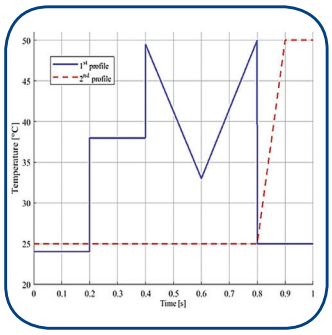
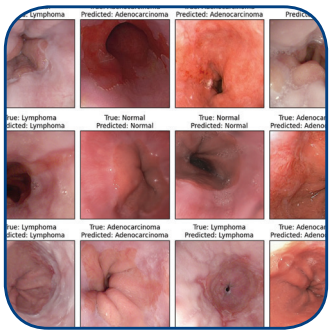
Volume 16, Number 5, 2025 ISSN 1847-6996

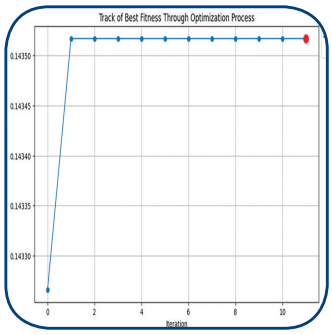
FERIT
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY **OSIJEK**

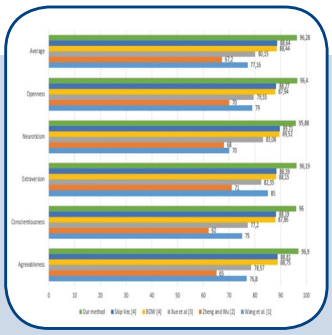
IJECES International Journal of Electrical and Computer Engineering Systems

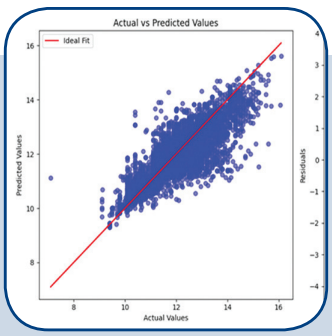
International Journal of Electrical and Computer Engineering Systems

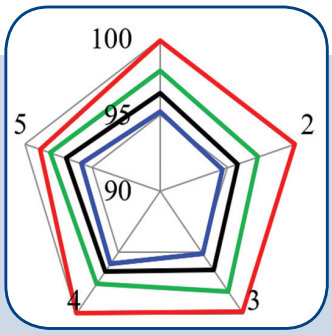












ISSN: 1847-6996

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 16, Number 4, 2025 | Pages 347 - 429

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

3D-based Convolutional Neural Networks for Medical Image Segmentation: A Review

Review Paper

Siti Raihanah Abdani
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
2022236678@student.uitm.edu.my

Nursuriati Jamil
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
lizajamil@computer.org

Syed Mohd Zahid Syed Zainal Ariffin*
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
zahidzainal@uitm.edu.my

Shafaf Ibrahim
College of Computing, Informatics, and Mathematics,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia
shafaf2429@uitm.edu.my

*Corresponding author

Abstract *– Medical image segmentation is essential for disease screening and diagnosis, particularly through techniques like anatomical and lesion segmentation that can be used to isolate critical regions of interest. However, manual segmentation is labor-intensive, costly, and susceptible to subjective bias, underscoring the need for automation. Deep learning, particularly convolutional neural networks (CNNs), has significantly advanced segmentation accuracy and efficiency. With the introduction of 3D imaging, research has evolved from 2D CNNs to 3D CNNs, which leverage inter-slice information to improve segmentation precision. This paper aims to provide a literature review of studies published between 2018 and 2024 on platforms such as Google Scholar and ScienceDirect, where the identified relevant research are "3D segmentation" and "3D medical imaging". This study outlines the key stages of 3D CNN segmentation that include preprocessing, region-of-interest extraction, and post-processing. Furthermore, this study emphasizes the application of 3D CNN architectures to complex lung imaging scenarios, such as lung cancer and COVID-19. Although 3D CNNs outperform 2D CNNs in preserving spatial continuity across slices, they present notable limitations. Key challenges include heavy computational and high memory demands, as well as a dependency on large annotated datasets, which are often scarce in medical imaging. Additionally, effective multiscale feature learning remains a challenging issue, with current architectures struggling to generalize the features of interest across several usage variations. To further improve the segmentation performance, future research should prioritize developing adaptive algorithms and fostering interdisciplinary collaboration between computer scientists and medical professionals to design efficient and scalable models, designed specifically for clinical applications. This future research direction will enhance diagnostic accuracy and segmentation quality in 3D medical imaging.*

Keywords*: Medical Imaging, Semantic Segmentation, Artificial Intelligence, Deep Learning, Diagnosis Tools*

Received: November 18, 2024; Received in revised form: December 17, 2024; Accepted: January 6, 2025

1. INTRODUCTION

Image segmentation, particularly in medical applications, is essential for accurately distinguishing and isolating regions of interest within medical images, thereby aiding in diagnosis and treatment planning. Although manual segmentation is often more precise, it is time-intensive and susceptible to subjectivity issues, prompting the need for automated approaches. A commonly used conventional technique is thresholding, where lung regions are identified as the largest blob, as described by Manikandan [1]. However, this method lacks robustness when applied to lung disease cases, such as those affected by pneumonia or COVID-19, due to substantial variability in lung image data. To address this limitation, adaptive techniques like watershed segmentation have been explored, as demonstrated by Navya and Pradeep [2]. Nonetheless, these methods are too dependent on basic assumptions, such as the use of Sobel edge operators, which may not effectively handle intensity variations across different CT scans. Similarly, preprocessing filters such as Wiener, mean, and median filters increase the computational load and do not ensure segmentation robustness [3]. The application of these filters is often followed by morphological masking as the post-processing step, which can inadvertently erode critical lung regions, thereby reducing segmentation accuracy.

Alternative methods, including those that combine super pixels and fuzzy clustering [4], have been proposed to enhance segmentation outcomes. However, super pixel-based clustering can be computationally demanding and frequently fails to achieve good pixel-level accuracy. This pixel grouping into super pixel blobs may obscure minor variations within these regions, potentially leading to the misclassification of lung areas. Such inaccuracies are particularly problematic for disease staging identification, where even slight errors can have substantial diagnostic implications. A review by Ker *et al.* [5] explores the application of machine learning, particularly convolutional neural networks (CNNs), in medical image analysis. It highlights the advantages of machine learning in handling large medical data by analyzing the data's hierarchical relationships without the extensive use of feature engineering.

Deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful approach for medical image segmentation [6-8] CNNs are capable of automatically learning and extracting features from large datasets, yielding more accurate and reliable segmentation results. The growing availability of large datasets and enhanced computational resources has led to the increasing adoption of deep learning methods in healthcare, where this approach provides robust solutions to the limitations of conventional machine learning techniques. Additionally, the deep learning approach can also be used to facilitate the generation of synthetic datasets through the generative adversarial network (GAN) approach, which helps mitigate privacy concerns that are associated with sensitive medical data [9-12].

Alongside advancements in segmentation techniques, improvements in 3D imaging technologies, such as computed tomography (CT) and magnetic resonance imaging (MRI), have also significantly enhanced medical imaging capabilities. These technologies enable a more detailed and precise assessment of pathologies, particularly small-scale anomalies like cancerous nodules, which may not be discernible in 2D imaging [13, 14]. By providing a comprehensive view of the anatomical structures, 3D imaging reduces the likelihood of mistaking small pathologies for noise and offers a more accurate representation of tissues that might otherwise appear as artifacts in 2D images [15].

Despite the potential of deep learning and advanced imaging technologies, several challenges persist in medical image segmentation. Deep learning models rely on large volumes of annotated data, and the quality of these ground truth data is dependent on the expertise of the annotators, which may lead to inconsistencies. Standardizing annotation practices and improving training for annotators are crucial for enhancing segmentation accuracy. Furthermore, medical segmentation tasks can be divided into anatomical and lesion segmentation. Anatomical segmentation involves delineating organs or structures, which can be complex due to visual similarities between different structures. In contrast, lesion segmentation focuses on identifying abnormal regions, which vary greatly in size, shape, and location across patients, adding to the difficulty of accurate segmentation.

Research on 3D CNNs for medical segmentation has also made substantial progress, with studies exploring both 2D slice-based methods and full 3D volumetric approaches. While 2D methods often overlook crucial inter-slice information, 3D approaches utilize the entire data volume to produce improved segmentation outcomes. Although existing reviews on 3D CNNs [16-18] discuss various facets of 3D segmentation, our work focuses specifically on the application and methodological workflows of 3D CNN segmentation in medical imaging. Additionally, this study examines the use of 3D deep learning methods in lung imaging, where the modification or the improvement of the backbone networks will be categorized and discussed. Further discussions are also added to address the network's limitations and challenges. Finally, future directions in this field are proposed, highlighting areas for continued research and development.

2. METHODOLOGY

This review was conducted by searching the Google Scholar and ScienceDirect databases for peer-reviewed journal articles and conference proceedings. Only English-language articles published between 2018 and 2024 were selected. The search terms are set to "3D segmentation" and "medical image," while exclusion criteria are set to omit books, newspapers, non-peer-reviewed articles, and any study that is not specifically focused on 3D image segmentation within medical applications. Initially, 114 articles were identified, but the selection was refined to focus on papers discussing 3D deep learning algorithms (specifically 3D CNNs) applied to medical image segmentation (see Fig. 2). Fig. 1 summarizes the literature review methodology.

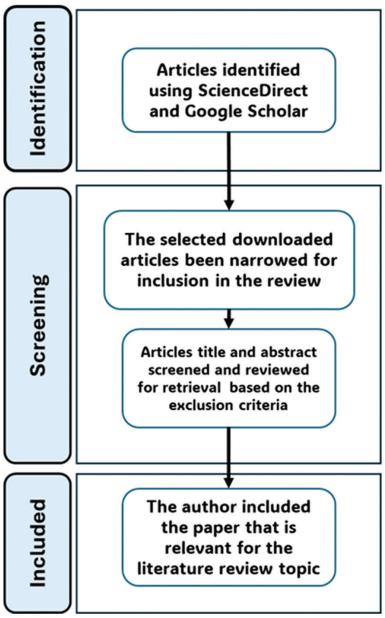


Fig. 1. Literature Review Methodology

3. MEDICAL IMAGE SEGMENTATION

This section begins with a general overview of the importance of medical image segmentation, followed by a discussion comparing the basic 2D CNN and 3D CNN methodologies. Figure 3 outlines the general steps in medical image segmentation using 3D CNNs, which include image pre-processing, region of interest (ROI) identification, 3D CNN segmentation, binary mask generation, and image post-processing. Based on Figure 2, each stage of the research methodology employed in the selected papers is discussed, except for image segmentation, which is later analyzed in greater detail specifically for the lung imaging. This process provides a comprehensive understanding of the overall approach to 3D CNN segmentation in medical imaging, with particular emphasis on lung imaging.

The development of automated segmentation algorithms has been extensively researched in various applications. Recent advancements in the field of medical image processing have led to the emergence of several segmentation models that can be categorized broadly into three classes: 1) conventional image processing-based algorithms, 2) machine learning-based algorithms, and 3) deep learning-based algorithms.

Conventional semantic segmentation algorithms, such as edge-based methods, are commonly employed to identify borders within an image. These methods rely on gradient-based edge detection operators, including Prewitt, Canny, Sobel, Roberts, and Laplacian filters. Despite its limitations, edge-based segmentation can be integrated with more advanced techniques to enhance its performance further. Besides that, local shape analysis has also been applied to segment lung pathologies [19]. In this approach, a set of predefined generic shapes representing local pathologies is compared with the tested input data using a geodesic distance metric. Another method proposed by Cui *et al.* [20] employs a more sophisticated technique involving predefined features through a boundary expansion approach. In this method, an initial seed representing the pathological region is defined, and color information is utilized to expand the region based on a fixed 20% threshold.

For machine learning-based category, it can be further divided into two approaches which are supervised and unsupervised learning. Unsupervised learning, particularly clustering methods, partitions data into distinct groups based on inherent feature similarities. Among these, the K-means algorithm is one of the most widely utilized clustering techniques. In contrast, supervised learning through classification tasks relies on a labeled training dataset, where each data point is associated with a specific target ground truth. One of the most used supervised algorithms is the K-nearest neighbor (K-NN) classifier.

For the third category, the deep learning-based approach mainly leverages Convolutional Neural Network (CNN), which is known for its robust feature extraction capabilities that have demonstrated exceptional performance in tasks such as natural image classification, object detection, and segmentation. As for the segmentation task, Fully Convolutional Neural Network (FCNN) is one of the earliest semantic segmentation models that is particularly well-suited for medical image segmentation tasks. These deep learning-based methods surpass traditional techniques in terms of robustness and accuracy, establishing themselves as the dominant approach in automatic medical image segmentation. Many medical image segmentation tasks have utilized the enhanced versions of the U-Net architecture, which is a symmetric network with skip connections between the encoder and decoder paths [21].

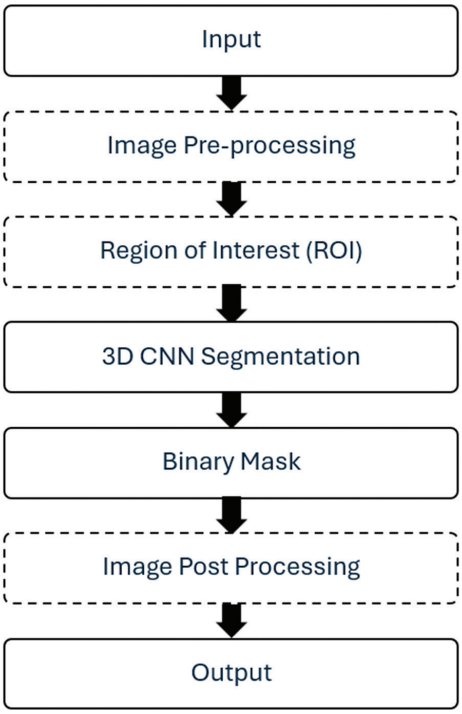


Fig. 2. The general flow of a 3D CNN segmentation. The dotted line boxes are not compulsory steps, while solid line boxes are compulsory steps.

4. 2D VS 3D CNN SEGMENTATION FOR 3D MEDICAL IMAGES

The application of deep learning-based segmentation to 3D medical images can be approached in two distinct ways. The first approach involves directly feeding 3D imaging data into a 3D CNN architecture. The second approach entails slicing the 3D imaging data into a series of 2D slices and inputting these individual 2D slices into a 2D CNN architecture. While considerable research has focused on the second approach due to its lower computational requirements, the 3D CNN approach holds particular advantages for segmentation tasks, especially when dealing with boundaries and edges. This is because 3D CNNs retain more spatial information by maximizing the interslice context, as compared to 2D methods, which may fail to capture important volumetric relationships across slices [22]. In

3D medical imaging, ROI often extends across multiple slices, making the interslice information critical. Additionally, 3D convolutional kernels can process data in all three spatial dimensions, as opposed to 2D convolutional kernels which can only analyze data in two-dimensional format [23].

The 2D U-Net architecture [24] which takes inspiration from the fully convolutional network consists of 23 layers in its symmetric encoder-decoder network. This architecture is typically divided into two segments: the down-sampling path (encoder) and the up-sampling path (decoder). During the down-sampling phase, convolutional and pooling layers are applied to the input image, generating feature maps at varying levels of abstraction. The up-sampling phase gradually restores the size of the feature maps by using deconvolutional layers. To recover the detailed information lost during the down-sampling process, the feature maps are merged with corresponding higher-resolution feature maps from the encoder side. These up-sampling procedures have been implemented in [25] to facilitate the reconstruction of 3D models.

However, since much of medical imaging data is inherently three-dimensional, the application of a 2D U-Net network can lead to the loss of critical spatial information. Moreover, the two-dimensional structure of the network results in the loss of contextual information during the down-sampling process [26] This limitation can reduce the network's sensitivity to fine border details, as they are usually not effectively restored during the up-sampling phase. Consequently, there is a need to employ a three-dimensional network for further optimization that may also contribute to the information loss. Additionally, the input data must undergo slicing, where the 3D data is divided into multiple 2D slices. This process may reduce the network's accuracy, as the correlation between adjacent slices will be lost.

