. Jayaswal, "Fast and Reliable Biometric Verification System Using Iris", Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies, 20 April 2018, pp. 456-460.

[17] F. Jan, "Non-circular iris contours localization in the visible wavelength eye images", Computers & Electrical Engineering, Vol. 62, 2017, pp. 166-177.

[18] J. A. Ridha, J. H. Saud, "Iris segmentation approach based on adaptive threshold value and circular hough transform", Proceedings of the International Conference on Computer Science and Software Engineering, Duhok, Iraq, 16-18 April 2020, pp. 32-37.

[19] M. Sardar, S. Mitra, B. U. Shankar, "Iris localization using rough entropy and CSA: A soft computing approach", Applied Soft Computing, Vol. 67, 2018, pp. 61-69.

[20] Q. W. Chai, J. W. Zheng, "Rotated black hole: a new heuristic optimization for reducing localization error of WSN in 3D terrain", Wireless Communications and Mobile Computing, 2021, pp. 1-13.

[21] A. Hatamlou, "Black hole: A new heuristic optimization approach for data clustering", Information Sciences, Vol. 222, 2013, pp. 175-184.

[22] R. Chen, X. R. Lin, T. H. Ding, "Iris segmentation for non-cooperative recognition systems", IET image processing, Vol. 5, No. 5, 2011, pp. 448-456.

[23] C. W. Tan, A. Kumar, „Unified framework for automated iris segmentation using distantly acquired face images", IEEE Transactions on Image Processing, Vol. 21, No. 9, 2012, pp. 4068-4079.

[24] C. Wang, Y. Zhu, Y. Liu, R. He, Z. Sun, „Joint iris segmentation and localization using deep multi-task learning framework", arXiv:1901.11195, 2019.

[25] V. Varkarakis, S. Bazrafkan, P. Corcoran, "A deep learning approach to segmentation of distorted iris regions in head-mounted displays", Proceedings of the IEEE Games, Entertainment, Media Conference, Galway, Ireland, 15-17 August 2018, pp. 1-9.

[26] Y. Chen, W. Wang, Z. Zeng, Y. Wang, "An adaptive CNNs technology for robust iris segmentation", IEEE Access, Vol. 7, 2019, pp. 64517-64532.

[27] H. Proença, J. C. Neves, "Segmentation-less and non-holistic deep-learning frameworks for iris recognition", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, California, 16-20 June 2019.

[28] B. Hassan, R. Ahmed, T. Hassan, N. Werghi, "Sip-segnet: A deep convolutional encoder-decoder network for joint semantic segmentation and extraction of sclera, iris and pupil based on periocular region suppression", arXiv:2003.00825, 2020.

[29] S. A. Sahmoud, "Enhancing iris recognition", Islamic University, Department of Computer Engineering, Deanery of Higher Studies, Gaza, Palestine, 2011, Master Thesis.

[30] J. Daugman, "New methods in iris recognition", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 37, No. 5, 2007, pp. 1167-1175.

[31] M. Garg, A. Arora, S. Gupta, "An efficient human identification through iris recognition system", Journal of Signal Processing Systems, Vol. 93, 2021, pp. 701-708.

[32] D. R. Hammou, S. A. Mahmoudi, R. Adjoudj, "Multi-Biometric Iris Recognition System Using Consensus Between Convolutional Neural Network Architectures", International Journal of Organizational and Collective Intelligence, Vol. 12, No. 1, 2022, pp. 1-30.

[33] F. Jan, S. Alrashed, N. Min-Allah, "Iris segmentation for non-ideal Iris biometric systems", Multimedia Tools and Applications, 2021, pp. 1-29.

[34] G. Huo, D. Lin, M. Yuan, "Iris segmentation method based on improved UNet++", Multimedia Tools and Applications, Vol. 81, No. 28, 2022, pp. 41249-41269.

[35] Y. Chen, H. Gan, H. Chen, Y. Zeng, L. Xu, A. A. Heidari, X. Zhu, Y. Liu, "Accurate iris segmentation and recognition using an end-to-end unified framework based on MADNet and DSANet", Neurocomputing, Vol. 517, 2023, pp. 264-278.

[36] R. Giacconi, "Black hole research past and future", In Black Holes in Binaries and Galactic Nuclei: Diagnostics, Demography and Formation: Proceedings of the ESO Workshop, Springer, Garching, Berlin, Germany, 6-8 September 1999, pp. 3-15.

[37] C. A. Pickover, "Black holes: A traveler's guide", 1st Editon, Wiley-VCH, 1997.

[38] S. C. Satapathy, N. S. M. Raja, V. Rajinikanth, A. S. Ashour, N. Dey, "Multi-level image thresholding using Otsu and chaotic bat algorithm", Neural Computing and Applications, Vol. 29, 2018, pp. 1285-1307.

[39] H. Hofbauer, F. Alonso-Fernandez, P. Wild, J. Bigun, A. Uhl, "A ground truth for iris segmentation", Proceedings of the 22nd international conference on pattern recognition, Stockholm, Sweden, 24-28 August 2014, pp. 527-532.

A Survey of Sentiment Analysis and Sarcasm Detection: Challenges, Techniques, and Trends

Review Paper

Ahmed Derbala Yacoub
Helwan University Faculty of Computers and Artificial Intelligence
Cairo EGYPT
ahmed_drbala_1073@fci.helwan.edu.eg

Salwa O. Slim
Helwan University Faculty of Computers and Artificial Intelligence
Cairo EGYPT
salwaosama@fci.helwan.edu.eg

Amal Elsayed Aboutabl
Helwan University Faculty of Computers and Artificial Intelligence
Cairo EGYPT
amal.aboutabl@fci.helwan.edu.eg

Abstract *– In recent years, more people have been using the internet and social media to express their opinions on various subjects, such as institutions, services, or specific ideas. This increase highlights the importance of developing automated tools for accurate sentiment analysis. Moreover, addressing sarcasm in text is crucial, as it can significantly impact the efficacy of sentiment analysis models. This paper aims to provide a comprehensive overview of the conducted research on sentiment analysis and sarcasm detection, focusing on the time from 2018 to 2023. It explores the challenges faced and the methods used to address them. It conducts a comparison of these methods. It also aims to identify emerging trends that will likely influence the future of sentiment analysis and sarcasm detection, ensuring their continued effectiveness. This paper enhances the existing knowledge by offering a comprehensive analysis of 40 research works, evaluating performance, addressing multilingual challenges, and highlighting future trends in sarcasm detection and sentiment analysis. It is a valuable resource for researchers and experts interested in the field, facilitating further advancements in sentiment analysis techniques and applications. It categorizes sentiment analysis methods into ML, lexical, and hybrid approaches, highlighting deep learning, especially Recurrent Neural Networks (RNNs), for effective textual classification with labeled or unlabeled data.*

Keywords: *Sarcasm detection, sentiment analysis, natural language processing, Deep learning, machine learning*

1. INTRODUCTION

The paper's objective is to thoroughly study all the related research to understanding people's feelings and identifying sarcasm, whether accomplished using a single language or multiple languages. We'll also compare these studies based on what they aim to solve, what methods they use, what the dataset used, how they test their results, who benefits from their work, what the outcomes are, and what's good and not so good about their approaches. We'll also investigate the problems in sentiment analysis and sarcasm detection, and we'll explore the ways researchers have come up with to tackle these problems. We'll attempt to identify which solutions work best and what's coming next in sentiment analysis and sarcasm detection. Sentiment analysis [1, 2], crucial for understanding public opinions on products and events, has become vital with the internet's growth, but analyzing vast online data poses challenges. The goal [3-11] is to assess sentiment, often on a positive-negative spectrum for written reviews. Sentiment analysis, situated at the intersections of computational linguistics, NLP, and data mining, utilizes techniques from these fields to extract sentiments from text. It explores research areas such as data mining and machine learning within the context of NLP [12,13]. Sarcasm [14-18], a statement conveying the opposite of its intended meaning and often used humorously or critically, poses a challenge for sentiment analysis

as it can invert the true sentiment of a statement. Despite being a popular research topic, sarcasm detection [19-26] is crucial as sentiment analysis can misinterpret sarcastic sentences, leading to inaccurate sentiment classifications. The difficulty lies in the nuanced nature of human emotions and expressions conveyed through text, making automatic sarcasm detection a challenging task within natural language processing (NLP) [27, 28]. In sentiment analysis, written opinions often mix several languages, making it difficult to fully capture the text messages, resulting in a more complex classification of the text's polarity. Sentiment analysis struggles with multilingual comments that mix different languages [29, 30]. In tackling this issue, a sentiment analysis tool is essential for sorting reviews in the intended language and discarding others. However, this might make data smaller and influence the accuracy of linear classifiers. Identifying the source language is essential for successful cross-lingual sentiment analysis, as it helps to use suitable resources for sentiment analysis in that language [31].

Some challenges in sentiment analysis and sarcasm detection are:

- Determining sentiment in a sentence becomes difficult with the presence of sarcasm, making it challenging to discern the intended meaning. Detecting sarcasm in text poses a challenge due to context absence, user personality, the lack of expressions and body language, diverse writing styles, noise in text, and limitations in accurately classifying input data for sentiment analysis.
- Classifying people's sentiments is hard when they use different languages, making it complex to decide if the message is positive or negative. Most sentiment analysis focuses on English, but as more people use their languages on social media like Facebook and Twitter, there's a need to handle reviews in different languages.
- Sentiment analysis becomes complicated in cases where contextual information and explicit sentiment words are absent in an implicit text and when individuals express their emotions in an obscure and implicit manner.
- While current word vectors primarily focus on word meaning in sentiment analysis, it's crucial to incorporate sentiment aspects in embeddings because words with different sentiments, like "I'm sad" and "I'm happy," can end up with similar vectors due to shared contexts.

For Methods, the study of sentiment analysis and sarcasm detection categorizes methods into three types: machine learning (ML), lexical-based, and hybrid approaches. Deep learning, especially recurrent neural networks (RNNs), is valuable for classifying textual information due to its flexibility and complex structure. Machine learning involves training a computer with algorithms to detect patterns in large datasets, with deep learning (DL) being a component that uses algorithms for advanced data abstraction. In ML, supervised learning relies on labeled training data to predict outcomes accurately, while unsupervised learning discovers patterns without labeled data. Popular ML techniques include Naive Bayesian, Support Vector Machine, Artificial Neural Networks, and the K-means algorithm.

The main contributions of this paper are:

- The survey paper compares 40 relevant studies on sentiment classification and detection of sarcasm, revealing significant variations in model performance and identifying gaps in the literature. It is a valuable resource, offering insights to guide future research and development in these areas.
- The paper evaluates the performance of diverse methods and algorithms in sarcasm detection and sentiment classification, comparing their accuracy and highlighting strengths and limitations to assist researchers in choosing suitable techniques for specific applications.
- The paper addresses the challenges of multilingual comments in sentiment analysis. It emphasizes the need for a sentiment analysis tool capable of processing multilingual and mixed-code datasets.
- The paper identifies future trends and expectations in sarcasm detection and sentiment classification. It discusses the potential of advanced deep learning techniques, such as transformers and attention mechanisms.

2. LITERATURE REVIEW

In the world of sentiment analysis and sarcasm detection for written communications, a growing area of study driven by the increasing need for a better understanding of textual data is appearing. This review thoroughly examines previous research in sentiment analysis and sarcasm detection. It covers studies focused on specific tasks, including Sentiment Analysis, Sarcasm Detection, Joint Sarcasm Detection and Sentiment Analysis, Multilingual Sentiment Analysis, and Multilingual Sarcasm Detection.

2.1. SENTIMENT ANALYSIS

Sharmin et al. [1] proposed an attention-based convolutional neural network (CNN) approach for sentiment analysis in the Bengali language. The authors use a dataset of Bengali movie reviews to train and evaluate their model. The literature supports the use of deep learning models, especially CNNs, for sentiment analysis tasks, and adding attention mechanisms to CNNs has been shown to improve their performance.

De Kok et al. [2] proposed a novel model for aspect-based sentiment analysis (ABSA) that analyzes sentiments or feelings related to attributes or characteristics of a product or service. The model utilizes ontology features to enhance the accuracy of ABSA models.

The authors use a dataset of hotel reviews to train and evaluate their model.

Elfaik et al. [3] proposed a deep learning approach for sentiment analysis of Arabic text using a deep bi-directional LSTM network with pre-trained word embedding. The authors utilize a dataset of Arabic tweets to train, assess, and compare their model with other state-of-the-art models.

Sarkar et al. [4] proposed a deep learning-based approach for sentiment analysis of Bengali tweets using the SAIL 2015 Bengali dataset, which contains 9,108 Bengali tweets labeled with positive, negative, and neutral sentiments. They employ a Bidirectional Long Short-Term Memory (BILSTM) network with an attention mechanism and incorporate external knowledge from SentiWordNet for improved performance. The proposed approach achieves an accuracy of 55%.

Chen et al. [5] proposed a new pre-processing method for implicit sentiment text classification that converts text features into picture features using FER-Net and TTP. The proposed method achieved an accuracy of 85% on the SMP2019 dataset. The study also introduces a new research direction in implicit sentiment analysis work by combining text classification and image analysis. Nonetheless, the proposed method has limitations in dealing with more detailed implicit sentiment text categorization tasks.

Zhao et al. [6] presented a new approach to sentiment analysis on social media platforms, which involves capturing the implicit text features hidden in user comments and the explicit sentiment network formed by the topological relationships of users to entities. The proposed approach comprises two primary components: a text representation module based on word graphs and a network embedding module for a heterogeneous social graph. Low-dimensional potential representations of network nodes are obtained by operating network embedding techniques, and non-consecutive and long-distance semantics are captured by adding feature learning technology. The original social network is divided into a user entity sentiment network and a user relationship structure network and acquires their node embedding representations individually using two separate autoencoders.

Huang et al. [7] proposed a novel model for sentiment analysis in Vietnamese by leveraging sentiment word embedding alongside transfer learning. The model utilizes a BiLSTM with an attention mechanism and generates sentiment vectors using Word2Vec. The model uses the Vietnamese Students' Feedback Corpus as the training dataset. The authors also demonstrate the effectiveness of using English corpus for transfer learning to improve the sentiment classification of Vietnamese. The proposed sentiment word vector can represent both the context and the sentiment information of words, which enhances the model performance. The accuracy achieved by the proposed model is 86%, which shows its potential in a successful sentiment analysis for the Vietnamese language.

Hussein et al. [8] proposed a method for Arabic sentiment analysis of multi-dialect text that utilizes machine learning techniques. The authors collected a dataset of tweets from different Arabic dialects and pre-processed it by removing stop words, non-alphanumeric characters, and URLs. They then used a feature extraction technique called Term Frequency-Inverse Document Frequency (TF-IDF) to convert the tweets into numerical vectors. The proposed model utilized three machine learning classifiers for sentiment analysis: Naive Bayes, Decision Trees, and Support Vector Machines (SVM). The experiments were conducted using a 10-fold cross-validation technique, and the results show that SVM achieved the highest accuracy of 87.2%, followed by Logistic Regression with 0.85 and Stochastic Gradient Descent with 85%. The authors also conducted experiments with different feature selection techniques and found that the Chi-square feature selection method provided the best results.

Ma et al. [9] proposed a novel model called Attentive LSTM with Common Sense Knowledge (ALCSK), which utilizes a pre-trained commonsense knowledge graph to enrich the aspect embedding layer and to enhance the accuracy of Targeted Aspect-Based Sentiment Analysis (TABSA) methods. The proposed model outperforms existing state-of-the-art methods on benchmark datasets, demonstrating its effectiveness in handling domain-specific knowledge and incorporating commonsense data.

Londhe et al. [10] propose a system that combines Long Short-Term Memory (LSTM), fuzzy logic, and incremental learning for sentiment analysis in restaurant reviews. The system effectively handles multi-category classification and captures cross-category correlations by incorporating binary classifiers and an ensemble approach. Adding fuzzy-encoded LSTM enhances model performance, achieving an accuracy of 81.04% and outperforming previous methods.

Li et al. [11] proposed an enhanced label propagation algorithm (LPA) referred to as the label attenuation propagation model (LAPM) for the automated assessment of emoji sentiment. The paper addresses the need for quantifying emoji sentiment and introduces the use of LPA in the emoji sentiment analysis field. LAPM is employed to compute sentiment uncertainty and emojis sentiment based on the Emoji Link Network (ELN) built for effectively managing and categorizing many social media emojis. The accuracy achieved by the proposed model is 85%, which is higher than the 74% accuracy achieved by LPA.

Hiyama et al. [12] proposed a neural network-based sentiment analysis model that uses an attention mechanism. The model consists of four layers: word-embedding, bidirectional LSTM, attention, and classification layers. The embedding layer converts the input into a fixed-size continuous vector. The bidirectional

LSTM layer produces contextual word vectors, while the attention layer assesses the significance of these contextual word vectors and computes a sentence vector. Finally, the classification layer utilizes the sentence vector to predict the sentiment polarity. The authors provide details on each layer and how they are used to enhance sentiment analysis results.

Liu et al. [13] proposed an algorithmic framework for cross-domain aspect-level sentiment analysis using the Amazon product review dataset, which comprises reviews for products across five domains: Books, DVD disks, Electronics, Kitchen appliances, and Videos. The proposed algorithmic framework utilizes the convolutional, adversarial, and BERT models. The authors Altered the CNN architecture by introducing a gated activation unit to improve its performance. The proposed model achieved an accuracy of 88.9%.

2.2. SARCASM DETECTION

Bagate et al. [14] proposed a model for detecting sarcasm in tweets without relying on specific clues such as hashtags. The dataset used for this research was collected from Kaggle and consisted of 51,188 tweets. The proposed model uses machine learning classification methodology and deep learning embedding techniques, which include data pre-processing steps such as tokenization, lower casing, stop word removal, and stemming. The model utilizes a stacking way that combines the outcomes of logistic regression and a Long Short-Term Memory (LSTM) recurrent neural network, feeding them into a light gradient boosting technique. The evaluation metrics used for the model are F1-score and confusion matrix. The proposed model achieves better results in sarcasm detection without the presence of hashtag 'sarcasm' compared to existing methods.

Zhang et al. [15] developed a method to enhance the performance of sarcasm detection with limited labeled data by applying adversarial training. The authors employed different neural network architectures, like Convolutional Neural Networks (CNN) and Hierarchical Recurrent Neural Networks (HRNN) and incorporated self-attention mechanisms into these models. For evaluation, the authors conducted evaluations using six different models on three datasets from the Internet Argument Corpus (IAC), Rhetorical Questions, and Hyperbole. The findings demonstrated that their adversarial training technique significantly improved sarcasm detection performance.

Zhang et al. [16] proposed a novel deep learning-based model for detecting sarcasm in Chinese text using contextual information from social networks. The model includes a Pair of concurrent modules, sketchy reading, and Intensive Reading, which use an attention mechanism and a hierarchical method to extract features from the input. The output from the two modules is then concatenated and passed through a fully connected layer to make the final prediction.

Tanya et al. [17] proposed a hybrid approach combining deep learning and traditional machine learning methods for detecting sarcasm in social media, using the News Headlines Dataset. The authors pre-process the dataset by removing usernames, stop words, and non-alphanumeric characters. To classify sarcasm patterns, authors used supervised machine learning techniques, hybrid neural network architectures, ensemble hybrid approaches, and standard models implementing word embedding. The results show that the deep learning model, specifically the RNN with LSTM, demonstrates the highest level of accuracy of 96%, followed by The RNN utilizing GRU with an accuracy of 95%, and the CNN-RNN hybrid model with an accuracy of 90%. These accuracies are much higher than those achieved by other supervised models.

Sharma et al. [18] proposed a hybrid method for sarcasm detection in social media. The model incorporates sentence embedding generation from Auto-encoder, bidirectional encoder representation transformer (BERT), and the universal sentence encoder (USE), aiming to accommodate a wide range of content types. It undergoes verification using three real-world social media datasets, achieving higher accuracy than previous state-of-the-art frameworks. The proposed model employs pre-processing, pre-training, and classification phases, utilizing three sentence-based techniques to provide the final categorization of sarcastic or non-sarcastic comments.

Kavitha et al. [19] developed a novel deep learning-based sarcasm detection classification model (DLE-SDC) that involves a pre-processing stage, Glove-based word vector representation, a combination of convolutional neural network and recurrent neural network (CNN-RNN) based classification, and an algorithm for teaching and learning based optimization (TLBO) for hyperparameter tuning. The model is validated using a benchmark dataset and evaluated based on precision, recall, accuracy, and F1 score.

Sengar et al. [20] introduce a novel method for sarcasm detection in text. It employs feature engineering techniques based on contrasts within sarcastic sentences, focusing on positive phrases followed by negative situations or vice versa in tweets. The dataset, sourced from Twitter, includes 151,900 sarcastic and 330,692 non-sarcastic tweets. The approach involves pre-processing, contrast sentiment feature extraction, pattern recognition, and N-grams feature extraction. A neural network with ReLU activation outperforms traditional machine learning models in accuracy and f1-score for sarcasm detection.

Rao et al. [21] developed a sarcasm detection model using Twitter data. The dataset used for the experiment is the Twitter headlines dataset.It contains around 30k tweets. The data is processed using various Python libraries like Word2Vec, NLTK, SciPy, TextBlob, Plotly, Matplotlib, and NumPy. The pre-processing steps involve data cleaning, tokenization, stemming, lemmatization,

and feature extraction using word embeddings. The sarcasm detector utilizes a long short-term memory (LSTM) network that ensures careful monitoring of the input data dependencies. The model can achieve an accuracy rate as high as 92% depending on the dataset and the number of epochs employed. The model classifies input data based on a prediction variable, represented as a two-dimensional array. A front-end engine was created for the RNN model, capable of detecting sarcasm in grammatically correct input and notifying users of incorrect input and spelling errors.

Bhardwaj et al. [22] proposed a deep learning model that uses BERT to pre-process the input sentence for sarcasm detection. The authors report an accuracy of 99% on the news headline dataset, which is a valuable improvement over the previous approach.

Ghosh et al. [23] focused on the class imbalance problem in sarcasm detection on Twitter. The authors extracted Indian tweets and manually classified them as sarcastic or non-sarcastic. They used N-gram features for tokenization and applied the Synthetic Minority Oversampling Technique (SMOTE) to augment the minority class data points. They used two classifiers, BNB and LRC, to Assess the performance of each and make a comparison before and after SMOTE. The results showed that SMOTE was effective when dealing with the imbalance in class issues and enhancing the classifier performance. BNB reached maximum accuracy with a low-level oversampling rate, while LRC achieved the same precision with a 50% oversampling rate.

Mohan et al. [24] address the challenge of detecting sarcasm in textual data, particularly in the context of customer feedback and social media. The paper proposes a hybrid model that combines Bidirectional Encoder Representations from Transformers (BERT) and Graph Convolutional Networks (GCN) for accurately detecting sarcasm. The model utilizes affective and adjacency graphs, BERT embeddings, and GCN to capture advanced structural and semantic patterns within sarcastic content. Experimental results showcase the outstanding performance of the proposed approach compared to baseline approaches, achieving high accuracy and f1 scores on publicly available sarcasm datasets.

Sharma et al. [25] introduce a hybrid ensemble approach designed for detecting sarcasm on social networking platforms. The model combines word-based and sentence-based models such as GloVe, Word2Vec, and BERT embeddings with a fuzzy logic module to overcome the limitations of individual methods and improve sarcasm detection accuracy. The study evaluates the model using real-world datasets from Twitter and Reddit, demonstrating its applicability and potential in practical and business contexts.

Kumar et al. [26] propose a hybrid deep-learning approach for detecting sarcasm in Hindi text and addressing challenges in languages with limited resources. The model utilized word-emojis embedding and demonstrated high accuracy and F-score in detecting sarcasm in Hindi tweets. The paper discusses tweet pre-processing, word embeddings for semantic relationships, and the representation of emojis as vectors (Emoji2vec). The hybrid model combines convolutional neural networks (CNNs) and long short-term memory (LSTM) networks to capture local and long-term features in the text.

Hasnat et al. [27] propose an automated system for detecting sarcasm in social media, specifically on Twitter. The study utilizes techniques such as Bag of Words, TF-IDF, and word embeddings to extract features from the text and employs a Long Short-Term Memory (LSTM) network for identifying sarcasm. The results demonstrate the effectiveness of the LSTM model with word embeddings, achieving a high accuracy score of 99.01%.

Ali et al. [28] introduce a novel deep-learning model called GMP-LSTM (Global Max Pooling with Long-Short Term Memory) for sarcasm identification in news headlines. The model replaces the LSTM layer with a GlobalMaxPool1D layer, incorporates dense layers with ReLU activation functions, and achieves promising accuracy in detecting sarcasm.

2.3. MULTILINGUAL SENTIMENT ANALYSIS

Khanvilkar et al. [29] proposed a system to determine the polarity of product reviews using ordinal classification and provide recommendations to users based on the achieved polarity. The methodology involves language standardization, preprocessing, applying machine learning algorithms such as SVM and Random Forest, determining polarity, and using sentiment analysis and user profile information for product recommendation.

Londhe et al. [30] proposed a hybridized model that combines the Social Eagle Algorithm and deep bidirectional long short-term memory (BiLSTM) to predict sentiment effectively. The model uses Transliteration to standardize the different languages and extract features from the data. The simulation results show that the proposed model outperforms current state-of-the-art methods with an accuracy of 91%, precision of 89%, recall of 91%, and F1 measure of 89%.

Manias et al. [31] performed a comparative study of two multilingual sentiment analysis methods, three multilingual deep learning classifiers, and a zero-shot classification method. The evaluation utilizes English dataset from Kaggle and another multilingual Twitter dataset. The multilingual deep learning classifiers demonstrated high performance, especially those pretrained and evaluated exclusively on monolingual data, the process of transferring inference when using these models with multilingual data. However, the zero-shot classification method fails to achieve high accuracies in monolingual data compared to multilingual data.

Kanfoud et al. [32] introduced SentiCode, a novel method for sentiment analysis across multiple languag-

es and domains. SentiCode employs a unified model using a singular level of abstraction and pseudocode, avoiding the need for separate models for each language. The approach generates a language-independent code, incorporating shared linguistic attributes like part-of-speech tags to extract sentiment-bearing words. The SentiCode vocabulary includes ADJ, ADV, NOUN, VERB, NOT, POS, and NEG. Experimental evaluations, encompassing English, Arabic, German, and Russian, demonstrated SentiCode's 70% accuracy, striking a balance between efficiency and computational time (around 67 seconds), making it well-suited for real-time applications.

Goel et al. [33] presented a study on developing a language-independent system for sentiment analysis of tweets using NLP and classification algorithms. The study utilizes a Twitter dataset comprising 2000 entries and applies preprocessing steps to filter the data. Additionally, the study employs the Google Translate API to develop a language-independent system. The Data modeling and training using Stanford NLP, then classification using the Naïve Bayes machine learning algorithm.

2.4. MULTILINGUAL SARCASM DETECTION

Rao et al. [34] proposed a sarcasm detection model for Amazon product reviews using machine learning. They collected a dataset, treated each review as a separate document, and applied preprocessing steps like tokenization and stemming. Feature extraction involved techniques such as TFIDF and n-grams. Classification employed Support Vector Machines, K Nearest Neighbors, and Random Forest algorithms. Results showed that Support Vector Machines achieved the highest accuracy at 67.58%, followed by Random Forest (62.34%) and K Nearest Neighbors (61.08%).

Khandagale et al. [35] proposed a system for detecting sarcasm in Hindi-English code-mixed tweets using machine learning and deep learning-based models. The authors compare the performance of various classification models and observe that the Random Forest and Logistic Regression classifiers yielded the highest F-score of 96%. The proposed system includes data pre-processing, feature extraction, feature selection, and classification models. The authors report that the Bi-LSTM encoder with an attention framework yields the best results among deep learning models with an F-score of 95%.

2.5. JOINT SARCASM DETECTION AND SENTIMENT ANALYSIS

Suhaimin et al. [36] proposed a sentiment analysis framework that integrates sarcasm detection. The proposed model comprises six modules: pre-processing, feature extraction, feature selection, initial sentiment classification, sarcasm detection, and final sentiment classification. The evaluation of this framework utilized Malay social media data, with results demonstrating that sarcasm detection leads to better sentiment classification performance. They use non-linear SVM for the initial sentiment classification module.

Modak et al. [37] extensively explore the correlation between sarcasm and sentiment analysis, utilizing Twitter data. They employ an innovative approach involving the fusion of K-means clustering, Principal Component Analysis (PCA), and Support Vector Machine (SVM) classifiers. The research methodology involves several phases: dataset collection and processing, feature extraction and reduction, clustering by K-means, and classification using SVM, Random Forest (RF), and K-nearest neighbor (KNN). The results show that the integration of PCA, K-means, and SVM classifier achieves a precision of 91.67%, recall of 92.67%, and accuracy of 91.90% for sarcasm detection. The study emphasizes the significance of considering textual information and sentiment diffusion patterns in analyzing sentiment in Twitter data.

M. Pal et al. [38] present a comprehensive study on sarcasm detection and sentiment analysis in Bengali. By collecting news headlines and comments from Twitter, the authors create datasets for both tasks. The sarcasm detection phase utilizes an LSTM model with GloVe embeddings, achieving high accuracy and performance. In the sentiment analysis phase, various classifiers undergo training using TF-IDF vectorization and n-gram features, with the best results obtained using the trigram feature and linear SVM.

Sait et al. [39] propose a novel approach called deep learning with natural language processing enabled SA (DLNLP-SA) for detecting sarcasm in textual data and its impact on sentiment analysis. The paper introduces the DLNLP-SA technique, which involves pre-processing, feature vector conversion using the N-gram feature extraction technique, and sarcasm classification using the MHSA-GRU (multi-head self-attention-based gated recurrent unit) model. The study also incorporates MFO (mayfly optimization) for hyperparameter optimization. The proposed DLNLP-SA model demonstrates superior performance compared to existing approaches, achieving an accuracy of 97.61% on a Twitter dataset.

Yin et al. [40] proposed a multi-task deep neural network for joint sarcasm detection and sentiment analysis (MT-SS). The model uses bidirectional gated recurrent units incorporating an attention network module to capture task-specific local feature representations and employing convolutional neural networks to capture global feature representations. The experiments utilized two datasets, and the proposed model outperforms existing approaches on both datasets.

3. COMPARATIVE ANALYSIS

This analysis provides insights into the strengths and weaknesses of different models, showcasing the impact of techniques, embeddings, and datasets on the performance of sentiment analysis and sarcasm detection models across various languages.

Table 1. Comparative Analysis of Models in Sarcasm Detection and Sentiment Analysis

Ref	lang	Dataset	Algorithm	Features	Acc.
[1]	Bangla	social media dataset	CNN-AM	Word embedding	72.06
[2]	English	restaurant reviews	SVM	ontology features	81.19
[3]	Arabic	Arabic Health Services Dataset	BiLSTM	Sent2seq-pre-trained word embedding	92.6
[4]	Bengali	SAIL 2015 Bengali dataset	BILSTM	SentiWordNet	55.73
[5]	Chinese	SMP2019 dataset	FER-Net	TTP (text to picture)	85.64
[6]	English	Weibo towards Movies dataset	non-linear layers	Graph-based	78.3
[7]	VN	Vietnamese Students Feedback	BiLSTM-AM	Word2Vec	86.64
[8]	Arabic	Multi-Dialect restaurant reviews	SVM	TF-IDF	87.7
[9]	English	SentiHood dataset	LSTM-AM	Commonsense Knowledge	89.32
[10]	English	restaurant reviews	LSTM, fuzzy logic	N-Grams	81.04
[11]	English	six microblogs' datasets	LAMP	Word embedding	85.2
[12]	Japanese - English	hotel reviews and movie reviews	BiLSTM-AM	word-embedding	87.2
[13]	English	Amazon product review dataset	CNN and Adversarial model	BERT	88.9
[14]	English	Twitter Dataset	LR -LSTM	XGBoost	73.1
[15]	English	generic sarcasm corpus	HRNN attention	Glove	75.3
[16]	Chinese	Chinese sarcasm dataset	AM and (MLP)	Encoder	69.53
[17]	English	News Headline Dataset	LSTM	TFIDF	96.11
[18]	English	news headlines dataset	Ensemble Model with Fuzzy Logic	Word2Vec, GloVe, and BERT	90.81
[19]	English	News Headlines Dataset	CNN-RNN	Glove	94.05
[20]	English	Twitter dataset	ILR and NN	Contrast Sentiment -N-grams	59
[21]	English	Twitter headlines dataset	LSTM	Word2Vec	92
[22]	English	News Headline Dataset	CNN-LSTM	BERT	99.63
[23]	Indian	786 Indian tweets	LRC	N-Grams	88
[24]	English	News Headlines Dataset	BERT and GCN	BERT	90.7
[25]	English	news headlines dataset	Ensemble Model with Fuzzy Logic	Word2Vec, GloVe, and BERT	90.81
[26]	Hindi	Twitter sarcasm dataset	CNN-LSTM	(CBOW) with Emoji Embedding	97.35
[27]	English	Twitter sarcasm dataset	LSTM	TFIDF and word embedding	99.01
[28]	English	News Headlines dataset	GMP-LSTM	word-embedding	92.5
[29]	Multi	user reviews	RF	Word embedding	95.03
[30]	multi	Twitter sentiment analysis dataset	BiLSTM	TF-IDF	91.57
[31]	Multi	Twitter dataset	MSC-VNN	Word embedding	84.31
[32]	multi	Amazon reviews	MLP	SentiCode-TF-IDF	70
[33]	Multi	Twitter dataset	NB	dictionary modeling	79.4
[34]	English	Amazon product reviews	SVM	TF-IDF, N-grams	67.58
[35]	Hindi-English	Code -mixed dataset	RF or LR	Sarcasm tokens, emoticons, n-grams	96
[36]	Malay-English	social media dataset	non-li SVM	(syntactic pragmatic- and prosodic)	90.5
[37]	English	tweets from Twitter	PCA-K-mean-SVM	component analysis model	91.9
[38]	Bengali	News headlines and News comments	LSTM	GloVe	91.94
[39]	English	Twitter dataset	MHSA-GRU	N-Grams	97.61
[40]	English	Sentiment and sarcasm dataset	Bi-GRU+AMCNN	Glove	92.01

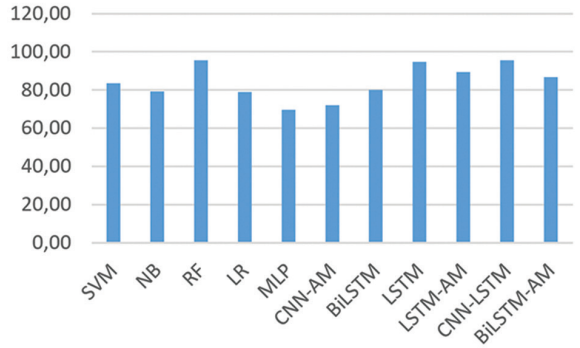


Fig. 1. Average Accuracy Comparison of Machine Learning Algorithms

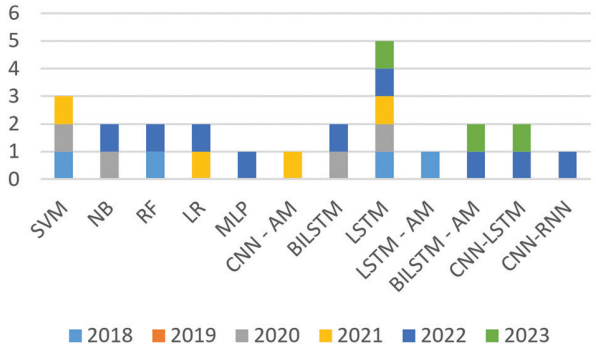


Fig. 2. Evolution of Machine and Deep Learning Algorithms Over the years

The performance of sentiment analysis and sarcasm detection models varies significantly due to factors like language, datasets, and algorithms. Integrating BERT word embedding techniques with LSTM or BiLSTM architectures consistently achieves high accuracy. However, accuracy depends on factors like feature engineering, word embedding algorithm, and model architecture. BERT-based models show promise in sentiment analysis and sarcasm detection. Future trends include the continued improvement of accuracy in sarcasm detection using advanced deep-learning models like BERT and LSTM. There is a focus on developing models capable of detecting sarcasm without explicit indicators. In sentiment analysis, we expect that progress in transformers and attention mechanisms will enhance performance. Multilingual sentiment analysis poses challenges, but we anticipate models that handle code-mixed and dialect-specific sentiment analysis. Integrating external knowledge sources and the increased use of transfer learning with pre-trained language models like BERT and GPT are expected to become more prevalent in sentiment analysis tasks. For sarcasm detection, notable models achieving the best accuracies include the one in [27], which employs LSTM for English Twitter sarcasm detection, achieving an impressive accuracy of 99.01%. This suggests that the model has a good capability to understand and classify sarcastic expressions in English tweets, and the model in [22] using CNN-LSTM with BERT for English News Headline sarcasm detection achieved a very high accuracy of 99.63%, showcasing the power of BERT embeddings in sentiment analysis tasks. For notable performance, models like the Hindi Twitter sarcasm detection model in [26] using CNN-LSTM with (CBOW) and Emoji Embedding achieved an accuracy of 97.35%, indicating its effectiveness in handling sarcasm detection in Hindi tweets, and the sarcasm detection model in [25] for English news headlines, using an ensemble model with fuzzy logic and combining Word2Vec, GloVe, and BERT embeddings, achieved an accuracy of 90.81%. For Multilingual Performance: The model in [35] for the Hindi-English code-mixed dataset, using Random Forest (RF) or Logistic Regression (LR) with features like sarcasm tokens, emoticons, and n-grams, achieved an accuracy of 96%, indicating effectiveness in handling multilingual sentiment analysis. For Dataset Influence: Performance varies across datasets, with some models achieving high accuracy on specific datasets, such as the model in [22] on the News Headline dataset. In sentiment analysis, a comprehensive evaluation of various models reveals intriguing insights. Notably, Random Forest with Word embedding excels with an impressive 95.03% accuracy in multi-user reviews [29], demonstrating its robustness in capturing nuanced sentiments. English Tweets from Twitter achieve high accuracy, notably with PCA-K-mean-SVM [9] and [30] BiLSTM with TF-IDF reaching 91.9% and 91.57% accuracy, respectively. These models showcase the effectiveness of diverse techniques in social media sentiment analysis. In multilingual analysis, the Non-linear SVM with syntactic pragmatic- and prosodic-based features achieves a commendable 90.5% accuracy in Malay-English social media [36]. A prominent trend highlights the consistent success of models leveraging advanced embeddings like Word2Vec, TF-IDF, and BERT. The versatility of techniques such as SVM, BiLSTM, and Random Forest, coupled with nuanced linguistic understanding, proves crucial for high accuracy in sentiment analysis across diverse datasets and languages. Tailoring models to specific linguistic nuances emerges as a good factor for optimal performance [9, 29, 30, 36].

4. CONCLUSION

This paper gives a thorough overview of sentiment classification, focusing on the challenges, especially in detecting sarcasm. Sentiment analysis is crucial for understanding people's opinions. Sarcasm makes Sentiment analysis difficult because it can lead to misclassification. The paper discusses different methods for sarcasm detection and their performance. It also explores the challenges of handling comments in multiple languages and the importance of language identification. The paper compares previous studies on sentiment analysis and sarcasm detection, noting variations in accuracy based on languages, datasets, and algorithms. Deep learning models like BERT and LSTM show promise for high accuracy. The paper identifies trends in sarcasm detection and sentiment analysis, including advanced deep learning, and addressing class imbalances. Transfer learning and pre-trained language models are expected to become more common. The key takeaway is that improving models with advanced techniques and addressing specific challenges will enhance sentiment analysis, aiding decision-making in various fields. Finally, practitioners should prioritize advanced embeddings (Word2Vec, TF-IDF, BERT) and versatile techniques (SVM, BiLSTM, Random Forest) for optimal sentiment analysis. Ethical considerations, including bias mitigation, transparent model decisions, and continuous monitoring, are essential to ensure responsible and effective sentiment analysis practices.

5. REFERENCES

[1] S. Sharmin, D. Chakma, "Attention-based convolutional neural network for Bangla sentiment analysis", AI & Society, Vol. 36, No. 1, 2021, pp. 381-396

[2] S. de Kok, L. Punt, R. van den Puttelaar, K. Ranta, K. Schouten, F. Frasincar, "Review-aggregated aspect-based sentiment analysis with ontology features", Progress in Artificial Intelligence, Vol. 7, No. 4, 2018, pp. 295-306.

[3] H. Elfaik, El H. Nfaoui, "Deep Bidirectional LSTM Network Learning-Based Sentiment Analysis for Arabic Text", Journal of Intelligent Systems, Vol. 30, No. 1, 2021, pp. 395-412.

[4] K. Sarkar, "Sentiment Analysis of Bengali Tweets Using Deep Learning", Proceedings of the International Conference on Computational Intelligence in Data Science, Chennai, India, 20-22 February 2020, pp. 71-84.

[5] M. Chen, K. Ubul, X. Xu, A. Aysa, Muhammat, Mahpirat, "Connecting Text Classification with Image Classification: A New Preprocessing Method for Implicit Sentiment Text Classification", Sensors, Vol. 22, No. 5, 2022.

[6] X. Zhao, Y. Liu, Z. Jin, "Multiplex network embedding for implicit sentiment analysis", Complex and Intelligent Systems, Vol. 7, No. 6, 2021, pp. 3179-3193.

[7] Y. Huang, S. Liu, L. Qu, Y. Li, "Correction to: Effective Vietnamese Sentiment Analysis Model Using Sentiment Word Embedding and Transfer Learning", Proceedings of the International Conference of Pioneering Computer Scientists, Engineers and Educators, Taiyuan, China, 18-21 September 2020, pp. 36-46.

[8] A. Hussein, I. Moawad, R. Badry, "Arabic Sentiment Analysis for Multi-dialect Text using Machine Learning Techniques", International Journal of Advanced Computer Science and Applications, Vol. 12, No. 12, 2021.

[9] Y. Ma, H. Peng, E. Cambria, "Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM", Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, No. 1, 2018.

[10] A. Londhe, P. V. R. D. P. Rao, "Dynamic Classification of Sentiments from Restaurant Reviews Using Novel Fuzzy-Encoded LSTM", International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 10, No. 9, 2022, pp. 112-124.

[11] D. Li, X. Luo, X. Wei, R. Xue, "Emojis Sentiment Analysis Based on Big Social Media Data", Proceedings of the International Conference on Applications and Techniques in Cyber Security and Intelligence, Shanghai, China, 11-13 July 2018, pp. 56-63.

[12] Y. Hiyama, H. Yanagimoto, "Word polarity attention in sentiment analysis", Artificial Life and Robotics, Vol. 23, No. 3, 2018, pp. 311-315.

[13] N. Liu, J. Zhao, "A BERT-Based Aspect-Level Sentiment Analysis Algorithm for Cross-Domain Text", Computational Intelligence and Neuroscience, Vol. 2022, 2022, p. 8726621.

[14] R. Bagate, S. Ramadass, "Sarcasm detection of tweets without sarcasm: Data science approach", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 33, 2021, pp. 993-1001.

[15] Q. Zhang, G. Chen, D. Chen, "Adversarial Training for Sarcasm Detection", Proceedings of the International Conference on Cognitive Computing, Seattle, WA, USA, 25-30 June 2018, pp. 43-54.

[16] L. Zhang, X. Zhao, X. Song, Y. Fang, D. Li, H. Wang, "A Novel Chinese Sarcasm Detection Model Based on Retrospective Reader", Proceedings of the International Conference on Multimedia Modeling, Phu Quoc, Vietnam, 6-10 June 2022, pp. 267-278.

[17] T. Sharma, N. Rani, A. Mittal, N. Rathee, "Sarcasm Detection in Social Media Using Hybrid Deep Learning and Machine Learning Approaches", Proceedings of the International Conference on Artificial Intelligence and Speech Technology, Delhi, India, 12-13 November 2022, pp. 455-464.

[18] D K. Sharma, B. Singh, S. Agarwal, H. Kim, R. Sharma, "Sarcasm Detection over Social Media Platforms Using Hybrid Auto-Encoder-Based Model", Electronics, Vol. 11, No. 18, 2022.

[19] K. Kavitha, S. Chittineni, "An Intelligent Metaheuristic Optimization with Deep Convolutional Recurrent Neural Network Enabled Sarcasm Detection and Classification Model", International Journal of Advanced Computer Science and Applications, Vol. 13, No. 2, 2022.

[20] C. Sengar, J. Nirmala, "Sarcasm Detection in Tweets as Contrast Sentiment in Words Using Machine Learning and Deep Learning Approaches", Proceedings of the International Conference on Machine Learning, Image Processing, Network Security and Data Sciences, Singapore, Silchar, India, 30-31 July 2020, pp. 73-84.

[21] R. P. Rao, S. Dayanand, K. R. Varshitha, K. Kulkarni, "Sarcasm Detection for Sentiment Analysis: A RNN-Based Approach Using Machine Learning", Proceedings of High Performance Computing and Networking, 2022, pp. 47-56.

[22] S. Bhardwaj, M. Prusty, "BERT Pre-processed Deep Learning Model for Sarcasm Detection", National

Academy Science Letters, Vol. 45, No. 2, 2022, pp. 203- 208.

[23] K. Ghosh, A. Banerjee, M. Bhattacharjee, S. Chatterjee, "Improved Twitter Sarcasm Detection by Addressing Imbalanced Class Problem", Vol. 165, 2021, pp. 135-145.

[24] A. Mohan, A. M. Nair, B. Jayakumar, S. Muraleedharan, "Sarcasm Detection Using Bidirectional Encoder Representations from Transformers and Graph Convolutional Networks", Procedia Computer Science, Vol. 218, 2023, pp. 93-102.

[25] D. Sharma, B. Singh, S. Agarwal, N. Pachauri, A. Alhussan, H. Abdallah, "Sarcasm Detection over Social Media Platforms Using Hybrid Ensemble Model with Fuzzy Logic", Electronics, Vol. 12, 2023, p. 937.

[26] A. Kumar, S. R. Sangwan, A. K. Singh, G. Wadhwa, "Hybrid Deep Learning Model for Sarcasm Detection in Indian Indigenous Language Using Word-Emoji Embeddings", ACM Transactions on Asian and Low-Resource Language Information Processing, Vol. 22, No. 5, 2023, p. 25.

[27] H. Saleem, A. Naeem, K. Abid, N. Aslam, "Sarcasm Detection on Twitter using Deep Handcrafted Features", Journal of Computing & Biomedical Informatics, Vol. 4, No. 2, 2023, pp. 117-127.

[28] R. Ali, T. Farhat, S. Abdullah, S. Akram, M. Alhajlah, A. Mahmood, M. Iqbal, "Deep Learning for Sarcasm Identification in News Headlines", Applied Sciences, Vol. 13, No. 9, 2023.

[29] G. Khanvilkar, D. Vora, "Sentiment Analysis for Product Recommendation Using Random Forest", International Journal of Engineering and Technology, Vol. 7, 2018, pp. 87-89.

[30] D. Londhe, A. Kumari, "Multilingual Sentiment Analysis Using the Social Eagle-Based Bidirectional Long Short-Term Memory", International Journal of Intelligent Engineering and Systems, Vol. 15, 2022, pp. 479-493.

[31] G. Manias, A. Kiourtis, A. Mavrogiorgou, D. Kyriazis, "Multilingual Sentiment Analysis on Twitter Data Towards Enhanced Policy Making", Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations, Hersonissos, Greece, 17-20 June 2022, pp. 325-337.

[32] R. Kanfoud, A. Bouramoul, "SentiCode: A new paradigm for one-time training and global prediction in multilingual sentiment analysis", Journal of Intelligent Information Systems, Vol. 59, 2022.

[33] P. Goel, V. Goel, A. K. Gupta, "Multilingual Data Analysis to Classify Sentiment Analysis for Tweets Using NLP and Classification Algorithm", Proceedings of Advances in Data and Information Sciences, 2020, pp. 271-280.

[34] M. V. Rao, Sindhu C., "Detection of Sarcasm on Amazon Product Reviews using Machine Learning Algorithms under Sentiment Analysis", Proceedings of the Sixth International Conference on Wireless Communications, Signal Processing and Networking, Chennai, India, 25-27 March 2021, pp. 196-199.

[35] K. Khandagale, H. Gandhi, "Sarcasm Detection in Hindi-English Code-Mixed Tweets Using Machine Learning Algorithms", Proceedings of the International Conference on Computing in Engineering & Technology: Applied Computational Technologies, Lonere, India, 12-13 February 2022, pp. 221-229.

[36] M. Suhairi, M. S. Md Suhaimin, M. Hanafi, M. Hijazi, R. Alfred, F. Coenen, "Modified framework for sarcasm detection and classification in sentiment analysis", TELKOMNIKA Indonesian Journal of Electrical Engineering, Vol. 13, 2019, pp. 1175-1183.

[37] S. Modak, A. Mondal, "Sentiment Analysis of Twitter Data Using Clustering and Classification", Proceedings of Third International Conference on Computing, Communications, and Cyber-Security, Vol. 421, 2023, pp. 651-664.

[38] M. Pal, R. Prasad, "Sarcasm Detection followed by Sentiment Analysis for Bengali Language: Neural Network & Supervised Approach", Proceedings of the International Conference on Advances in Intelligent Computing and Applications, Kochi, India, 1-3 February 2023, pp. 1-7.

[39] A. R. W. Sait, M. K. Ishak, "Deep Learning with Natural Language Processing Enabled Sentimental Analysis on Sarcasm Classification", Computer Systems Science and Engineering, Vol. 44, 2023, pp. 2553-2567.

[40] C. Yin, Y. Chen, W. Zuo, "Multi-Task Deep Neural Networks for Joint Sarcasm Detection and Sentiment Analysis", Pattern Recognition and Image Analysis, Vol. 31, 2021, pp. 103-108.

An improved algorithm of generating shortening patterns for polar codes

Original Scientific Paper

Issame El Kaime
Ibn Tofail University
National School of Applied Sciences, Advanced Systems Engineering Laboratory
Issame.elkaime@uit.ac.ma
Kenitra, Morocco

Reda Benkhouya
Ibn Tofail University
Faculty of Sciences, MISC Laboratory, Ibn Tofail University, Kenitra, Morocco
reda.benkhouya@gmail.com
Kenitra, Morocco

Abdessalam Ait Madi
Ibn Tofail University
National School of Applied Sciences, Advanced Systems Engineering Laboratory
abdessalam.aitmadi@uit.ac.ma
Kenitra, Morocco

Soufiane Hajbi
Ibn Tofail University
Faculty of Sciences, LCS Laboratory, Ibn Tofail University, Kenitra, Morocco
abdessalam.aitmadi@uit.ac.ma
Kenitra, Morocco

Abstract *– The rate matching in polar codes becomes a solution when non-conventional codewords of length N≠2n are required. Shortening is employed to design arbitrary rate codes from a mother code with a given rate. Based on the conventional shortening scheme, length of constructed polar codes is limited. In this paper, we demonstrate the presence of favorable and unfavorable shortening patterns. The structure of polar codes is leveraged to eliminate unfavorable shortening patterns, thereby reducing the search space. We generate an auxiliary matrix through likelihood and subsequently select the shortening bits from the matrix. Unlike different existing methods that offer only a single shortening pattern, our algorithm generates multiple favorable shortening patterns, encompassing all possible favorable configurations. This algorithm has a reduced complexity and suboptimal performance, effectively identifying shortening patterns and sets of frozen symbols for any polar code. Simulation results underscore that the shortened polar codes exhibit performance closely aligned with the mother codes. Our algorithm addresses this security concern by making it more difficult for an attacker to obtain the information set and frozen symbols of a polar code. This is done by generating multiple shortening patterns for any polar code.*

Keywords*: Polar codes, Bit error rate, Rate matching, Shortening*

1. INTRODUCTION

Polar codes are linear block correcting codes that have gained significant attention in recent years, thanks to their capacity-achieving properties and low decoding complexity. Polar codes were first proposed by Arikan in 2009 [1] and have since been extensively studied and used in various communication systems. Polar codes were subsequently refined by Arikan to incorporate systematic polar coding [2]. This evolution retains the inherent low-complexity nature of polar codes while additionally conferring notable advantages in terms of bit error rate performance. Polar codes can achieve the theoretical limit of the channel capacity. This is in contrast to other error-correcting codes, such as turbo codes [3] and low-density parity-check codes [4] which can only achieve close to the capacity limit. However, researchers have explored concatenating polar codes with convolutional codes to leverage the performance advantages of both error codes [5]. When building a polar code, the standard approach is to use a 2x2 polarization kernel matrix with a length as a power of two. Nevertheless, in practical applications, the code length could either be a power of two or differ from it. To tackle this challenge, rate-matching techniques such as puncturing, shortening, and repetition have been utilized to modify the encoder output. These techniques provide the flexibility to adjust the code length according to the specific requirements of the application. The study of rate-matching techniques for polar codes has been a widely researched topic in the literature. Some of the most notable approaches include puncturing and shortening techniques.

A study in [6] delves into the fundamentals of puncturing and shortening within polar codes, utilizing binary domination as the core investigative basis. In [7],

a novel class of symmetric puncturing patterns was introduced, accompanied by an efficient technique for generating these patterns. More recently, research by Hong et al. in [8] demonstrated the feasibility of achieving channel capacity with punctured polar codes.

Furthermore, in [9], a circular-buffer rate matching (CB-RM) approach based on a two-stage polarization by El Khamy et al. was presented. Wanqi et al. introduced a concise algorithm for selecting shortening patterns (SP). This approach involves creating an auxiliary matrix through bit-reversal permutation, followed by the selection of shortening bits from this matrix. This technique streamlines the process of generating shortened polar codes with ease [10]. a novel encoding hardware architecture for the general construction of polar codes in [11], which supports puncturing and shortening modes. In [12], a proposed algorithm called information set approximation puncturing (ISAP) introduces guard bits before known information bits. The algorithm then calculates the required puncturing pattern based on the actual transmission block length. Tianwei Han and alt. introduced an enhanced puncturing scheme for polar codes, where the initial log-likelihood rate (LLR) of the punctured bits is set to infinity to improve decoding performance [13].

Several other shortening algorithms have been proposed in recent years. One particular algorithm that exemplifies this is the mapping shortening (MS) algorithm, which takes into account the effect that the shortening has on the capacity of split channels [14]. The MS algorithm has been shown to outperform existing shortening algorithms in terms of channel capacity and BER and FER performance, especially under high code rates [2].

The distinction between the existing method and our proposed approach lies in their capacity to generate multiple SP. While the current method is limited to producing a single pattern. This versatility empowers our algorithm to adapt more flexibly to diverse communication scenarios, amplifying the potential for tailored and optimal performance. Furthermore, the proposed algorithm demonstrates a higher level of generality by encompassing and enhancing the capabilities of existing methods, such as Bit-Reversal Shortening (BRS) and Wang-Liu Shortening (WLS). Notably, our approach generates the same patterns that these methods produce. This integration of previously established methods not only showcases the adaptability of our algorithm but also consolidates various techniques under a unified framework. Beyond performance, our proposed method also introduces a crucial dimension of security enhancement. By dynamically altering the position of frozen bits within the generated patterns, our algorithm strengthens the security of the transmission. In scenarios where an eavesdropper intercepts the signal and is aware of the code's length, they are thwarted from accurately identifying the positions of both the information bits and the frozen bits. This fortification adds an extra layer of protection against potential attacks, safeguarding the integrity and confidentiality of transmitted data.

This paper is organized as follows. Section II provides a review of the background of polar codes and emphasizes the significance of SP in polar code design. Section III describes the details of the proposed algorithm and its functionality. Section IV then discusses the simulation results, along with a comparison to conventional shortening algorithms. Finally, section V presents the conclusions.

2. BACKGROUND

Polar codes are a class of error-correcting codes that have gained popularity in recent years due to their superior performance compared to other codes. They were introduced by Arıkan in 2009 [1], and are now integrated into multiple communication standards, including 5th generation wireless systems (5G) [15], The exceptional error correction performance exhibited by Polar codes renders them well-suited for upholding data integrity within satellite communication links [16]. Polar codes have been subject to investigation for the purpose of error correction within storage systems, notably in flash memory, where errors may arise due to physical limitations. This exploration is motivated by the capacity-achieving attributes inherent to Polar codes, which render them an appealing choice for mitigating errors within storage devices [17]. Furthermore, within the field of quantum communication, Polar codes have garnered attention for their potential applications in error correction and the transmission of information through quantum channels [18]. Additionally, the applicability of Polar codes has extended to wireline communication systems, including optical fiber networks, as they offer a mechanism to alleviate transmission errors and enhance the overall performance of such systems [19].

A polar code, in the context of information theory, is a linear block error-correcting code. Its construction involves iteratively concatenating a short kernel code, which transforms the physical channel into outer virtual channels. As the number of iterations increases, the virtual channels tend to polarize, exhibiting either high or low reliability [1]. Consequently, the data bits are allocated to the most reliable channels, leveraging the polarization effect for effective error correction. In a parallel spirit, the integration of polar coding has been harnessed across diverse applications, exemplified by the expansion of Golay codes through the strategic utilization of polar code techniques [20].

2.1. RATE MATCHING IN POLAR CODES

Rate matching is a crucial process in polar codes that ensures compatibility between the source data rate and the channel transmission rate. It involves adapting the number of bits from the coded output to fit the available channel resources. Rate-matching techniques

in polar codes are designed to optimize transmission efficiency and reliability. Polar code implementations encompass diverse rate-matching methods, such as puncturing, shortening, and repetition [21].

i. Puncturing

Puncturing is a process in polar codes where S bits are removed from the original codeword with length N, resulting in a shorter codeword of length Ns. These punctured bits are not transmitted, and, as a result, the decoder on the receiving end lacks any stochastic information about them. Hence, their initial Log-Likelihood Ratios (LLR) is typically set to zero.