To address these challenges, Çiçek *et al.* [27] proposed the 3D U-Net, which is an extension of the original U-Net architecture by incorporating 3D convolutional and pooling layers in the encoder side and 3D deconvolutional layers in the decoder side. However, the 3D U-Net only utilized three down-sampling layers due to the high computational cost, which limits its ability to extract deep-layer image features. This restriction has resulted in reduced accuracy for certain medical image segmentation tasks. To overcome this challenge, Milletari *et al.* [28] introduced another model, the V-Net, which incorporates residual connections to enable deeper network architectures. Subsequently, over the years, numerous modifications and enhancements have been proposed that significantly improve segmentation accuracy. 3D CNN-based segmentation methods have been successfully applied across a range of medical imaging applications, including head and neck [29], heart [30], lung [31], kidney [32], liver [33], brain [34, 35], and multi-organ segmentation [36], as illustrated in Fig. 3.

The main contribution of this review paper is the discussion of several categories of these model modifications, highlighting their contributions to performance improvements. In subsequent sections, we discuss the limitations of these modifications and propose future research directions. These insights aim to guide future researchers in medical image processing and provide valuable perspectives for healthcare professionals or clinicians.

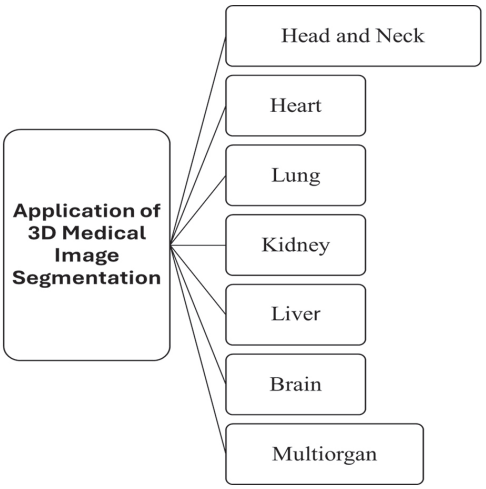


Fig. 3. 3D CNN segmentation applications in the medical imaging field

5. THE SUMMARY OF THE INCLUDED PAPERS

Based on the findings from the included studies, it was observed that not all studies incorporated pre-processing steps as part of their methodology as can be seen in Figure 2, despite its potential to enhance image quality and facilitate better feature extraction. This variability in methodology highlights differing approaches, with some studies relying entirely on the robustness of their 3D CNN models for effective segmentation. Generally, many studies employed single-stage segmentation pipelines where the 3D CNN directly processes the input images. In contrast, there are also studies that utilized a two-stage pipeline, which includes a preliminary Region of Interest (ROI) extraction step. This additional ROI extraction stage allows the model to focus on specific areas of the image, potentially improving segmentation performance by reducing irrelevant and noisy data.

Studies that incorporated pre-processing techniques alongside 3D CNNs will be discussed in detail in Section 5.1, with an emphasis on how these techniques contributed to improved model performance and segmentation accuracy. In contrast, studies that utilized ROI extraction as part of their two-stage pipeline are analyzed in Section 5.2, highlighting the role of this additional step in optimizing the segmentation process.

Moreover, Section 5.3 explores the contribution of post-processing techniques, which are integral steps in refining the segmentation outputs of 3D CNNs. This respective section details how post-processing methods,

such as smoothing, morphological operations, or filtering, are employed to enhance the quality of segmentation results. By structuring the review analysis in this manner, this paper aims to present a comprehensive evaluation of the segmentation pipeline, specifically that pertains to 3D CNN methodologies.

There are several significant trends that have been identified with regard to the design of 3D CNN architectures to address specific challenges of medical imaging dataset characteristics, such as low contrast, noise, or irregular anatomical structures. The adaptability of these architectures reflects their targeted approach to overcoming image-related limitations, emphasizing the importance of architectural customization in achieving effective segmentation outcomes.

In Table 1, we also categorize the segmentation methods based on their backbone networks. The backbone network serves as a reference framework for understanding how these approaches are classified. The identified backbone networks can be generally grouped into 3D U-Net, 3D FCN, 3D CNN, V-Net, and others. It was observed that most of the segmentation models are based on 3D U-Net, highlighting its popularity in medical image segmentation tasks. This preference is likely due to its symmetric encoder-decoder structure, which is particularly effective for capturing multi-scale contextual information and maintaining spatial precision which is crucial for medical imaging applications.

In addition to 3D U-Net, Fully Convolutional Network (FCN) architectures are also frequently employed. FCN eliminates the fully connected layers on the decoder side, enabling pixel-wise predictions and making them suitable for dense segmentation tasks. Meanwhile, vanilla Convolutional Neural Networks (CNNs), which are the foundational architecture for image analysis, have been adapted to 3D applications for volumetric segmentation but often lack the multi-scale feature aggregation of U-Net variants. Furthermore, V-Net is another prominent backbone utilized in medical segmentation. It is a 3D extension of the U-Net design, incorporating residual connections to enhance gradient flow during training, which is particularly beneficial for deeper networks.

In summary, while 3D U-Net remains the dominant choice due to its proven effectiveness, architectures like FCN, vanilla CNN, and V-Net provide additional options, catering to specific requirements of segmentation tasks. Given the focus of this review on 3D CNN-based segmentation methods, a substantial portion of the discussion is dedicated to lung imaging applications. These methods serve as a representative example of the capabilities and variations inherent in 3D CNN-based approaches, making them an ideal case for an in-depth analysis of semantic segmentation strategies. This section will explore the categories and limitations of 3D CNN architectures, as well as highlight the available public lung imaging databases that are commonly used in this field.

Table 1. Summary of the included papers

Reference	Pre-Processing	ROI	3D Backbone Network	Post-Processing
Zhang *et al.* [21]	x	✓	UNET	x
Xu *et al.* [37]	x	x	VNET	x
Shi *et al.* [38]	×	✓	UNET	x
Li *et al.* [39]	✓	x	UNET	x
Jin *et al.* [40]	✓	x	UNET	✓
González Sánchez *et al.* [41]	✓	x	UNET	x
Dalvit Carvalho da Silva *et al.* [42]	x	✓	UNET	x
Ren *et al.* [29]	✓	✓	CNN	x
Nikan *et al.* [43]	✓	x	FCN	x
Gao *et al.* [44]	✓	x	UNET	x
López-Linares Román *et al.* [30]	✓	x	VNET + FCN	x
Chen *et al.* [45]	✓	x	UNET	x
Brahim *et al.* [46]	✓	✓	UNET	x
Zhang *et al.* [47]	✓	x	UNET	x
Yang *et al.* [48]	x	x	UNET	x
Xiao *et al.* [49]	✓	✓	UNET	x
Wang *et al.* [50]	x	✓	VNET	x
Wang *et al.* [51]	x	x	UNET	x
Hussain *et al.* [52]	✓	x	UNET	x
Hossain *et al.* [31]	✓	x	CNN	✓
Zhao *et al.* [32]	✓	✓	UNET	✓
Yang *et al.* [53]	✓	x	CNN	x
Kang *et al.* [54]	✓	✓	UNET	✓
Yang *et al.* [55], [55]	x	✓	FCN	✓
Zheng *et al.* [56]	✓	✓	UNET	x
Xu *et al.* [57]	✓	x	CNN	x
Meng *et al.* [58]	✓	x	CNN	✓
Hu *et al.* [59]	✓	x	CNN	✓

Reference	Pre-Processing	ROI	3D Backbone Network	Post-Processing
Deng *et al.* [60]	✓	x	CNN	x
Alalwan *et al.* [33]	✓	x	UNET	x
Qayyyum *et al.* [61]	x	x	CNN	x
Subramaniam *et al.* [62]	x	x	UNET	x
Sharrock *et al.* [63]	✓	x	VNET	x
Saleem *et al.* [64]	✓	x	UNET	x
Niyas et *al.* [65]	✓	x	UNET	x
Liang *et al.* [66]	✓	x	UNET	x
Li *et al.* [35]	✓	x	UNET	x
Radiuk *et al.* [67]	✓	x	UNET	x
Lin *et al.* [68]	x	x	UNET	x
Feng *et al.* [36]	✓	✓	CNN	✓
Yousefi *et al.* [69]	✓	x	UNET	x
Souadih *et al.* [70]	x	✓	CNN	✓
Dai *et al.* [71]	✓	x	CNN	x
Chen *et al.* [72]	✓	x	CNN	x
Chao *et al.* [73]	✓	x	CNN	x
Baldeon *et al.* [74]	✓	x	CNN	✓
Liu *et al.* [75]	✓	x	CNN	x
Hua *et al.* [76]	x	x	UNET	x
Wang *et al.* [77]	✓	✓	UNET	x
Ao *et al.* [78]	✓	x	CNN	x
Ding *et al.* [79]	✓	x	UNET	x
Xiao *et al.* [80]	x	✓	CNN	x
Yang *et al.* [81]	✓	x	CNN	x
Yang *et al.* [82]	✓	x	UNET VNET	x
Chen et al. [83]	x	x	UNET INCEPTION RESNET	x
Bose et al. [84]	✓	x	UNET	x
Singh et al. [85]	✓	✓	CNN	x

5.1. IMAGE PRE-PROCESSING

According to Table 1, it appears that most studies for semantic segmentation of medical imaging do include pre-processing steps as part of their methodology. To enhance the effectiveness of the training process, 3D medical images are typically preprocessed before being fed into a CNN model. This preprocessing step helps in improving the input data quality due to the presence of unknown noise within the patient's body, which may introduce artifacts. These artifacts can result in unnatural intensity variations, significantly affecting image quality. The outlier voxels generated by these artifacts can negatively impact the performance of deep-learning models during the training process [32]. As a result, several preprocessing techniques have been proposed, including voxel intensity normalization and data augmentation.

Voxel value normalization is commonly applied to CT scan images, as each type of tissue in the scan corresponds to a distinct Hounsfield unit (HU) value. Normalizing the HU scale or applying window clipping enhances the features of the target organ, thereby improving the quality of the training process [29], [30]. Each organ would return different HU scale clipping, for example, head and neck values are in the range of [-200 200] [29], while a lung CT scan would be in the range of [-1000 400] and a kidney CT scan would be in the range of [-100 30] [49].

Data augmentation is another widely used technique to address the challenge of limited training data, a common problem in medical image research [33, 41, 43, 45, 54]. Image augmentation involves generating synthetic data to supplement the existing real dataset, which can be achieved through both simple and complex data generation methods. Simple augmentation techniques include basic image processing operations such as translation, rotation, zooming, and flipping [86]. In contrast, more complex augmentation methods may involve the use of Generative Adversarial Networks (GANs) to generate new data based on specific conditions [87]. Additionally, for brain imaging, skull stripping techniques have been employed to improve segmentation accuracy, as demonstrated in studies by [63], [64], [65], and [66].

5.2. REGIONS OF INTEREST EXTRACTION

Instead of feeding raw input data directly into the 3D CNN architecture, some researchers have chosen to apply ROI extraction approach before the 3D CNN segmentation as you can see in the figure 3. In this method, only a subset of the raw data, specifically the extracted ROI, is input into the 3D CNN model [32].

The primary goal of this approach is to reduce the complexity of the segmentation process and lower computational costs. ROI extraction also serves as an initialization step for subsequent segmentation stages. A notable research trend involves the use of a two-stage CNN segmentation process, where ROI extraction typically focuses on anatomical lesions.

Table 2. ROI extraction methodologies

Type of ROI	Purpose of ROI	Reference
Automatic	Region Selection/ organ localization	[21, 36, 38, 40, 44, 46, 50, 55, 56, 70, 77]
	Organ segmentation	[42]
Manual	Region selection	[29, 85]
	Fixed region selection	[31]
	Statistical calculation	[80]

For instance, Zhang et al. [21] implemented a two-stage segmentation approach, where the first stage involves a coarse ROI extraction, followed by a refinement stage that produces the binary output maps. In their study, automated ROI extraction is performed using a 3D-DMFNet, which detects the femur region and removes irrelevant areas, thus reducing memory usage for the latter refinement stage, which is carried out by the 3D ResUNet model. Similarly, Jin et al. [40] performed both localization and segmentation of the frontal vertebrae slices, utilizing the intensity patterns of the vertebrae for RoI extraction via the U-Net architecture. The concept of employing organ localization methods is commonly applied as a coarse-to-fine segmentation approach, where the organ is first localized, followed by lesion segmentation using a series of CNN networks. A limitation of this approach is that the accuracy of lesion segmentation is heavily dependent on the input from the automated ROI extraction process.

In contrast, Ren et al. [26, 29] employed manual annotation for ROI extraction, utilizing multi-atlas-based segmentation methods. While the studies in [28, 31] relied on the researchers' prior knowledge of the lung's location, opting for fixed region selection on each slice. This also applied in [85], the authors proposed a method to manually enhance sharp edges and shapes around the anomalous region of CT scans before inputting them into the 3D CNN. Additionally, due to the small size of the hippocampus, other research has focused on statistical location-based methods, performing cropping based on calculated regions [80].

5.3. IMAGE POST-PROCESSING

Although, in theory, post-processing should not be required for the CNN model since they are designed to leverage all relevant information to generate optimal results, current network architectures are unable to explicitly enforce certain output constraints, such as 3D connectivity and shape conformity. Therefore, further research is needed to integrate such constraints into the design of network structures. Additionally, overfitting remains a concern, which makes post-processing steps essential for rule-based methods. In this study, a simple 3D connectivity analysis was employed to remove small, isolated regions. Gaussian smoothing was also applied to improve specific cases, while the probability output from the network was utilized to assess the reliability of the segmentation maps, enabling case-specific post-processing adjustments [36].

Jin et al. [40] proposed methods to reduce false positives by excluding small predictions (i.e., those under 200 voxels) and refining the segmentation through binary conversion and connected component analysis. Their approach incorporated mask padding and applied an optimal threshold of 0.75. They also used morphological operations, such as dilation and erosion, to eliminate noise and small patches, resulting in a more than 50% reduction in false positives across various models [31].

Zhao et al. [32] employed a post-processing technique based on kidney anatomy, retaining only those tumor components connected to the kidneys, which significantly enhanced the segmentation performance. In a related study [28], segmentation results were binarized, with a focus on the two largest connected components. Morphological operations were applied to improve accuracy, particularly for small tumors, leading to an improvement of 1.77% and 2.82% in renal tumor segmentation for different training models.

In a study by Yang et al. [55], input volumes were limited to 64 slices, requiring the division of regions into smaller sub-volumes. The segmentation process was refined using majority voting and a 3D conditional random field (CRF) algorithm to correct misclassifications. Similarly, Meng et al. [58] employed fully connected CRFs (FC-CRFs) to refine segmentation boundaries, utilizing CT values and category labels for improved accuracy. In the work of Hu et al. [59], morphological operations were applied to align segmented liver tissues with manual annotations. However, challenges remained in distinguishing organs with similar intensity values, as highlighted by Souadih et al. [70]. In their approach, prior anatomical knowledge combined with mathematical morphology was used to accurately locate the sphenoid sinus, with final segmentations confirmed through largest connected component analysis [74].

6. 3D CNN SEGMENTATION FOR LUNG IMAGING

This section explores various 3D deep-learning techniques applied to lung imaging. Segmenting lung regions is a critical step in the screening and diagnosis of lung-related diseases, such as COVID-19, pneumonia, lung cancer nodules, and other medical conditions [82]. The analysis highlights the unique segmentation challenges posed by each lung disease and how 3D CNN-based algorithms are designed to address these issues, as depicted in Fig. 4 and Table 1.

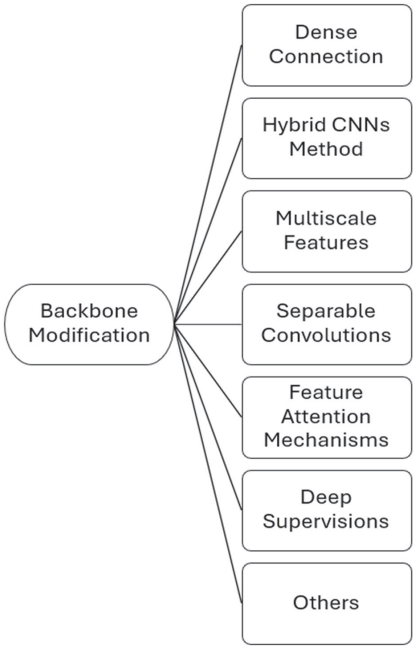


Fig. 4. 3D CNN modifications of segmentation models applied to lung imaging

The reviewed studies categorized 3D CNN segmentation approaches based on their architectural backbones, including U-Net, FCN, CNN, and V-Net, with U-Net being the most frequently employed. The analysis of these models will focus on the modification steps applied to these backbone architectures that aim to enhance segmentation performance and overcome inherent model limitations, often referred to as algorithmic advantages. This section emphasizes lung imaging as a representative use case, as techniques applied to this application are applicable to other medical imaging scenarios. The backbone modifications discussed include dense connections, hybrid CNN methods, multiscale features, separable convolutions, feature attention mechanisms, deep supervisions, and others. Some studies fall into multiple modification categories, as researchers often combined and tailored their models to meet specific objectives or segmentation goals.

6.1. BACKBONE MODIFICATIONS

Many lung lesions are small in size, presenting challenges for segmentation models like U-Net, which is known to be less effective for fine-grained cases. As a result, several studies proposed significant backbone modifications to address these issues. For instance, in [49], a 3D-UNet architecture was enhanced with a 3D-Res2Net module. This hierarchical connection network improves multi-scale feature extraction, capturing finer details and reducing the likelihood of vanishing or exploding gradient problems. The inclusion of 3D-SE blocks recalibrates channel weights, which further optimizes feature representation. The modified architecture, termed as 3D-Res2UNet, achieved a Dice coefficient of 95.30% on the LUNA16 dataset, surpassing the baseline 3D-UNet (89.12%) and 3D-UNet+CRF (93.25%).

In [83], a multiscale block called MSCblock replaced 3D convolution blocks within U-Net. Inspired by the Inception-ResNet architecture, this approach combined parallel convolutional layers of different kernel sizes and identity mappings, enhancing the multiscale feature capability of the model with a more efficient training process. The MSDS-Unet [48] integrated ResNet modules at each block of a 3D U-Net, enabling the network to capture inter-slice continuity and learn richer feature representations. A two-pathway deep supervision mechanism improved gradient flow, leading to better segmentation performance. These enhancements addressed key challenges like vanishing gradients and insufficient feature representation, making the network robust for complex tasks such as lung tumor segmentation.

Other notable modifications include the SegSEUNet architecture [47], which incorporated Recombination and Recalibration Modules (RRM) with SegSE blocks. This embedding enhanced both spatial and channel recalibration, focusing more on tumor-relevant regions while suppressing irrelevant features. SegSEUNet achieved a Dice coefficient of 0.806 ± 0.120, outperforming traditional SE blocks (0.740 ± 0.144).

The study in [50] proposed an adaptation of V-Net using Parametric ReLU (PReLU) activations and CoordConv layers, which incorporated positional awareness that is critical for pulmonary lobe segmentation. The model achieved an average Dice coefficient of 0.947, significantly surpassing the baseline V-Net model (0.795).

Finally, in [52], a modified 3D U-Net with residual connections was employed for volumetric segmentation. This approach stabilized gradient flow and effectively learned from sparse expert-annotated data, improving the model's Dice scores from 0.730 ± 0.066 (baseline) to 0.763 ± 0.069, when it is combined with gradient-based active learning strategies.

6.1.1. DENSE CONNECTIONS

The integration of a dense Conditional Random Field (CRF) framework significantly improved the segmentation model's ability to delineate precise tumor boundaries. For instance, in [47], the CRF refined segmentation probability maps across scales, mitigating boundary inaccuracies and enhancing spatial consistency. The Dice coefficient improved from 0.842 ± 0.082 to 0.851 ± 0.071, and the Positive Predictive Value (PPV) increased from 0.900 ± 0.107 to 0.917 ± 0.101. Dense connections within the 3D-Res2Net module also ensured efficient gradient flow, enabling superior performance for small and irregular nodules.

6.1.2. HYBRID CNN METHODS

The hybrid CNN modifications come in various strategies such as cascading more than one CNN and combining multiple parallel CNNs, which have been

proposed to address the main limitations of a single CNN model. One common approach is coarse-to-fine segmentation, where a coarse segmentation model provides input for a fine segmentation network. For example, in [83], the authors employed a lightweight 3D CNN to capture long-range contextual information and a 2D CNN for fine-grained semantic details. The two networks were fused using a hybrid feature fusion module, which improved computational efficiency and segmentation accuracy. The proposed Hybrid Segmentation Network (HSN) achieved a mean Dice score of 0.898, outperforming standalone 3D CNNs (0.844) and 2D CNNs (0.751).

Another coarse-to-fine approach was proposed in [49], where lung parenchyma was first segmented to isolate the region of interest, followed by a detailed segmentation of lung nodules using a 3D-Res2UNet. This method reduced the influence of surrounding tissues, leading to improved segmentation accuracy for small lesions.

Another hybrid method is the pseudo-3D approach, where 2D feature maps are stacked and processed using 3D convolutions. For instance, in [31], the LungNet framework used stacked 2D slices fused via 3D convolutions, achieving a Dice coefficient of 70.39, outperforming traditional U-Net and LungNet models while maintaining computational efficiency.

6.1.3. MULTISCALE FEATURES

The variation in object sizes and shapes in medical images necessitates the implementation of multi-scale feature extraction. In [84], the D3MSU-Net architecture employed dense dilated convolutions with varying dilation rates to expand the receptive field without increasing the size of the parameters. This design effectively captured multi-scale spatial features, enhancing segmentation accuracy for diverse biomedical datasets. Similarly, MSDS-Unet [48] used multi-scale deeply supervised learning, combining features at different scales to handle heterogeneous tumor characteristics, particularly for small and big-sized tumors.

A multi-scale strategy was also employed in [41], where image cubes of varying dimensions were processed through separate SegSEUNet models. The resultant output maps were further refined using a dense CRF method, resulting in improved segmentation performance. Ablation studies revealed that removing the multi-scale strategy reduced the Dice coefficient from 0.851 ± 0.071 to 0.820 ± 0.115, highlighting its effectiveness.

6.1.4. SEPARABLE CONVOLUTIONS

Deep learning architectures often face challenges due to high computational costs, requiring the development of efficient methods to mitigate these issues. Separable convolution has emerged as one of the main techniques used to reduce computational cost and the number of parameters. For instance, the S3D method proposed in [83] replaces a full 3D convolution with two consecutive layers: a 2D convolution to capture spatial features and a 1D convolution to extract temporal features. This approach effectively decouples the learning process into spatial (inter-slice) and temporal (intra-slice) components. Compared to models utilizing full 3D convolutions, the S3D approach demonstrates superior performance, achieving a 1.1% improvement in Dice evalution.

6.1.5. FEATURE ATTENTION MECHANISMS

Attention mechanisms play a critical role in the segmentation model, particularly in recalibrating feature maps for tumor regions. The SegSE block [47], which is a novel extension of SE blocks, adds spatial recalibration for voxel-specific attention. This mechanism's performance surpasses conventional channel-only recalibration in SE blocks, making it more suitable for segmentation tasks. Comparative studies in the paper demonstrate that SegSE blocks yield better performance than CBAM and SE mechanisms, with a Dice coefficient improvement from 0.740 ± 0.144 (SE) and 0.751 ± 0.179 (CBAM) to 0.806 ± 0.120.

Another approach used the 3D-SE blocks, which are integrated into the Res2Net modules that act as attention mechanisms, enhancing feature map focus by reassigning channel-wise weights. This mechanism improves the model's sensitivity to small or irregular lung nodules, resulting in better segmentation accuracy even for edge features. For example, the proposed network accurately segments smooth ellipse-like and jagged edges, contributing to its high Dice score as shown in [49].

6.1.6. DEEP SUPERVISIONS

Deep supervision is a core innovation in the MS-DS-Unet [48] architecture. By integrating multi-level supervision mechanisms, the network incorporates direct side outputs from hidden layers alongside auxiliary tasks. This approach ensures an effective learning process across different stages of the network. The use of hard fusion and soft fusion strategies combines local and global losses, resulting in more accurate segmentation labeling. Furthermore, the deep supervision mechanism allows the network to better handle multi-scale features and provides consistent improvements in segmentation accuracy.

6.1.7. OTHERS

A unique contribution in [50] is the use of CoordConv layers, which enhance the conventional convolution operator by adding three additional channels that correspond to the x, y, and z coordinates. These added channels enable the model to leverage spatial location as a "soft constraint," significantly reducing errors in segmenting lobes with overlapping or indis-

tinct boundaries. The inclusion of CoordConv layers improved the overall Dice coefficient from 0.795 (baseline) to 0.916. For example, the left-upper lobe Dice coefficient increased from 0.859 to 0.958 with this modification.

A key innovation of another study in [52] is the introduction of gradient-based sample weighting mechanisms to address the noise in machine-generated pseudo-annotations. The first mechanism evaluates gradient similarity, which reflects the alignment of gradients between pseudo-labeled data and expert-annotated validation data, emphasizing sample trustworthiness. The second mechanism assesses gradient magnitude to measure the informativeness of training samples by identifying those that provide new information to the model. By combining these strategies, the model dynamically prioritizes the most reliable and informative samples during the training process. This approach increased the Dice score from 0.607 (using only gradient similarity) to 0.616 when both strategies were employed on a challenging dataset.

To reduce dependency on extensive expert annotation, the method in [52] incorporates a noisy teacher-based active learning strategy. Machine-generated pseudo-labels from the noisy teacher are used to annotate unlabeled data, while a query function adaptively selects the most informative samples for training. By combining gradient similarity and magnitude weights, the model eliminates less trustworthy samples, ensuring a more accurate optimization process. This strategy significantly enhanced segmentation performance, with Dice scores improving from 0.590 (semi-supervised learning alone) to 0.621 when active learning was applied to the Challenge data. While the model's performance on the Benchmark dataset with the active learning strategy managed to further improve the Dice score from 0.756 ± 0.085 to 0.763 ± 0.069.

6.2. CURRENT RESEARCH LIMITATION / CHALLENGES IN LUNG IMAGING

Based on the previous discussion, it is evident that various deep learning-based 3D CNN segmentation methods have demonstrated promising outcomes in generating medical imaging segmentation maps. At the same time, it can be concluded that researchers have introduced diverse approaches to enhance the performance of basic algorithms. Additionally, it should be noted that there are several limitations observed in the field of medical image segmentation, particularly when dealing with challenges such as small lesion size that often causes class imbalance, and poor image quality, which is normally encountered in certain modalities like CT scans.

Firstly, architectural constraints within these algorithms pose a huge challenge. The absence of self-adaptive mechanisms often restricts the model's ability to achieve optimal performance across diverse datasets. Additionally, certain 3D CNN architectures, such as those proposed in [47], exhibit deficiencies in capturing fine contour details, leading to inaccuracies in segmenting complex anatomical structures.

A significant challenge also lies in multiscale feature learning. Many diseases exhibit multiscale characteristics, requiring the models to effectively capture features across varying scales. Despite efforts to integrate multiscale modules, current methods often struggle to accurately learn details across scales, particularly in detecting small tumors, where features may be subtle and highly variable. While other multiscale techniques such as waterfall connections have also been explored [49], their utility remains largely confined to specialized applications, such as small tumor detection, rather than providing generalizable solutions applicable across a broad range of clinical scenarios. Another prominent limitation of this 3D network is the high computational demands to efficiently execute the deep CNN models. As these models grow increasingly complex with many layers, coupled with advanced modules such as squeeze blocks and multiscale pathways, the computational burden and training time of this 3D model has increased significantly. The requirement for extensive computational resources may render effective deployment in clinical environments impractical due to the limited access to high-performance computing infrastructure [28, 31, 47, 49].

Data scarcity in the medical field also presents a significant challenge. High-performing deep learning models require large and well-annotated datasets for optimal training processes. However, several factors such as privacy concerns, the labor-intensive nature of annotation, and the limited availability of publicly accessible datasets often impede the development of robust models. This data shortage issue can lead to overfitting and diminished generalizability problems, thereby reducing the algorithm's effectiveness across diverse patient populations [28, 88].

In conclusion, despite notable advances in the use of 3D CNNs for medical image segmentation, the field still faces several challenges, including architectural limitations, difficulties in multiscale feature extraction, high computational demands, and constrained data availability. Overcoming these obstacles will necessitate continued research into adaptive and resource-efficient algorithms, potentially benefiting from increased collaboration between the fields of computer science and medicine. This topic will be discussed in the next section.

6.3 FUTURE RESEARCH RECOMMENDATIONS IN LUNG IMAGING

From the findings in section 6.2, there are several future research directions that should be explored, which are further discussed in the following subsections.

6.3.1. Challenges in Medical Image Segmentation Dataset

A primary challenge in medical image segmentation lies in the availability of data. Due to strict privacy concerns surrounding patient information, open access to medical datasets remains limited. It is imperative to revisit and refine protocols for data protection to facilitate the use of anonymized datasets without compromising patient confidentiality. Addressing this issue could significantly benefit the research community. Additionally, there is a notable scarcity of volumetric data necessary for training robust deep-learning models. Collaborative efforts between healthcare institutions, domain experts, and image-processing researchers are essential to expand the availability of such data. Establishing training programs for postgraduate students under the guidance of clinical investigators, who are experts in specific diseases, may also support data collection efforts.

Another pressing issue is the labor-intensive and time-consuming process of creating annotated ground truth data. Semi-supervised learning techniques and transfer learning can be leveraged to mitigate this limitation. Pretrained deep learning models, for instance, can effectively reduce the demand for large annotated datasets by utilizing knowledge transfer across related domains.

6.3.2. Advancements in Network Architecture

Currently, most network architectures for medical image segmentation are heavily based on U-Net, which has demonstrated excellent performance in various applications. However, exploring alternative backbones, such as HRNet, could reveal additional potential. Moreover, reconsidering the parameter size within these architectures is also crucial. Increasing complexity by simply adding more parameters is not always efficient. Strategies like pyramid pooling and dilated (atrous) convolutions have emerged as promising alternatives. Dilated convolutions, in particular, help address multiscale challenges, as diseases often present lesions of varying sizes and shapes depending on their stage. However, careful investigation of dilation rates is necessary to avoid the "gridding" effect that arises when large dilation rates are used.

As networks grow more complex, the associated increase in computational cost must also be considered. Depthwise separable convolutions offer a potential solution by significantly reducing the number of parameters, which is particularly advantageous for 3D medical imaging applications. While reduced parameters will lower computational demands, researchers must ensure that model performance and accuracy are not compromised. Balancing these trade-offs may involve integrating techniques like attention mechanisms or deep supervision to maintain existing model performance.

6.3.3. Generalization of Deep Learning Models

A significant limitation of current deep learning models is their generalizability across various conditions. Most models are developed and tested using data from a single source, which limits their ability to generalize across different conditions that may be encountered when varying imaging instruments and configurations are used to capture the images. Expanding studies to include multicenter datasets could greatly enhance model robustness and applicability.

Additionally, many current research often focuses on the segmentation of a single organ or modality. Broadening this scope to include multiple organs or multimodal imaging data for specific diseases could yield more versatile and generalized models. Encouraging healthcare institutions to collect multimodal datasets for particular organs or diseases would further support this research direction and open new avenues for automated medical screening and diagnosis.

6.4 PUBLIC LUNG CLINICAL DATASETS

Most of the research studies utilized public datasets and a few of them mixed with private datasets. Usually, public datasets are the preferred dataset for comparison purposes so that the generalizability capability of the tested algorithms can be compared fairly [47, 48, 50, 88, 89] shown in table 2. It is also important to consider privacy concerns in the medical field, which limit the availability of certain datasets. However, recent trends show that the use of private datasets has become increasingly important. Most of the public datasets are not too big in numbers, highlighting the need for a hybrid approach of combining public and private datasets to support a more effective training process of deep learning models. This strategic combination approach also helps address the challenges of overfitting and class imbalance, ultimately enabling the models to produce better generalization capability in medical research and applications.