The impact of puncturing on channel polarization is particularly significant compared to other rate-matching schemes. When certain bits are punctured, it leads to noticeable degradation of symmetric capacities in certain split channels, reducing them to zero [6].

In the next part, we discuss another technique called "shortening", which is a form of rate matching that optimizes the length of polar codes to match the available resources.

ii. Shortening

The novel puncturing scheme described in [13] serves as the second-rate matching technique, which forms the foundation of the shortening process. When a coded bit is shortened, its corresponding input bit channel is transferred to the set of frozen bits, while another bit is moved to the set of information bits. Specifically, the selection of shortened bits (at the output of the polar encoder kernel) is designed such that when a 0 is inputted to the corresponding input bit channels, the shortened bits also become 0.

If the ratio of K to E exceeds 7/16, execute shortening by eliminating bits from the trailing end of the sub-block [22]. This guarantees that the decoder does not lose any information due to the shortened bit, as the LLR for the shortened bit is set to infinity, indicating that it represents a 0. In a bit-reversal shortening scheme, the coded bits are shortened starting from the end in a reverse bit order.

It is important to carefully select the SP to achieve the desired balance between code length reduction and maintained error correction performance. By identifying and exploiting good SP, the overall efficiency of the polar codes can be optimized while still ensuring reliable communication. Good SP ensure that the shortened codes maintain a high level of error correction capability, effectively preserving the reliability of the original codes. On the other hand, "bad" SP refer to those configurations in which the performance of the shortened polar codes is significantly lower compared to the original (or "mother") codes.

In the upcoming section, we delve into another technique known as "repetition," which serves as a form of rate matching.

iii. Repetition

If the available resource G is slightly larger than the length $N=2^{\lfloor \log_2 G \rfloor -1}$ of polar codes N, it would be beneficial to use a polar code of length N along with repetition. This approach helps avoid performance degradation caused by excessive puncturing compared to selecting a polar code of length $2^{\lfloor \log_2 G \rfloor}$. Additionally, it reduces the complexity of encoding and decoding by approximately half [6].

During transmission, $G - N$ bits in the codeword can be generated by repeating certain bits in the original message X. Before decoding at the receiver, the LLRs of the repeated bits are combined with those of the corresponding original bits. This LLR combining technique enhances the effective bitwise channels associated with the repeated bits. Choosing a repetition pattern is a crucial factor that influences the decoding performance [6]. The upcoming section delves into a comprehensive discussion of the proposed method for generating favorable SP for polar codes. The details of this approach are thoroughly explained.

3. THE PROPOSED ALGORITHM

In this section, the primary objective is to demonstrate the existence of both favourable and unfavourable SP. Subsequently, we examine the common characteristics shared among all good SP. We then formulate the characteristics of good SP in mathematical terms. Finally, we present the algorithm for generating SP. The algorithm outlines the step-by-step process through which these patterns are generated.

3.1. GOOD AND BAD SHORTENING PATTERNS

Through our analysis, we have identified the presence of two distinct types of SP. The first type demonstrates performance closely resembling that of the original mother code, suggesting that the shortened polar codes retain similar levels of error correction capabilities. This suggests that these patterns effectively preserve the reliability and performance of the original code. However, the second type of shortening pattern demonstrates a significant deviation from the mother code's performance. The decoding performance of the shortened polar codes associated with these patterns is noticeably poorer compared to the original code. As a result, these patterns are considered less favourable in terms of error correction capabilities.

The clear distinction between these two types highlights the existence of good and bad SP. It emphasizes the need to identify and use SP that closely align with the desired performance requirements and error correction objectives. By opting for SP that closely resemble the mother code's performance, we can effectively maintain the reliability and efficiency of the polar codes. After retrieving good and bad SP, we proceed next to the analysis of the common characteristics among all the good SP.

3.2. THE COMMON CHARACTERISTICS AMONG ALL GOOD SHORTENING PATTERNS

The objective of this part is to identify and analyze the shared characteristics among all good SP. Subsequently, we aim to develop a mathematical model capable of generating good SP systematically. By examining multiple instances of good SP, we can observe recurring features that contribute to their effectiveness. Through a comprehensive analysis, our aim is to identify the shared characteristics exhibited by these patterns. Once these common characteristics are identified, our next goal is to construct a mathematical model that can reliably generate these patterns.

The process of shortening in polar coding involves assigning values, which are known to the decoder in advance, to bits that are not transmitted. As a result, the shortened bits possess bit-channels with likelihoods of ∞. After extensive analysis, we have made a notable observation during the decoding stage. Specifically, we have identified a distinctive characteristic in the matrix of likelihood ratios. It has come to our attention that all good SP exhibit the same row value in the position where the shortening occurs. This finding suggests a strong correlation between the row value and the effectiveness of SP. Regardless of the specific shortening position, we consistently observe this shared characteristic among the identified good patterns. By recognizing this common characteristic, we can further optimize the design and implementation of SP within the decoding process. Understanding the relationship between the row value and the effectiveness of SP opens up possibilities for more efficient and reliable decoding techniques, ultimately enhancing the overall performance of the system. After identifying the shared characteristics of good SP, the following paragraph focuses on modeling this concept with the aim of automatically generating good SP.

3.3. FORMULATING THE MATHEMATICAL CHARACTERISTICS OF GOOD SHORTENING PATTERNS

Based on the observations made in the previous section, we have identified a common characteristic among all good SP. To formalize this idea mathematically, we propose a modeling approach that extends the concept of decoding but with a flexible operation. In the initialization stage of the likelihood matrix, we set the values of elements in the first column to be arbitrary, except for the position corresponding to the shortened bit, where the value is set to infinity. Subsequently, the operation for computing the remaining elements follows a similar logic to SC as described in Algorithm 2. The objective is to compute all the elements of the matrix.

During the computation stage, if at least one element in the shortening row is different from infinity, it indicates that the proposed pattern is not good. On the other hand, if all the bits in the shortening row are infinity, it signifies that the proposed shortening pattern is good. By leaning on this mathematical model, we can evaluate the quality of different SP based on the values in the likelihood matrix.

3.4. PROPOSED METHOD STEPS

Our algorithm aims to find suitable SP for polar codes with a code length of N (where N is power of two) and a specified number of shortened bits S. The first step of the algorithm involves generating all possible SP, denoted as Pi. Each pattern Pi is a vector of length S, where the elements are natural numbers lower than N-1, and N-1 itself must be an element of the Pi.

After generating the SP, the algorithm proceeds by selecting the first pattern, denoted as Pi, to examine. Then, the algorithm initializes the likelihood matrix, M, which has a size of N*(n+1), where n=log_2 (N).

In the first column of matrix M, arbitrary numbers are assigned to the elements, except for the elements $m_{(i1)}$ ($i \in Pi$) corresponding to the shortened bits. For those elements, their values are set to infinity, as shown in equation (1).

$$m_{(i1)} = \{\infty \; if \; i \in \; P_i \quad arbitry \; \; if \; i \notin \; \; P_i \tag{1}$$

The first value of matrix M is given by equation (2), where "nan" represents the placeholder for unspecified values, and "inf" represents infinity:

$$\begin{pmatrix} m_{01} & nan & nan & nan \\ inf & nan & nan & nan \\ m_{03} & nan & nan & nan \\ m_{04} & nan & nan & nan \\ inf & nan & nan & nan \\ m_{06} & nan & nan & nan \\ m_{07} & nan & nan & nan \\ inf & nan & nan & nan \end{pmatrix} \tag{2}$$

We start calculating the elements for the matrix M using Algorithm 3 until all members have been assigned a value or until one member in the shortening row reaches a non-infinite value. The shortened row is a specific row in the matrix M. This row refers to the positions of the shortening bits; it is visually emphasized with a red highlight in (2). The bad rate matching patterns refer to those patterns in which at least one element of the shortened row in the corresponding matrix M is different from infinity, as illustrated in (3) where $m_{43} \neq inf$.

$$\begin{pmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ inf & inf & inf & inf \\ m_{20} & m_{21} & m_{22} & inf \\ m_{30} & m_{31} & m_{32} & inf \\ inf & inf & inf & \boxed{m_{43}} \\ m_{50} & m_{51} & m_{52} & m_{53} \\ m_{60} & m_{61} & m_{62} & m_{63} \\ inf & inf & inf & inf \end{pmatrix} \tag{3}$$

The effective SP are characterized by having all elements of the shortening row in the corresponding matrix equal to infinity, as showed in (4).

$$\begin{pmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ inf & inf & inf & inf \\ m_{20} & m_{21} & m_{22} & m_{23} \\ m_{30} & m_{31} & m_{32} & m_{33} \\ inf & inf & inf & inf \\ m_{50} & m_{51} & m_{52} & m_{53} \\ m_{60} & m_{61} & m_{62} & m_{63} \\ inf & inf & inf & inf \end{pmatrix} \quad (4)$$

For a detailed understanding of the approach, the proposed Algotihm 3 gives a step-by-step process to identify favorable SP. For simplicity reasons, the main Algorithm 3 uses Algorithm 2 which in turn calls the Algorithm 1. Note that Algorithm 1 determines the position of the first "1" in the binary representation of the input number, which is required by Algorithm 2 to compute elements of the LLR matrix M.

Algorithm 1: Get active LLR

```
Input = (i,n)
Mask=2n
C=1
for k =1 to n
    If Mask ==0 and i ==0
        C=C+1
        mask >>= 1
    Else
        Break
Return the min between C and n
```

Algorithm 2: Perform LLR update

```
Input :N,M
M=Matrix of LLR
N=the mothercode block length
Output:
    Shortening patterns state
n= int(log2(N))
a= n - activeLLR(l, n)
for s = a to n
    bl = 2^(s+1)
    bs =the remainder of the euclidean division between bl and 2
    for j from 1 to n step=bs
        if (j % bl) < bs
            top= M[j, s]
            btm=M[j+bs, s]
            if M[j, s]== inf
                M[j, s + 1] = M[j + bs, s]
                return "bad"
            else
                M[j, s + 1] = inf
        else
            btm = M[j, s]
            top = M[j - bs, s]
            bit = B[j - bs, s + 1]
            if M[j, s]== inf
                if M[j - bs, s]!=inf
                    M[j, s + 1] = inf-M[j - bs, s]
                    if M[j, s + 1]!= inf:
                        return "bad"
                else
                    M[j, s + 1]=M[j, s]
                    if M[j, s + 1]!= inf:
                        return "bad"
                else
                    M[j, s + 1] = M[j, s]
```

Algorithm 3: Identify favourable shortening pattern

```
Input: N, S
    N=the mothercode block length
    S= the number of shortened bits
n= int(log2(N))
P= Combination of (S-1) elements from 0 to (N-2)
Append N-1 to each P_j in P
For P_j in P
    M={mij,N*(n+1)}
    m(i,1)=inf if i ∈ P_j
    m(i,1)=arbitray number if i ∉ P_j
    For i=1 to N
        For l in Bit-reversal (i,n)
            if update_llrs(l,n,M)=="bad":
                Return bad
        return Good
```

For making the proposed algorithm easy to follow, the sequencing of steps is illustrated visually in Fig. 1.

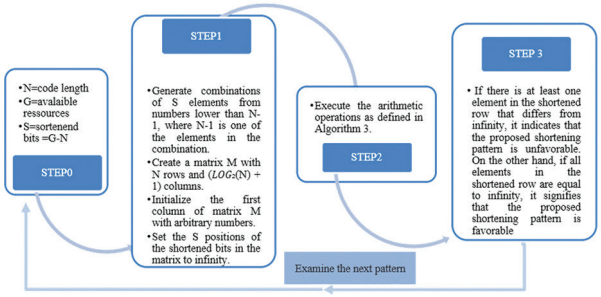


Fig 1. The simplified process of the proposed procedure

3.5. COMPARAISON OF LDPC AND POLAR CODES

LDPC codes outperformed polar codes for longer messages, while polar codes outperformed LDPC codes for shorter messages. This is because LDPC codes are better at correcting burst errors, which are common in long messages, while polar codes are better at correcting random errors, which are more common in short messages [23]. In 5G and beyond wireless communication systems, data channels typically carry long messages, such as streaming video and file transfers. Control channels, on the other hand, typically carry short messages, such as signaling and configuration information. Therefore, LDPC codes are

a better choice for data channels, while polar codes are a better choice for control channels [23]. Gaussian noise dominates the BER error at lower SNR values because it is a more fundamental source of noise in electronic circuits. Phase estimation error dominates the BER error at higher SNR values as it is caused by imperfections in the oscillators used to generate and synchronize the carrier signals.

In addition to the above, here are some other factors to consider when choosing between LDPC and polar codes:

- Complexity: LDPC codes are more complex to design and decode than polar codes.
- Latency: LDPC codes can have higher latency than polar codes.
- Power consumption: LDPC codes can consume more power than polar codes.

Overall, LDPC and polar codes are both promising channel coding techniques for 5G and beyond wireless communication systems [23]. The best choice of code for a particular application will depend on a number of factors, including the required code rate, the channel conditions, the complexity constraints of the system, and the latency and power consumption requirements

3.6. HARDWARE IMPLIMENTATION

The proposed shortening pattern generation algorithm can be implemented in hardware on an FPGA or ASIC. This would allow for high-throughput SP generation, which is important for applications such as real-time polar coding.

One way to implement the proposed algorithm in hardware is to use a parallel processing architecture. This would allow multiple SPs to be generated simultaneously, which would improve the throughput of the algorithm.

Another way to improve the performance of the algorithm in hardware is to use pipelining. This would allow the algorithm to start processing the next input while the previous input is still being processed.

The proposed algorithm is also relatively memory-efficient, which makes it suitable for implementation in hardware. The algorithm only requires a small amount of memory to store the mother code and the shortening patterns that have been generated.

The proposed algorithm can be implemented in hardware using PYNQ. PYNQ is an open-source framework that makes it easy to develop and deploy Python applications on FPGAs.

There are several advantages to implementing the proposed algorithm in hardware using PYNQ:

- Performance: FPGAs can provide significantly higher performance than software implementations of the proposed algorithm.
- Efficiency: FPGAs can be configured to efficiently implement the proposed algorithm, which can lead to reduced power consumption and increased battery life.
- Flexibility: PYNQ allows users to develop and deploy Python applications on FPGAs, which provides a high degree of flexibility and control.

3.7. WIRETAP CHANNEL SECURITY

A wiretap channel is a communication channel where an attacker can eavesdrop on the communication between the sender and the receiver. Wiretap channels are a major security concern, as they can allow attackers to steal sensitive information or to disrupt communication altogether.

Conventional shortening schemes for polar codes suggest only a single shortening pattern, which is known to both the receiver and attackers. The proposed shortening algorithm addresses this security concern by generating multiple favorable shortening patterns for any polar code, encompassing all possible favorable configurations. This makes it more difficult for an attacker to determine the information set and frozen symbols of a polar code, even if they are able to eavesdrop on the communication. if an attacker is able to eavesdrop on the communication, they will only be able to see the shortened polar code. However, they will not be able to determine which shortening pattern was used, as the sender has generated multiple shortening patterns. This makes it much more difficult for the attacker to exploit any weaknesses in the coding scheme. The proposed algorithm can be used to generate polar codes with a high minimum distance, which would make it more difficult for an attacker to decode the code.

4. RESULTS AND DISCUSSION

In this section, we apply the proposed algorithm to generate good SP for different code lengths. We then evaluate the performance of these patterns in terms of BER and FER.

Fig. 2 and Fig. 3 represent respectively the curves of Bit Error Rate (BER) and Frame Error Rate (FER) versus energy per bit to noise power spectral density ratio (Eb/No) for all possible shortening polar codes with a code length of $N_s = 30$ and information length of $K = 25$. The analysis of Fig. 2 and Fig. 3 reveals the existence of two distinct types of SP.

The first type of SP exhibits performance that closely resembles that of the mother code, However, the second type of SP demonstrates a significant deviation from the mother code's performance. The decoding performance of the shortened polar codes associated with the second type of patterns is noticeably worse compared to the original.

It is essential to emphasize that the selected parameters for this example (N = 32, K = 25, Ns = 30) serve to demonstrate the evaluation process. Nevertheless, the analysis can be applied to polar codes of varying lengths

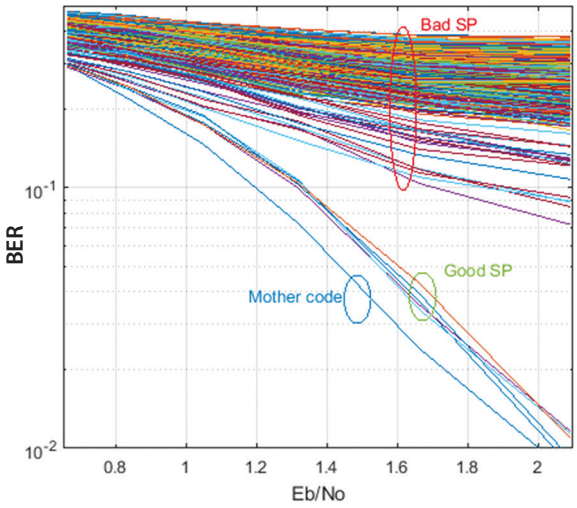


Fig. 2. Comparative Analysis of BER vs. Eb/No for various Rate Matching Patterns where Ns=30, K=25.

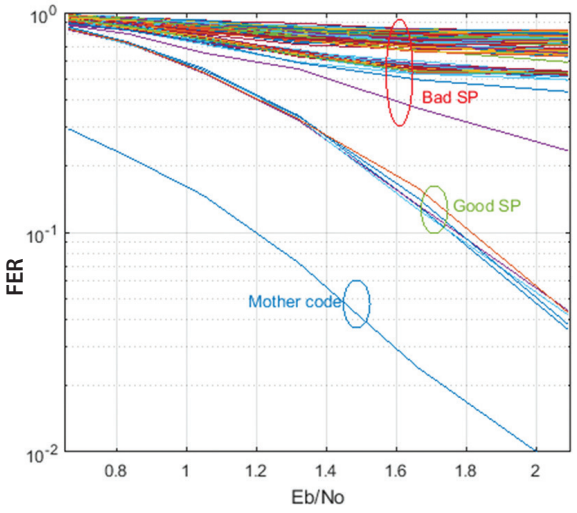


Fig. 3. Comparative Analysis of FER vs. EB/NO for various Rate Matching Patterns where Ns=30, K=25.

To confirm the last proposition, we conducted another test using all SP of polar code with a length of 15 and an information length of 12. From Fig. 4, which represents the Bit Error Rate (BER) versus Eb/No for all possible SP, we observe the existence of two distinct pattern types. The first type of patterns exhibited performances that were close to those of the mother codes. However, the second pattern showed significantly lower performance compared to the mother codes.

Based on the last figures, we have confirmed the existence of good and bad SP within polar codes. These findings underscore the criticality of thoughtful design and careful selection of patterns when utilizing polar codes. Fig. 5 depicts the BER of the proposed SP for a polar code length of 13. In order to provide a more comprehensive evaluation of our algorithm's performance, we also considered bad SP for comparison. The purpose of evaluating these bad patterns was to highlight the difference in performance between them and the proposed patterns.

Therefore, we conducted simulations of the Bit Error Rate (BER) for the proposed SP with a polar code length of 13, as depicted in Fig. 5.

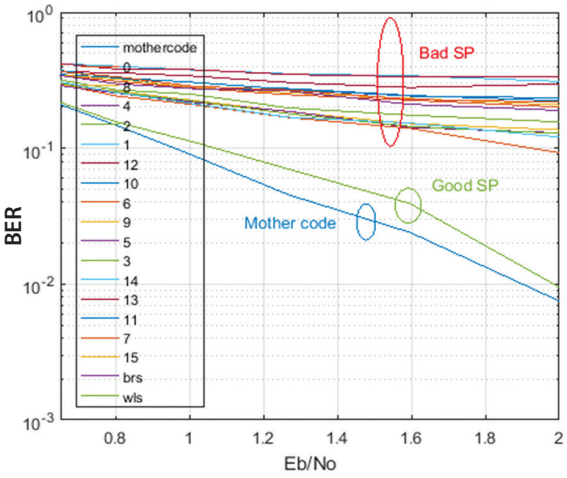


Fig 4. Comparative Analysis of BER vs. Eb/No for various Rate Matching Patterns where Ns=15, K=12.

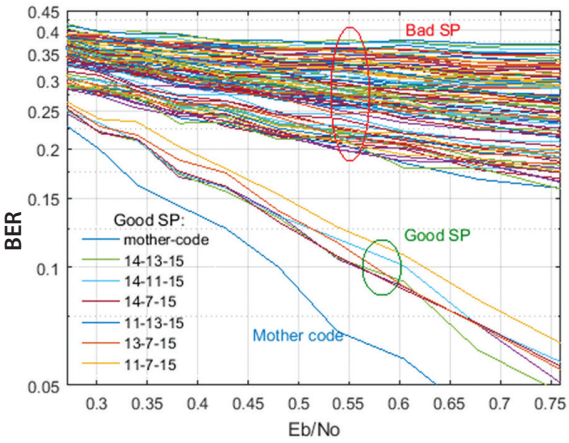


Fig 5. Comparative Analysis of BER vs Eb/No for various Rate Matching Patterns where Ns=13, K=7

The results in Fig. 5 indicate that the patterns produced by the proposed algorithm and mentioned in Table 1 demonstrate favorable BER values across various Eb/No, where Table 1 presents the set of good SP generated by our algorithm for different code lengths. This suggests that the proposed algorithm effectively generates SP that exhibit good BER performance.

Table 1 Good shortening patterns for different code lengths

Polar Code length	Good shortening patterns	BRS	WLS
2046	[1023, 2047], [1535, 2047], [1791, 2047], [1919, 2047], [1983, 2047], [2015, 2047], [2031, 2047], [2039, 2047], [2043, 2047], [2045, 2047], [2046, 2047]	[1023,2047]	[2046,2047]

1022	[511, 1023], [767, 1023], [895, 1023], [959, 1023], [991, 1023], [1007, 1023], [1015, 1023], [1019, 1023], [1021, 1023], [1022, 1023]	[511, 1023]	[1022,1023]
63	[63]	[63]	[63]
31	[31]	[31]	[31]
30	[15, 31], [23, 31], [27, 31], [29, 31], [30, 31]	[15,31]	[30, 31]
14	[7, 15], [11, 15], [13, 15], [14, 15]	[7,15]	[14,15]
6	[3, 7], [5, 7], [6, 7]	[3, 7]	[6, 7]
5	[3, 5, 7], [3, 6, 7], [5, 6, 7]	[5 3 7]	[5, 6, 7]
7	[7]	[7]	[7]
13	[7, 11, 15], [7, 13, 15], [7, 14, 15], [11, 13, 15], [11, 14, 15], [13, 14, 15]	[11, 7, 15]	[13, 14, 15]
61	[31, 47, 63], [31, 55, 63], [31, 59, 63], [31, 61, 63], [31, 62, 63], [47, 55, 63], [47, 59, 63], [47, 61, 63], [47, 62, 63], [55, 59, 63], [55, 61, 63], [55, 62, 63], [59, 61, 63], [59, 62, 63],[61, 62, 63]	[47, 31, 63]	[61, 62, 63]
29	[15, 23, 31], [15, 27, 31], [15, 29, 31], [15, 30, 31], [23, 27, 31], [23, 29, 31], [23, 30, 31],[27, 29, 31], [27, 30, 31], [29, 30, 31]	[23,15,31]	[29, 30, 31]
15	[15]	[15]	[15]

Results depicted in Fig. 6 present BER for all SP of polar codes with a code length of Ns=30. These results demonstrate that certain patterns exhibit BER values very close to those of the original codes. These specific patterns are also generated by our algorithm and are listed in Table 1.

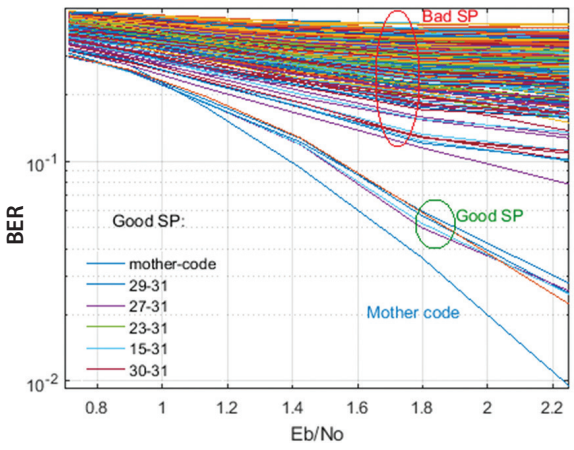


Fig 6. Comparative Analysis of BER vs. Eb/No for various Rate Matching Patterns where Ns=30,K=27

In contrast, the suboptimal patterns, which were not generated by our algorithm, exhibit poor performance. The results presented in all figures demonstrate that the patterns generated by the proposed algorithm present good performance. These findings reinforce the effectiveness of our algorithm in producing high-quality SP that maintain the performance of the mother codes, even when dealing with diverse code lengths.

To provide further evidence of the effectiveness of our algorithm in generating good SP for polar codes, we conducted a comparison with two existing methods: Bit-Reversal Shortening (BRS) and Wang-Liu Shortening (WLS). The Fig. 7 illustrates the comparison of BER of the SP generated by the proposed algorithm and those generated by the two existing methods (BRS and WLS).

The results provide compelling evidence that show that the performance of the proposed patterns is very similar to the patterns generated by the algorithms of WLS and BRS. Based on Fig. 7., it is evident that the patterns proposed by the BRS method are 6-5-7, while the WLS method suggests the pattern 5-3-7.

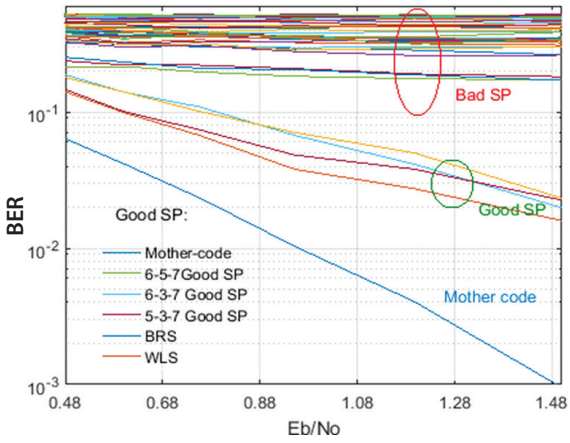


Fig 7. Comparative Analysis of BER vs. Eb/No for various Rate Matching Patterns where Ns=5, K=3.

Remarkably, our algorithm not only reproduces these patterns but also generates various others like 6-3-7.

A comparison between the patterns listed in Table 1 and the SP produced by the BRS and WLS methods for different codes length reinforces our algorithm's capacity. From Table 1, it is evident that the SP provided by both algorithms BRS and WLS are also suggested by the proposed method.The results demonstrate the effectiveness of the proposed algorithm in generating high-quality SP for polar codes.

5. CONCLUSION

The proposed algorithm outperforms existing algorithms in terms of flexibility for practical applications because the ability to generate multiple SPs allows it to select the SP that is best suited for the specific application requirements. This can lead to improved performance over existing algorithms, which typically generate a single SP.

The proposed algorithm outperforms existing algorithms in terms of security due to its the ability to

generate multiple SPs and dynamically change the position of frozen bits. This can make it more difficult for eavesdroppers to intercept and decode polar code transmissions. Additionally, the proposed algorithm introduces randomness in pattern selection, which further enhances confidentiality.

Finally, the proposed algorithm outperforms existing algorithms in terms of complexity as it does not require the computation of the reliability of each channel. This reduces the complexity of the algorithm and makes it more suitable for high-throughput applications. The difference in performance between the proposed and existing methods is due to the fact that the proposed algorithm tests all possible shortening patterns, while existing methods only provide one.

However, it is important to note that the proposed algorithm requires further improvement in the future to develop mechanisms for synchronizing the employed patterns between senders and receivers.

6. REFERENCES

[1] E. Arikan, "Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels", IEEE Transactions on Information Theory, Vol. 55, No. 7, 2009, pp. 3051-3073.

[2] E. Arikan, "Systematic Polar Coding", IEEE Communications Letters, Vol. 15, No. 8, 2011, pp. 860-862.

[3] D. Divsalar, F. Pollara, "On the Design of Turbo Codes", Telecommunications and Data Acquisition Progress Report, Vol. 42, 1995.

[4] R. Gallager, "Low-density parity-check codes", IEEE Transactions on Information Theory, Vol. 8, No. 1, 1962, pp. 21-28.

[5] A. Agarwal, S. N. Mehta, "PC-CC: An advancement in forward error correction using polar and convolutional codes for MIMO-OFDM system", Journal of King Saud University - Computer and Information Sciences, Vol. 32, No. 8, 2020, pp. 917-927.

[6] M. Jang et al. "Rate matching for polar codes based on binary domination", arXiv:1901.02287, 2019.

[7] L. Chandesris, V. Savin, D. Declercq, "On Puncturing Strategies for Polar Codes", arXiv:1701.04746, 2017.

[8] S.-N. Hong, D. Hui, I. Marić, "Capacity-Achieving Rate-Compatible Polar Codes", IEEE Transactions on Information Theory, Vol. 63, No. 12, 2017, pp. 7620-7632.