Table 2. Public Lung Clinical Datasets

Public Datasets	Segmentation Tasks	Studies that utilize the dataset
NSCLC-Radiomics	Lung Tumor	[31, 47]
LIDC	Lung Tumor	[47]
LUNA 16	Lung Nodule	[36, 48, 88]
COVID-19 – Ma et al.	Covid-19 Lesion	[89]
COVID-19 Challenge	Covid-19 Lesion	[89]

7. CONCLUSION

This review provides a valuable foundation for those new to the application of 3D CNNs in medical image segmentation. It offers the public health communities and computer science researchers, a clearer understanding of both the advantages and limitations of automated segmentation, particularly within the context of lung

disease segmentation tasks. While no single "optimal" method currently exists for segmenting medical images, this paper presents a comprehensive overview of recent advancements in 3D CNN research, serving as a basis for future progress in the field. However, it is crucial to recognize that the deployment of 3D CNN models on real-world datasets remains a significant challenge. To address this, there is an urgent need to amass larger datasets for model training and to explore the potential of synthetically generated data. Furthermore, the development of more robust algorithms that are capable of effectively addressing the multiscale problem is very crucial, given that the variations in lesion and organ sizes across different disease stages differ significantly. This underscores the importance of collaboration between image processing researchers and medical professionals to refine the developed 3D CNN models, ensuring they are aligned with the objectives of having effective and efficient support tools for screening and diagnosis purposes. By fostering such interdisciplinary collaboration, significant strides can be made in improving the accuracy and efficacy of medical image analysis in three dimensions.

8. ACKNOWLEDGEMENT

This work was supported by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme under Grant FRGS/1/2022/ICT02/UITM/02/8.

9. REFERENCES

[1] K. Manikandan, "Blob based segmentation for lung CT image to improving CAD performance", Proceedings of the International Conference on Recent Trends in Information Technology, Chennai, India, 10-12 April 2014, pp. 1-6.

[2] K. T. Navya, G. Pradeep, "Lung nodule segmentation using adaptive thresholding and watershed transform", Proceedings of the 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, Bangalore, India, 18-19 May 2018, pp. 630-633.

[3] B. B. Verma, S. Saraswat, V. Saraswat, R. Misra, "Interstitial Lung Disease Patterns Classification using Hybrid Features Set and Multi Level Segmentation Implemented by Machine Learning Algorithm", Proceedings of the 8th International Conference on Communication and Electronics Systems, Coimbatore, India, 1-3 June 2023, pp. 1811-1815.

[4] O. W. Herng, A. S. A. Nasir, F. N. Shaari, A. S. A. Sukor, "Superpixels-Based Automatic Density Peaks and Fuzzy Clustering Approach in COVID-19 Lung Segmentation", Proceedings of the IEEE 2nd National Biomedical Engineering Conference, Melaka, Malaysia, 5-7 September 2023, pp. 152-157.

[5] J. Ker, L. Wang, J. Rao, T. Lim, "Deep learning applications in medical image analysis", IEEE Access, Vol. 6, 2017, pp. 9375-9389.

[6] P. Blanc-Durand et al. "Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network", European Journal of Nuclear Medicine and Molecular Imaging, Vol. 48, 2021, pp. 1362-1370.

[7] S. Shabani, M. Homayounfar, V. Vardhanabhuti, M. A. N. Mahani, M. Koohi-Moghadam, "Self-supervised region-aware segmentation of COVID-19 CT images using 3D GAN and contrastive learning", Computers in Biology and Medicine, Vol. 149, 2022, p. 106033.

[8] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, S. R. Abdani, "A review of deep learning techniques for lung cancer screening and diagnosis based on CT images", Diagnostics, Vol. 13, No. 16, 2023, p. 2617.

[9] S. Ebrahimkhani, M. H. Jaward, F. M. Cicuttini, A. Dharmaratne, Y. Wang, A. G. S. de Herrera, "A review on segmentation of knee articular cartilage: from conventional methods towards deep learning", Artificial Intelligence in Medicine, Vol. 106, 2020, p. 101851.

[10] S. Modak, E. Abdel-Raheem, L. Rueda, "Applications of deep learning in disease diagnosis of chest radiographs: A survey on materials and methods", Biomedical Engineering Advances, Vol. 5, 2023, p. 100076.

[11] C. W. Wang, M. A. Khalil, N. P. Firdi, "A survey on deep learning for precision oncology", Diagnostics, Vol. 12, No. 6, 2022, 1489.

[12] S. R. Abdani, M. A. Zulkifley, M. I. Shahrimin, N. H. Zulkifley, "Computer-assisted pterygium screening system: A review", Diagnostics, Vol. 12, No. 3, 2022, p. 639.

[13] Z. Zhou, F. Gou, Y. Tan, J. Wu, "A cascaded multistage framework for automatic detection and segmentation of pulmonary nodules in devel-

oping countries", IEEE Journal of Biomedical and Health Informatics, Vol. 26, No. 11, 2022, pp. 5619-5630.

[14] Y. Liu, L. W. L. Yip, Y. Zheng, L. Wang, "Glaucoma screening using an attention-guided stereo ensemble network", Methods, Vol. 202, 2022, pp. 14-21.

[15] X. Xiong, D. H. Steel, B. Obara, "Evaluation of 2D and 3D deep learning approaches for automatic segmentation of the retinal external limiting membrane in spectral domain optical coherence tomography images", Proceedings of the 7th International Conference on Intelligent Informatics and Biomedical Science, Vol. 7, Nara, Japan, 24-26 November 2022, pp. 341-345.

[16] A. Dhillon, G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection", Progress in Artificial Intelligence, Vol. 9, No. 2, 2020, pp. 85-112.

[17] S. P. Singh, L. Wang, S. Gupta, H. Goli, P. Padmanabhan, B. Gulyás, "3D deep learning on medical images: a review", Sensors, Vol. 20, No. 18, 2020, p. 5097.

[18] Y. He, H. Yu, X. Liu, Z. Yang, W. Sun, S. Anwar, A. Mian, "Deep learning based 3D segmentation: A survey", arXiv:2103.05423, 2021.

[19] S. Diciotti, S. Lombardo, M. Falchini, G. Picozzi, M. Mascalchi, "Automated segmentation refinement of small lung nodules in CT scans by local shape analysis", IEEE Transactions on Biomedical Engineering, Vol. 58, No. 12, 2011, pp. 3418-3428.

[20] H. Cui, X. Wang, D. Feng, "Automated localization and segmentation of lung tumor from PET-CT thorax volumes based on image feature analysis", Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, USA, 28 August - 1 September 2012, pp. 5384-5387.

[21] X. Zhang, Y. Zheng, X. Bai, L. Cai, L. Wang, S. Wu, J. Huang, "Femoral image segmentation based on two-stage convolutional network using 3D-DMFNet and 3D-ResUnet", Computer Methods and Programs in Biomedicine, Vol. 226, 2022, p. 107110.

[22] B. Asadi, J. H. Choi, P. Greer, J. Welsh, S. Chalup, J. Simpson, "PO-1686 MRI-Based Synthetic CT Generation for Head and Neck: 2D vs 3D Deep Learning", Radiotherapy and Oncology, Vol. 170, 2022, pp. S1485-S1486.

[23] S. Niyas, S. J. Pawan, M. A. Kumar, J. Rajan, "Medical image segmentation with 3D convolutional neural networks: A survey", Neurocomputing, Vol. 493, 2022, pp. 397-413.

[24] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation", Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, October 5-9, 2015, Part III, pp. 234-241.

[25] N. Elloumi, A. Ben Makhlouf, A. Afli, B. Louhichi, M. Jaidane, J. M. R. Tavares, "CT Images Segmentation Using a Deep Learning-Based Approach for Preoperative Projection of Human Organ Model Using Augmented Reality Technology", International Journal of Computational Intelligence and Applications, Vol. 22, No. 02, 2023, p. 2350006.

[26] M. Tabassum, A. Al Suman, C. Russo, A. Di Ieva, S. Liu, "A Deep Learning Framework for Skull Stripping in Brain MRI", Proceedings of the 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, Sydney, Australia, 24-27 July 2023, pp. 1-4.

[27] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation", Proceedings of the 19th International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17-21 October 2016, Part II, pp. 424-432.

[28] F. Milletari, N. Navab, S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation", Proceedings of the Fourth International Conference on 3D Vision, Stanford, CA, USA, 25-28 October 2016, pp. 565-571.

[29] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, Q. Wang, "Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images", Medical Physics, Vol. 45, No. 5, 2018, pp. 2063-2075.

[30] K. L.-L. Román et al. "3D pulmonary artery segmentation from CTA scans using deep learning with realistic data augmentation", Lecture Notes in Computer Science, Vol. 11040, 2018, Springer, pp. 225-237.

[31] S. Hossain, S. Najeeb, A. Shahriyar, Z. R. Abdullah, M. A. Haque, "A pipeline for lung tumor detection and segmentation from CT scans using dilated convolutional neural networks", Proceedings of ICASSP 2019 - IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 12-17 May 2019, pp. 1348-1352.

[32] W. Zhao, D. Jiang, J. Pena Queralta, T. Westerlund, "Multi-scale supervised 3D U-Net for kidneys and kidney tumor segmentation", arXiv:2004.08108, 2020.

[33] N. Alalwan, A. Abozeid, A. A. ElHabshy, A. Alzahrani, "Efficient 3D deep learning model for medical image semantic segmentation", Alexandria Engineering Journal, Vol. 60, No. 1, 2021, pp. 1231-1239.

[34] J. Ker, S. P. Singh, Y. Bai, J. Rao, T. Lim, L. Wang, "Image thresholding improves 3-dimensional convolutional neural network diagnosis of different acute brain hemorrhages on computed tomography scans", Sensors, Vol. 19, No. 9, 2019, p. 2167.

[35] P. Li, W. Wu, L. Liu, F. M. Serry, J. Wang, H. Han, "Automatic brain tumor segmentation from Multiparametric MRI based on cascaded 3D U-Net and 3D U-Net++", Biomedical Signal Processing and Control, Vol. 78, 2022, p. 103979.

[36] X. Feng, K. Qing, N. J. Tustison, C. H. Meyer, Q. Chen, "Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images", Medical Physics, Vol. 46, No. 5, 2019, pp. 2169-2180.

[37] J. Xu, J. Liu, D. Zhang, Z. Zhou, C. Zhang, X. Chen, "A 3D segmentation network of mandible from CT scan with combination of multiple convolutional modules and edge supervision in mandibular reconstruction", Computers in Biology and Medicine, Vol. 138, 2021, p. 104925.

[38] D. Shi, Y. Pan, C. Liu, Y. Wang, D. Cui, Y. Lu, "Automatic localization and segmentation of vertebral bodies in 3D CT volumes with deep learning", Proceedings of the 2nd International Symposium on Image Computing and Digital Medicine, Chengdu, China, 13-14 October 2018, pp. 42-46.

[39] W. Li, Y. M. Tang, Z. Wang, K. M. Yu, S. To, "Atrous residual interconnected encoder to attention decoder framework for vertebrae segmentation via 3D volumetric CT images", Engineering Applications of Artificial Intelligence, Vol. 114, 2022, p. 105102.

[40] L. Jin et al. "Deep-learning-assisted detection and segmentation of rib fractures from CT scans: Development and validation of FracNet", EBioMedicine, Vol. 62, 2020.

[41] J. C. G. Sánchez, M. Magnusson, M. Sandborg, Å. C. Tedgren, A. Malusek, "Segmentation of bones in medical dual-energy computed tomography volumes using the 3D U-Net", Physica Medica, Vol. 69, 2020, pp. 241-247.

[42] R. Dalvit Carvalho da Silva, T. R. Jenkyn, V. A. Carranza, "Development of a convolutional neural network based skull segmentation in MRI using standard tesselation language models", Journal of Personalized Medicine, Vol. 11, No. 4, 2021, pp. 310.

[43] S. Nikan, K. Van Osch, M. Bartling, D. G. Allen, S. A. Rohani, B. Connors, S. K. Agrawal, H. M. Ladak, "PWD-3DNet: a deep learning-based fully-automated segmentation of multiple structures on temporal bone CT scans", IEEE Transactions on Image Processing, Vol. 30, 2020, pp. 739-753.

[44] Y. Gao, R. Huang, Y. Yang, J. Zhang, K. Shao, C. Tao, Y. Chen, D. N. Metaxas, H. Li, M. Chen, "FocusNetv2: Imbalanced large and small organ segmentation with adversarial shape constraint for head and neck CT images", Medical Image Analysis, Vol. 67, 2021, p. 101831.

[45] J. Chen, H. Li, G. He, F. Yao, L. Lai, J. Yao, L. Xie, "Automatic 3D mitral valve leaflet segmentation and validation of quantitative measurement", Biomedical Signal Processing and Control, Vol. 79, 2023, p. 104166.

[46] K. Brahim, A. Qayyum, A. Lalande, A. Boucher, A. Sakly, F. Meriaudeau, "A 3D network based shape prior for automatic myocardial disease segmentation in delayed-enhancement MRI", IRBM, Vol. 42, No. 6, 2021, pp. 424-434.

[47] B. Zhang, S. Qi, Y. Wu, X. Pan, Y. Yao, W. Qian, Y. Guan, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images", Computer Methods and Programs in Biomedicine, Vol. 222, 2022, p. 106946.

[48] J. Yang, B. Wu, L. Li, P. Cao, O. Zaiane, "MSDS-UNet: A multi-scale deeply supervised 3D U-Net for automatic segmentation of lung tumor in CT", Computerized Medical Imaging and Graphics, Vol. 92, 2021, p. 101957.

[49] Z. Xiao, B. Liu, L. Geng, F. Zhang, Y. Liu, "Segmentation of lung nodules using improved 3D-UNet neural network", Symmetry, Vol. 12, No. 11, 2020, p. 1787.

[50] W. Wang, J. Chen, J. Zhao, Y. Chi, X. Xie, L. Zhang, X. Hua, "Automated segmentation of pulmonary lobes using coordination-guided deep neural networks", Proceedings of the IEEE 16th International Symposium on Biomedical Imaging, Venice, Italy, 8-11 April 2019, pp. 1353-1357.

[51] L. Wang, X. Ye, D. Zhang, W. He, L. Ju, Y. Luo, Z. Ge, "3D matting: A benchmark study on soft segmentation method for pulmonary nodules applied in computed tomography", Computers in Biology and Medicine, Vol. 150, 2022, p. 106153.

[52] M. A. Hussain, Z. Mirikharaji, M. Momeny, M. Marhamati, A. A. Neshat, R. Garbi, G. Hamarneh, "Active deep learning from a noisy teacher for semi-supervised 3D image segmentation: Application to COVID-19 pneumonia infection in CT", Computerized Medical Imaging and Graphics, Vol. 102, 2022, p. 102127.

[53] E. Yang, C. K. Kim, Y. Guan, B. B. Koo, J. H. Kim, "3D multi-scale residual fully convolutional neural network for segmentation of extremely large-sized kidney tumor", Computer Methods and Programs in Biomedicine, Vol. 215, 2022, p. 106616.

[54] L. Kang, Z. Zhou, J. Huang, W. Han, "Renal tumors segmentation in abdomen CT images using 3D-CNN and ConvLSTM", Biomedical Signal Processing and Control, Vol. 72, 2022, p. 103334.

[55] G. Yang, G. Li, T. Pan, Y. Kong, J. Wu, H. Shu, L. Luo, J. L. Dillenseger, J. L. Coatrieux, L. Tang, X. Zhu, "Automatic segmentation of kidney and renal tumor in CT images based on 3D fully convolutional neural network with pyramid pooling module", Proceedings of the 24th International Conference on Pattern Recognition, Beijing, China, 20-24 August 2018, pp. 3790-3795.

[56] R. Zheng, Q. Wang, S. Lv, C. Li, C. Wang, W. Chen, H. Wang, "Automatic liver tumor segmentation on dynamic contrast enhanced MRI using 4D information: deep learning model based on 3D convolution and convolutional LSTM", IEEE Transactions on Medical Imaging, Vol. 41, No. 10, 2022, pp. 2965-2976.

[57] M. Xu, Y. Wang, Y. Chi, X. Hua, "Training liver vessel segmentation deep neural networks on noisy labels from contrast CT imaging", Proceedings of the IEEE 17th International Symposium on Biomedical Imaging, Iowa City, IA, USA, 3-7 April 2020, pp. 1552-1555.

[58] L. Meng, Y. Tian, S. Bu, "Liver tumor segmentation based on 3D convolutional neural network with dual scale", Journal of Applied Clinical Medical Physics, Vol. 21, No. 1, 2020, pp. 144-157.

[59] P. Hu, F. Wu, J. Peng, P. Liang, D. Kong, "Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution", Physics in Medicine & Biology, Vol. 61, No. 24, 2016, p. 8676.

[60] Z. Deng, Q. Guo, Z. Zhu, "Dynamic regulation of level set parameters using 3D convolutional neural network for liver tumor segmentation", Journal of Healthcare Engineering, Vol. 2019, No. 1, 2019, p. 4321645.

[61] A. Qayyum, A. Lalande, F. Meriaudeau, "Automatic segmentation of tumors and affected organs in the abdomen using a 3D hybrid model for computed tomography imaging", Computers in Biology and Medicine, Vol. 127, 2020, p. 104097.

[62] P. Subramaniam, T. Kossen, K. Ritter, A. Hennemuth, K. Hildebrand, A. Hilbert, J. Sobesky, M. Livne, I. Galinovic, A. A. Khalil, J. B. Fiebach, "Generating 3D TOF-MRA volumes and segmentation labels using generative adversarial networks", Medical Image Analysis, Vol. 78, 2022, p. 102396.

[63] M. F. Sharrock, W. A. Mould, H. Ali, M. Hildreth, I. A. Awad, D. F. Hanley, J. Muschelli, "3D deep neural

network segmentation of intracerebral hemorrhage: development and validation for clinical trials", Neuroinformatics, Vol. 19, 2021, pp. 403-415.

[64] H. Saleem, A. R. Shahid, B. Raza, "Visual interpretability in 3D brain tumor segmentation network", Computers in Biology and Medicine, Vol. 133, 2021, p. 104410.

[65] S. Niyas, S. C. Vaisali, I. Show, T. G. Chandrika, S. Vinayagamani, C. Kesavadas, J. Rajan, "Segmentation of focal cortical dysplasia lesions from magnetic resonance images using 3D convolutional neural networks", Biomedical Signal Processing and Control, Vol. 70, 2021, p. 102951.

[66] J. Liang, C. Yang, L. Zeng, "3D PSwinBTS: an efficient transformer-based Unet using 3D parallel shifted windows for brain tumor segmentation", Digital Signal Processing, Vol. 131, 2022, p. 103784.

[67] P. Radiuk, "Applying 3D U-Net architecture to the task of multi-organ segmentation in computed tomography", Applied Computer Systems, Vol. 25, No. 1, 2020, pp. 43-50.

[68] M. Lin, Q. Cai, J. Zhou, "3D Md-Unet: A novel model of multi-dataset collaboration for medical image segmentation", Neurocomputing, Vol. 492, 2022, pp. 530-544.

[69] S. Yousefi, H. Sokooti, M. S. Elmahdy, F. P. Peters, M. T. M. Shalmani, R. T. Zinkstok, M. Staring, "Esophageal gross tumor volume segmentation using a 3D convolutional neural network", Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16-20 September 2018, pp. 343-351.

[70] K. Souadih, A. Belaid, D. B. Salem, "Automatic segmentation of the sphenoid sinus in CT-scans volume with DeepMedics 3D CNN architecture", Medical Technologies Journal, Vol. 3, No. 1, 2019, pp. 334-346.

[71] W. Dai, B. Woo, S. Liu, M. Marques, C. Engstrom, P. B. Greer, S. Crozier, J. A. Dowling, S. S. Chandra, "CAN3D: Fast 3D medical image segmentation via compact context aggregation", Medical Image Analysis, Vol. 82, 2022, p. 102562.

[72] J. Chen, Z. Wan, J. Zhang, W. Li, Y. Chen, Y. Li, Y. Duan, "Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet", Computer Methods and Programs in Biomedicine, Vol. 200, 2021, p. 105878.

[73] X. Cao, H. Chen, Y. Li, Y. Peng, Y. Zhou, L. Cheng, T. Liu, D. Shen, "Auto-DenseUNet: Searchable neural network architecture for mass segmentation in 3D automated breast ultrasound", Medical Image Analysis, Vol. 82, 2022, p. 102589.

[74] M. B. Calisto, S. K. Lai-Yuen, "EMONAS-Net: Efficient multiobjective neural architecture search using surrogate-assisted evolutionary algorithm for 3D medical image segmentation", Artificial Intelligence in Medicine, Vol. 119, 2021, p. 102154.

[75] Y. Liu, Z. Zhang, J. Yue, W. Guo, "SCANeXt: Enhancing 3D medical image segmentation with dual attention network and depth-wise convolution", Heliyon, Vol. 10, No. 5, 2024.

[76] X. Hua et al. "A lightweight multi-scale multi-angle dynamic interactive transformer-CNN fusion model for 3D medical image segmentation", Neurocomputing, Vol. 608, 2024, p. 128417.

[77] S. Wang, Y. Wang, Y. Peng, X. Chen, "MSA-Net: Multi-scale feature fusion network with enhanced attention module for 3D medical image segmentation", Computers and Electrical Engineering, Vol. 120, 2024, p. 109654.

[78] Y. Ao, W. Shi, B. Ji, Y. Miao, W. He, Z. Jiang, "MS-TCNet: An effective Transformer–CNN combined network using multi-scale feature learning for 3D medical image segmentation", Computers in Biology and Medicine, Vol. 170, 2024, p. 108057.

[79] W. Ding, S. Geng, H. Wang, J. Huang, T. Zhou, "FDiff-Fusion: Denoising diffusion fusion network based on fuzzy learning for 3D medical image segmentation", Information Fusion, Vol. 112, 2024, p. 102540.

[80] Z. Xiao, Y. Zhang, Z. Deng, F. Liu, "Light3DHS: A lightweight 3D hippocampus segmentation method using multiscale convolution attention and vision transformer", NeuroImage, Vol. 292, 2024, p. 120608.

[81] F. Yang, F. Wang, P. Dong, B. Wang, "HCA-former: Hybrid Convolution Attention Transformer for 3D Medical Image Segmentation", Biomedical Signal Processing and Control, Vol. 90, 2024, p. 105834.

[82] J. Yang, H. Li, H. Wang, M. Han, "3D medical image segmentation based on semi-supervised learning using deep co-training", Applied Soft Computing, Vol. 159, 2024, p. 111641.

[83] W. Chen, H. Wei, S. Peng, J. Sun, X. Qiao, B. Liu, "HSN: hybrid segmentation network for small cell lung cancer segmentation", IEEE Access, Vol. 7, 2019, pp. 75591-75603.

[84] S. Bose, R. S. Chowdhury, R. Das, U. Maulik, "Dense dilated deep multiscale supervised U-network for biomedical image segmentation", Computers in Biology and Medicine, Vol. 143, 2022, p. 105274.

[85] S. P. Singh, L. Wang, S. Gupta, B. Gulyas, P. Padmanabhan, "Shallow 3D CNN for detecting acute brain hemorrhage from medical imaging sensors", IEEE Sensors Journal, Vol. 21, No. 13, 2020, pp. 14290-14299.

[86] E. Elizar, M. A. Zulkifley, R. Muharar, "Scaling and cutout data augmentation for cardiac segmentation", Proceedings of the International Conference on Data Science and Applications, Vol. 2, 2023, pp. 599-609.

[87] M. A. Zulkifley, S. R. Abdani, N. H. Zulkifley, "COVID-19 screening using a lightweight convolutional neural network with generative adversarial network data augmentation", Symmetry, Vol. 12, No. 9, 2020, p. 1530.

[88] Y. Feng, P. Hao, P. Zhang, X. Liu, F. Wu, H. Wang, "Supervoxel based weakly-supervised multi-level 3D CNNs for lung nodule detection and segmentation", Journal of Ambient Intelligence and Humanized Computing, 2023, pp. 1-11.

[89] X. Wang, P. Teng, P. Lo, A. Banola, G. Kim, F. Abtin, J. Goldin, M. Brown, "High throughput lung and lobar segmentation by 2D and 3D CNN on chest CT with diffuse lung disease", Proceedings of Image Analysis for Moving Organ, Breast, and Thoracic Images: Third International Workshop, Fourth International Workshop, and First International Workshop, Granada, Spain, 16-20 September 2018, Proceedings 3, pp. 202-214.

Optimizing Gastric Cancer Classification with QCNN and Fine-Tuning

Original Scientific Paper

Aroop D S*
Department of Computer Science and Engineering
Karunya Institute of Technology and Science, Coimbatore
aroopd@karunya.edu.in

Sujitha Juliet D
Department of Computer Science and Engineering
Karunya Institute of Technology and Science, Coimbatore
sujitha@karunya.edu

Shamila Ebenezer
Department of Computer Science and Engineering
Karunya Institute of Technology and Science, Coimbatore
Shamila_cse@karunya.edu

*Corresponding author

Abstract *– Cancer ranks as one of the primary contributors to morbidity and mortality worldwide, standing as the second leading cause of death on a global scale. According to data from the National Cancer Registry Program of the Indian Council of Medical Research, over 1300 individuals in India lose their lives daily as a result of cancer-related causes. Gastric cancer is among the top five most prevalent cancers globally, after cancer in the lung, breast, colorectum, and prostate, highlighting the importance of accurate classification for effective treatment strategies. In this study, a novel approach utilizing a Quadratic Convolutional Neural Network combined with Extreme Learning and Fine-Tuning technique, a deep learning architecture specifically designed to capture intricate patterns and features within medical imaging data. Fine tuning technique is used to enhance the model's generalization capability and adaptability to diverse datasets. Through extensive experimentation and validation on a comprehensive dataset comprising gastric cancer images, the proposed approach achieves an impressive accuracy of 94%. The findings indicate the efficacy of the proposed approach for classifying gastric cancer. With its high accuracy and robust performance, the developed QCNN model holds promise for assisting clinicians in accurate diagnosis and prognosis of gastric cancer patients, ultimately contributing to improved patient outcomes and personalized treatment strategies.*

Keywords: *Deep learning, Digestive system, Quadratic Convolutional Neural Network, Endoscopy, Gastric cancer, Extreme Learning, Fine tuning*

Received: October 18, 2024; Received in revised form: February 10, 2025; Accepted: February 11, 2025

1. INTRODUCTION

Gastric cancer, also referred to as stomach cancer, is a malignant tumor that develops in the stomach [1]. Various factors contribute to its development, including Helicobacter pylori infection, dietary habits, smoking, genetic predispositions, and familial history [2]. The structure and function of the stomach provide important context for understanding gastric cancer and its impact on the body. The stomach is a vital organ in the digestive system, connecting the esophagus to the small intestine [3]. Structurally, it is J-shaped with distinct regions: the cardia, fundus, body, and pylorus. Lymph nodes surround it, aiding in immune response. The pylorus serves as a connection to the duodenum, with the pyloric sphincter facilitating stomach emptying [4]. The greater curvature forms a convex lateral surface, while the lesser curvature creates a concave medial border. This complex anatomy supports the stomach's role in digestion and underscores its importance in overall health. Understanding the intricate anatomy of the stomach is crucial in recognizing the early symptoms and improving the prognosis of gastric cancer through timely detection and intervention.

Early symptoms may be nonspecific, making early detection challenging. Gastric cancer prognosis heavily relies on the stage of diagnosis, highlighting the critical

importance of timely detection for effective treatment and improved outcomes [5]. Timely identification of gastric cancer is crucial for improving patients' survival rates. The disease often progresses asymptomatically or with mild symptoms in its initial stages, leading to delayed diagnosis and treatment initiation. As gastric cancer advances, it becomes more difficult to treat, with limited therapeutic options and poorer outcomes [6]. Therefore, early detection through screening programs enables the identification of tumors at an earlier, more treatable stage, facilitating curative interventions such as surgery, chemotherapy, or radiation therapy. Improved prognosis associated with early detection underscores the significance of developing accurate and efficient diagnostic methods for gastric cancer.

Convolutional neural network (CNN) have transformed medical image analysis by allowing automated interpretation of diagnostic images with exceptional accuracy and efficiency [7]. CNN excel at learning hierarchical representations of image features directly from raw data, eliminating the need for handcrafted features or domain-specific knowledge. Through the use of large annotated datasets, CNNs can identify subtle patterns and abnormalities in medical images, aiding in disease detection, classification, and prognosis. In the context of gastric cancer detection, CNNs offer a promising approach for analyzing endoscopic images, histopathological slides, and radiological scans to assist clinicians in identifying suspicious lesions. Gastric cancer classification has been extensively studied due to its critical role in improving patient outcomes. Various deep learning approaches, such as CNNs, ResNet, and EfficientNet, have demonstrated significant potential in analyzing endoscopic and histopathological images. However, these methods often face limitations such as overfitting on small datasets, difficulties in capturing intricate patterns unique to gastric cancer, and reduced generalizability across diverse patient populations. These challenges highlight the need for more advanced and adaptable models capable of addressing these limitations while maintaining high accuracy in classification.

The main goal of this research is to create and assess a new method for detecting gastric cancer, which integrates a Quadratic Convolutional Neural Network (QCNN) with Extreme Learning and Fine Tuning techniques [8-9]. By doing so, we aim to improve the sensitivity, specificity, and overall accuracy of gastric cancer detection while also addressing the shortcomings of current diagnostic methods. Specifically, the study seeks to:

1. Investigate the feasibility and effectiveness of QCNN in analyzing gastric cancer-related imaging data.
2. Explore the integration of Extreme learning and Fine tuning techniques to optimize model performance and generalization.
3. The performance of the method was thoroughly evaluated on a diverse dataset comprising gastric cancer images.
4. Compare the diagnostic accuracy of the QCNN approach with current deep learning models for gastric cancer detection.

2. LITERATURE REVIEW

In their study, Lee et al. [10] developed a multilayer feedforward neural network using a scaled conjugate gradient backpropagation technique. The World Health Organization (WHO) recognizes cancer as a heterogeneous disease with various subtypes, highlighting the critical importance of early prognosis and diagnosis to improve survival rates. Therefore, there is an increasing need in cancer research to facilitate subsequent clinical management of patients. The authors identified 10 amino acid biomarkers in saliva and extracted 19 fingerprint Raman bands generated by these biomarkers, which can effectively differentiate between cancer patients and healthy individuals. Back propagation was employed to minimize the network error, while scaled conjugate gradient backpropagation was utilized for training the artificial network classifier. The approach yielded an accuracy of 92.27.

Qiu et al. [11] aimed to enhance the efficiency of GC diagnosis by utilizing DL algorithms to aid in diagnosing gastric cancer. Lesion samples in the images were annotated by multiple endoscopists with extensive clinical experience. The acquired training set was input into a CNN for training, resulting in the algorithm model DLU-Net identified with an overall accuracy of 94.1%. A cascaded deep learning model was suggested by Teramoto et al. [12] to identify the invasive location and categorize endoscopic images. Two different U-Net models are used to segment the images labeled as cancer based on the amount of invasion by stomach cancer.

Deep CNN was utilized by Xie et al. [13] to achieve automatic categorization of pathological images related to stomach cancer since DCNN is capable of efficiently extracting deep characteristics from images. A CNN architecture was developed by Hatami et al. [14] for the identification of stomach cancer. The authors were motivated by the concept of the fire module to decrease the architecture's size and improve the model's classification accuracy. The findings indicate that this model has an 89% classification accuracy.

A model based on the Deeplab v3+ neural network was proposed by Wang et al. (2021) [15] to increase the effectiveness of gastric cancer. With a 92.76% accuracy rate and a 91.66% Dice coefficient, the model outperforms the SegNet and Faster-RCNN models by over 12%. Additionally, the model's parameter scale is significantly lowered. An approach called U-Net R-CNN was proposed by Teramoto et al. in (2021) [16] based on a semantic segmentation method. In order to identify stomach cancer, U-Net was presented as a semantic segmentation technique. The primary constraint of the strategy is the limited quantity of images used for training.