[9] M. El-Khamy, H.-P. Lin, J. Lee, I. Kang, "Circular Buffer Rate-Matched Polar Codes", IEEE Transactions on Communications, Vol. 66, No. 2, 2018, pp. 493-506.

[10] W. Li et al. "A shortening pattern selection method of shortened Polar codes", Proceedings of the 19th International Conference on Optical Communications and Networks, Qufu, China, 23-27 August 2021, pp. 1-3.

[11] W. Song, Y. Shen, L. Li, K. Niu, C. Zhang, "A General Construction and Encoder Implementation of Polar Codes", arXiv:1903.02899, 2019.

[12] J. Zhao, W. Zhang, Y. Liu, "A Novel Puncturing Scheme of Low Rate Polar Codes Based on Fixed Information Set", IEEE Communications Letters, Vol. 25, No. 7, 2021, pp. 2104-2108.

[13] T. Han, M. Fu, M. Zhang, G. Hu, "Puncturing scheme for polar codes based on channel reliability estimation", Telecommunication Systems, Vol. 82, 2023, pp. 499-508.

[14] H. Zhu, Y. Zhao, "A Mapping Shortening Algorithm for Polar Codes", IEEE Access, Vol. 7, 2019, pp. 87110-87117.

[15] N. Radha, M. Maheswari, "An empirical analysis of concatenated polar codes for 5G wireless communication", Telecommunication Systems, 2023, pp. 1-24. (in press)

[16] A. C. Vaz, N. T. Hegde, A. A. Shetty, "Performance analysis of Polar Codes in a Visible Light Communication System", Journal of Physics: Conference Series, Vol. 2466, 2023, p. 012006.

[17] N. K. Gerrar, S. Zhao, L. Kong, "Error correction in data storage systems using polar codes", IET Communications, Vol. 15, No. 148, 2021, pp. 1859-186.

[18] X. Chen, W. Qian, K. Tao, Y. Wang, W. Wang, Y. Wang, "Design of Polar Turbo Product Codes for Optical Fiber Communications Based on FPGA", Proceedings of the Asia Communications and Photonics Conference, Shenzhen, China, 5-8 November 2022, pp. 968-971.

[19] T. Ahmad, F. Khan, "On the Coding Gain of Polar Codes for High Speed Communications over AWGN Channels", Proceedings of the 17th International Bhurban Conference on Applied Sciences

and Technology, Islamabad, Pakistan, 14-18 January 2020, pp. 690-694.

[20] D. Khebbou, I. Chana, H. Ben-Azza, "Decoding of the extended Golay code by the simplified successive-cancellation list decoder adapted to multi-kernel polar codes", TELKOMNIKA (Telecommunication Computing Electronics and Control), Vol. 21, No. 3, 2023.

[21] M. S. Abed, "Novel Hybrid Model Investing in 5G Network Optimization Under Suzuki Fading Channel", International Journal of Electronics and Telecommunications, Vol. 69, No. 3, 2023.

[22] P. D, A. Puthuraya, G. Srinivasan, H. Kumar, K. N. Reddy, "Polar Coding in 5G systems", International Research Journal on Advanced Science Hub, Vol. 5, No. 5, 2023, pp. 175-180.

[23] S. Höt, "Performance analysis of LDPC and Polar codes for message transmissions over different channel models", 1st Edition, Wiley, 2019.

Wideband circularly polarized fractal antenna with SSRR metasurface for 5G applications

Original Scientific Paper

Maalim Al-Abbasi
University Malaya (UM),
Faculty of Engineering, Department of Electrical Engineering
Wilayah Persekutuan, Kuala Lumpur, Malaysia
maalim@siswa.um.edu.my

Tarik Abdul Latef
University Malaya (UM),
Faculty of Engineering, Department of Electrical Engineering
Wilayah Persekutuan, Kuala Lumpur, Malaysia
tariqlatef@um.edu.my

Abstract *–In this study, a 3.5 GHz wideband circularly polarised antenna design is presented. The suggested antenna is constructed using square and flower fractal shapes within the patch. The square fractal shapes and flower fractals are implemented inside the slot patch and in the top left and right corners of the patch. The key advantage of using this method is to increase the bandwidth of the antenna besides reduce the size and improve the axial ratio. Then a Square Split Ring Resonator (SSRR) metasurface is added above the antenna for enhancement of gain and back lobe. The metasurface is added on the front of the antenna in order to increase the gain directivity and decrease the sidelobes and back lobe. Using computer simulation software (CST), the proposed metasurface antenna is simulated and then constructed on FR4 substrate with dimensions of (ε_r=4.4 and h=1.6 mm). Operating bandwidth between 1.79 GHz and 5 GHz is achieved, the metasurface antenna's performance attained a fractional bandwidth of 91%. An improved gain of 3.05 dB is observed when metasurface is applied. A measured AR below 3 dB is achieved over the obtained bandwidth. The proposed fractal was reduced in size by 40.43% as compared to conventional antennas. The antenna recommended is suitable for use in the fifth generation (5G) applications.*

Keywords*: ircularly polarized, fractal antenna, metasurface, metamaterial, 5G, Axial ratio, bandwidth, CST software*

1. INTRODUCTION

The innovation of mobile and wireless communication applications toward fifth generation (5G) technology requires higher capacity and data rates compared to 4G technology [1-2]. For 5G mobile applications, a lower band of 3.5 GHz is recommended [3]. Antenna array networks were then suggested to increase the capacity and directivity of the proposed 5G mobile systems. Antennas can provide multiple beams with high gain and capacity toward the target. However, the antenna at this frequency (sub-6 GHz) is quite large and inflexible. At the suggested 5G bands, the critical design points for antennas are in terms of gain, size, backward radiation, sidelobes, and bandwidth. These terms are significantly important to providing a high-performance antenna for 5G bands. The alignment of the transmitter and receiver systems is necessary to reduce polarization mismatch losses when employing linearly polarized (LP) antennas in both transmission and reception systems. However, it is not required for a circularly polarized antenna (CP).

Typically, two orthogonal components of an electric field vector with equal amplitude are needed for circularly polarized antennas. This imposes a limitation on the bandwidth effectiveness of these antennas [4, 5]. On the other hand, axial ratio (AR) is a parameter that determines the polarization purity of CP. In this way, the perfect circular polarization, as seen at the poles of the Poincare sphere, The AR-enhanced circularly polarized antennas' bandwidth is increased using a variety of techniques. The first one is the radiator's built-in perturbation element [6]. Although it's a simple technique to produce circular polarization, it has a very narrow bandwidth. The second technique involves using a radiator with several feeds and external circuitry to create a phase shift between the feeds. This approach results in more wideband antennas, but their bulk is bigger, and their performance is constrained by the phase

shifter [6]. Wideband CP radiation can also be generated using crossed dipoles [7]. The stacked patch technique is another popular method for increasing the bandwidth and AR of CP microstrip antennas [8, 9-12].

It is also feasible to enhance the CP slot antenna's bandwidth by combining the CP slot antenna with a patch antenna. A circular slot antenna fed by a microstrip line with a parasitic patch antenna is presented in [10]. The inverted L-shaped feeding line achieves the CP radiation of the circular slot, similar to the CP slot antenna examined in [11]. The 3 dB AR bandwidth can be expanded to 45% when the parasitic patch is excited as well. This produces an additional CP resonance, which is then combined constructively with the CP resonances from the slot and the patch. The fundamental design of these CP patch antennas consists of the corners-truncated square patch [12], rectangular patch [13], elliptical patch [14], and slot- loaded square patch [15]. These fundamental CP patch antennas have a very small bandwidth, usually between 1% and 2%. CP patch antennas' bandwidth is increased using a variety of techniques, especially the AR bandwidth. However, aside from the size and complexity constraints, all of these designs still have restricted bandwidth and a 15% maximum AR improvement. Furthermore, the antenna designers typically used many radiators or complicated polarizers to feed radiators in circularly polarized microstrip antennas in order to achieve high gain, wide impedance bandwidth, wide axial ratio bandwidth, and beam width performance. As a result, the microstrip patch antenna's size increases and it is no longer appropriate for use with small, portable wireless devices. Metamaterials have therefore been suggested as a solution to these problems. According to definitions, metamaterials are artificial materials with special electromagnetic characteristics not found in natural materials [16]. On the other hand, metamaterial structures have demonstrated potential electromagnetic properties that are applicable to industry and microwave sub-6 GHz bands. Furthermore, a material that derives its properties from its structure rather than its composition directly is referred to as a metamaterial. Permittivity and permeability are the fundamental characteristics of metamaterials constructions [17]. Other properties such as refractive index, characteristic impedance, etc. can be defined by permittivity and permeability. Most natural materials exhibit a permeability equal to free space permeability $\mu 0$ and permittivity greater than the permittivity of free space $\varepsilon 0$ [18]. The goal of a metamaterial approach to antenna design is to improve antenna efficiency, bandwidth, weight, directivity, and other factors while overcome the limitations of conventional design techniques. It is well known that there is an inverse relationship between antenna size and radiation frequency. To reduce the size of an antenna while maintaining radiation efficiency, metamaterial structures are used to modify the dispersion relation or the near field characteristic of the antenna. [19, 20].

In response to the demands of portable communications devices, this paper presents a circularly polarized using a fractal antenna at 3.5 GHz with metasurface superstrate above the patch by using metamaterial structures implementation as superstrate placed on the top of the antenna patch with air gap distance with high performance. Recently metasurface antennas have been widely applied for increase gain and bandwidth enhancement for microstrip antenna due to the features of low profile, wide bandwidth, and easy implementation. Various metasurface antennas have been developed for compact size, wideband, low profile, or circular polarization applications [21-24]. Typically, a metasurface layer is built using an array of uniform periodic unit cells in various shapes, such square, rectangle, H-shaped, etc. After that, the metasurface is positioned above or below the main radiating element. [25].

This work contributes a new method to create a compact, high-gain, single-feed, single-layer microstrip antenna utilizing a fractal square array shaped with a modified slot ground to increase the return loss and avoid mismatching during fabrication and measurement process. When compared to previous procedures used on patches of a comparable size, this method is naturally simple and achieves a wideband axial ratio. The suggested antenna is designed to achieve a wideband of more than 2 GHz and an axial ratio of the same bandwidth. The antenna is created using the fractal square iteration technique. Then it is integrated into an array of fractal square iterations inside the patch. The suggested antenna is constructed from a FR-4 substrate with dimensions of (ε_r=4.4) and a height of h=1.6 mm. The antenna can be employed in future 5G applications because the measured findings were in good agreement with the simulated outcomes.

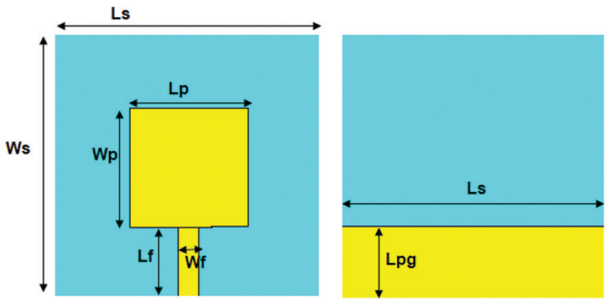
2. ANTENNA DESIGN WITH METASURFACE

This section focused on realizing a wideband circularly polarized fractal antenna with SSRR metamaterial as metasurface superstrate placed above the patch with high performance which can be used in the characterization of 5G base station.

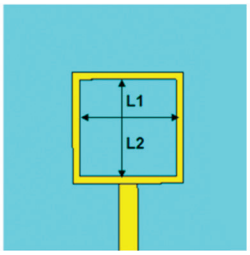
2.1. FRACTAL ANTENNA DESIGN

The design of the fractal antenna with square iterations is realized in four steps as follows; firstly, the required formulas used to find the length and width of the patch are provided in [8].

Hence, when the frequency is resonated at 3.5 GHz with ε_r=4.4 and thickness of 1.6 mm, the square length (L_p) is about 27.4 mm. A cut is made inside the square patch in order to study the effects of the slot square on the return loss before inserting the fractal shape. Fig. 1 depicts the basic and slotted structure of the antenna. The layout of the antenna in Fig.1 has a key benefit of simplicity in modifying its structure. While fractal iteration shape could be easily applied on the antenna structure.



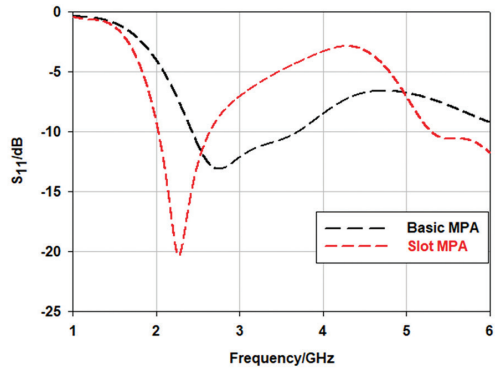
(a)



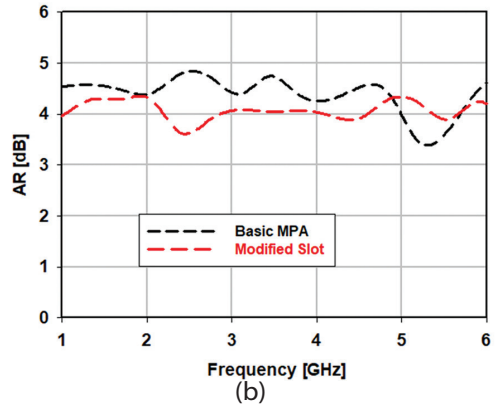
(b)

Fig. 1. (a) Front and ground view (b) Modified slot patch

The simulated response of the return loss and axial ratio for both shapes are presented in Fig. 2. Fig. 2(a) clearly shows that the slot square effects the return loss by shifting the frequency to the lower bands compared to the standard MPA. In Fig. 2(b), the axial ratio is above the required 3 dB level at 3.5 GHz for both antennas. Secondly, four square fractal shapes connected as a plus topology are added to the modified slot patch.



(a)



(b)

Fig. 2. Parametric study, (a) Return loss, (b) Axial ratio

The following equation is used to calculate the length of the first iteration of the square:

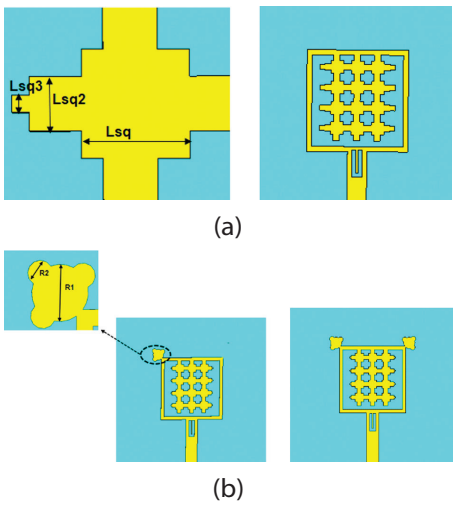
$$L_i = L_{i-1} * S \tag{1}$$

where the iteration is i = 1, 2, 3, 4,..., and S is the scaling factor reducing the size. Fig. 3 shows the proposed fractal antenna design process. The proposed fractal shape is inserted into the slot square shape as a matrix fractal of 3 × 4, where the fractal matrix is connected in the upper line of the modified square antenna, as observed in Fig. 3(a). In order to increase the impedance bandwidth and return loss, a U-shaped slot is added in the feed line, as in Fig. 3 (a). This helps the suggested antenna cover more spectrum and enhance the matching impedance.

Finally, as shown in Fig. 3(b), a fractal shape is added to the top corner of the proposed antenna to increase bandwidth and improve AR. The modified shapes with two flower shapes added to the antenna's left corner, followed by the same shapes added to the right corner as in Fig. 3 (b), with the same partial ground plane. Fig. 4 shows the parametric study on return loss and axial ratio as a better enhancement is achieved for both responses. Finally, the partial ground is modified with two rectangular slots cut distributed on both left and right sides of the feed line. Also, a T-shape slot cut is made on the partial ground to broaden the bandwidth and enhance the return loss order to cover more sub-6 GHz bands as illustrated in Fig. 5. The final dimensions for the suggested fractal antenna with amended ground plane are presented in Table 1.

Table 1. The final dimension for the modified ground fractal antenna

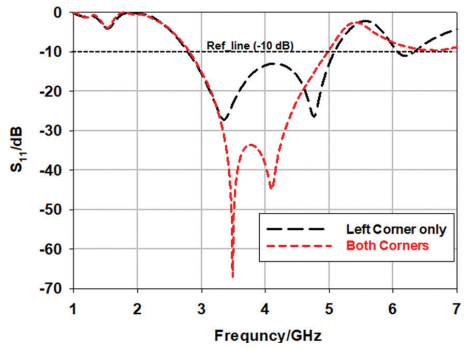
Parameters	Value (mm)	Parameters	Value (mm)
Ls	50	Lsq^2	1.75
Ws	50	Lsq^3	0.8
Lp	27.4	*R1*	4.7
Wp	27.4	*R2*	2.04
Lf	21	*Lf1*	58.8
Wf	5.3	*Lf2*	25.6
Lpg	70	*Lc1*	8
L1	25.2	*Lc2*	5.6
L2	25.2	*Ls1*	7
R1	1.5	*R2*	0.75
Lsq	3.5	Sub-6 GHz	
Slot dimension		4 × 0.56 mm²	



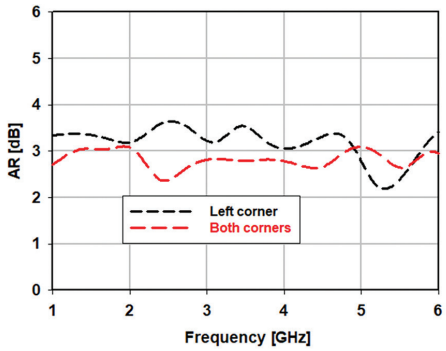
(a)

(b)

Fig. 3. (a) Fractal square proposed for fractal antenna. (b) The modified antenna with flower shape fractal added to the corner of the fractal MPA



(a)



(b)

Fig. 4. Parametric study by adding fractal and flower shape, (a) Return loss enhancement, (b) Axial ratio enhancement

Fig. 6 shows the performance of the optimized fractal MPA with modified partial ground in terms of return loss (S_{11}), radiation pattern, gain, and axial ratio (AR). At 4.2 GHz, a maximum return loss of -46 dB is achieved, resulting in an impedance bandwidth of 4 GHz. As shown in Fig. 6 (a), A return loss of below -10 dB is obtained over the whole bandwidth. A radiation pattern at a specific frequency of 4 GHz is plotted in Fig. 6 (b). 3.62 dB of directivity is reported. Fig. 6 (c) illustrates the gain value over the obtained bandwidth that ranges from 1.8 dB to 5.6 dB, with a 1.8 minimum value gain and a 5.6 maximum value gain at 1.8 GHz and 5.8 GHz, respectively. According to Fig. 6(d), an excellent circular polarization is indicated by the fact that the axial ratio is below 3 dB throughout the whole bandwidth.

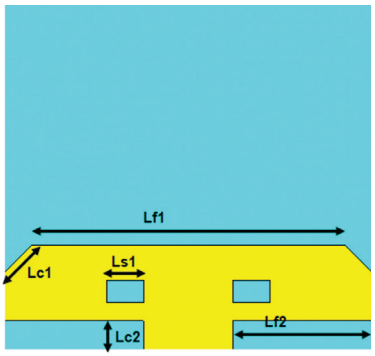
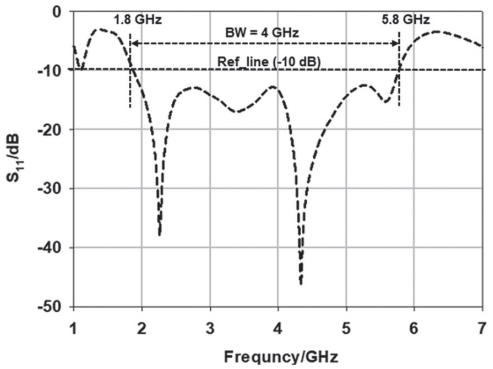
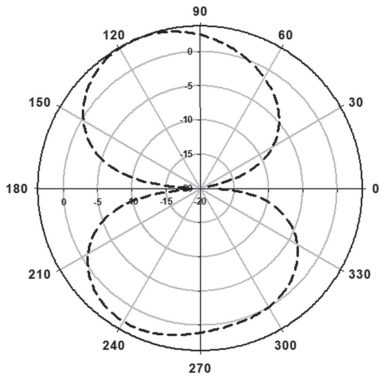


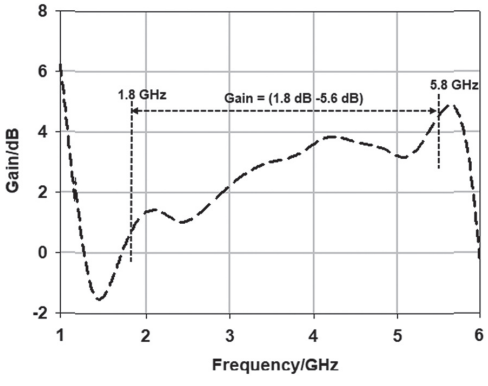
Fig. 5. Modified slot of Partial ground plane



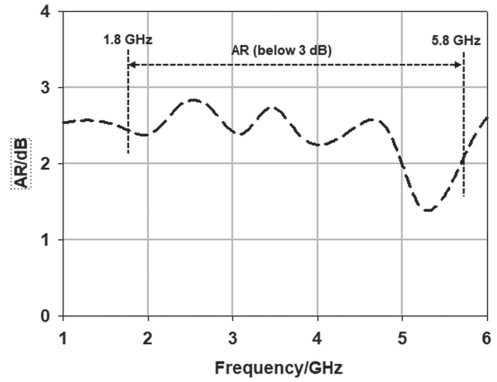
(a)



(b)



(c)



(d)

Fig. 6. The performance of the suggested fractal antenna with modified partial ground plane. (a) Return loss. (b) Radiation pattern. (c) Gain. (d) AR.

2.2. DESIGN OF METASURFACE WITH ANTENNA

Resonant frequency (f) is a property of metasurface. Calculating the resonant frequency takes into account the transmission line's capacitance (C) and inductance (L) properties. The frequency can be changed to a higher or lower frequency or in the direction of the desired frequency by adjusting the capacitance and inductance values. Generally, metasurface resonant frequency (f) is determined theoretically by its inductance (L) and capacitance (C). The resonant frequency varies by changing the values of inductance and capacitance The resonant frequency can be determined by [13]:

$$f = \frac{1}{2\pi\sqrt{LC}} \tag{2}$$

A dielectric substrate has a lattice etched into it to filter the element's conductivity. This creates a complete metasurface made up of several unit cells. Additionally, it enables the metasurface to pass or reject particular signals at a particular frequency. Therefore, in order to maximize the antenna's gain and bandwidth, a metasurface is created using a square split ring resonator (SSRR).

Four SSRR cells compose the unit cell, as seen in Fig. 7. The four SSRR's frequency is determined by [26]:

$$f_0 = \frac{1}{2\pi\sqrt{L\left[\left(2l-\frac{g}{2}\right)C+\frac{\varepsilon_0 wh}{2g}\right]}} \tag{3}$$

The final resonator width is (SS), and the cut in each SSR is (g). The capacitance and the inductance can be determined as:

$$C = \frac{\sqrt{\varepsilon_e}}{c_0 Z_0} \tag{4}$$

$$L = 0.00508l\left(2.303\log_{10}\frac{4l}{w} - \theta\right) \tag{5}$$

$$l = \frac{\lambda}{N} \quad \text{N= 4, 10, 20, …., N-1} \tag{6}$$

Hence, the values of the unit cell are as follows: (L=7mm, WW=0.5mm, S=0.4 mm, g=0.5mm, and SS=1 mm). Fig. 8 displays the performance of the SSRR unit cell in terms of S-parameters, permittivity (ε), and permeability (μ).

Metamaterials can be epsilon negative materials (ENG), mu negative materials (MNG), or double negative materials (DNG) [27]. Adding a number of SSRR's array (SSRR, SSRR2, SSRR3) to the form shifts the return loss (S11) below 3.5 GHz, as illustrated in Fig. 8 (a). The transmission coefficient lags behind the return loss at a lower frequency, as shown in Fig. 8(b), when compared to a higher frequency greater than 3.5 GHz. At sub-6 GHz the permeability is positive, While the permittivity is negative between 3 and 4 GHz as depicted in Fig. 8 (c) and Fig. 8 (d), respectively. As a result, the suggested metamaterial unit cell elements used in this paper, namely SSRR (the simplest realization of an LC circuit as a metamaterial element) acts as an ENG.

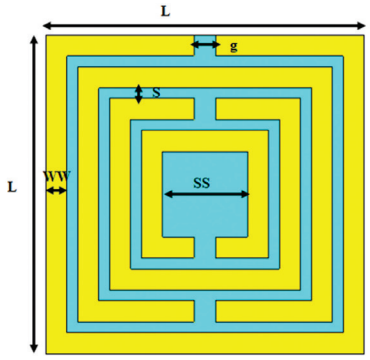


Fig. 7. SSRR unit cell design

Hence, metasurface is filtered by consisting of lattice of elements conductivity etched on a substrate dielectric. That allows metasurface unit cells to configure for selective frequencies to reject and pass through signals of certain frequency. As a result, the SSRR unit cell is integrated to create a 5 × 5 metasurface configuration, as seen in Fig. 9. Due to the suggested metasurface's symmetry, the cross-polarization and co-polarization of the normal x-polarized EM wave will be identical. Co-polarization between 4.2 GHz and 4.45 GHz is less than -10 dB. As shown in Fig. 10 (a), co-polarization is still less than -10 dB at frequency ranges over 6 GHz. The cross-polarization value in the mid-band (at 3.5 GHz) is within -1 dB. The Perfect conductor resonator (PCR) achieved a value more than 90% at 3.5 GHz and 5.8 GHz, as in Fig. 10 (b). As a result, the metasurface confirms the current surface increase in order to improve the gain.

To verify the high gain property shown in Fig. 11, A recommended fractal antenna is positioned above the suggested 5 × 5 metasurface. In the proposed configuration, a metasurface layer with unit cells faces the antenna from above and is separated from the bottom fractal antenna layer by an air separation gap between the layers (d).

The comparison of the simulated findings for the antenna's return loss, gain, axial ratio, and radiation pattern with respect to the metasurface at d = 15 mm is seen in Fig. 12. The return loss for both antennas with metasurface at 3.5 GHz is found to be -23.5 dB with a

fractional bandwidth of 3 GHz, as depicted in Fig. 12 (a). Fig. 12 (b) shows that the gain increases to 7.33 dB when the SSRR metasurface is used, which is comparable to the fractal antenna's gain of 2.33 dB. In both cases, the axial ratios remained constant along the bandwidth with a value below 3 dB, as illustrated in Fig. 12(c). Without the application of a metasurface, a circularly polarised radiation pattern can be observed compared to a directed radiation pattern when a metasurface is employed, as in Fig. 12 (d). Hence, the antenna with the metasurface demonstrated excellent impedance bandwidth, high gain, and a wide axial ratio in the simulation. The antenna measurement outcomes with the metasurface utilized will be discussed in the next section.