A strategy for identifying and classifying gastric cancer areas from gastrointestinal endoscopic images was developed by Shibata et al. in 2020 [17] using Mask R-CNN. The results suggest that the sensitivity per image was 96.0%.

The CNN-based approach was introduced by Li et al. [18] to assess stomach mucosal lesions detected by M-NBI. CNN's diagnostic accuracy for early-stage stomach cancer was 90.91%. GoogLeNet, a deep neural network architecture, was employed by Horiuchi et al. (2020) [19] to identify stomach cancer. With 220 of the 258 images properly diagnosed, the accuracy was 85.3%. The nature of this investigation is retrospective. Table 1 provides an overview of recent research on the identification of stomach cancer, highlighting a range of strategies and techniques.

Table 1: Summary of recent works on gastric cancer detection

Author, Reference & Year	Methodology	Remarks
Lee et al. [10] 2021	Multilayer feedforward neural network back propagation technique	Achieved 92.27% accuracy in cancer detection based on saliva biomarkers and Raman spectroscopy. Performance may vary based on the number of neurons and hidden layers in the neural network.
Qiu et al. [11] 2022	CNN	Achieved 94.1% accuracy in identifying different types of lesions. Limited to the analysis of gastro scopic images.
Teramoto et al [12] 2022	Cascaded deep learning model and U-Net models.	Limited to endoscopic images collected from a single facility.
Xie et al. [13] 2023	DCNN with Adapted GoogLeNet and AlexNet models for gastric cancer diagnosis.	Improved sensitivity using GoogLeNet and AlexNet models. Also significantly reduces the computational burden.
Hatami et al. [14] 2020	CNN incorporated with fire module architecture for increased accuracy.	Achieved 89% classification accuracy on a dataset of gastric disease images. Limited to classification of gastric diseases observed through endoscopy.
Teramoto et al. [16] 2021	Developed a U-Net R-CNN model for object detection in gastric cancer images, combining semantic segmentation.	Limited by a small number of images collected from a single facility, affecting generalizability.
Shibata et al. [17] 2020	Utilized Mask R-CNN for detection and segmentation of early gastric cancer regions from endoscopic images.	The suggested approach was implemented utilizing the information gathered from a solitary establishment.
Li et al. [18] 2020	CNN-based system using narrow-band imaging (M-NBI).	Achieving 90.91% accuracy in diagnosing early gastric cancer. Limited the study to non-polypoid and non excavated lesions, restricting the applicability of the CNN system.
Horiuchi et al. [19] 2020	Employed GoogLeNet for diagnosing MENBI images of lesions undergoing endoscopic sub mucosal dissection (ESD) treatment.	Achieved an accuracy of 85.3% in identifying lesions. Retrospective study design with limited clarity in some images, potentially impacting diagnostic accuracy.

3. METHODOLOGY

A novel approach is introduced for identifying and classifying gastric cancer through the application of deep learning techniques from stomach endoscopy images. Three main steps make up the methodology: feature extraction, classification, and preprocessing. A visual illustration of the suggested methodology is shown in Fig. 1. For classification, the QNN is utilized as the primary classifier, leveraging its ability to effectively classify complex patterns in medical images.

Additionally, extreme learning and fine-tuning techniques are applied to further enhance the performance of the classifier, refining its ability to accurately detect and categorize gastric cancer.

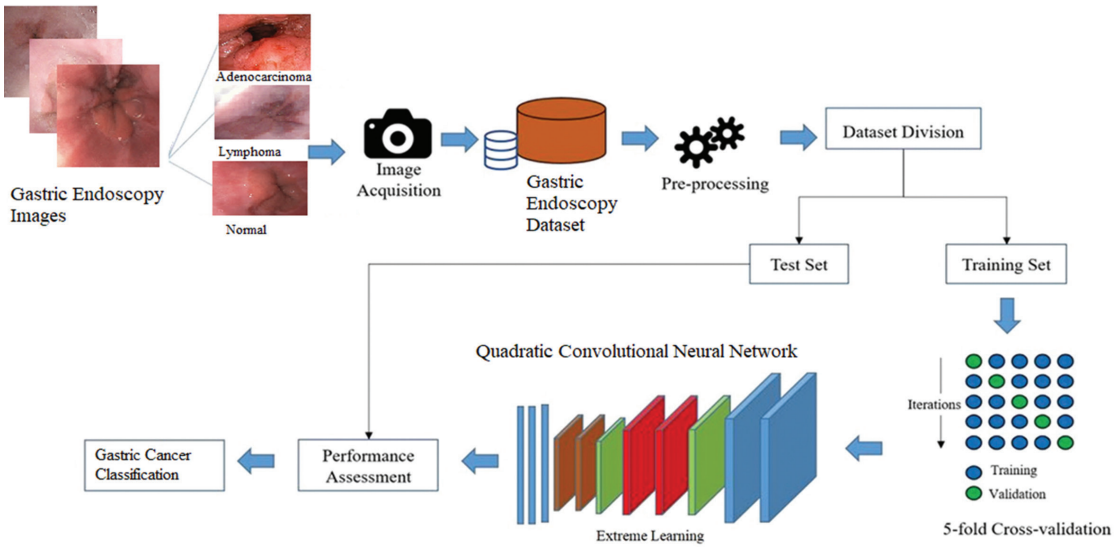


Fig. 1. Flow diagram of the proposed work

3.1. DATASET

The dataset provides a useful tool for the suggested deep learning model for the detection and categorization of stomach cancer. The endoscopic images from the Fujita Health University Hospital database comprised the dataset used in this investigation. These images are standardized to a size of 256×256 pixels and are represented in the RGB color space. The dataset includes images representing two types of gastric cancer and healthy control images. Each category within the dataset presents a diverse range of endoscopic views and pathological conditions, providing a comprehensive representation of gastric abnormalities for analysis and classification. With images standardized in size and color space, and categorized into relevant groups, the dataset facilitates consistent and reliable analysis, enabling researchers to effectively assess the performance and generalizability of the developed classification system across different pathological conditions and patient populations. Fig. 2 presents sample images from the gastric cancer dataset, providing visual examples of the types of endoscopy images used in the study for classification purposes.

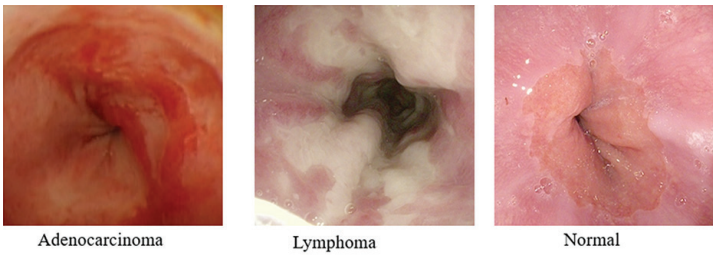


Fig. 2. Sample images from the dataset

3.2. DATA PREPROCESSING AND AUGMENTATION

In the data preprocessing and augmentation phase, two key techniques are applied to the images obtained from the gastric endoscopy dataset, which are image augmentation and image resizing. Image augmentation is a critical step aimed at enriching the diversity and quantity of images in the dataset, which is essential for effectively training deep neural networks. This technique involves applying various image processing operations such as flipping, rotation, and cropping to generate a new augmented version of the original image. Flipping involves mirroring the image horizontally or vertically, thereby introducing variations in orientation. Rotation entails rotating the image by a certain degree, which simulates different viewing angles. Cropping involves extracting a portion of the image, which can help focus on specific regions of interest. By introducing these modifications, the dataset is expanded with a wider range of perspectives and variations. The augmentation procedure helps to increase the model's capacity to generalize to previously unobserved data and enhances validation accuracy by exposing the model to a more comprehensive set of scenarios and conditions.

Image resizing is employed to standardize the size of the endoscopy images to 256x256 pixels and 3 color channels (RGB). This operation is crucial for reducing the computational complexity of the deep learning process. Larger images require processing a higher number of pixels, which increases the computational time and complexity. By resizing the images to a uniform size, the computational burden is mitigated, facilitating more efficient processing and analysis by the deep neural network. Uniform image sizes ensure consistency in the input data, which is essential for achieving reliable and reproducible results. The combination of image augmentation and image resizing optimizes the dataset for training and enhances the efficiency of the subsequent classification task. Preprocessing is employed to enhance the quality of the images, ensuring optimal input for subsequent analysis. Feature extraction aims to extract relevant information from the images that can distinguish between cancerous and non-cancerous tissues. Fig. 3 depicts data visualization, facilitating insights into the distribution and characteristics of the data.

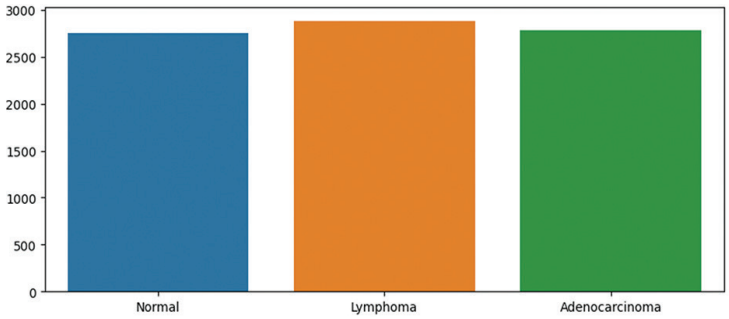


Fig. 3. Data visualization

3.3. QUADRATIC CONVOLUTIONAL NEURAL NETWORK

QNN is a type of neural network architecture specifically designed for image classification tasks [20]. It builds upon the traditional CNN architecture by incorporating quadratic convolutional layers, which introduce additional non-linearity to the network. By applying quadratic filters to the input image, the convolutional layers in a QNN enable the network to identify more intricate patterns and correlations in the data. These quadratic filters enable the network to model non-linear interactions between image features, enhancing its ability to discriminate between different classes.

QNNs typically include pooling layers, fully connected layers, and activation functions, similar to traditional CNN architectures as shown in Figure 4. The use of quadratic convolutional layers distinguishes QNNs from standard CNNs, offering potentially improved performance for certain image classification tasks. However, training and optimizing QNNs may require additional computational resources and careful parameter tuning due to their increased complexity compared to traditional CNNs.

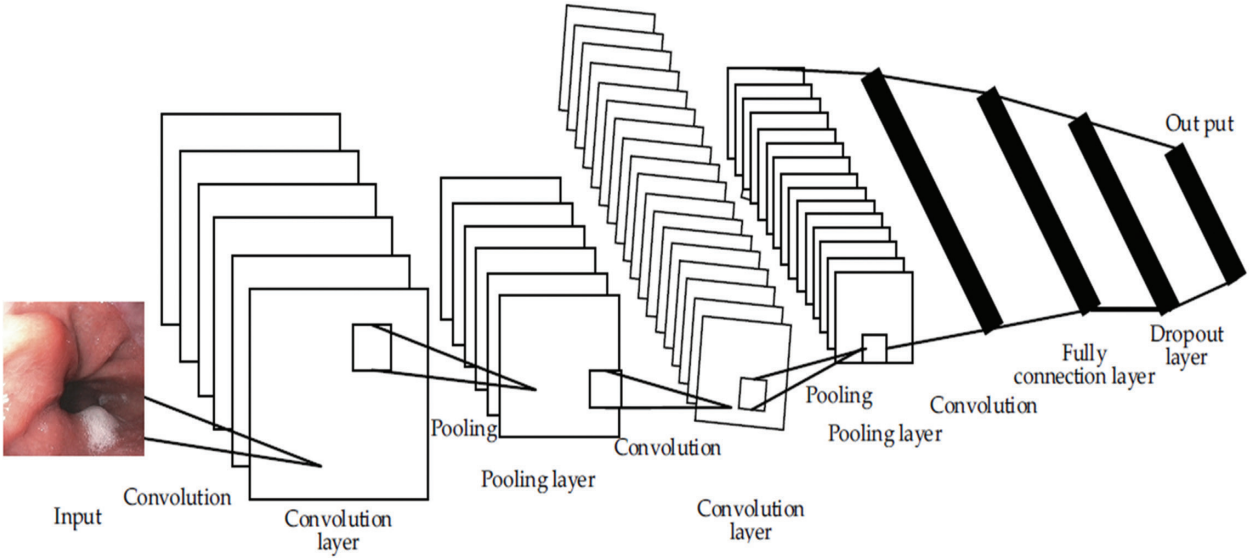


Fig. 4. Architecture of the QCNN model

The input vector be S^t with dim as dimension and the transpose be $\{\}^t$. Linear function of the neuron be

$$F(S)=WS+B \tag{1}$$

Weights are represented by $W=\{w_1, w_2,,,,,,w_{dim}\}$ and the bias be 'b'.

A quadratic function for a neuron can be described as a mathematical expression that incorporates quadratic terms, representing a non-linear relationship between the neuron's input and output.

$$Q(S)=S'^t\, W_q\, S' \tag{2}$$

S'^t denotes the augmented vector represented by $S'^t=\{S^t \mid 1\}=\{s_1, s_2,,,,s_{dim},1\}$ and the weights are

$$\text{Weight}_C = \begin{bmatrix} W'_{1,1} & W'_{1,2} & \cdot & \cdot & W'_{1,\dim+1} \\ W'_{2,1} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ W'_{\dim,1} & \cdot & \cdot & \cdot & W'_{\dim+1,\dim+1} \end{bmatrix} \tag{3}$$

For image classification, we construct networks using a mix of quadratic and linear neurons, with the convolutions layers created utilizing linear neurons for later stages of classification, and quadratic neurons are employed for extracting picture representation. Two factors guided the decision: first, the network can learn complicated representations when higher-order functions are used to extract picture data; second, the kernel size restricts the neurons' input dimension, resulting in networks with manageable parameters.

Let $S_{m,n}{}^t=\{s_1, s_2, ,,,s_{N^2})\}$ represent the pixels in the image's receptive field that the kernel spans at position (m,n). Next, the quadratic neuron's output is calculated as

$$QD(S_{m,n}) = S'^t{}_{m,n}\, W_q\, S'_{m,n} \tag{4}$$

A quadratic kernel has $P(n^2+1)^2$ parameters, while a CNN layer with linear neurons made up of P kernels of size $n\times n$ has $P(n^2+1)$ parameters; however, n is typically limited to smaller values (1, 3, 5). thereby making it possible for us to construct quadratic networks with a controllable rise in the number of parameters.

3.4. EXTREME LEARNING

Extreme Learning begins by randomly initializing the input weights and biases of the hidden neurons. These weights and biases are typically drawn from a random distribution. Once the parameters are initialized, the algorithm proceeds with forward propagation.

Given an $N \times P$ sized data matrix (X)where P is the number of samples, the hidden layer output (H) is computed using a non-linear activation function ($g(.)$). Mathematically, this can be expressed as:

$$H=g(W\cdot X+b) \quad (5)$$

$W\cdot X$ represents the weighted sum of inputs to the hidden layer, and $g(.)$ is typically a sigmoid, tanh, or ReLU function applied element-wise. Following the computation of the hidden layer output, ELM proceeds to compute the output weights(B) using a linear regression approach. This is achieved by solving a linear system, expressed as:

$$B = H^{+}.T \quad (6)$$

where H^{+} is the Moore-Penrose pseudo-inverse of Hand Tis the target output matrix. In matrix notation, this equation can be represented as:

$$B = (HT.H)^{-1}\ H^{T}.T \quad (7)$$

Fig. 5 presents a flowchart depicting the Extreme Learning process, outlining the sequential steps involved in implementing this technique.

ELM offers a streamlined approach to training neural networks. By randomly initializing the input weights and biases and utilizing a fixed, single hidden layer, ELM achieves fast training speeds, particularly advantageous for large datasets. Despite its simplicity, ELM's performance hinges on the quality of randomly chosen parameters and the representativeness of the training data.