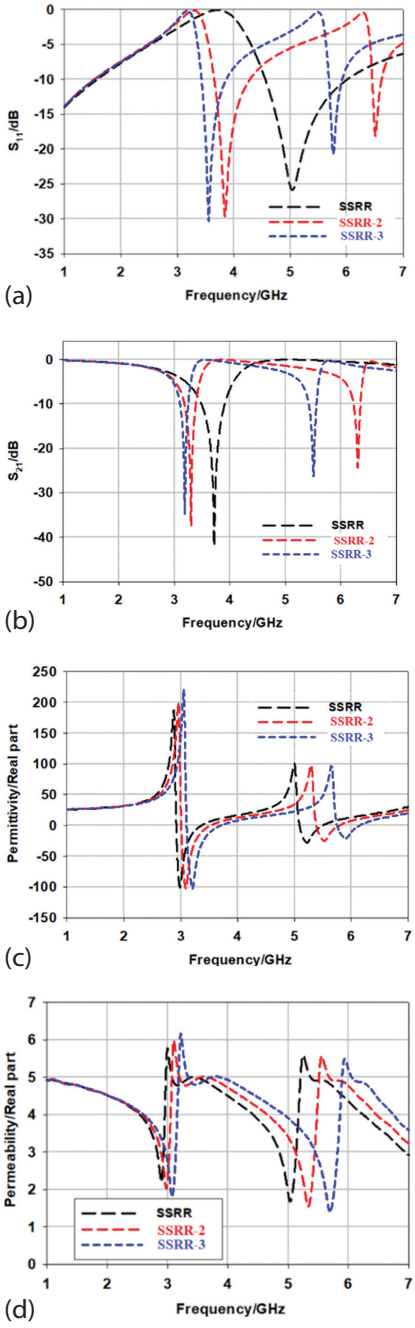


Fig. 8. Performance of the SSRR unit cell. (a) Return loss. (b) Transmission. (c) Permittivity. (d) Permeability

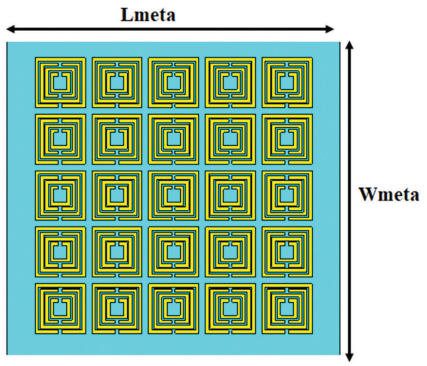


Fig. 9. The suggested configuration of 5 × 5 metasurface

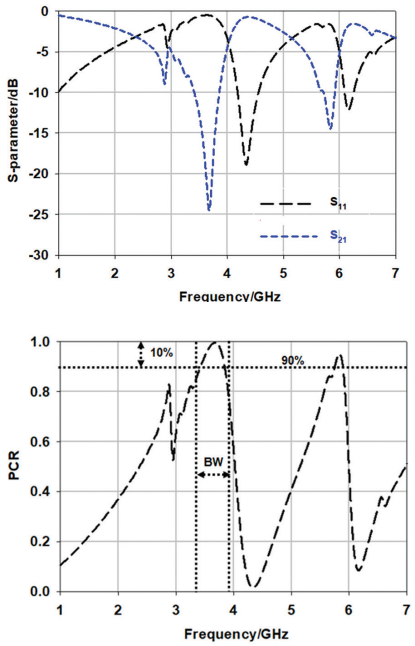


Fig. 10. Simulated response of 5 × 5 metasurface (a) return loss and transmission, (b) PCR

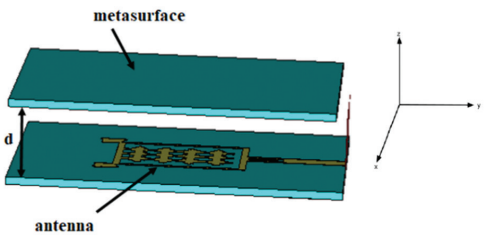
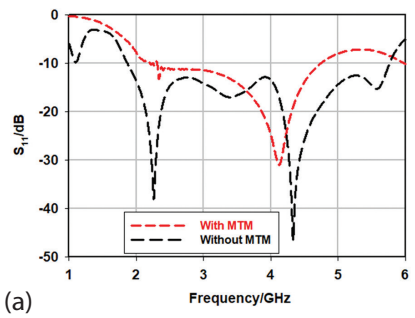
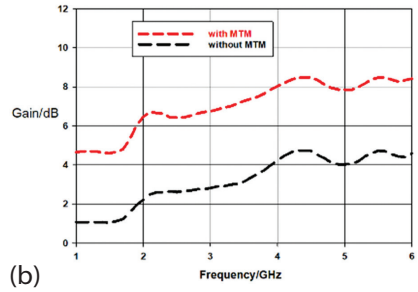
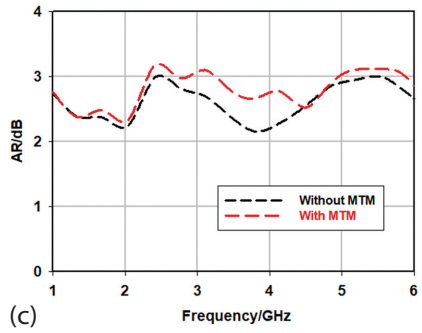


Fig. 11. The suggested metasurface positioned above the antenna

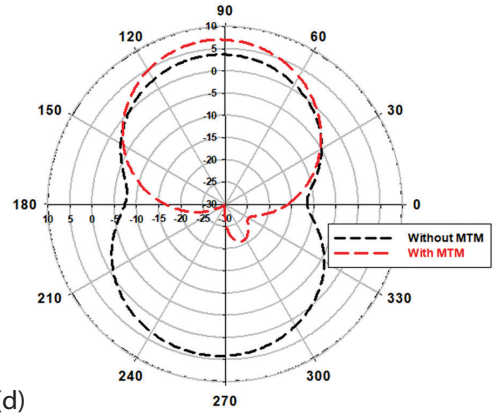




(b)



(c)



(d)

Fig. 12. Comparison of suggested antenna with and without Metasurface. (a) Return loss. (b) Axial ratio. (c) Radiation pattern. (d) Gain

3. RESULTS AND DISCUSSION

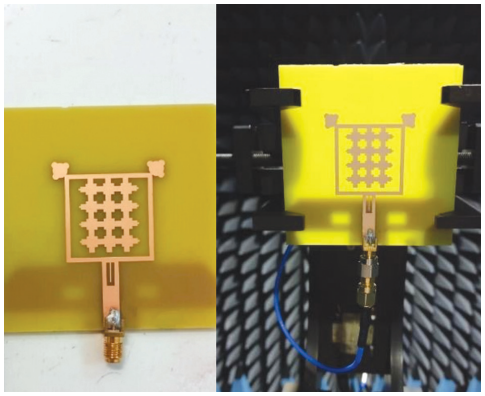
In this section fabrication and measurement results of fractal and metasurface antenna are introduced and compared with simulated results.

3.1. ANTENNA RESULTS

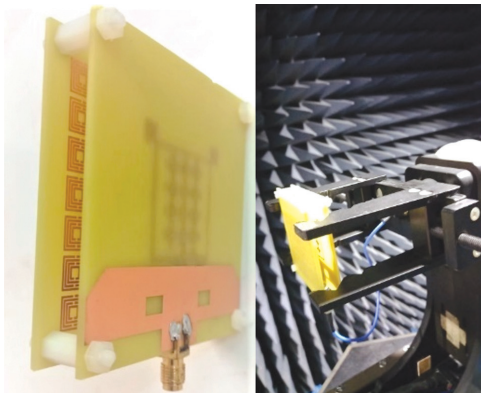
The printed fractal antenna in Fig. 13 is depicted with dimensions of 50 mm × 50 mm highlighting the compact size of the fractal antenna. Using the Keysight (Agilent Technologies) N9925A vector network analyser (VNA), the measurement in terms of S-parameters is carried out. Fig. 14 compares the printed fractal antenna's performance to simulation findings in terms of return loss, radiation pattern, gain, and AR. Fig. 14 (a) displays the printed antenna's measured return loss. Compared to the simulated result, the fractal antenna has a return loss of -12.45 dB at 3.5 GHz with a bandwidth of 3.16 GHz. The suggested antenna's simulated and measured radiation patterns in both directions are shown in Fig. 14(b).

In comparison to the simulated one, the measured prototype's major lobe is seen to be tilted by 3 degrees. According to the measurements, the prototype's sidelobes are slightly low. The measured and simulated realized gain for the proposed antennas is plotted in Fig. 14 (c). In comparison to the simulated gain of 3.05 dB, the measured gain at 3.5 GHz is 2.61 dB, indicating a slight loss of 0.5 dB. The primary cause of the decrease is the unwanted interference from the surrounding environment.

The plot of the measured and simulated antenna AR over frequency is observed in Fig. 14(d). The measured AR of the proposed antennas is well above 80% within the operating frequency. The measured AR is 2.9 dB at 3.5 GHz. Due to high radiation efficiency, the gap between the measured and simulated AR is slightly smaller in the 1.79 GHz to 5 GHz bandwidth range.

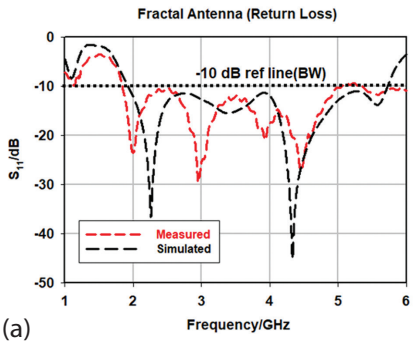


(a)



(b)

Fig. 13. The printed prototypes. (a) fractal antenna, (b) metasurface antenna



(a)

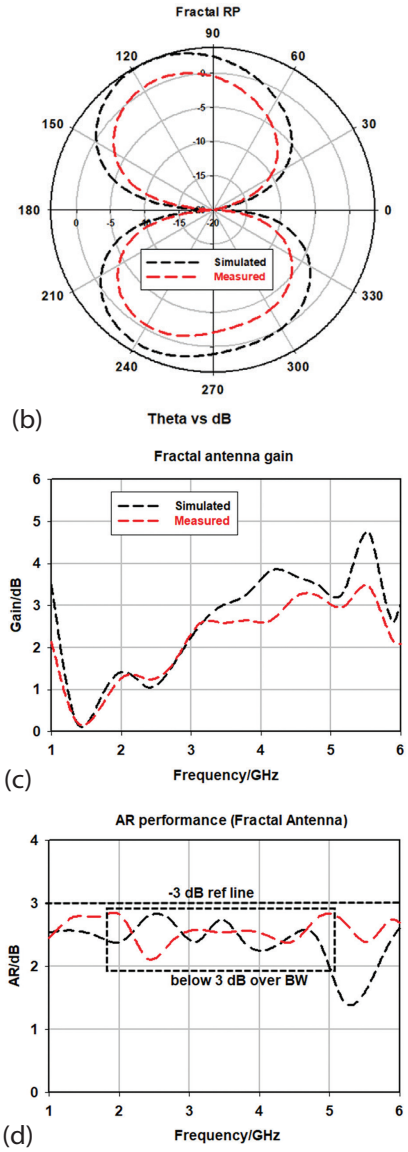


Fig. 14. The measured performance of the printed fractal antenna: (a) Return loss, (b) Radiation pattern, (c) Gain, (d) AR

3.2. METASURFACE ANTENNA RESULTS

Fig. 15 (a) displays the measured return loss performance. At 4 GHz, the measured S11 is -23.58 dB versus a simulated value of -30.2 dB. With a loss of 100 MHz of simulated bandwidth (3 GHz), the measured fractional bandwidth is 2.99 GHz. As seen in Fig. 15 (b), the measured radiation pattern and the simulated one agree well with a 3-degree erroneous tilt. The observed gain response is contrasted with the simulated one in Fig. 15 (c). The measured gain, in comparison to the simulated 8 dB gain, is 7.2 dB with a loss of 0.8 dB. This is due to interfering sources within the radiation pattern chamber. In Fig. 15 (d), the AR performance is compared. The AR value is greater than 82% over the operational frequencies. For instance, at 4 GHz, there is only a 0.5 dB difference between the measured and simulated AR. Table 2 compare the measured results of metsurface antenna with the simulated ones. Table 3 includes a comparison between different metamaterial-based circularly polarized antennas.

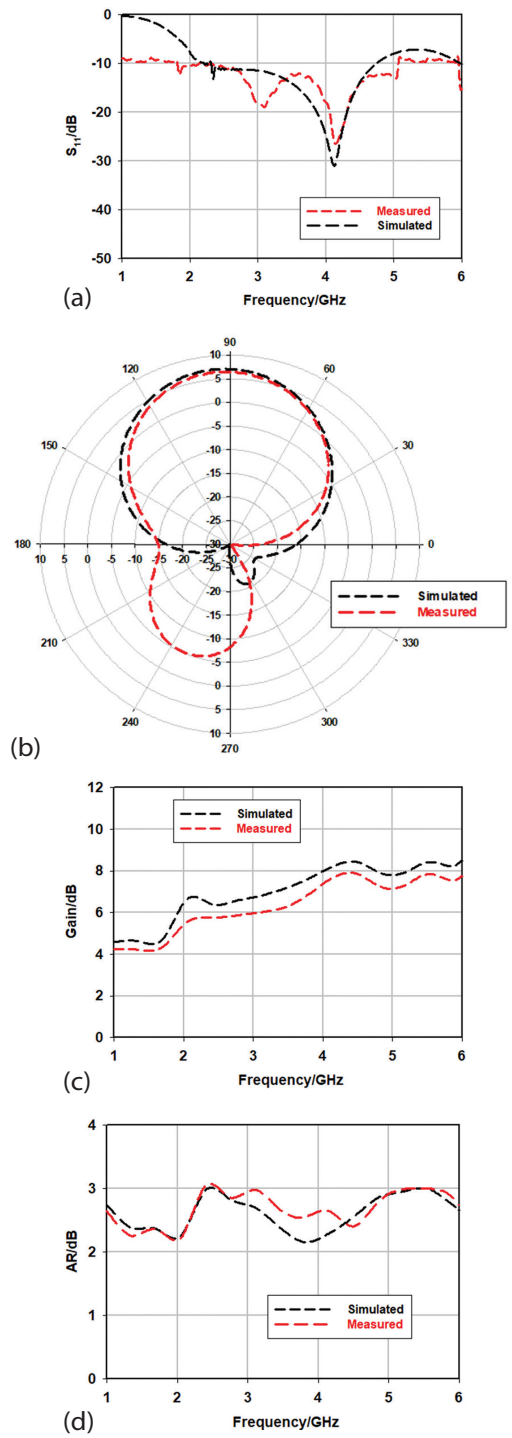


Fig. 15. Performance of the measured printed metasurface antenna: (a) Return loss, (b) Radiation pattern, (c) Gain, (d) AR

Table 2. comparison between the measured results of metsurface antenna with the simulated ones

Parameters/metasurface	Simulated	Measured
Return loss (S11)	-15.25	-12.45
Bandwidth (GHz)	3.8	3.16
Gain (dB)	3.05	2.61
AR (dB)	2.7	2.9

Table 3. Comparison between different metamaterial-based circularly polarized antennas

Comparison with related works					
Ant	**Feeding method &resonant freq**	**Gain dBic**	**Impedance bandwidth**	**AR band**	**Observations And complexity**
[28]	Coaxial probe feed 3 GHz	4.15	2.92-3.06 GHz	3.11-3.17 GHz	Crossbar fractal tree slot and three-turn complementary spiral resonators
[29]	CPW feed 1.95 &2.61 GHz	6.9 1.1	1.94 t-1.965 GHz 2.545-2.685 GHz	2.60 - 2.62 GHz	(CRLH)composite right/left-handed unit cell is designed as a radiation element
[30]	Coaxial probe feed Around 3 GHz	0.4-0.95	1.475-1.483 GHz 1.481-1.489 GHz	1.479-1.486 GHz	Structure is quite complicated; the gain is also quite low
[31]	Coaxial probe feed Around 3 GHz	6.26 6.97	2.875-2.961 GHz 3.812-3.836 GHz	2.879-2.901 GHz 3.822-3.829 GHz	Provides a very narrow operational bandwidth
This work	Microstrip feed 3.5 GHz	6	1.79 – 5 GHz	1.82-5 GHz	Simple method using square fractal shapes and flower fractals

4. CONCLUSION

This study discusses a 3.5 GHz wideband circularly polarized fractal antenna, the square and flower fractal shapes were used to create the fractal antenna. Then metamaterials structures are developed as metasurface to increase the gain and reduce the side lobes of the proposed antennas. The metamaterial designed individually at the same frequency type and is selected based on the current works on enhancement specifically parameters such as size, bandwidth, and gain. The performance of the proposed design validated through measurement process and it was much in line with the simulation's findings. The designed antenna achieved great S-parameters, bandwidth, and AR performance. A high fractional bandwidth of 91% is provided by the antenna which operates in the 1.79 GHz to 5 GHz frequency range. By 40.43%, the antenna had successfully acquired a favorable profile of compact size. The suggested antenna is appropriate for use in a future antenna array system for 5G.

5. REFERENCES:

[1] Y. He, S. Lv, L. Zhao, G.-L. Huang, X. Chen, W. Lin, "A compact dual-band and dual-polarized millimeter-wave beam scanning antenna array for 5G mobile terminals", IEEE Access, Vol. 9, 2021, pp. 109042-109052.

[2] I. Syrytsin, S. Zhang, G. F. Pedersen, A. S. Morris, "Compact quad-mode planar phased array with wideband for 5G mobile terminals", IEEE Transactions on Antennas and Propagation, Vol. 66, No. 9, 2018, pp. 4648-4657.

[3] N. O. Parchin, J. Zhang, R. A. Abd-Alhameed, G. F. Pedersen, S. Zhang, "A planar dual-polarized phased array with broad bandwidth and quasi-endfire radiation for 5G mobile handsets", IEEE Transactions on Antennas and Propagation, Vol. 69, No. 10, 2021, pp. 6410-6419.

[4] Y. Kumar, S. Singh, "Microstrip Fed Multiband Hybrid Fractal Antenna for Wireless Applications", Applied Computational Electromagnetics Society Journal, Vol. 31, No. 3, 2016, pp. 327-332.

[5] M. I. Sabran, S. K. A. Rahim, P. J. Soh, C. Y. Leow, G. A. E. Vandenbosch, "A Simple Electromagnetically Fed Circularly Polarized Circular Microstrip Antenna", Applied Computational Electromagnetics Society Journal, Vol. 30, No. 11, 2015, pp. 1180-1187.

[6] M. Majidzadeh, J. Nourinia, C. Ghobadi, "Compact CPW-Fed Antenna with Circular Polarization Characteristics in WLAN Frequency Band", Applied Computational Electromagnetics Society Journal, Vol. 28, No. 10, 2013, pp. 938-943.

[7] Y. Liu et al. "A K-band broadband circularly polarized slot antenna based on L-shaped waveguide cavity", IEEE Antennas and Wireless Propagation Letters, Vol. 20, No. 9, 2021, pp. 1606-1610.

[8] Z. Zhou, Z. Wei, Z. Tang, Y. Yin, "Design and analysis of a wideband multiple-microstrip dipole antenna with high isolation", IEEE Antennas and Wireless Propagation Letters, Vol. 18, No. 4, 2019, pp. 722-726.

[9] X. Shuai, S. Xiao, "A novel dual-band circularly polarized slot antenna with fractal slot geometry: Shuai and Xiao", Microwave and Optical Technology Letters, Vol. 59, No. 2, 2017, pp. 451-456.

[10] S. Maity, K. R. Barman, S. Bhattacharjee, "Silicon-based technology: Circularly polarized microstrip patch antenna at ISM band with miniature structure using fractal geometry for biomedical application", Microwave and Optical Technology Letters, Vol. 60, No. 1, 2018, pp. 93-101.

[11] K. S. Kola, A. Chatterjee, D. G. Patanvariya, "Design of a wideband right-handed circularly polarized ar-

ray of miniaturized mushroom-shaped antennas for direct broadcast satellite application", Microwave and Optical Technology Letters, Vol. 62, No. 11, 2020, pp. 3542-3555.

[12] E. Wang, M. Liu, D. Lin, J. Wang, "A compact circular polarized patch fractal antenna for global navigation satellite systems", Microwave and Optical Technology Letters, Vol. 64, No. 3, 2022, pp. 520 – 524.

[13] H. A. Atallah, A. B. Abdel-rahman, K. Yoshitomi, R. K. Pokharel, "Mutual Coupling Reduction in MIMO Patch Antenna Array Using Complementary Split Ring Resonators Defected Ground Structure", Applied Computational Electromagnetics Society Journal, Vol. 31, No. 7, 2016, pp. 737-743.

[14] G. Mansour, E. Nugoolcharoenlap, R. H. Mahmud, T. Tippo, P. Akkaraekthalin, D. Torrungrueng, "Circularly Polarized Elliptical Patch Array Antennas for GPS", Proceedings of the Research, Invention, and Innovation Congress, Bangkok, Thailand, 1-3 September 2021.

[15] M. Asad Rahman, E. Nishiyama, I. Toyoda, "A Circular Polarization Antenna with Square Ring Slot Loaded on Ground Plane", Proceedings of the International Symposium on Antennas and Propagation, Phuket, Thailand, 30 October - 2 November 2017.

[16] A. Sihvola, "Metamaterials in electromagnetics", Metamaterials, Vol. 1, No. 1, 2007, pp. 2-11.

[17] J. Undrakonda, R. K. Upadhyayula, "A Novel Miniaturized Isotropic Patch Antenna for X -Band Radar Applications Using Split Ring Resonators", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 1, 2023, pp. 21-27.

[18] C. R. Simovski, "Material Parameters of Metamaterials (a Review)", Optics and Spectroscopy, Vol. 107, No. 5, 2009, pp. 726-753.

[19] A. K. Vallappil, B. A. Khawaja, M. K. A. Rahim, M. N. Iqbal, H. T. Chattha, "Metamaterial-Inspired Electrically Compact Triangular Antennas Loaded with CSRR and 3 × 3 Cross-Slots for 5G Indoor Distributed Antenna Systems", Micromachines, Vol. 13, No. 2, 2022.

[20] K. Jairath, N. Singh, M. Shabaz, V. Jagota, B. Kumar Singh, "Performance Analysis of Metamaterial-Inspired Structure Loaded Antennas for Narrow Range Wireless Communication", Scientific Programming, Vol. 2022, 2022.

[21] F. H. Lin, Z. N. Chen, "Low-profile wideband metasurface antennas characteristics mode analysis", IEEE Transactions on Antennas and Propagation, Vol. 65, No. 4, 2017.

[22] Z. Liang, J. Ouyang, F. Yang, "Low-profile wideband circularly polarized single-layer metasurface antenna", Electronics Letters, 2018, Vol. 54, No. 24, 2018, pp. 1362-1364.

[23] W. Lin, Z. N. Chen, X. Qing, "Metamaterial-based low profile broadband mushroom antenna", IEEE Transactions on Antennas and Propagation, Vol. 62, No. 3, 2014, pp. 1165-1172.

[24] Q. Chen, H. Zhang, "Dual-patch polarization conversion metasurface-based wideband circular polarization slot antenna", IEEE Access, Vol. 6, 2018, pp. 74772-74777.

[25] T. Tu Le, H. H. Tran, A. A. Althuwayb, "Wideband Circularly Polarized Antenna Based on a Non-Uniform Metasurface", Applied Science, Vol. 10, No .23, 2020.

[26] Y. Dong, T. Itoh, "Metamaterial-Based Antennas", Proceedings of the IEEE, Vol. 100, No. 7, 2012.

[27] G. V. Eleftheriades, K. G. Balmain, "Negative-refraction metamaterials: fundamental principles and applications", John Wiley & Sons, 2005.

[28] H.-X. Xu, G.-M. Wang, J.-G. Liang, M. Q. Qi, X. Gao, "Compact Circularly Polarized Antennas Combining Meta-Surfaces and Strong Space-Filling Meta-Resonators", IEEE Transactions on Antennas and Propagation, Vol. 61, No. 7, 2013, pp. 3442-3450.

[29] C. Zhou, G. Wang, Y. Wang, B. Zong, J. Ma, "CPW-Fed Dual-Band Linearly and Circularly Polarized Antenna Employing Novel Composite Right/Left-Handed Transmission-Line", IEEE Antennas and Wireless Propagation Letters, Vol. 12, 2013, pp. 1073-1076.

[30] B.-C. Park, J.-H. Lee, "Omnidirectional Circularly Polarized Antenna Utilizing Zeroth-Order Resonance of Epsilon Negative Transmission Line", IEEE Transactions on Antennas and Propagation, Vol. 59, No. 7, 2011, pp. 2717-2721.

[31] S.-T. Ko, B.-C. Park, J.-H. Lee, "Dual-Band Circularly Polarized Patch Antenna with First Positive and Negative Modes", IEEE Antennas and Wireless Propagation Letters, Vol. 12, 2013, pp. 1165-1168.

Recyclability and ecological-economic analysis of a simple photovoltaic panel

Original Scientific Paper

Goran Rozing
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Department of Core Courses
Kneza Trpimira 2B, Osijek, Croatia
goran.rozing@ferit.hr

Hrvoje Glavaš
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Department of Power Engineering
Kneza Trpimira 2B, Osijek, Croatia
hrvoje.glavas@ferit.hr

Dina Jukić
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Department of Core Courses
Kneza Trpimira 2B, Osijek, Croatia
dina.jukic@ferit.hr

Matej Žnidarec
J. J. Strossmayer University of Osijek,
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Department of Power Engineering
Kneza Trpimira 2B, Osijek, Croatia
matej.znidarec@ferit.hr

Abstract *– Photovoltaic industry has displayed an exponential expansion rate over the past two decades. Currently, 1185 GW of the global electricity demand is derived from photovoltaic (PV) panels with more to come in the next two decades. With the growing installment rate, a concern regarding managing panels that reached the end of their life-span grows. Despite the fact that the approximate life-span of a panel is 25 years, oftentimes this interval is shortened due to weather conditions, low-quality installations, and poor maintenance. Therefore, a global attention is focused on research and law-regulation of proper recycling and disposal of the PV waste. In this contribution, recyclability of one PV panel is determined upon subjecting it to the simplest recycling process, which is manual mechanical separation using common hand-tools. In addition, an ecological-economic analysis of the panel, which yields an absolute profit of the panel is presented.*

Keywords: *photovoltaic panel, recyclability analysis, ecological-economic analysis, waste management*

1. INTRODUCTION

Since the industrial revolution, the enormous rate of the world's energy demand has resulted in a high consumption of fossil fuels, which has had a detrimental impact on the environment due to a concerning release of greenhouse gases into the atmosphere. Negative environmental impact of such consumption was/is on the urgent agenda of major global organizations as recently confirmed by the United Nation Secretary-General's opening remarks at the press conference on climate where it was stated that "the era of global warming has ended; the era of global boiling has arrived" [1]. Hence, renewable energy resources gained attention and are being developed by generating electric power with nearly zero pollution [2].

However, at the moment, renewable energy supplies short to quarter of electricity in the world. After reflections and absorptions, around 100 000 terawatts of the total irradiated energy from the Sun reaches the Earth's surface, which is a tremendous quantity of "free" energy. Despite this potential, solar energy is only the third contributor in the global clean energy network, after wind and hydro energy [3]. Of the total energy produced globally, 29.9% is produced from renewable sources. Hydropower is the biggest contributor with 15.1% followed by wind and solar energy which amount to 15.1% combined [4]. In spite of the wind and solar energy lower figures, the rate of wind turbines and solar panels instalments is quick, and the outlook seems positive.

Recently, an un-heard rate of expansion was displayed by the photovoltaic industry. Over the last 16 years, power derived from PV panels installed globally increased by over 60 times at the exponential rate, i.e. from 9.1 GW in 2007 to over 300 GW in 2016 [5], whereas the current

state of the art is determined at around 1185 GW [4] of the global electricity demand. This remarkable achievement is possible due to severally combined reasons: decreased production costs of PV panels, increased PV panel efficiency, and improved PV panel dependability. The advancement of PV solar cells is probably the key of the PV industry success. PV cells are essentially a semiconductor that has been engineered to effectively absorb irradiated solar energy and transform it into the electrical energy. PV cells are connected and protected from the environment in PV modules, which are the fundamental components of PV systems. Therefore, the photovoltaic installation capacity would have reached 4500 GW by 2050 is of no surprise [6].

Consequently, it is predicted that solar waste will have accounted for up to 14% of the total generation capacity by 2023, with the assumption that the PV panel lifespan is 25 years. A further projection is 78 million tons of waste by 2050, which is 80% of the overall production capacity [7]. As a result, the use of photovoltaic modules is a substantial environmental, social, and economic challenge in the years to come, particularly in light of the advocated notion of sustainable production and the European Green Deal [8]. The European Green Deal, a document outlining a precise set of targeted steps to combat climate change, was released by the European Commission in 2019. By 2050, Europe must stop producing net greenhouse gas emissions in order to achieve the target. To separate economic progress from resource waste, the development paradigm itself must be modified. It is an ambitious initiative that will have an impact on tens of millions of people which the major institutions of Europe will work tirelessly for years. Europe will be the first continent to attain climate neutrality if the accord is successfully implemented. Some advancements have been made due to the European Union issuing the Guideline 2012/19/EU (replacing the earlier 2002/96/EU) [9]. According to this Guideline, end-of-life photovoltaic panels must be regarded as electrical and electronic equipment waste (WEEE). On the other hand, based on observations of global market realities, there is currently no clear plan for handling waste from the PV industry sector. For example, India is the fifth solar PV power producer in the world, but has no laws which dictate safe disposal of solar energy waste [10].

Aggravating circumstance is a lack of knowledge of the precise quantity and kind of this type of waste that has previously been transferred and put in national landfills. Furthermore, there is no specific statute dealing with the guidelines for processing solar power waste under the applicable laws and there is no regulation over the usage or import of this kind of waste. Also, the conflict between Russia and Ukraine seems to raise a lot of questions about the Green Deal. Due to the scarcity of natural gas from Russia, there are some intentions to start using coal to generate energy, but there is no significant plan with a renewable solution in place for the upcoming years [11].

Therefore, the motivation for this research is a growing concern about managing photovoltaic panels that reached the end of their life-span. For this purpose, two analyses were conducted and are presented in this paper. Firstly, we analyzed the recyclability of a simple photovoltaic panel that was manually mechanically separated in a manner and with tools that the average user would be able to perform. Secondly, we analyzed the cost and income from recycling the panel to determine if recycling is the optimal option at the current state of the art.

2. PV PANEL LIFE-SPAN AND RECYCLING

Solar power solutions can be roughly divided into two types: solar thermal and photovoltaic.

Solar thermal applications include residential hot water heaters, cooking, solar water purification, molten salt energy storage, solar power transmission, fuel production, and Concentrated Solar Power (CSP). The CSP employs mirrors, which are often parabolic troughs or dishes, to direct and track Sun rays on the working fluid. Expanding and vaporizing heated fluid powers a turbine. The conversion of solar energy into heat, which can be easily stored, is a major benefit of CSP. In the case of photovoltaic systems, converted electric energy is more challenging to store despite advancements in the battery technology.

Solar PV panels convert energy directly. Panels are most effective in direct solar radiation, but also operative in diffuse cloudy solar radiation, which makes them a compelling green solution. When Sun rays strike a solar panel, which is still the most frequently made of 99.9999% purity silicon, it releases electrons, i.e., electric current. Not long ago, the only solar panels on the market were constructed of pure silicon, which is expensive and requires a lot of energy to produce. A variety of innovative solar panels has been created as a result of wafer technology research, including those made of cadmium telluride, indium-copper-gallium alloys, and perovskites.

As briefly mentioned in the introductory part, the approximate life-span of a photovoltaic panel is 25 years. More often than not, a panel life-span is shortened due to occurrences as [12]:

- extreme weather conditions;
- catastrophes caused by nature;
- building blazes;
- high rate of low-quality rooftop panel installations;
- financial incentives to remove panels before they reach the end of their life-span;
- discouragement to replace one rooftop panel rather than the entire system;
- improper installation sites;
- poor or no maintenance;
- moisture penetration;
- damage during transit or installation.

In comparison to other electric energy sources, solar PV panels are thought to produce an excessively large quantity of waste per unit of energy [13]. However, due to the rising demand for renewable energy to reduce carbon dioxide emissions, PV panels are to be used more frequently, which is in contrast with the waste management hierarchy that lists prevention of waste as a preferable method above reducing, reusing, recycling, and lastly disposing of the waste. Coincidently, the same reasoning eliminates reducing as a method of PV waste management. The reusing method should get more attention in the future as there is a potential market for second-hand PV panels. After a life-span of operation, a panel's efficiency is lowered to over 80% compared to its initial efficiency due to the natural decay of the photovoltaic cells caused by solar irradiation [14]. This is because the efficiency of the panels declines at a rate of roughly 1% per year. Note that [14] refers to the natural decay and excludes the waste produced by the circumstances previously listed in this section. That said, even if reused, PV panels will eventually become a waste, which leaves recycling as an optimal option. To put in a perspective, half of carbon footprint of producing a photovoltaic panel is due to converting raw materials into cells, cables, and frame, but if a panel is recycled, almost 80% of raw material can be recovered [10].

Currently, PV panels are subjected to one out of three recycling processes: mechanical separation, thermal treatment, and chemical treatment [15].

Mechanical separation is disassembling of PV panels by removing encircling aluminum frames, junction boxes, and imbedded cables. Panels, junction boxes, and cables are the components of PV modules that are shredded and crushed to check each component's individual toxicity and the module's overall toxicity before disposal. The aluminum frame component can be recovered through secondary metallurgy after being detached from the module. Iron, silicon, and nickel are considered small quantity elements that are typical components of aluminum alloys and are further reused.

By thermal and/or chemical treatment, a solar panel is delaminated, its chemical components are separated, and the pure silicon inside is extracted. The outcomes of research in this field are encouraging, but there is still a lack of efficiency on a broad scale. Additionally, this recycling method still faces difficulties with regard to its commercial feasibility due to the creation and required treatment of dangerous gases.

3. RECYCLABILITY ANALYSIS

In this paper, a simple PV panel was disassembled to elements via mechanical separation using basic tools to prepare the PV panel for a recyclability evaluation method. This method analyzes a product as an assembly composed of a certain number of elements [16], and assigns to them dimensionless coefficient R which indicates product's (re)processability potential.

The coefficient R, simply called recyclability, can assume values that separate the product in three categories:

- desirable recyclability $1.00 \geq R \geq 0.75$;
- need for product reconstruction or selective disassembly $0.74 \geq R \geq 0.50$;
- down-cycling procedures (lower quality recyclates) or disposal $R \leq 0.49$.

The photovoltaic panel from the study is model MODULE SL30 AA, manufactured by SOLE. Figure 1 (top) shows the studied panel before mechanical treatments, and Fig. 1 (bottom) shows the panel disassembled using pliers, screwdriver, scalpel, spatula, and blow-dryer. The steps done in manual mechanical separation of the panel into elements are listed in Table 1, as well as the time duration of each step. The next process in recycling would be a chemical separation of metal from the anti-reflective (AR), P, and N layers, but it was omitted in this paper.

Fig. 1. PV panel MODULE SL30 AA under study, before (top) and after (bottom) a manual mechanical separation

The recyclability analysis of the product includes determination of elementary indicators and the calculation of complex indicators. The quantitative results are expressed as ratios of the sum of the recyclability of individual parts and the ideal (maximum) recyclability.

Table 1. Manual mechanical separation of the PV panel from the study.

Procedure	Duration (s)	Cumulative duration (s)	Tool
Separating aluminum frame	87.13	87.13	Pliers
Separating fuse box	31.14	118.27	Screwdriver
Removing screws from the fuse box	44.58	162.85	Screwdriver
Removing wires	204.04	366.89	Pliers
Removing protective foil	359.66	726.55	Scalper
Removing layers (AR, P, N)	2880.31	3606.86	Spatula Blow-dryer

Evaluation of recyclability depends on the product structure, type of material, and types of connections [17]. Finally, the recyclability R is determined as:

$$R = \sum_{i=1}^{n} m_i b_i r_i \frac{1}{M\, r_{max}}, \quad (1)$$

where m_i is the mass of the i-th element in grams, b_i is the number of repetitions of the i-th element in the product, r_i is the recyclability rating of the i-th element, M is the total mass of the product in grams, and r_{max} the highest recyclability rating. Upon determination of the type of the material elements of the PV panel are made of, mass of each element was measured with the analytic scale Kern ALS 220-4N, sensitivity 0,01 g. The results of the recyclability analysis are shown in Table 2. Note that M in Eq. 1 is a sum of mi (Table 2). Based on the data presented in Table 2, and according to Eq. 1, photovoltaic panel recyclability was obtained to be $R = 0.73$. This suggests that the panel should be reconstructed rather than recycled.

Table 2. Recyclability analysis of the PV panel.

Element	Material	Mass (g)	Repetition	Rating of recyclability	Recyclability of the element
i	v_i	m_i	b_i	r_i	$m_i b_i r_i$
Frame	Aluminum	141	1	5	705
Screw	Steel	0.5	4	5	10
Fuse box	Polymer	35	1	2	70
Wire	Copper	22	1	5	110
Foil	Unknown	184	1	2	368
AR+P+N	Hazardous waste	157	1	0	0
Glass cover	Glass	2200	1	4	8800

4. ECOLOGICAL-ECONOMIC ANALYSIS

The ecological-economic analysis model is based on the calculation of cost and income from recycling, and for the assumed method of disposal, it provides the user with data on the absolute profit.

Total recycling cost T is a sum of disassembly T_{da}, shredding T_s, recycling T_r, and disposal costs T_d [18]:

$$T = T_{da} + T_s + T_r + T_d. \quad (2)$$

Total income from recycling P is a sum of income from the sale of recycled materials P_s, income from energy savings P_{es}, and income from emission reduction P_{er} [18]:

$$P = P_s + P_{es} + P_{er}. \quad (3)$$

Finally, the profit from recycling DA is simply the difference between the total income and total costs:

$$DA = T - P. \quad (4)$$

The cost of disassembling of the analyzed PV panel depends on the duration of this process, which is 1.002 hours, as determined from Table 1, and the cost of required labor per hour, whose average is € 10. Hence, disassembling cost is obtained as € 10.02. Upon disassembly, the remaining parts were fairly small (Fig. 1), and further shredding was not required, so $T_s = 0$. Recycling and disposal costs are shown in Table 3, and are predetermined by the cost of recycling/disposing of individual material per unit of mass.

Table 3. Costs of recycling and disposing of the analyzed PV panel

Recycling			
Element	**Cost per kg**	**Mass (kg)**	**Tr**
Aluminum	0.65	0.141	0.0917
Copper	0.65	0.225	0.1463
Polymer	0.65	0.219	0.1424
Glass	0.65	2.200	1.4300
	Total	2.785	1.8103
Disposing			
Element	**Cost per kg**	**Mass (kg)**	**Td**
Hazardous waste	0.40	0.157	0.0628

Therefore, according to Eq. 2, the total recycling cost of the PV panel is $T = 11.89$ €, as shown in Fig. 2.

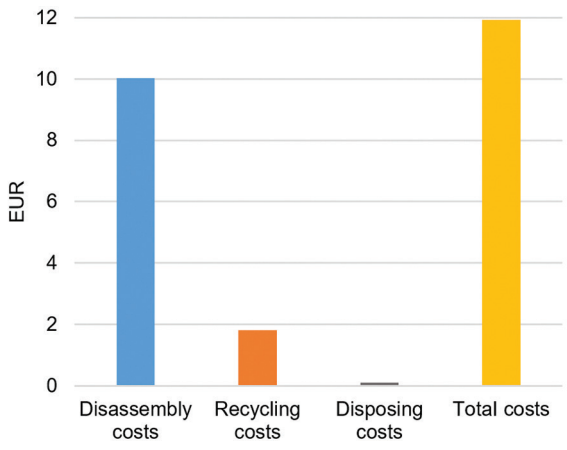


Fig. 2. Costs of recycling.

Table 4. Income analysis.

Recyclate	Mass	Price	Income from sale	Energy savings	Income from energy savings	Emission reduction	Emission fee	Income from emission reduction	Total income
r	M_r (kg)	P_r (€)	$P_s=M_rP_r$	$/\backslash e$ (€/kg)	$Pes=Mr/\backslash e$	E	p_e	$P_{er}=M_rEp_e$	P
Aluminum	0.141	4.20	0.59	1.200	0.17	4.5	0.80	0.51	1.27
Copper	0.225	6.20	1.40	0.300	0.07	1.2	0.80	0.22	1.69
Polymer	0.219	0.18	0.04	0.020	0.05	0.9	0.80	0.16	0.25
Glass	2.200	2.00	4.40	0.600	1.32	1.2	0.80	2.11	7.83

Plugging the data into Eq. 3 gives the total income from recycling one PV panel P = 11.45 €. The income from recycling, according to the type of the material, is shown in Fig. 3.

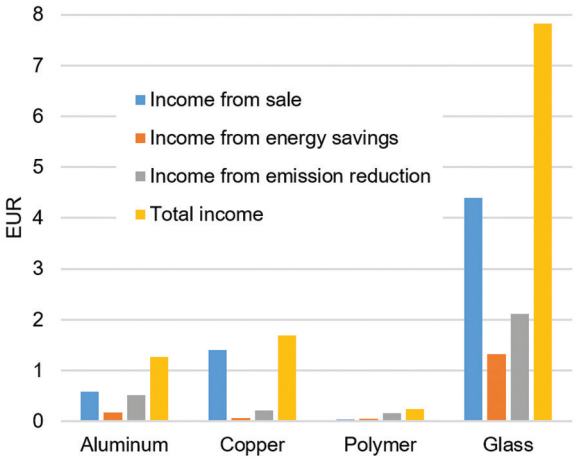


Fig. 3. Income from recycling

The obtained data on the total costs were compared with the total income from recycling, and the profit was calculated according to Eq. 4. Data analysis indicates that the profit from recycling one photovoltaic panel is negative. A user would make a loss of € 0.44 to manually recycle one PV panel on his own. This outcome is due to the high cost of disassembly caused by the long process of disassembling the PV panel. In addition, the amount of hazardous waste to dispose is large, relatively to the total mass of individual elements of the PV panel that can be recycled.

5. CONCLUSIONS

In this paper, recyclability and ecological-economic analyses of one simple photovoltaic panel are presented. To prepare the PV panel for the analyses, a manual mechanical separation of the panel into the building elements was performed using common hand tools and blow-dryer. Disassembling process resulted with 7 different components which took 6 operations that lasted for over 1 hour in total. The most challenging operation was removing the anti-reflective, P, and N layers from the panel, for which a spatula and blow-dryer was used.

The recyclability analysis showed that the PV panel recyclability is 73%, which indicated that this product should be either reconstructed or further selectively disassembled, rather than recycled after a mechanical separation only.

By applying the model of the ecological-economic analysis, it was determined that there is a negative profit from recycling the PV panel. The analysis showed that there is € 0.44 loss for recycling and disposing of one panel. The loss is the consequence of a long procedure of disassembling the panel into components, high share of hazardous waste, which requires disposing, in comparison to materials that can be recycled, and a certain share of unknown polymers that were not labeled on the product.

Therefore, to increase the recyclability of PV panels, manufacturers should adhere to the design guidelines for increasing the recyclability of electronic devices and equipment. One way to achieve this is by using well-known polymers that have a label of their type, which facilitates the sorting of parts and achieves a higher price for recycling material. Additionally, the use of simpler components would make the construction more modular and would result in easier and faster disassembly of PV panels.

Note that the estimates of prices were used in this paper, therefore, a certain error may have been introduced into the outcome of the ecological-economic analysis. Considering that the method of manual disassembly of PV panels did not give a positive profit, the chosen method of disassembly is not ideal. It is recommended to use a combination of a manual and strictly mechanical disassembly to increase the process speed.

6. REFERENCES

[1] United Nations, "Secretary-General's Opening Remarks at Press Conference on Climate", https://www.un.org/sg/en/content/sg/speeches/2023-07-27/secretary-generals-opening-remarks-press-conference-climate (accessed: 2023)

[2] P. Nema, R. K. Nema, S. Rangnekar, "A Current and Future State of Art Development of Hybrid Energy

System Using Wind and PV-Solar: A Review", Renewable and Sustainable Energy Reviews, Vol. 13, No. 8, 2009, pp. 2096-2103.

[3] T. Z. Ang, M. Salem, M. Kamarol, H. S. Das, M. A. Nazari, N. Prabaharan, "A Comprehensive Study of Renewable Energy Sources: Classifications, Challenges and Suggestions", Energy Strategy Reviews, Vol. 43, 2022, p. 100939.

[4] REN21, "Renewables 2023: Global Status Report", UN Environment Programme, https://www.ren21.net/gsr-2023 (accessed: 2023)

[5] T. Letcher, V. M. Fthenakis, "A Comprehensive Guide to Solar Energy Systems: With Special Focus on Photovoltaic Systems", Academic Press, 2018.

[6] M. Niekurzak, W. Lewicki, H. H. Coban, A. Brelik, "Conceptual Design of a Semi-Automatic process line for Recycling Photovoltaic Panels as a Way to Ecological Sustainable Production", Sustainability, Vol. 15, No. 3, 2023, p. 2822.

[7] Statista, "Projected Cumulative Volume of Solar Photovoltaic Panel Waste Worldwide from 2016 to 2050", https://www.statista.com/statistics/565743/solar-pv-panels-waste-projection-worldwide/ (accessed: 2023)

[8] EUR-Lex-52019DC0640-EN, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52019DC0640 (accessed: 2023)

[9] "Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on Energy Efficiency, Amending Directives 2009/125/EC And 2010/30/EU and Repealing Directives 2004/8/EC and 2006/32/EC", https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02012L0027-20200101&from=NL#:~:text=This%20Directive%20establishes%20a%20common,way%20for%20further%20energy%20efficiency (accessed: 2023)

[10] J. Malaviya, "Need for Legal Compliance for Safe Solar Energy Waste Disposal", Proceedings of the European Photovoltaic Solar Energy Conference and Exhibition 2020, Online: Virtual Conference and Exhibition, 7-11 September 2020, pp. 1204-1208.

[11] F. Simon, "With Russian Gas Gone, Coal Makes EU Comeback as 'Traditional Fuel'", https://www.euractiv.com/section/coal/news/with-russian-gas-gone-coal-makes-eu-comeback-as-traditional-fuel/ (accessed: 2023)

[12] P. Majewski, W. Al-Shammari, M. Dudley, J. Jit, S. H. Lee, K. Myoung-Kug, K. Sung-Jim, "Recycling of Solar PV Panels-Product Stewardship and Regulatory Approaches", Energy Policy, Vol. 149, 2021, p. 112062.

[13] S. F. Baldwin, "Quadrennial Technology Review: An Assessment of Energy Technologies and Research Opportunities", USA Department of Energy, 2015.

[14] V. Sharma, S. Chandel, "Performance and Degradation Analysis for Long Term Reliability of Solar Photovoltaic Systems: A Review", Renewable and Sustainable Energy Reviews, Vol. 27, 2013, pp. 753-767.

[15] M. S. Chowdhury, K. S. Rahman, T. Chowdhury, N. Nuthammachot, K. Techato, M. Akhtaruzzaman, S. K. Tiong, K. Sopian, N. Amin, "An Overview of Solar Photovoltaic Panels' End-of-Life Material Recycling", Energy Strategy Reviews, Vol. 27, 2020, p. 100431.

[16] M. Kljajin, M. Opalić, A. Pintarić, "Recikliranje električnih i elektroničkih proizvoda", Mechanical Engineering Faculty in Slavonski Brod, Slavonski Brod, Croatia, 2006. (in Croatian)

[17] H. Martens, D. Goldman, "Recyclingtechnik: Fachbuch fur Lehre und Praxis", Springer, 2016.

[18] A. Pintarić, "Assessment of Recyclability for Electrical Waste", Proceedings of the 13th International Symposium on Waste Management, Zagreb, Croatia, 6-7 November 2015, pp. 347-355.

Comprehensive Analysis of Power System: Exploring Load Factor, Power Balance, Active Load Variation, and Increment Factors with Iterative Implications

Original Scientific Paper

Agron Bislimi
UBT, University for Business and Technology
St. Calabria n.n. 10000, Prishtina, Kosovo
agron.bislimi@ubt-uni.net

Abstract *– The contemporary power landscape is marked by escalating complexity driven by surging energy demands, the integration of renewable sources, and the imperative for heightened system performance. In response to these challenges, this paper presents a thorough investigation. Firstly, the load factor Lambda (λ) is expounded upon, emphasizing its significant impact on system stability due to load characteristics. Static load models are employed for voltage stability studies, providing insights into resource allocation and optimization by comparing actual energy consumption to the maximum potential consumption. Secondly, the paper delves into the power balance, a critical aspect that scrutinizes the equilibrium among generation, consumption, and distribution, underscoring its pivotal role in ensuring system stability. Within a 6-node network, the concept of power balance, denoted as "load λ," is illustrated by tracking its variations across iterations, representing the disparity between Power Generator (PG) and Load (PL). To maintain balance during continuous power flow (CPF) analysis, power supply can be finely adjusted using the load factor lambda (λ). Thirdly, the integration of renewable sources introduces active load variation, underscoring the necessity to comprehend load fluctuations over time for ensuring grid reliability. Active load variation is demonstrated based on the number of iterations. Additionally, the paper elucidates the increment factor τ, explaining its impact on the number of correction iterations by selecting the size of the step factor. The graphical representation of the increment factor for iterations 8 and 11 in a network with load nodes provides further clarification. This paper explores the intricate interactions of load factors, power balance, active load changes, and increment factors within power systems. The presented findings contribute to the enhancement of the reliability, efficiency, and sustainability of modern power systems.*

Keywords: *Power System Analysis, Load factor, Power balance, Active Load Variation, Increment Factor*

1. INTRODUCTION

This paper presents a comprehensive investigation into various critical dimensions that collectively influence the behavior and performance of modern power systems. With a focus on addressing energy needs and the integration of renewable sources, our analysis encompasses key aspects of power systems, including Load Factor (λ), Power Balance, Active Load Variation, and Increment Factors (τ). These dimensions play a pivotal role in guiding decisions for the efficient, reliable, and sustainable operation of power systems. Efficiency and resource allocation within power systems are crucial considerations, and our paper quantifies efficiency by comparing actual energy consumption to potential usage. Emphasizing the significance of maintaining equilibrium among generation, consumption, and distribution, we delve into the role of this balance in ensuring system stability and an uninterrupted power supply. In exploring load patterns, particularly with the integration of renewable energy, our study investigates their fluctuations and the ensuing impact on power system reliability. Furthermore, we examine how small adjustments in increment factors can significantly influence power system responses under varying conditions, taking into account the iterative nature of these adjustments. The paper aims to provide a holistic understanding of the interconnected nature of load factors, power balance, active load variation, and increment factors within contemporary power systems. By doing so, it seeks to contribute to informed decision-making that optimizes the reliability, efficiency, and sustainability of power systems in the ever-evolving energy landscape.

The Load Factor (λ), examined in [1-3], plays a pivotal role in evaluating energy efficiency and resource allocation. As investigated in [4-7], the Power Balance highlights its crucial significance in maintaining system equilibrium. Active Load Variation, explored in [8, 9], [10], delves into the fluctuations in load patterns and their consequential impact on system stability and reliability. Increment Factor, examined in [11-13], contributes to understanding its role in power system responses to varying conditions. These cited references collectively form the foundation for a comprehensive analysis of these dimensions within power systems. References [14-16] collectively contribute to the advancement of power system analysis. In [14, 15], a detailed analysis assesses the scalability and performance of a proposed formulation on a substantial 2383-bus Polish test network. Reference [16] investigates maximum loadability and transfer capability in a Power System (PES) through systematic load condition adjustments. For a detailed understanding of security, reliability in power system stability, including its definition and classification, refer to [17-20].

In [1], load characteristics play a crucial role in system stability. CPF analysis incrementally raises the load using the load factor Lambda (λ) before each iteration, beginning from base load values Po, Qo. Accurate step factor adjustments are essential for corrector convergence along the power-voltage (PV) curve, identifying the stability limit by the sign change of dLambda. Notably, convergence slows near the critical point, particularly in its vicinity. Moreover, corrector convergence is swifter with the continuous parameter λ than with the continuous parameter V.

In [2], the calculation of network reliability involves simulating interruptions of network components (lines, transformers, buses) and analyzing the effects of failures, including checking for violations of voltage, current, and reactive power limits.

In [3], incorporating the continuation technique into load flow analysis requires integrating the load parameter lambda (λ) within the equations. This is especially pertinent for models with constant load characteristics. Both the load (indexed as L) and generation (indexed as G) at a node depend on the load parameter.

To replicate load fluctuations, adjustments are applied to load values PLi and QLi. Changes in active power generation follow the equation: $PGi = PGio\ (1 + KGi\ \lambda)$, where $PGio$ represents the active generation at the node during the base scenario. The constant KGi is utilized to calibrate the degree of generation adjustment [4].

The selection of the projected step factor's magnitude is tailored to align with the existence of a load flow solution, as dictated by the provided continuous parameter. If an iteration fails to converge with the current step factor size, the magnitude is systematically decreased. This corrective process is reiterated until a solution is successfully obtained [5].

In [6], the characterization of nodes within an electrical network identifies three primary types: reference node (slack node), PQ (load node), and PV (generator node). The reference node plays a crucial role in maintaining the power system's equilibrium.

In a steady-state balanced power system, as outlined in [7], requirements include meeting generation demands, maintaining bus voltage levels, operating generators within specified power limits, and preventing overloading of transmission lines and transformers.

In references [8] and [9], the analysis focuses on the stability of voltage in electrical power networks, emphasizing the importance of maintaining voltage within permissible limits. This is particularly significant as networks may encounter voltage fluctuations due to the continuous expansion of renewable energy sources. The paper further investigates the causes of voltage drop and explores measures taken by consumers, actions implemented in the network, and generation-related strategies to mitigate these effects.

In [10], a novel method is presented for controlling DC bus voltage in a hybrid energy storage system, contributing to enhanced stability and reliability for renewable energy systems.

In [11], the iterative process involves selecting the size of the projected step factor in a manner that aligns with the existence of a load flow solution based on the continuous parameter. If an iteration fails to converge with the current step factor size, an adaptive approach is employed, systematically reducing the magnitude until a successful solution is achieved.

In [12], an analysis of the transmission system's role in voltage stability is presented, focusing on two fundamental concepts: the maximum power deliverable to the loads and the relationship between load power and network voltage. Failure to meet these essential criteria can lead to voltage instability.

In [13], the CPF algorithm's step length control is crucial for efficiently identifying the nose point. This control offers a choice between a constant or adaptive approach, providing flexibility to overcome convergence challenges as the system nears its loading limit. As the curve approaches the loading point's pinnacle, step sizes naturally decrease, refining the trajectory toward maximum loading conditions.

In [14, 15], an exhaustive examination was conducted, presenting a detailed and comprehensive analysis of the scalability and performance assessment of a proposed formulation. This evaluation was executed on a sizable 2383-bus Polish test network, maintaining loading conditions from a 30-bus test case. The findings showcased the robustness of the formulation, with a particular focus on the LU transversality condition and the behavior of geig. The study emphasizes the significance of the network size in influencing the convergence rates for gLU and $geig$, while gQR and gsvd demonstrated a more

consistent performance, less dependent on the network size. Moreover, adjustments in the Newton step size (α) were imperative for convergence in larger test cases.

In [16], the study investigates the maximum loadability and transfer capability in a Power System (PES) through systematic adjustments of load conditions at each bus. This allows for loadability assessment under three scenarios: Scenario 1 explores maximum loadability without technical constraints, Scenario 2 introduces voltage magnitude constraints, and Scenario 3 adds constraints on generation power output, line thermal limits, and voltage magnitude. The paper focuses on assessing interconnection feasibility, aiming to determine the maximum surplus capacity (estimated at around 200 MW) available for export to the Democratic People's Republic of Korea (DPRK), without compromising the security and stability of the PES grid. The analysis considers strict constraints on voltage levels, active and reactive generations, and line limits for system reliability. Results in [16] reveal a maximum loadability margin (λmax) of 2.8854 and a Total Transfer Capability (TTC) of 57.71 MW, offering crucial insights into operational limits and potential interconnection opportunities.

In [17], key power system security analysis methods, voltage stability investigation, and their integration into reliability analysis are highlighted. The real power system operating near capacity, along with the evaluation of measures to enhance reliability, is discussed for a clear overview.

In [21], the focus is on the selection and control of the step size, a crucial factor for the success of continuous power flow. Flexibility in step size control is paramount to adapting to the power system, thereby facilitating strong convergence. The introduction of the load parameter lambda results in the formation of the nonlinear equation $F(V,\delta_2)\ \lambda,) = 0$.

In [22], the paper introduces a cost-effective power factor (PF) metering system using an Arduino microcontroller, emphasizing the importance of power factor in preventing energy costs and outages.

The remainder of the paper is structured as follows: Section 2 provides a detailed explanation of the methods used. Section 3 presents the Results, which are divided into four subsections: Subsection 3.1 analyzes the load factor λ. Subsection 3.2 examines the balance of power as a function of the number of iterations. Subsection 3.3 demonstrates active load as a function of the number of iterations. Subsection 3.4 illustrates the increment factor τ and explained with an example. Section 4 presents the Conclusions.

2. METHODS: LOAD FACTOR, POWER BALANCE, AND ITERATION-RELATED PARAMETERS

2.1 THE LOAD FACTOR (λ) IN CPF.

Four operating states, namely Normal, Vulnerable, Disturbed, and Recovered, are conventionally established to ensure the safety of the electrical network. Transitions between these states also play a significant role.

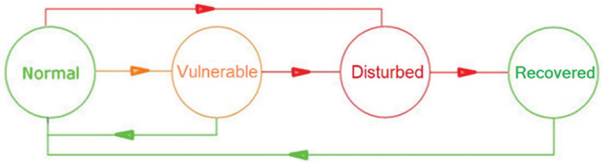


Fig. 1. The states of an energy network

A steady-state balanced power system necessitates adherence to the following conditions: Generation must satisfy both demand (load) and losses, bus voltage magnitudes should remain close to their rated values, generators must operate within specified limits for real and reactive power, and transmission lines and transformers must avoid overloading.

The stability of the system relies significantly on load characteristics. Studies on voltage stability often utilize static load models, including constant power, constant current, or constant impedance models. Additionally, load models may incorporate polynomial functions with voltage magnitude as a parameter [1].

The interplay between the concepts of power system reliability, security, and stability is extensively discussed in both theoretical frameworks and practical applications, as elaborated upon in more detail in reference.

In general, an electrical network is considered safe when the generation capacity surpasses the cumulative load demand, transmission elements operate within specified load thresholds, node voltages adhere to acceptable limits, the network exhibits resilience against generator failures ensuring continued functionality, the network accommodates transmission line losses without significant repercussions, and stability prevails within the network even under short-circuit conditions.

The load characteristic in voltage stability analysis is modeled by a function that expresses the active and reactive power in terms of the voltage magnitude (V) and an independent variable (λ), referred to as the load change parameter: $P = P(\lambda, V)$, $Q = Q(\lambda, V)$. Security refers to the network's capacity to withstand unforeseen disturbances, such as short circuits or unexpected component losses [18-20].

It effectively communicates that in the context of CPF analysis, the Load Factor (λ) plays a crucial role in evaluating system behavior and stability. It also highlights the impact of varying load factors on different aspects of power systems in CPF scenarios.