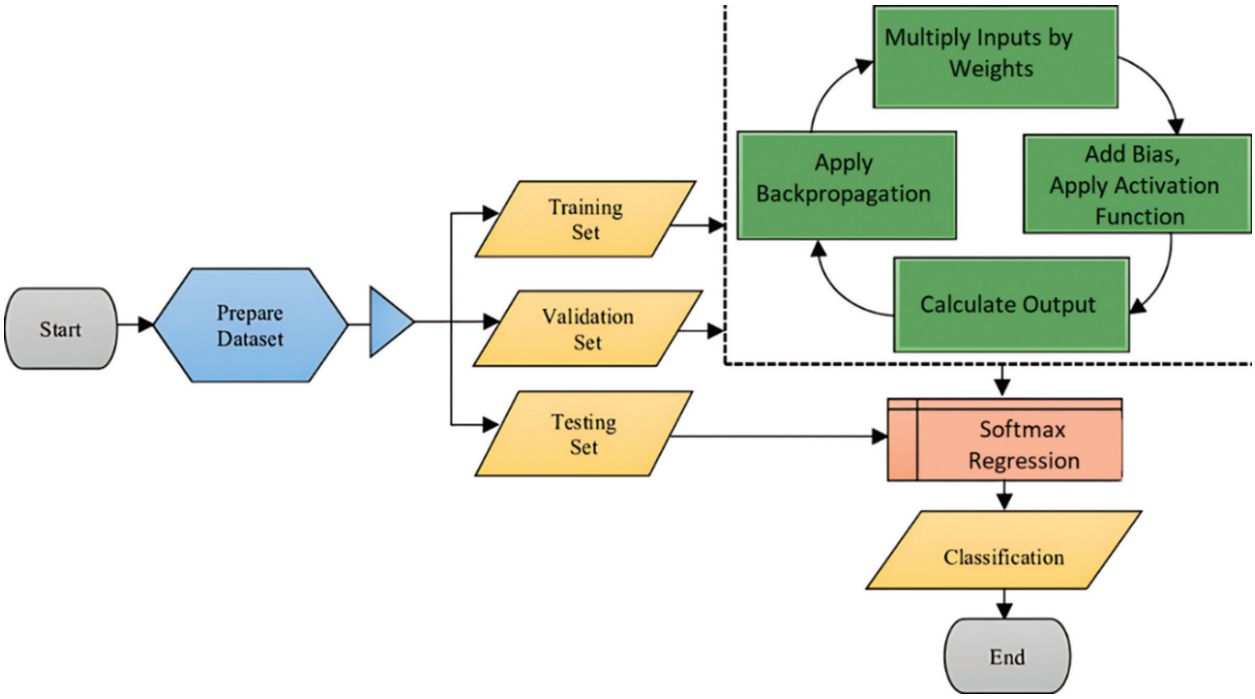


Fig. 5. Flowchart of Extreme learning

3.5. PROPOSED ARCHITECTURE

The proposed research introduces a novel system for identifying and classifying gastric cancer from stomach endoscopy images, employing deep learning techniques. Three primary steps make up the methodology: feature extraction, classification, and preprocessing. Pre-processing is the process of improving the endoscopic image quality in order to make further analysis easier. Feature extraction aims to derive meaningful characteristics from the images, representing the disease-affected regions effectively. These features serve as inputs for training and testing the classification model, which is pivotal in accurately categorizing the presence and type of gastric cancer in the images. Fig. 6 displays the proposed model architecture for gastric cancer classification.

The heart of the methodology lies in the feature extraction process, as it directly influences the efficacy of the classification system. Extracting a large number of features from each image could lead to computational inefficiencies during classifier training. To address this, QCNN is proposed as the classifier. However, QCNNs typically demand vast amounts of training data, which may not always be readily available, especially in the context of gastric endoscopy images. In cases where the dataset is limited, methods such as extreme learning and fine-tuning become valuable for enhancing the classifier's performance.

The dataset comprises gastric endoscopy images from various subjects, encompassing different categories, including two types of cancers and healthy controls. By leveraging this dataset, the proposed methodology aims to train a QCNN framework, updating its parameters using the available training set.

This approach enables the classification system to learn and distinguish between different cancer types and

healthy tissue accurately, contributing to improved diagnosis and treatment of gastric cancer. Table 2 provides an overview of the hyperparameters used in the model.

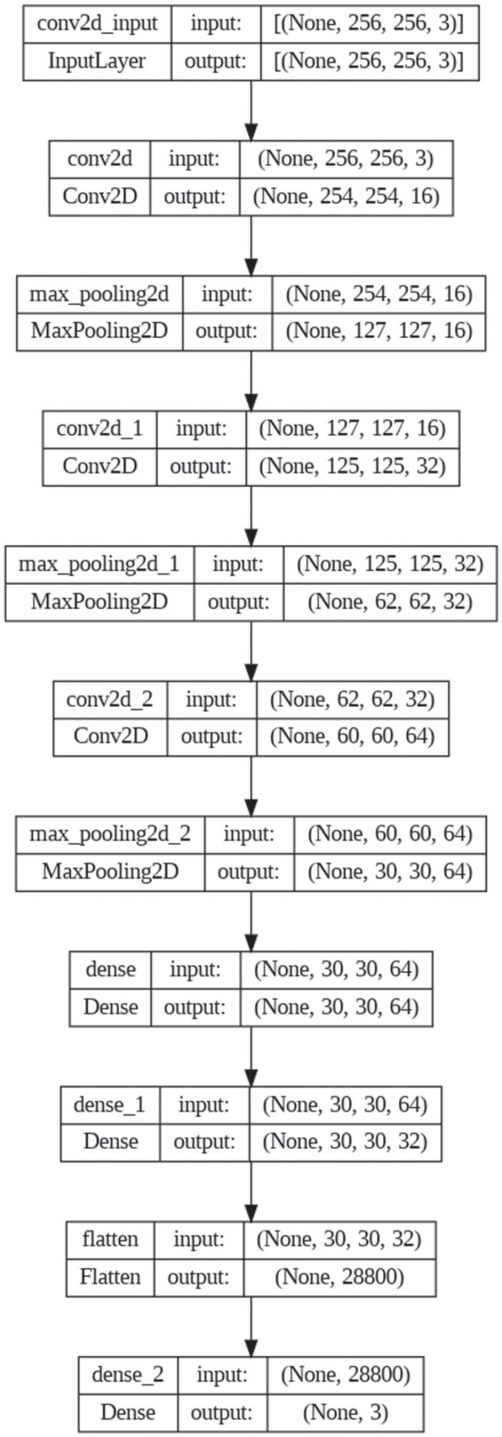


Fig. 6. Proposed model architecture

Table 2. Hyper parameters

Parameters	Values
Optimizer	Adam
Activation Function	Relu, softmax
Loss	Categorical crossentropy
Batch Size	128
Number of epochs	25

3.6. FINE TUNING

Fine-tuning, it is generally accepted that all model layers should be copied, with the exception of the final layer. This new layer should have the same number of neurons as in the new target domains. Fine-tuning a portion of the network, usually the last layer, enables the network to adjust to the characteristics of the target domain, leading to enhanced performance across various classification tasks. This is because the early layers of the network extract features that are applicable to a wide range of image recognition tasks. A detailed overview of the suggested approach for classifying stomach cancer is shown in Table 3.

Table 3. Model summary

Total Parameters	116,227
Trainable Parameters	116,227
Non- Trainable Parameters	0

3.7. HARDWARE AND SOFTWARE SETUP

The proposed study utilized the Google Colaboratory platform in conjunction with the Microsoft Windows 10 operating system to establish a robust computational environment. The modeling process involved the application of the Python programming language, leveraging the Keras package and TensorFlow backend for training. The conceptualized models were specifically configured to accept preprocessed and augmented datasets, ensuring precise decision-making capabilities. To assess the efficacy of the proposed model, evaluate the predictions of the model on the test dataset.

4. RESULT AND DISCUSSION

Performance indicators that are essential for assessing the model's efficacy, especially in classification tasks, include accuracy, recall, precision, and F1-score. Table 4 gives an idea about the performance parameters used in the study. Accuracy is a straightforward metric, often used when the class distribution is balanced. While accuracy offers a view of performance of the model, its adequacy might be limited in scenarios where there is an imbalance in the class distribution.

Accuracy by itself might not be sufficient to fully comprehend a model's performance, particularly in cases of imbalanced datasets in which one class predominates over the other. Contrarily, precision calculates the percentage of accurate positive predictions made out of all positive forecasts.

Recall quantifies the percentage of real positive cases among all actual positive cases that the model correctly identifies as true positives. It aids in determining how well the model is able to locate every positive occurrence without overlooking any. When false negatives are expensive, recall becomes crucial. The f1-Score provides a balance between precision and recall by taking the har-

monic mean of these two criteria. It is particularly helpful in cases of unequal class distribution. F1-Score is a more accurate indicator of a model's overall performance, particularly in situations where the class distribution is not uniform or when simultaneous optimization of precision and recall is required. Table 5 provides a comprehensive classification report for the system.

Table 4. Performance parameters

Parameters	Equation
Accuracy	(TP+TN)/(TP+TN+FP+FN)
Precision	(TP)/(TP+FP)
Recall	(TP)/(TP+FN)
F1-Score	2*(Precision*Recall)/ (Precision+Recall)
TP=True Positive TN=True Negative FP=False Positive FN=False Negative	

Table 5. Classification report for the system

	precision	recall	f1-score	support
Adenocarcinoma	0.98	0.96	0.97	522
Lymphoma	0.93	0.94	0.93	590
Normal	0.93	0.94	0.93	572
accuracy			0.94	1684
macro avg	0.95	0.95	0.95	1684
weighteg avg	0.95	0.94	0.94	1684

An accuracy plot of a proposed system displays the performance of the system over different iterations, epochs, or other training iterations. It shows how the accuracy of the model evolves during the training process. As shown in Fig. 7, the x-axis usually represents the number of iterations or epochs, while the y-axis represents the accuracy achieved by the model on the training or validation data. This plot is essential for understanding how well the proposed system learns from the data over time. It helps in diagnosing potential issues such as overfitting or underfitting.

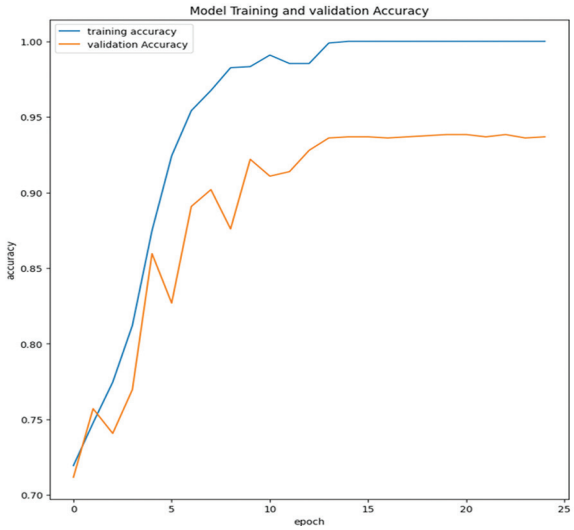


Fig. 7. Accuracy plot of the proposed system

A loss plot of a proposed system illustrates how the loss function decreases (or increases) over the course of training epochs or iterations, as shown in Fig. 8. The loss function quantifies how well the model is performing; typically, lower values indicate better performance. During the initial stages of training, the loss is typically high as the model's parameters are randomly initialized, and it makes random predictions. As training progresses, the model adjusts its parameters to minimize the loss, aiming to improve its predictions. The loss plot should show a decreasing trend over time. Fluctuations in the loss values may occur due to various factors such as the complexity of the dataset, learning rate, and batch size.

Monitoring the loss plot is crucial for assessing the training progress and diagnosing potential issues like overfitting or underfitting.

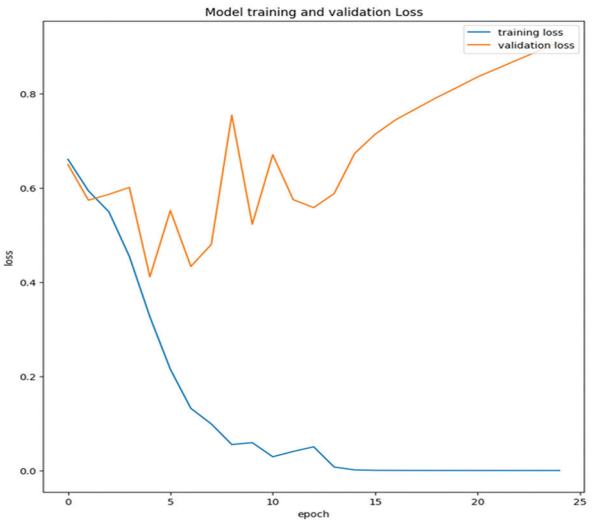


Fig. 8. Loss plot of the proposed system

As seen in Fig. 9, the confusion matrix offers a thorough analysis of the right and wrong predictions the model made on a dataset. In the confusion matrix, it is demonstrated that 501 adenocarcinoma images were correctly classified as adenocarcinoma, 552 lymphoma images were correctly classified as lymphoma, and 538 images were correctly classified as normal.

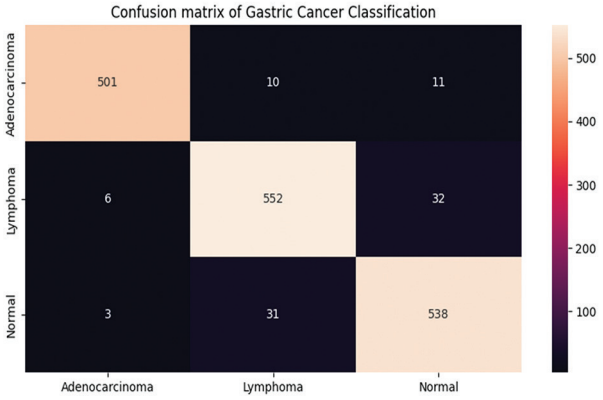


Fig. 9. Confusion matrix of the proposed system

Fig. 10 illustrates the classification output of the proposed gastric cancer classification system, showcasing the model's predictions for a sample set of endoscopy images.

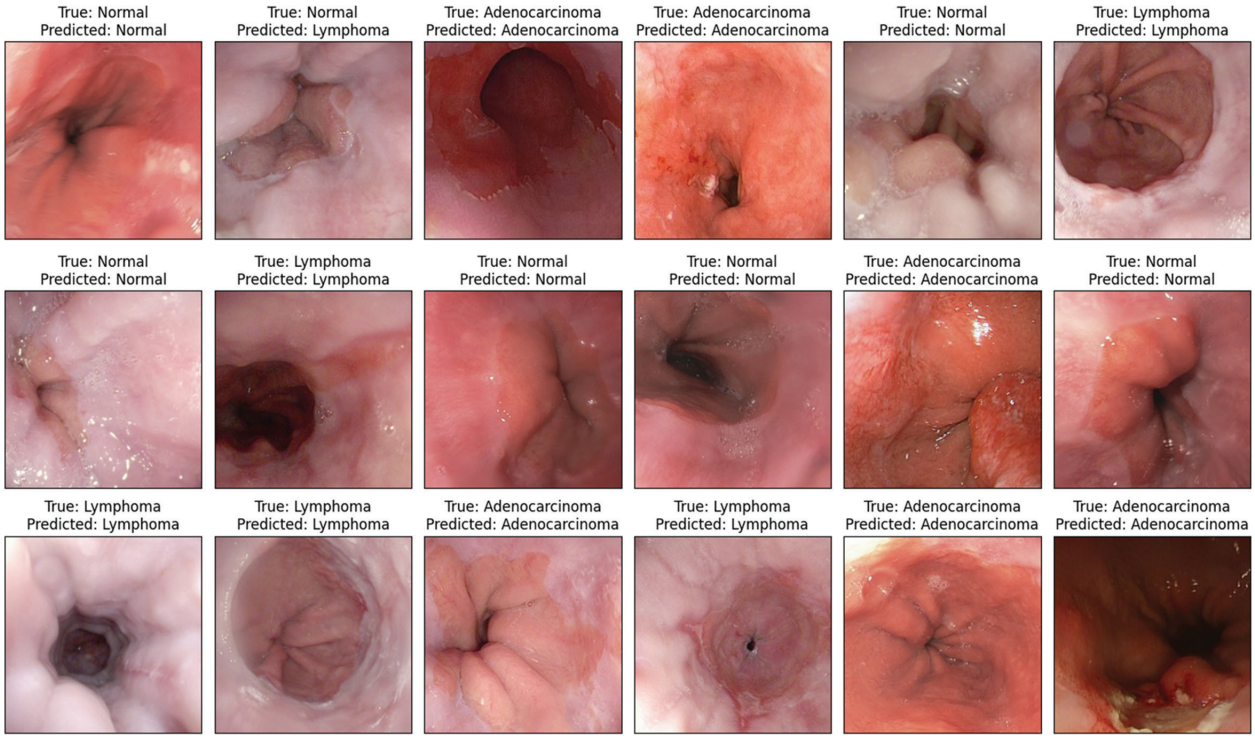


Fig. 10. Classification output of the proposed system

Table 6. A comparison between the suggested system and existing methods

Sl No	Author & Year	Methodology	Accuracy
1	Gong et al [23]	Deep learning based clinical decision support system	81.5 %
2	Yao et al [27]	YOLO	85.15 %
3	Du et al [22]	ENDOANGEL-MM	86.54 %
4	Zhou et al [24]	EfficientNet	88.3 %
5	Li et al [26]	Deep learning based ENDOANGEL-LA	88.76 %
6	Liu et al [21]	2 DCNN	90.8 %
7	Jin et al [25]	Mask R-CNN	90.25 %
	Proposed System		**94 %**

5. CONCLUSION

With gastric cancer ranking as the fifth most frequent cancer globally, cancer continues to pose a serious threat to global health. Accurate classification is essential for effective treatment strategy and improving patient outcomes. The study presents a novel approach utilizing a QCNN combined with extreme learning and fine-tuning techniques for the classification of gastric cancer from stomach endoscopy images. The proposed methodology achieves an impressive accuracy of 94% through extensive experimentation and validation on a comprehensive dataset comprising gastric cancer images. The results demonstrate the effectiveness and potential of the approach for gastric cancer classification. By leveraging QCNN architecture specifically designed to capture intricate patterns and features within medical imaging data, with a fine-tuning technique to enhance generalization capability, the model delivers robust performance. The technique of concatenating QCNN structures with extreme learning proves to be efficient in achieving peak classification rates. The developed QCNN model holds significant promise for assisting clinicians in accurate diagnosis and personalized treatment strategies, ultimately contributing to better patient outcomes in the fight against gastric cancer.