In CPF analysis, the Load Factor represents the degree of load demand on the system relative to its rated capacity. It directly affects voltage profiles and load distributions. A higher load factor places the system closer to its capacity limits, raising concerns about voltage stability and system reliability. Understanding the

interplay between load factor variations and system stability is paramount in CPF analysis. Voltage stability is a central concern in CPF studies. Load factor variations can lead to voltage instability or even collapse. By examining the correlation between load factor increments and voltage stability margins, we gain insights into critical load factor thresholds that signal potential voltage instability.

2.2 BALANCE OF POWER AS A FUNCTION OF THE NUMBER OF ITERATIONS IN CPF

Understanding power balance in CPF scenarios is crucial for comprehending the system's behavior as conditions evolve. The power balance undergoes changes with an increasing number of iterations in CPF analysis. Investigating this evolution provides valuable insights into how power distribution adjusts in response to changing conditions. Observing power balance trends over iterations aids in predicting convergence behavior and helps identify scenarios where the balance is compromised. Through iterative analysis, the Balance of Power helps gauge the system's stability during CPF iterations. Deviations from balanced power conditions can signify instabilities or impending voltage issues. By correlating power balance variations with system response, we gain a deeper understanding of how the system adapts and stabilizes under different loading scenarios. Electric power systems, by their nature, require a delicate equilibrium, where electricity demand aligns precisely with supply at all times.

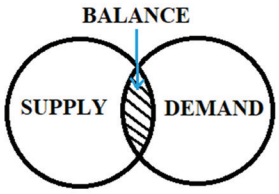


Fig. 2. Supply-Demand balance

In electrical power systems, maintaining a delicate equilibrium between supply and demand is crucial to prevent grid failures. Employing adaptive mechanisms becomes imperative to counteract fluctuations and ensure the system's stability. System operators play a pivotal role in this process, continuously monitoring relevant parameters to prevent any disturbances beyond specified threshold limits [7, 22].

2.3 ACTIVE LOAD IN THE FUNCTION OF THE NUMBER OF ITERATIONS.

In the realm of CPF analysis, a profound comprehension of how active load variations influence system behavior is paramount for accurate predictions and decision-making. Active load variations exert a direct impact on power distribution and voltage profiles. Observing the influence of active load changes on system response over iterations unveils discernible patterns, facilitating effective load management strategies. The analysis of active load enhances our ability to maintain balanced power conditions, preventing undue stress on components.

The resilience of the system in CPF analysis is significantly influenced by active load variations. A sudden surge in active load can lead to overloading and voltage instability. The study of the intricate relationship between active load, iteration count, and system response enables the identification of critical points where interventions such as load shedding or redistribution may be imperative to ensure the integrity of the system [5, 12].

2.4 THE INCREMENT FACTOR τ

The Increment Factor Tau (τ) dictates the rate at which power flow adjustments occur during CPF iterations. A well-chosen tau value ensures convergence while preventing oscillations. The selection of tau significantly impacts the stability of CPF analysis, necessitating a careful balance between rapid convergence and system stability. We define the Increment Factor τ as a critical parameter controlling the rate of power flow adjustments during CPF iterations, influencing iteration convergence and stability. This passage outlines the strategy employed to select appropriate values for the Increment Factor τ.

In the Predictor step, the main goal is to calculate the tangent vector component, represented as 't.' The estimated predictor can be computed as follows

$$\begin{bmatrix} \underline{\delta}_i \\ \underline{V}_i \\ \lambda_i \end{bmatrix} = \begin{bmatrix} \underline{\delta}_{(i-1)} \\ \underline{V}_{(i-1)} \\ \lambda_{(i-1)} \end{bmatrix} + \tau \begin{bmatrix} d\underline{\delta} \\ d\underline{V} \\ d\lambda \end{bmatrix} \quad (1)$$

Tau (τ) is symbolically represented as a scalar, serving as the step size factor in Equation 2.

$$\underline{It}_{ip} = \underline{it}_{(i-1)} + \tau * \underline{t} \quad (2)$$

$\underline{It}_{ip}$ refers to values after the predictor step, while $\underline{it}_{(i-1)}$ represents initial values before the predictor step, where 'i' is the iteration number. The estimated predictor step values require subsequent correction [8, 9].

3. RESULTS: LOAD FACTOR, POWER BALANCE, AND ITERATION-RELATED ANALYSIS

3.1. THE LOAD FACTOR λ.

During the course of analyzing the CPF, the load increases with the load factor λ before each CPF iteration, starting from the base load (P_0, Q_0), for example, in the 17-node network with λ=0.15.

$P_0 5$=1.00; $Q_0 5$=0.80; K=1.0

$P5=P_0 5\ (1+K\lambda)$*100= 1.0*1.15*100 = 115 MW,
$Q5=Q_0 5\ (1+K\lambda)$ *100= 0.8*1.15*100 = 92 MW.

K is a calibration factor that can be assumed differently for active and reactive loads, and it can also vary

for individual nodes. However, in the present investigations, it is consistently set to 1.0.

To ensure that the balance (supply–load) remains stable during the CPF analysis, the supply can also be increased proportionally with the load factor. However, this adjustment did not occur in all calculated networks. In cases where supplies were kept constant, the load increased and shifted to the reference.

Table 1. Active and Reactive Power values for iterations 1,14,19, and 32 at Node 5 in the 17-node Network.

Iteration	P5 (MW)	Q5 (MVAr)
Initial ($\lambda = 0$ p.u.)	100	80
1st ($\lambda = 0.15$ p.u.)	115	92
14th ($\lambda = 3.1817$ p.u.)	418.17	334.536
19th ($\lambda = 3.1547$ p.u.)	415.47	332.376
32nd ($\lambda = 0$ p.u.)	100	80

In Table 1, it is observed that the load reaches its maximum (critical point) in iteration 14 with a load factor (λ) of 3.1817. By iteration 32, the base load is once again attained, but the operating point is now on the lower part of the PV-curve.

3.2. BALANCE OF POWER AS A FUNCTION OF NUMBER OF ITERATIONS

In Fig. 3, during the first iteration, the load (PL) is 10 MW. Since the generated power (PG) remains constant for all iterations, there is a surplus of 90 MW in the grid, sourced from the slack node.

As lambda increases, the load on the grid rises to 30 MW in iteration 2. The power difference between generation and load gradually decreases, reaching zero by iteration 5. Starting from iteration 6, the grid load surpasses the generation, and the difference is supplied from the reference node.

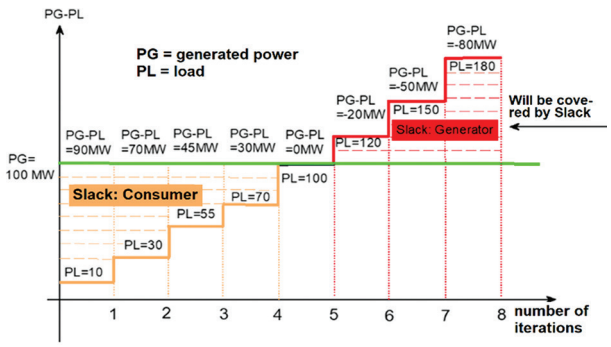


Fig. 3. Power balance PG-PL as a function of the number of iterations. Orange: Slack is a consumer; Red: Slack is a generator.

3.3 ACTIVE LOAD IN THE FUNCTION OF THE NUMBER OF ITERATIONS

Fig. 4 presents the active load ('Load') in the 6-node network as a function of the number of iterations.

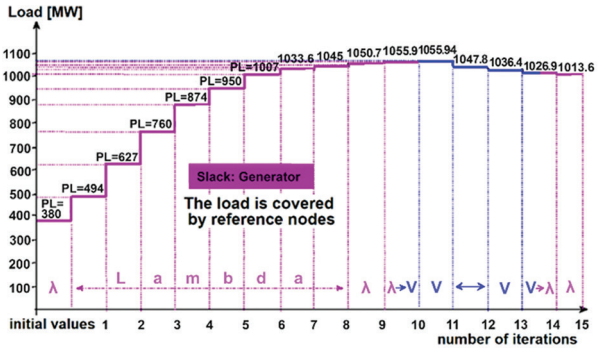


Fig. 4. Network load as a function of the number of the iterations in the 6-node network (Load)

The network consists of 3 load nodes with no additional supply except for the reference node (slack). In Iteration 1, this covers a grid load of 380 MW. From Iteration 1 to 9, the parameter lambda (λ) is treated as a continuous parameter (depicted in purple). In Iteration 10, there is a transition from the parameter lambda (λ) to V ($\lambda \rightarrow V$).

From Iteration 11 to 13, the voltage V is employed as the continuous parameter (depicted in blue). In Iteration 14, there is a shift from parameter V to λ ($V \rightarrow \lambda$) once again (depicted in purple, representing the lower part of the curve). The load reaches its maximum in Iteration 11, marking the critical point, and the descending part of the load curve corresponds to the lower part of the PV curve. Fig. 3 displays only 15 out of the total 24 iterations.

3.4. INCREMENT FACTOR τ

There is a tendency for the corrector to converge more slowly the closer it gets to the critical point. Furthermore, the corrector converges more slowly with the continuous parameter V than with the continuous parameter λ. Slow convergence also occurs when the condition to change the continuous parameter λ to V is fullfilled by multiple nodes simultaneously. Therefore, in order to achieve convergence of the corrector in each part of the PV curve, the size of predicted step factor must be adjusted in each CPF iteration. When the critical point approaches, it reduces; after it exceeds (in the lower part of the PV curve), it can be increased again. For example, in the case of a 6-node network near the critical point, with a step size factor τ=0.01, approximately 8 corrector steps are needed for the iteration to converge.

If the step size factor is reduced to τ=0.0075, only 5 correction steps are required, as illustrated in brown in Fig. 5.

The increment factor for iterations 8 and 11 in the 6-node network labeled 'Load'

In Fig. 5, the impact of the selection of the step factor size on the number of correction iterations is illustrated. Ideally, each iteration should converge within 4-5 iterations for a given step factor size. The brown color in Fig. 5 indicates the increased number of iterations due to the larger step factors.

For a step size factor of 0.03, 5 correction steps are required for convergence. Increasing the step size from τ=0.03 to τ=0.035 necessitates 6 corrector iterations. In iteration 11, voltage V4 serves as a continuous parameter. With an increment factor of τ=0.025, only 2 correction steps are necessary. If the step size factor is then increased to τ=0.035, the corrector converges in just 3 iterations. It's important to note that if the selected step factor is too small, while reducing the number of correction iterations, it significantly increases both the number of CPF iterations and calculation time.

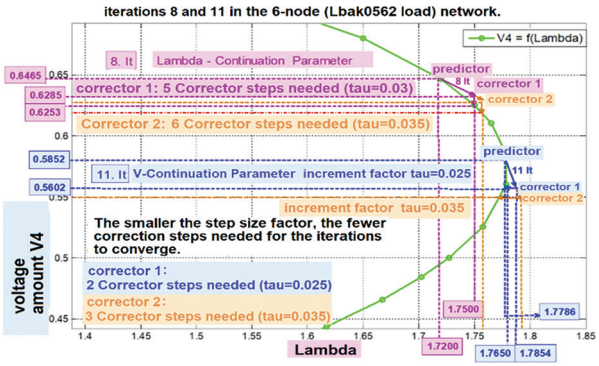


Fig. 5. Step size factor (τ) for iterations 8 and 11 in a 6-node network labeled 'Load'.

In the CPF iteration 8, the parameter lambda is used as a continuous parameter.

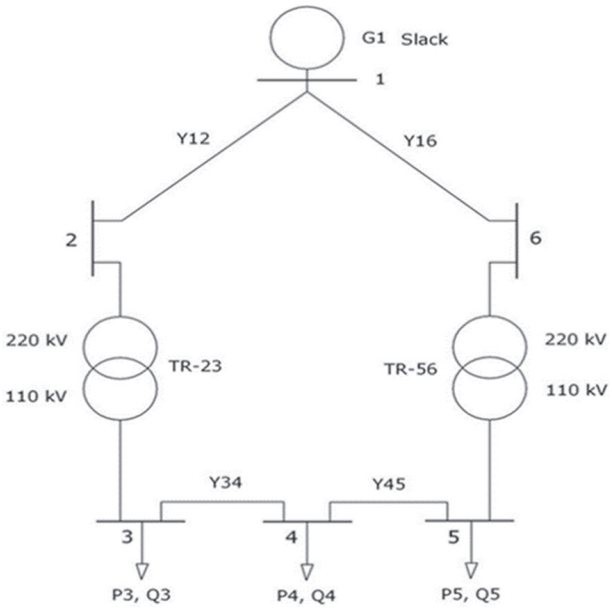


Fig. 6. 6-node network labeled 'Load'

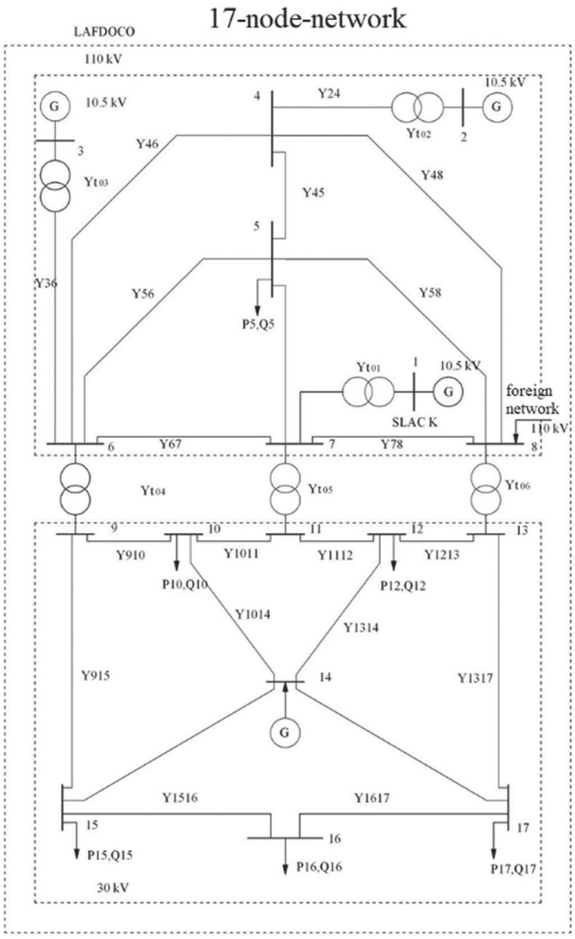


Fig. 7. 17-Node Network

4. CONCLUSIONS

In this paper, we investigate the interdependencies shaping power system behavior. Our comprehensive exploration, spanning load factor, power balance, active load variation, and increment factor, yields valuable insights that contribute to an enhanced understanding of power system operations.

Firstly, the load factor Lambda (λ) is explained. The examination of load factor λ emphasizes its significance in evaluating system efficiency and utilization. By comparing actual energy consumption to the maximum potential consumption, it provides a clear assessment of resource management, essential for optimizing power generation and distribution.

Next, the practical demonstration of power balance in the network is explained and demonstrated over several iterations.The balance is depicted by the difference between Power Generator (PG) and Load (PL). To maintain balance during CPF analysis, supplies can be increased with the load factor lambda (λ). When supplies are constant, load increases shift to the reference node. This investigation highlights the pivotal role of power balance in maintaining system stability

Third, the study of active load variation has highlighted the challenges posed by the integration of re-

newable sources. Understanding load fluctuations over time is crucial for adapting grid strategies and ensuring a reliable electricity supply. Practically, through an example in the 6-node network, "$Load$" is calculated using the CPF program, demonstrating the active load as a function of the number of iterations.

Finally, the analysis delves deeper into the increment factor τ, illustrating the influence of the step size factor on correction iterations. Graphical explanations for iterations 8 and 11 in a 6-node network (Load) are provided. Increased step size requires more correction steps for convergence, while a too-small factor reduces correction iterations but significantly extends CPF iterations and calculated time. This analysis underscores the crucial role of increment factors τ in optimizing power system performance.

The interaction among load factors, power balance, active load variation, and increment factors unveils complex system connections. This understanding facilitates the development of strategies to improve the reliability, efficiency, and sustainability of modern power systems. The insights gained contribute to optimizing power system performance, supporting a smooth transition toward a future of reliable and sustainable energy supply in the evolving energy landscape.

5. REFERENCES

[1] A. Bislimi, "Influence of voltage stability problems on the safety of electrical energy networks", Institute for Electrical Systems and Energy Economics, Vienna University of Technology, Austria, 2012, PhD Thesis.

[2] G. Theil, "Outage data analysis-the base for high voltage network reliability assessment", Proceedings of the IEEE Bologna Power Tech Conference, Bologna, Italy, 23-26 June 2003.

[3] V. Ajjarapy, C. Christy, "The Continuation Power Flow: A Tool for steady state Voltage stability analysis", IEEE Transactions on Power Systems, Vol. 7, No. 1,1992, pp. 416-423.

[4] V. Ajjarapy, "Computation Techniques for Voltage Stability Assessment and Control", Springer, 2006.

[5] P. Kundur, "Power system stability and control", McGraw-Hill, 1994.

[6] S. J. Chapman, "Electric Machinery and power system fundamentals", McGraw-Hill Education, 2001.

[7] J. D. Clover, M. S. Sarma, T. J. Overbye, "Power system analysis and design", Fourth edition, Thomson, 2008.

[8] A. Bislimi, "Analysis of Convergence Behavior and Derivation of Divergence Indicator in Continuation Power Flow Iterations", International Journal on Energy Conversion, Vol. 11 No. 3, 2023.

[9] A. Bislimi, "Illustration of the voltage stability by using the slope of the tangent vector component", International Journal of Electrical and Computer Engineering Systems, Vol. 14 No. 6, 2023, pp. 725-732.

[10] H. Guentri, A. Dahbi, T. Allaoui, S. Aoulmit, A. Bouraiou "Development of a Control Strategy for the Hybrid Energy Storage Systems in Standalone Microgrid", International Journal of Electrical and Computer Engineering Systems, Vol. 14 No. 5, 2023, pp. 575-584.

[11] A. B. Neto, L. R. A. G. Filho, D. A. Alves, "Continuation Power Flow: A Parameterization Technique and Adaptive Step Size Control", Proceedings of IEEE URUCON, Montevideo, Uruguay, 24-26 November 2021, pp. 75-79.

[12] T. Cutsem, C. Vournas, "Voltage stability of electric power systems", Norwell, Kluwer, 1998.

[13] P. S. Nirbhavane, L. Corson, S. M. H. Rizvi, A. K. Srivastav, "Three-phase Continuation Power Flow Tool for Voltage Stability Assessment of Distribution Networks with Distributed Energy Resources", IEEE Transactions on Industry Applications, Vol. 57, No. 5, 2021, pp. 5425-5436.

[14] A. Mazhar, A. Dymarsky, K. Turitsyn. "Transversality enforced Newton–Raphson algorithm for fast calculation of maximum loadability", IET Generation, Transmission & Distribution, Vol. 12, No. 8, 2018, pp. 1729-1737.

[15] D. Baluev, A. Mazhar, E. Gryazina. "State of the art approach for comprehensive power system security assessment—Real case study", International Journal of Electrical Power & Energy Systems, Vol. 155, 2024, p. 109594.

[16] R. D. Zimmerman, C. E. Murillo-Sánchez, R. J. Thomas, "MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education," IEEE Transactions on Power Systems, Vol. 26, No. 1, 2011, pp. 12-19.

[17] G. Theil, "Security and reliability analysis of high voltage transmission systems", e & i Elektrotechnik

und Informationstechnik, Vol. 121, 2004, pp. 435-439. (in German)

[18] P. Kundur et al. "Definition and Classification of Power System Stability IEEE/CIGRE Joint Task Force on Stability Terms and Definitions", IEEE Transactions on Power Systems, Vol. 19, 2004, pp. 1387-1401.

[19] A. Bose, "Definition and Classification of Power System Stability ", Electra, Vol. 208, 2003, pp. 75-79.

[20] M. Shahidehpour, W. Tinney, Y. Fu, "Impact of Security on Power Systems Operation", Proceedings of the IEEE, Vol. 93, No. 11, 2005, pp. 2013-2025.

[21] Y. Mansour, "Suggested techniques for voltage stability analysis", IEEE Power Engineering Subcommittee Report 93TH0620-5- PWR, 1993.

[22] A. M. T. I. Al-Naib, B. A. Hamad, "A Cost-Effective Method for Power Factor Metering Systems", International Journal of Electrical and Computer Engineering Systems, Vol. 13 No. 5, 2022, pp. 409-415.

Design and simulation of a building-based off-grid photovoltaic microgrid using PVsyst: A case study

Case Study

Anjana Saikia
Tezpur University,
Research Scholar, Department of Electronics and Communication Engineering
Napam, Sonitpur, India
anjana.saikia113@gmail.com

Santanu Sharma
Tezpur University,
Faculty of Engineering, Department of Electronics and Communication Engineering
Napam, Sonitpur, India
santanu.sharma@gmail.com

Abstract *– In the absence of a main or central grid, an off-grid renewable energy-based system could be a viable solution to address the electricity demand of a particular region by utilizing the available renewable energy sources (RES) of that area. This also leads society to a step toward sustainable energy development. A planned RES-integrated microgrid system not only fulfills the energy crisis and reduces electricity costs but also plays a significant role in the conservation of fossil fuels. Therefore, the design and simulation of RES integrated systems are indispensable as this type of analysis determines the vital parameters of the system e.g., the system performances including electricity output, energy conversion efficiency, yearly energy potential, and the system losses before being involved in practical implementation. Further, the cell temperature of the Photovoltaic (PV) panel based on regional weather conditions, the system life cycle based on the uncertainty factor of weather, total system installation cost with a full account of the annual savings and payback etc. can also be determined by designing and simulating such systems. This paper presents a design of a 40 kW off-grid photovoltaic (PV) microgrid system according to the load requirements at the Department of Electronics and Communication Engineering (ECE), Tezpur University, India using PVsyst software. The proposed design has been created to meet the daily peak load demand of 37 kW of the department. The energy available through the generating units of the proposed designed PV system is 45.46 MWh/Year and the system performance ratio in terms of efficiency over the year is 0.814. The overall performance analysis and simulation results would be helpful in determining the efficiency and viability of the proposed PV system while implementing it practically in the near future.*

Keywords: *Load estimation, Photovoltaic (PV), PVsyst software, Renewable Energy System (RES), Sustainable development*

1. INTRODUCTION

In today's world, economic, industrial, and social developments define the overall progress of a country. However, the integration of different RES with electric power systems is also a crucial factor in the economic growth and development of a country. As conventional power generating systems are not sufficient to fulfill all the power uncertainties and have shown serious impacts on the environment, significant importance should be given to the penetrations of non-conventional generating sources to the small independent electric power systems such as renewable energy-based microgrids [1]. According to the annual report of the World Bank, till 2021, approximately 770 million people of the total world population still live without access to electricity, whereas in India, 13% of Indian households still do not have access to central grid-connected electrification [2]. Hence, it seems that people of places that are not included in the central electricity grid system yet are mainly dependable on battery-based systems or renewable energy-based systems [3]. India is a tropical country with a significant amount of solar insolation rate and excellent solar energy potential with 300

sunny days [4]. India has a total RES capacity of 106.37 GW as per the report of the central authority of Electricity India in February 2022, among them 50.7 GW is the total solar capacity i.e., 48% of the total RES capacity [5]. Fig.1 shows the breakups of various RES capacities in India. Hence in India, photovoltaic microgrid has the potential to overcome the load demand along with the high cost and high power loss that would be charged by the main grid while extending their feeders. Moreover, the long life of Photovoltaic panels (25 years) makes the PV-based system more affordable for common people providing clean and green energy with zero environmental impact, faster electrification with low maintenance costs and reliable power supply with low energy loss. Literature shows a wide number of researches in the field of PV-based systems highlighting their energy assessments, performance analysis, and improved efficiencies over a long period. A PV system designed for cities in Nigeria is found in [6] showing the economic viability of the solar energy potential for that place. In [7] technical and economic assessment of a grid-connected PV system is mentioned for household application. In [8] review of the life cycle assessment of the PV system is designed for sustainability and environmental performance.

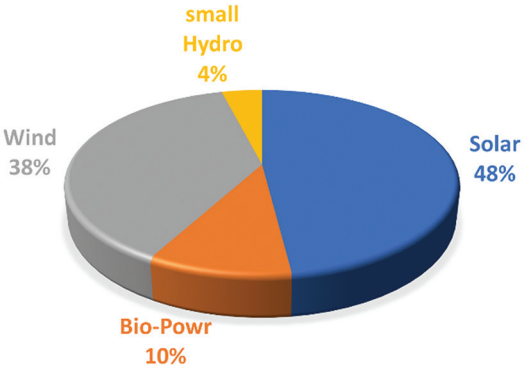


Fig. 1. Various RES capacities in India [5]

The design and monitoring of PV systems are explained in [9]. In [10], the design and cost analysis of a 1 kWp Photovoltaic system based on actual performance in an Indian Scenario has been explained. In [11] solar potential assessment using PVsyst software in the Northern zone of Morocco is performed and multiple geographical locations have been analyzed. In [12] design and simulation of a standalone solar PV system using PVsyst Software is evaluated but the required energy according to the load demand is a little less. A solar energy potential and performance analysis of a standalone PV system are explained in [13]. Performance analysis of a 700 KWp PV plant in Afghanistan is presented using PVsyst software in [14]. The design analysis of a grid-connected PV system has been performed in [15]. Case studies of two regions with improved design procedures to obtain structural optimization and load flow-based parametrizes have been marked in [16]. A design and simulation of a 5 MW PV system using PVsyst and HOMER software has been compared in [17]. A hybrid microgrid design and simulation is also performed in [18-20].

Microgrid is designed as a small distribution system and its distribution feeders are short-ranged compared to other conventional systems [21]. Generally, the loads connected to the small-scale microgrid feeders are unbalanced and hence it causes a higher R/X ratio of the feeder and a low short circuit capacity problem resulting in jeopardized stability of the system [22]. Hence to overcome this problem accurate estimation of loads is required while designing and sizing a microgrid. Moreover, the intermittent nature of RES such as PV, wind, etc. used in microgrids causes higher uncertainty. Hence, proper design and sizing are also necessary to defeat the uncertain characteristics of RES-based power plants.

This work aims to design a realistic PV microgrid that can be implemented practically soon. The key action involves accurate estimation of the load and simulation using PVsyst software to establish the viability of the proposed system. The proposed work presents the design and simulation of a 40 kW off-grid PV system for the departmental building of Electronics and Communication Engineering, at Tezpur University in India. From the simulation results, it can be observed that the proposed system has 81.4% efficiency and is able to fulfil the load demand of the departmental building. Detailed data can be obtained on the basis of solar potential, electrical energy generation and loss analysis of the PV panels and inverters over the year based on various environmental factors including irradiance, temperature, wind velocity etc.

The paper is organized as stated: In the first section a brief introduction of the PV-based systems and a throwback of their energy assessments using PVsyst software has been discussed. In the second section, the system design and the detail methodology for the case study have been explained following the results and discussion in the third section. The fourth section explains the conclusion and future work.

2. SYSTEM DESIGN

A photovoltaic system is dependent on the solar potential of a certain location [23]. However, the inclination, orientation, and quality parameters of the PV panels and the selection of the inverters along with the Maximum Power Point Tracking (MPPT) controller involved in the system also play a significant role in the design and sizing of the system. In this section, the methodology for designing, sizing, and simulating a 40 kW PV off-grid microgrid for the Department of Electronics and Communication Engineering, Tezpur University (India) using PVsyst software is discussed. The design of the overall system is based on the following factors:

- Critical load estimation: For load estimation, the department of ECE, Tezpur University (latitude 26.64° N and longitude 92.80° E and 61 meters above the sea level) is chosen and all the linear

and nonlinear loads with single phase and 3-phase loads (i.e., the three-phase motors with different power factors used in the departmental Workshop and the Laboratories) are taken into account and total energy consumption per day in kWh/day has been calculated.

- From the state-of-the-art design and sizing of the Microgrid system, based on the load estimation, the sizing of PV panels and inverters has been estimated using PVsyst software, and the solar irradiation of the mentioned geographical area is calculated. PVsyst is a software that enables the optimization and design of any PV-based system. This software has been considered for meteorological databases; thus, it helps to estimate the annual energy generation of a PV-based system, and the yearly solar irradiance data. Moreover, the accurate estimated load profile data are achieved along with the daily power demand.

2.1. SYSTEM DESIGN

A PV-based off-grid microgrid system consists of PV panels, MPPT controllers, inverters and energy storage devices and the loads have been described in Fig 2.

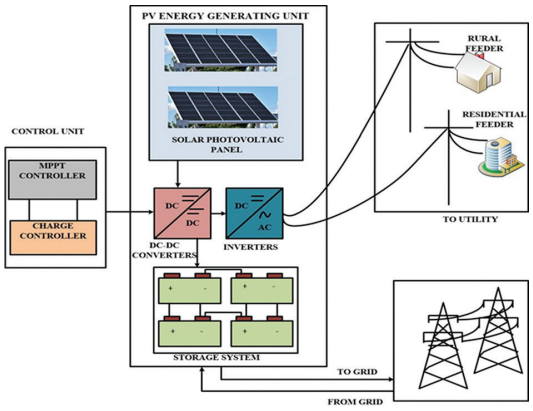


Fig. 2. Schematic diagram of a PV microgrid system

2.2. LOAD ESTIMATION OF THE SYSTEM

Load estimation is a challenging task for any distribution system operation and planning problems. A proper load estimation requires detailed parameters of the connected loads and the time interval over which the load variation is to be estimated [24]. To estimate the overall load demand, the load data of the department of ECE, Tezpur University is collected manually. Types and quantities of electrical appliances, their operating time in 24 hours duration and the maximum power consumption in watt (W) has been noted. With this information gathered, the total load data is estimated as shown in Table 1. The total energy consumed by the system loads per day is 366.97 kWh/day and the total power consumption per day is 36.985 kW. Hence to sufficiently fulfil the estimated load demand a 40 kW microgrid design and sizing is proposed with the provision of some spare capacity to use if future expansion is required.

Table 1. Load estimation in the Dept. of Tezpur University

Sl No.	Appliances	Quantity	Power (W)	Total power (kW)	Run time Day (hr)	kWh in Day	Run time Night (hr)	kWh in Night	Total Energy Consumption (kWh/Day)
1.	LEDs	21	15	0.315	4	1.2	12	3.7	4.9
2.	Tubes	100	55	5.5	8	44	12	66	110
3.	Fans	96	100	9.6	8	76.8	5	48	124.8
4.	AC	9	1500	13.5	6	81	0	0	81
5.	Motors	6	747.08	4.4	1	4.4	0	0	4.4
6.	Pcs	36	80	2.8	8	23	5	14.4	37.4
7.	Printers	6	45	0.27	1	0.27	0	0	0.27
8.	Other Appliances	10	60	0.6	6	3.6	1	0.6	4.2
	Total			**36.985**		**234.2**		**132.7**	**366.97**

2.3. PVSYST SOFTWARE

PVsyst is a simulation software for configuring a PV system. The users can design and size a standalone, grid-connected, and PV-pumped-based system in a user-friendly environment. The main input variables for the simulations are the geographical location and the load requirement of the system. Further, the PV panels and the inverter selection and sizing are to be done accordingly. The software stores location-wise meteorological databases from all around the globe. The post-simulation results can be obtained including the solar potential data for a particular location, performance ratios and loss analysis of the system both in tabular and diagram form. The step-by-step system design methodology is explained in Fig 3. The solar potential of the selected area can be generated by providing the geographical location of that area using latitude, longitude, and altitude. Plant orientation has fixed the tilt and azimuth angle associated with PV Panels. Based on the solar potential and the load requirement of the selected area, the selection and sizing of the PV panels, inverters, and MPPT controllers are provided. The system specification of the proposed design is systemized in Table 2.

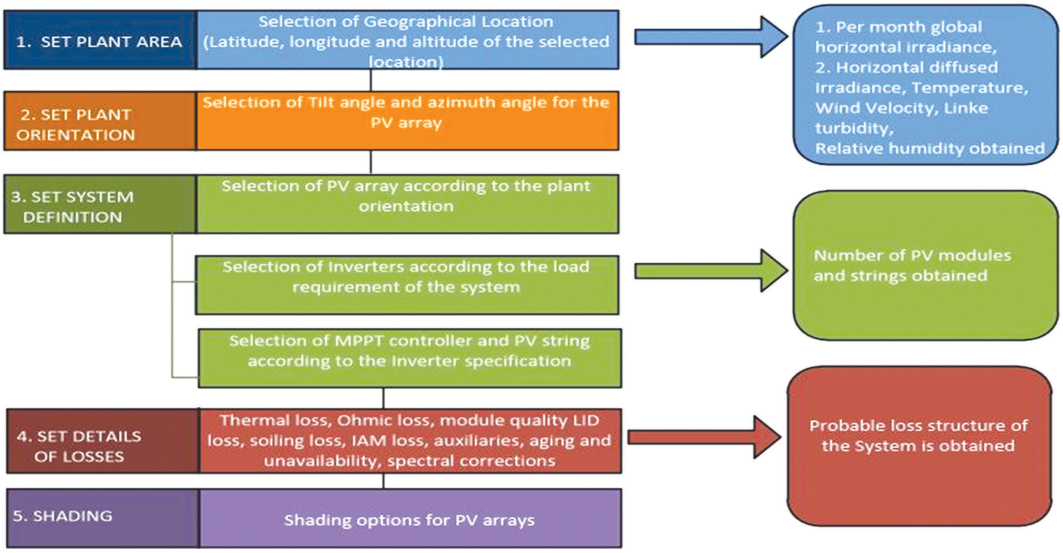


Fig. 3. Step-by-step procedures for PVsyst simulation Analysis

Table 2. The system specification

Category	Parameters	Value
Geographical location	latitude	26.70° N
	longitude	92.83° E
	Altitude	48m
	Time Zone	UTC+5.5
	Albedo	0.2
PV field Orientation (Fixed)	Tilt/ Azimuth angle	26/0°,26/0°
System input	PV module	120 Units (STP-335-A60-Wnhb)
	Module Capacity	335 Wp per module at 50°C
	Technology	Si-Poly
	Nominal Power	40.2 kW
	Module Area	202 m^2
	V_{mpp} (50°C)	611 V
	I_{mpp} (20°C)	29 A
	Number of inverters	02
	Inverter type	Solar Inverter RPI M20A
	Inverter Rating	200-820 V, 50 Hz

2.3.1. Geographical Location

In Table 2, the system input parameters for geographical location are categorized. The Department of Electronics and Communication Engineering, Tezpur University is located at the latitude 26.70° N and longitude 92.83° E and 48 m altitude (location height from the sea level) in India using the time zone of UTC+5.5 and albedo 0.2.

2.3.2. Tilt and Azimuth Angle

Tilt angle and azimuth angle decide the orientation and inclination of the PV panels in the PV generating unit. Tilt angle is the angle of inclination of the PV panel where it is to be mounted. Tilt angle depends on the latitude of the location and is taken as approximately equal to it. It can be seasonal considering the changing position of the sun with the season. The azimuth angle is the angle by which the PV panel can be mounted as south-facing for the northern hemisphere and north-facing for the Southern hemisphere. The optimized value of both these angles is required to get optimum energy from the PV panels. In the proposed system tilt angle is taken as 26° and the azimuth angle is taken as 0° for two strings of PV panels.

2.3.3. PV specification

Panels are selected to be 120 units with an output of 335 W. The panels are made of Si-polycrystalline technology and module no is STP-335-A60-Wnhb with an MPPT voltage of 611 V and current of 29 A. The 120 units of PV panels with the mentioned specification are connected in two parts with 2 Strings × 60 series connected panels.

2.3.4. Inverter Specification

For the 40 kWp PV microgrid the inverter required is of 2 numbers with a power rating of 200-820 V,50/60 Hz. The inverter model is Solar Inverter RPI M20A. The DC power input at MPPT is 17.45 kW and the maximum AC power generated is 21 kW for both the inverter.

3. RESULTS

The report provides a detailed performance analysis of the proposed 40 kW PV microgrid through diagrams and tables generated by PVsyst software. To satisfy 40 kWp of load the system required 120 units of PV panels of 335 Wp capacity as generating units and two numbers of inverters having the capacity of 200-820 V,50 Hz with a conversion efficiency of 96%. To get optimized energy output from the PV-generating units, two MPPT controllers are used with each inverter. Applying all the necessary conditions the simulation has been performed in the software. The software calculated the solar irradiance data for the proposed system. The yearly global horizontal irradiance is 1408.4 kWh/m^2 on average as mentioned in Table 3. The monthly diffused horizontal irradiance, ambient temperature, global incident irradiance, effective array energy, the energy available on the load side and performance ratios for all over the year can be obtained from Table 3.

Table 3. Measured data of various irradiances, ambient temperatures, array energy, energy available on the load side and performance ratio

Month	Global Horizontal Irradiance (kWh/m²)	Diffused Horizontal Irradiance (kWh/m²)	Ambient Temperature (℃)	Global Incident Irradiance (kWh/m2)	Effective Array Energy (MWh)	The energy available in the load (MWh)	Performance ratio
January	93.0	48.75	17.05	103.5	3.420	3.330	0.838
February	95.0	59.68	19.76	101.6	3.322	3.235	0.829
March	135.2	85.76	23.24	139.6	4.509	4.389	0.819
April	121.2	78.58	24.40	121.8	3.917	3.809	0.814
May	127.1	88.57	26.71	124.6	3.980	3.870	0.809
June	135.7	75.42	27.78	131.7	4.173	4.053	0.802
July	120.1	84.23	28.64	117.2	3.713	3.609	0.802
August	149.2	85.34	28.91	147.9	4.680	4.549	0.801
September	117.4	69.36	27.80	120.1	3.805	3.699	0.802
October	115.2	62.69	26.09	122.2	3.895	3.791	0.808
November	105.2	50.21	21.67	117.7	3.815	3.717	0.822
December	94.1	48.05	18.40	106.5	3.496	3.406	0.833
Year	1408.4	836.65	24.23	1454.2	46.726	45.456	0.814

The average Incident radiation data on a collected plane is 1454 kWh/m², temperature is 24°C and effective array energy is 46.7 MWh over the year. Fig. 4 shows the daily produced useful energy over the year for the designed system is 3.24 kWh/kWp/day and results show the highest energy generation is in the months of March and August while the PV array losses and system losses are 0.65 kWh/kWp/day and 0.09 kWh/kWp/day respectively.

The performance ratio is represented in Fig 5. The lowest performance ratio is for the month of August at 80.1% and the highest is 83.8% for the month of January. The overall performance ratio over the year is 81.4%.

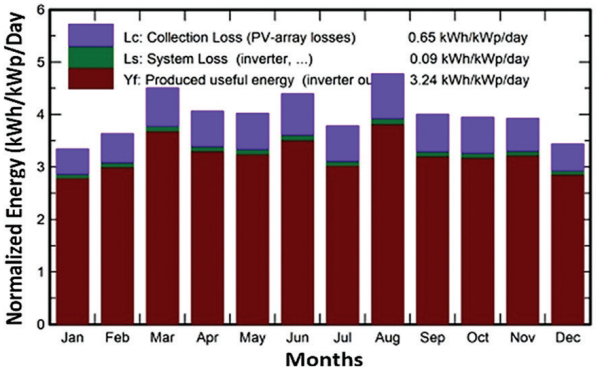


Fig. 4. Graphical analysis of normalized energy distribution over the year

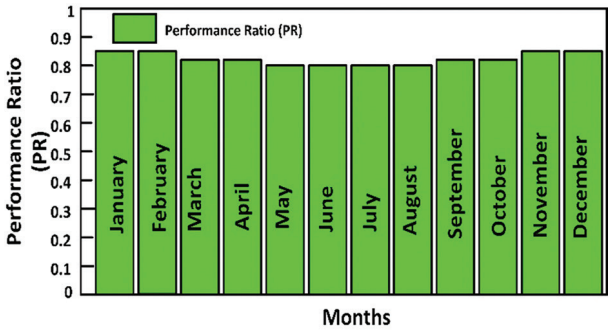


Fig. 5. Performance Ratio Analysis

The hourly consumption of the energy is shown in Fig. 6 in 24-hour format. The designed system has a peak load demand in the 15th hour measured at approximately 25 kW according to the load demand of the system and lowest in the 6th to 10th hours showing nearly 0 kW.

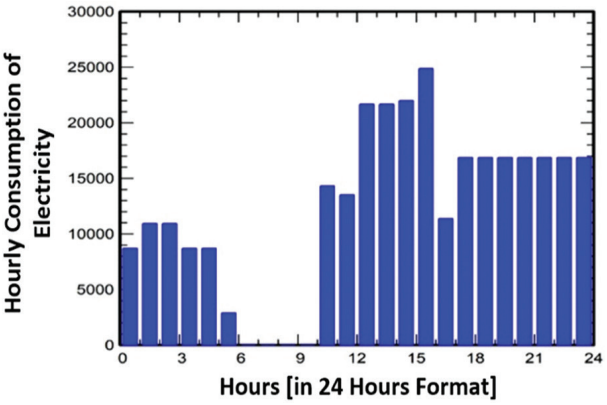


Fig. 6. Hourly consumption of energy

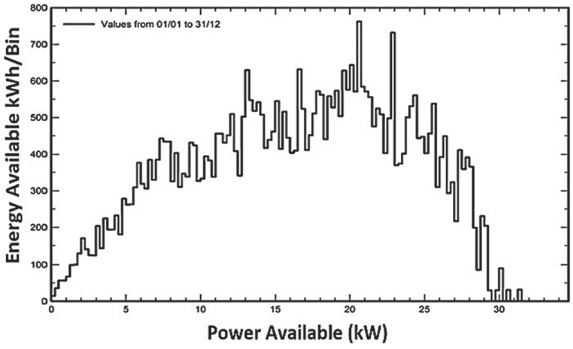


Fig. 7. Daily energy and power available on the inverter side

The yearly energy available in kWh/Bin versus the available power of the inverter side in kW is graphically represented in Fig. 7 showing the yearly energy available in the PV panel versus the energy available to inject to the grid.

Fig 8 explains the graphical analysis of the energy produced over the year including the PV array losses and system losses. The various losses due to wiring, soil factor, and temperature mismatch in the system are 0.82%, 3% and 5.89% of the total generated energy respectively.

The other losses of the system are calculated by the software according to the geographical location and sizing of the system. After the analysis of various losses present in the system, the energy obtained in the load side is analyzed by the software and the value is 1362 kWh while the energy generated by the generating unit is 1408 kWh.

From the above analysis, it has been clear that the data simulated in this work is detailed data of the designed PV system using PVsyst simulation software to implement the proposed designed PV microgrid practically. The insertion of more PV panels leads the system to generate more PV energy to inject into the central grid system. Accordingly, all the measured data of irradiances, available array energies, performance ratios and their average value every month over a year can be evaluated.

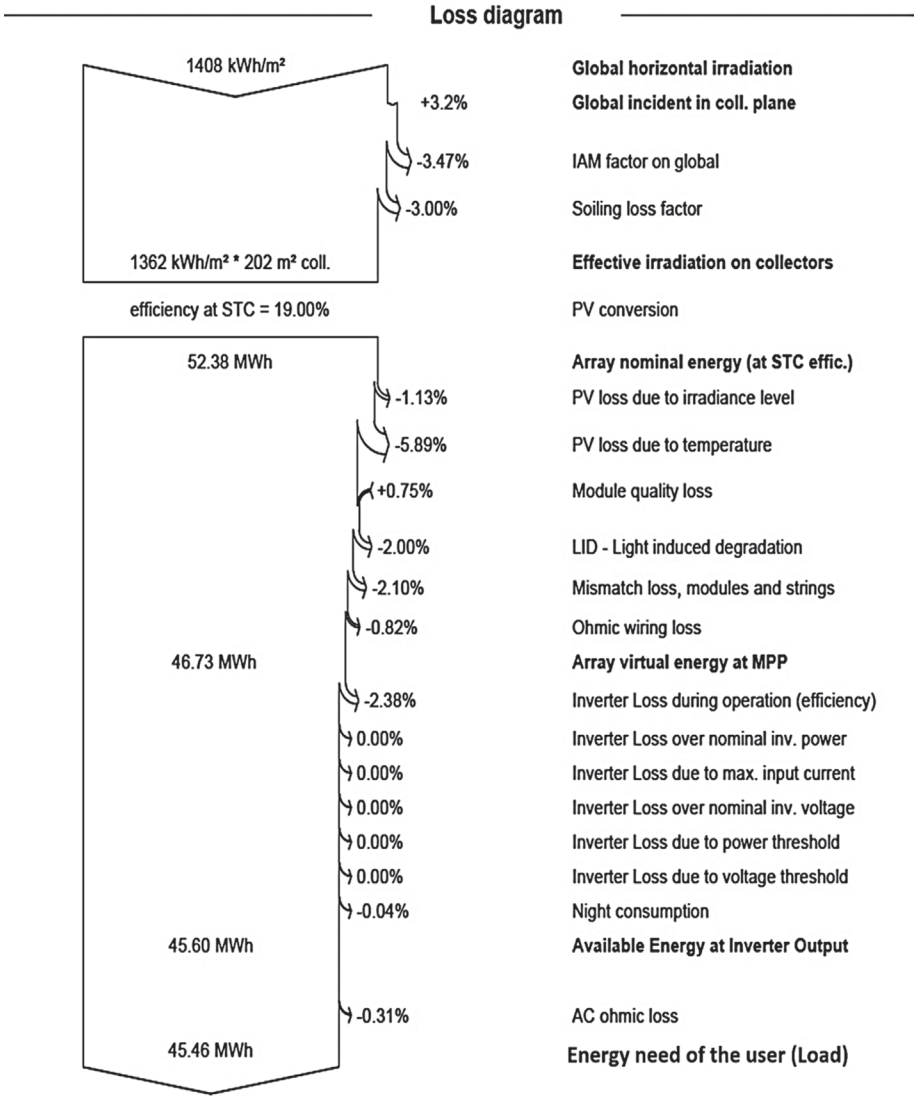


Fig. 8. Loss analysis of the proposed designed system

4. CONCLUSION

The design and simulation of a 40 kW photovoltaic microgrid system using PVsyst software are presented in this paper. The results are based on the solar potential and load analysis of the aforementioned area. The PVsyst software-based performance ratio for the system is 0.814 which shows that the mentioned area is well-fitted for a PV-fed system and it acquires a good solar potential. The PVsyst software result can provide reliable sizing for the departmental building at the time of installation and can be helpful while choosing the ratings and quantities of the PV panels, inverters, MPPT controllers and batteries to fulfil the load demand. This kind of analysis can help the researcher and industrialist design a PV-based microgrid while installing it in different locations. In the future, grid-connected PV systems with large capacities would be preferred to obtain an energy assessment by designing via PVsyst software with a better performance ratio to fulfill the power demand of the grid.

List of abbreviations:

Abbreviations	Definitions
AC	Alternative Current
DC	Direct Current
GW	Gigawatt
Hz	Hartz
IAM	Incident angle losses
I_{mpp}	Current at maximum power point
LID	Light-induced degradation
MPPT	Maximum Power Point Tracking
kW	Kilowatt
kWh	Kilowatt-hour
MW	Megawatt
MWh	Megawatt hour
PV	Photovoltaic
RES	Renewable Energy Sources
V_{mpp}	Voltage at maximum power point
W	Watt

5. REFERENCES

[1] F. S. Al-Ismail, "DC Microgrid Planning, Operation, and Control: A Comprehensive Review", IEEE Access, Vol. 9, 2021, pp. 36154-36172.

[2] The world bank annual report, "From Crisis to Green, Resilient, and Inclusive Recovery Contents", IBRD, IDA, World Bank Group, 2021.

[3] A. Sharma, S. Sharma, "Review of power electronics in vehicle-to-grid systems", Journal of Energy Storage, Vol. 21, 2019, pp. 337-361.

[4] S. K. Suman, J. Ahamad, "Solar Energy Potential and Future Energy of India: An Overview", International Journal of Engineering Science and Computing, Vol. 8, No. 5, 2018, pp. 17575-17579.

[5] Central Electricity Authority, "Executive Summary on Power Sector Government of India", Ministry of Power, February 2022, pp. 1-4.

[6] C. O. Okoye, O. Taylan, D. K. Baker, "Solar energy potentials in strategically located cities in Nigeria: Review, resource assessment, and PV system design", Renewable and Sustainable Energy Reviews, Vol. 55, 2016, pp. 550-566.

[7] Y. Cui, J. Zhu, F. Meng, S. Zoras, J. McKechnie, J. Chu, "Energy assessment and economic sensitivity analysis of a grid-connected photovoltaic system", Renewable Energy, Vol. 150, 2020, pp. 101-115.

[8] J. Peng, L. Lu, H. Yang, "Review on life cycle assessment of energy payback and greenhouse gas emission of solar photovoltaic systems", Renewable and Sustainable Energy Reviews, Vol. 19. Elsevier Ltd, pp. 255-274, 2013.

[9] M. Aghaei, N. M. Kumar, A. Eskandari, H. Ahmed, A. K. V. De Oliveira, S. S. Chopra, "Solar PV systems design and monitoring", Photovoltaic Solar Energy Conversion: Technologies, Applications and Environmental Impacts, 2020, pp. 117-145.

[10] S. Ahsan, K Javed, A. S. Rana, M. Zeeshan", Design and cost analysis of 1kW Photovoltaic system based on actual performance in Indian scenario", Recent Trends in Engineering and Material Sciences, 2016.

[11] B. Belmahdi, A. El Bouardi, "Solar potential assessment using PVsyst software in the northern zone of Morocco", Procedia Manufacturing, Vol. 46, 2020, pp. 738-745.

[12] R. Kumar, C. S. Rajoria, A. Sharma, S. Suhag, "Design and simulation of standalone solar PV system using PVsyst Software: A case study", Materials Today: Proceedings, Vol. 46, No. 11, 2021, pp. 5322-5328.

[13] A. Shrivastava, R. Sharma, M. K. Saxena, V. Shanmugasundaram, M. L. Rinawa, Ankit, "Solar energy capacity assessment and performance evaluation of a standalone PV system using PVSYST", Materials Today: Proceedings, Vol. 80, No. 3, 2021, pp. 3385-3392.

[14] M. Baqir, H. K. Channi, "Analysis and design of solar PV system using Pvsyst software", in Materials Today: Proceedings, Vol. 48, No. 5, 2022, pp. 1332-1338.

[15] T. Hasarmani, R. Holmukhe, S. Tamke, "Performance Analysis of Grid Interfaced Photovoltaic Systems for Reliable Agri-Microgrids using PVsyst", Proceedings of the International Conference on Information and Communications Technology, Yogyakarta, Indonesia, 2019, pp. 894-898.

[16] S. M. Bhagavathy, G. Pillai, "PV microgrid design for rural electrification", Designs, Vol. 2, No. 3, 2018, pp. 1-22.

[17] S. A. D. Mohammadi, C. Gezegin, "Design and Simulation of Grid-Connected Solar PV System Using

PVSYST, PVGIS and HOMER Software", International Journal of Pioneering Technology and Engineering, Vol. 1, No. 1, 2022, pp. 36-41.

[18] M. B. T. Wijata, A. A. A. Suryawan, K. Astawa, I. G. K. Sukadana, I. M. Parwata, "Design Simulation of Micro-Grid Hybrid Solar Power Plant as a Power Supply", Natural Sciences Engineering and Technology Journal, Vol. 3, No. 2, 2023, pp. 232-240.

[19] S. Alfieri, S. Piccini, M. Kermani, "Feasibility study of Nearly Zero Energy Building in a real Microgrid case study", Proceedings of the IEEE International Conference on Environment and Electrical Engineering and the IEEE Industrial and Commercial Power Systems Europe, Genova, Italy, 2019, pp. 1-6.

[20] J. Lacea, E. Querikiol, E. Taboada, "Balancing Energy Trilemma Using Hybrid Distributed Rooftop Solar PV (DRSP)/Battery/Diesel Microgrid: A Case Study in Gilutongan Island, Cordova, Cebu, Philippines", Energies, Vol. 14, No. 21, 2021, p. 7358.

[21] G. A. Shahgholian, "Brief review on microgrids: Operation, applications, modeling, and control", International Transactions on Electrical Energy Systems, Vol. 31, No. 6, 2021, p. e12885.

[22] R. Kljajić, P. Marić, H. Glavaš, M. Žnidarec, "Microgrid Stability: A Review on Voltage and Frequency Stability", Proceedings of the IEEE 3rd International Conference and Workshop in Óbuda on Electrical and Power Engineering, Budapest, Hungary, 2020.

[23] W. S. Ebhota, P. Y. Tabakov, "Assessment of solar PV potential and performance of a household system in Durban North, Durban, South Africa", Clean Technologies and Environmental Policy, Vol. 24, 2022, pp. 1241-1259.

[24] T. Y. Salihu, M. F. Akorede, A. Abdulkarim, A. I. Abdullateef, "Off-grid photovoltaic microgrid development for rural electrification in Nigeria", Electricity Journal, Vol. 33, No. 5, 2020.

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia.

About this Journal

The International Journal of Electrical and Computer Engineering Systems publishes original research in the form of full papers, case studies, reviews and surveys. It covers theory and application of electrical and computer engineering, synergy of computer systems and computational methods with electrical and electronic systems, as well as interdisciplinary research.

Topics of interest include, but are not limited to:

- Power systems
- Renewable electricity production
- Power electronics
- Electrical drives
- Industrial electronics
- Communication systems
- Advanced modulation techniques
- RFID devices and systems
- Signal and data processing
- Image processing
- Multimedia systems
- Microelectronics
- Instrumentation and measurement
- Control systems
- Robotics
- Modeling and simulation
- Modern computer architectures
- Computer networks
- Embedded systems
- High-performance computing
- Parallel and distributed computer systems
- Human-computer systems
- Intelligent systems
- Multi-agent and holonic systems
- Real-time systems
- Software engineering
- Internet and web applications and systems
- Applications of computer systems in engineering and related disciplines
- Mathematical models of engineering systems
- Engineering management
- Engineering education

Paper Submission

Authors are invited to submit original, unpublished research papers that are not being considered by another journal or any other publisher. Manuscripts must be submitted in doc, docx, rtf or pdf format, and limited to 30 one-column double-spaced pages. All figures and tables must be cited and placed in the body of the paper. Provide contact information of all authors and designate the corresponding author who should submit the manuscript to https://ijeces.ferit.hr. The corresponding author is responsible for ensuring that the article's publication has been approved by all coauthors and by the institutions of the authors if required. All enquiries concerning the publication of accepted papers should be sent to ijeces@ferit.hr.

The following information should be included in the submission:

- paper title;
- full name of each author;
- full institutional mailing addresses;
- e-mail addresses of each author;
- abstract (should be self-contained and not exceed 150 words). Introduction should have no subheadings;
- manuscript should contain one to five alphabetically ordered keywords;
- all abbreviations used in the manuscript should be explained by first appearance;
- all acknowledgments should be included at the end of the paper:
- authors are responsible for ensuring that the information in each reference is complete and accurate. All references must be numbered consecutively and citations of references in text should be identified using numbers in square brackets. All references should be cited within the text;
- each figure should be integrated in the text and cited in a consecutive order. Upon acceptance of the paper, each figure should be of high quality in one of the following formats: EPS, WMF, BMP and TIFF;
- corrected proofs must be returned to the publisher within 7 days of receipt.

Peer Review

All manuscripts are subject to peer review and must meet academic standards. Submissions will be first considered by an editor-in-chief and if not rejected right away, then they will be reviewed by anonymous reviewers. The submitting author will be asked to provide the names of 5 proposed reviewers including their e-mail addresses. The proposed reviewers should be in the research field of the manuscript. They should not be affiliated to the same institution of the manuscript author(s) and should not have had any collaboration with any of the authors during the last 3 years.

Author Benefits

The corresponding author will be provided with a .pdf file of the article or alternatively one hardcopy of the journal free of charge.

Units of Measurement

Units of measurement should be presented simply and concisely using System International (SI) units.

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003

Published: semiannually

Copyright

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.

Postal Address

Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b
31000 Osijek, Croatia

IJECES Copyright Transfer Form

(Please, read this carefully)

This form is intended for all accepted material submitted to the IJECES journal and must accompany any such material before publication.

TITLE OF ARTICLE (hereinafter referred to as "the Work"):

COMPLETE LIST OF AUTHORS:

The undersigned hereby assigns to the IJECES all rights under copyright that may exist in and to the above Work, and any revised or expanded works submitted to the IJECES by the undersigned based on the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the complete Work and all incorporated parts of the Work. Otherwise he/she warrants that necessary permissions have been obtained for those parts of works originating from other authors or publishers.

Authors retain all proprietary rights in any process or procedure described in the Work. Authors may reproduce or authorize others to reproduce the Work or derivative works for the author's personal use or for company use, provided that the source and the IJECES copyright notice are indicated, the copies are not used in any way that implies IJECES endorsement of a product or service of any author, and the copies themselves are not offered for sale. In the case of a Work performed under a special government contract or grant, the IJECES recognizes that the government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official government purposes only, if the contract/grant so requires. For all uses not covered previously, authors must ask for permission from the IJECES to reproduce or authorize the reproduction of the Work or material extracted from the Work. Although authors are permitted to re-use all or portions of the Work in other works, this excludes granting third-party requests for reprinting, republishing, or other types of re-use. The IJECES must handle all such third-party requests. The IJECES distributes its publication by various means and media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various collections, databases and other publications. The IJECES publisher requires that the consent of the first-named author be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes. If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IJECES publisher assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

Authors of IJECES journal articles and other material must ensure that their Work meets originality, authorship, author responsibilities and author misconduct requirements. It is the responsibility of the authors, not the IJECES publisher, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

- The undersigned represents that he/she has the authority to make and execute this assignment.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
- The undersigned agrees to indemnify and hold harmless the IJECES publisher from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

_____ _____

Author/Authorized Agent **Date**

CONTACT
International Journal of Electrical and Computer Engineering Systems (IJECES)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek
Kneza Trpimira 2b
31000 Osijek, Croatia
Phone: +38531224600,
Fax: +38531224605,
e-mail: ijeces@ferit.hr