6. REFERENCES:

[1] E. Morgan et al. "The current and future incidence and mortality of gastric cancer in 185 countries, 2020–40: a population-based modelling study", eClinicalMedicine, Vol. 47, 2022.

[2] L. Yang et al. "Gastric cancer: Epidemiology, risk factors and prevention strategies", Chinese Journal of Cancer Research, Vol. 32, No. 6, 2020, p. 695.

[3] A. Tobias, N. M. Sadiq, "Physiology, gastrointestinal nervous control", StatPearls Publishing, 2019.

[4] M. Hsu, A. O. Safadi, F. Lui, "Physiology, stomach", StatPearls Publishing, 2023.

[5] F. H. Wang et al. "The Chinese Society of Clinical Oncology (CSCO): clinical guidelines for the diag-

nosis and treatment of gastric cancer, 2021", Cancer Communications, Vol. 41, No. 8, 2021, pp. 747-795.

[6] T. H. Patel, M. Cecchini, "Targeted therapies in advanced gastric cancer", Current Treatment Options in Oncology, Vol. 21, No. 9, 2020, p. 70.

[7] N. Ketkar, J. Moolayil, "Convolutional neural networks", Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch, 2021, pp. 197-242.

[8] P. Mantini, S. K. Shah, "CQNN: Convolutional Quadratic Neural Networks", Proceedings of the 25th International Conference on Pattern Recognition, Milan, Italy, 10-15 January 2021, pp. 9819-9826.

[9] T. Alshalali, D. Josyula, "Fine-tuning of pre-trained deep learning models with extreme learning machine", Proceedings of the International Conference on Computational Science and Computational Intelligence, Las Vegas, NV, USA, 12-14 December 2018, pp. 469-473.

[10] S. A. Lee, H. C. Cho, H. C. Cho, "A novel approach for increased convolutional neural network performance in gastric-cancer classification using endoscopic images", IEEE Access, Vol. 9, 2021, pp. 51847-51854.

[11] W. Qiu, J. Xie, Y. Shen, J. Xu, J. Liang, "Endoscopic image recognition method of gastric cancer based on deep learning model", Expert Systems, Vol. 39, No. 3, 2022, p. e12758.

[12] A. Teramoto, T. Shibata, H. Yamada, Y. Hirooka, K. Saito, H. Fujita, "Detection and characterization of gastric cancer using cascade deep learning model in endoscopic images", Diagnostics, Vol. 12, No. 8, 2022, p. 1996.

[13] K. Xie, J. Peng, "Deep learning-based gastric cancer diagnosis and clinical management", Journal of Radiation Research and Applied Sciences, Vol. 16, No. 3, 2023, p. 100602.

[14] S. Hatami, R. Shamsaee, M. H. Olyaei, "Detection and classification of gastric precancerous diseases using deep learning", Proceedings of the 6th Iranian Conference on Signal Processing and Intelligent Systems, Mashhad, Iran, 23-24 December 2020, pp. 1-5.

[15] J. Wang, X. Liu, "Medical image recognition and segmentation of pathological slices of gastric cancer based on Deeplab v3+ neural network", Computer Methods and Programs in Biomedicine, Vol. 207, 2021.

[16] A. Teramoto, T. Shibata, H. Yamada, Y. Hirooka, K. Saito, H. Fujita, "Automated detection of gastric cancer by retrospective endoscopic image dataset using u-net r-cnn", Applied Sciences, Vol. 11, No. 23, 2021, p. 11275.

[17] T. Shibata, A. Teramoto, H. Yamada, N. Ohmiya, K. Saito, H. Fujita, "Automated detection and segmentation of early gastric cancer from endoscopic images using mask R-CNN", Applied Sciences, Vol. 10, No. 11, 2020, p. 3842.

[18] L. Li et al. "Convolutional neural network for the diagnosis of early gastric cancer based on magnifying narrow band imaging", Gastric Cancer, Vol. 23, 2020, pp. 126-132.

[19] Y. Horiuchi et al. "Convolutional neural network for differentiating gastric cancer from gastritis using magnified endoscopy with narrow band imaging", Digestive Diseases and Sciences, Vol. 65, 2020, pp. 1355-1363.

[20] Y. Jiang, F. Yang, H. Zhu, D. Zhou, X. Zeng, "Nonlinear CNN: improving CNNs with quadratic convolutions", Neural Computing and Applications, Vol. 32, 2020, pp. 8507-8516.

[21] L. Liu et al. "Diagnosis and segmentation effect of the ME-NBI-based deep learning model on gastric neoplasms in patients with suspected superficial lesions—a multicenter study", Frontiers in Oncology, Vol. 12, 2023, p. 1075578.

[22] H. Du et al. "A deep-learning-based system using multi-modal data for diagnosing gastric neoplasms in real-time (with video)", Gastric Cancer, Vol. 26, No. 2, 2023, pp. 275-285.

[23] E. J. Gong et al. "Deep learning-based clinical decision support system for gastric neoplasms in real-time endoscopy: development and validation study", Endoscopy, Vol. 55, No. 8, 2023, pp. 701-708.

[24] B. Zhou, X. Rao, H. Xing, Y. Ma, F. Wang, L. Rong, "A convolutional neural network-based system for detecting early gastric cancer in white-light en-

doscopy", Scandinavian Journal of Gastroenterology, Vol. 58, 2023, pp. 157-162.

[25] J. Jin et al. "Automatic detection of early gastric cancer in endoscopy based on Mask region-based convolutional neural networks (Mask R-CNN) (with video)", Frontiers in Oncology, Vol. 12, 2022, p. 927868.

[26] J. Li et al. "Development and validation of a feature extraction-based logical anthropomorphic diagnostic system for early gastric cancer: A case-control study", eClinicalMedicine, Vol. 46, 2022, p. 101366.

[27] Z. Yao, T. Jin, B. Mao, B. Lu, Y. Zhang, S. Li, W. Chen, "Construction and multicenter diagnostic verification of intelligent recognition system for endoscopic images from early gastric cancer based on YOLO-V3 algorithm", Frontiers in Oncology, Vol. 12, 2022, p. 815951.

Efficient Privacy-Utility Optimization for Differentially Private Deep Learning: Application to Medical Diagnosis

Original Scientific Paper

Rafika Benladgham*
Abu Baker Belkaid University of Tlemcen, Faculty of Sciences, Department of Computer Sciences, LRIT
Tlemcen, Algeria
rafika.benledghem@univ-tlemcen.dz

Fethallah Hadjila
Abu Baker Belkaid University of Tlemcen, Faculty of Sciences, Department of Computer Sciences, LRIT
Tlemcen, Algeria
fethallah.hadjila@univ-tlemcen.dz

Adam Belloum
University of Amsterdam, Informatics Institute
Amsterdam, The Netherlands
a.s.z.belloum@uva.nl

*Corresponding author

Abstract *– The optimization of differentially private deep learning models in medical data analysis using efficient hyper-parameter tuning is still a challenging task. In this context, we address the fundamental issue of balancing privacy guarantees with model utility by simultaneously optimizing model parameters and privacy parameters across two primary medical datasets, with additional validation on PathMNIST. Our framework encompasses both tabular data (Wisconsin Breast Cancer dataset) and medical imaging (BreastMNIST and PathMNIST), implementing four distinct optimization approaches: Grid Search, Random Search, Bayesian Optimization, and Bat Algorithm. Through extensive experimentation, we demonstrate a promising performance: achieving 93.62% accuracy with strong privacy guarantees (ε = 0.5) for tabular data, and 74.91% accuracy for medical imaging, with the Bat Algorithm discovering an unprecedented privacy level (ε = 0.293). Further validation on PathMNIST histopathology images demonstrated the framework's scalability, achieving 44.71% accuracy with privacy guarantees (ε = 2.603). Our comparative analysis reveals that different medical data types require distinct optimization strategies, with Bayesian Optimization excelling in tabular data applications and Random Search providing efficient solutions for image processing. The experiments with PathMNIST histopathology images provided valuable insights into the framework's behavior with complex medical data, revealing configuration-dependent performance variations and computational trade-offs. Our framework incorporates Pareto analysis and visualization techniques to enable systematic exploration of privacy-utility trade-offs, while early stopping mechanisms optimize privacy budget utilization. This comprehensive approach, validated across diverse medical imaging complexities and data modalities, establishes practical guidelines for implementing privacy-preserving machine learning in healthcare settings while highlighting the importance of balanced optimization strategies and computational efficiency in secure and efficient medical data analysis.*

Keywords*: differential privacy, deep learning, medical data analysis, privacy-utility optimization, hyper-parameter tuning*

Received: November 20, 2024; Received in revised form: January 30, 2025; Accepted: March 7, 2025

1. INTRODUCTION

The rapid digital transformation of healthcare has led to an unprecedented accumulation of sensitive medical data, from structured tabular data to complex medical imaging [1]. While this data surge offers immense opportunities for advancing medical research and improving diagnostic accuracy through machine learning, it also introduces a critical challenge: balancing data utility and privacy protection [2]. This challenge is particularly acute in healthcare, where advancing research and safeguarding patient confidentiality must coexist. Consequently, there is an urgent need for robust privacy-preserving mechanisms that do not compromise the analytical capabilities of machine learning models.

Among privacy protection approaches, differential privacy (DP) stands out for its mathematically sound privacy assurances [3, 4]. Within deep learning applications, differentially private stochastic gradient descent (DP-SGD) has become the predominant DP implementation method [5]. This approach combines two key mechanisms: gradient clipping and noise addition to ensure privacy protection during model training [5]. The process involves first limiting individual gradients by clipping them to maintain a specific $\ell 2$ norm threshold, followed by incorporating Gaussian noise into the averaged gradient before model parameter updates [5, 6]. By implementing these modifications, DP-SGD achieves bounded sensitivity for each training update, thus establishing privacy guarantees through controlled noise introduction into the learning process [5, 7].

However, optimizing hyper-parameters in differentially private models is inherently more complex than in non-private settings. Private hyperparameter optimization requires tuning additional parameters, including the clipping norm and noise scale, which are highly sensitive and make the process intricate and demanding [8]. Earlier studies have focused on fine-tuning privacy parameters to match non-private model performance or achieving acceptable performance levels while ensuring privacy assurances [8]. Despite these efforts, significant gaps remain in optimizing hyperparameters for differentially private deep learning models.

Traditional HP.O methods, such as grid search (G.S) and random search (R.S), are non-adaptive, evaluating hyperparameters from fixed or randomly generated sets [9]. While simple to apply, they are computationally intensive and poorly suited for high-dimensional search spaces, especially when additional privacy parameters (e.g., noise multiplier, clipping norm) must be tuned [9]. While adaptive methods like Bayesian optimization (B.O) use a probabilistic model to link hyperparameters to performance metrics and have become the preferred choice over non-adaptive methods due to their superior performance and scalability [10, 11], they often struggle to dynamically adjust privacy parameters during training, which is critical for balancing utility and privacy in real-time applications [10, 11]. Furthermore, existing approaches lack the ability to effectively navigate the complex trade-offs between exploration and exploitation in private HPO, limiting their scalability and performance in privacy-sensitive domains like healthcare [9]. These limitations underscore the need for innovative optimization techniques that can handle the unique challenges of differentially private deep learning. Inspired by its adaptive exploration and exploitation capabilities, we propose the Bat Algorithm as a novel solution for dynamic parameter tuning, addressing these gaps and enabling more efficient and scalable privacy-preserving models.

Recent advancements have demonstrated the effectiveness of swarm intelligence algorithms, such as the Bat Algorithm, in navigating complex search spaces [12]. Inspired by bats' echolocation and social behaviors, the Bat Algorithm dynamically adjusts search patterns to identify optimal hyper-parameters. Its ability to balance local and global search capabilities and adaptive frequency tuning makes it particularly well-suited for fine-tuning hyper-parameters in complex scenarios [12]. In differentially private deep learning, where privacy parameters like noise multiplier and clipping norm are critical, the Bat Algorithm offers a promising approach for dynamic parameter tuning. This leads to a compelling research question: *Can the Bat Algorithm be integrated to dynamically adjust privacy parameters during training, further improving the efficiency of differentially private model optimization*?

In this work, we address the challenge of HPO in differentially private deep learning by focusing on four explicit hyperparameters (learning rate, batch size, privacy budget, and maximum gradient norm) and two implicit ones (noise multiplier, and training epochs via early stopping). Our contributions are fourfold:

1. **Novel Application of the Bat Algorithm:** We propose and evaluate the Bat Algorithm for HP.O in differentially private deep learning, marking its first application in this domain.
2. **Comprehensive Comparison:** We systematically compare the Bat Algorithm against baseline methods (G.S, R.S, and B.O), providing consistent and reproducible results.
3. **Real-World Validation:** We validate our approach on real-world medical datasets (Breast Cancer Wisconsin, BreastMnist), demonstrating its practical applicability in privacy-sensitive healthcare applications.
4. **Generalizability and Scalability:** To further demonstrate the generalizability, feasibility, and scalability of our framework, we extend our evaluation to the PathMNIST dataset, which is more complex in terms of both data structure (histopathology images) and model architecture (ResNet-50). This extension rigorously tests the applicability of our framework to larger and more complex datasets, further validating its potential for real-world deployment in privacy-sensitive medical applications.

The remainder of this paper is organized as follows: Section 2 outlines the methodology and experimental setup, Section 3 presents results and analysis, Section 4 discusses findings, and Section 5 concludes with implications and future research directions.

2. RELATED WORK

The intersection of differential privacy and deep learning has been an active area of research, particularly in optimizing the balance between privacy guarantees and model utility.

2.1. DIFFERENTIAL PRIVACY IN MACHINE LEARNING

Recent advancements in Differential privacy (DP) have significantly expanded theoretical foundations and

practical applications of privacy-preserving techniques. Kulynych et al. [13] introduced an attack-aware noise calibration framework that moves beyond traditional ε-based approaches, demonstrating improved model accuracy while maintaining strong privacy guarantees. Complementing this work, Lu [14] established crucial relationships between noise addition strategies in stochastic gradient descent (SGD) and their impact on the model performance. In the domain of privacy budget management, Thantharate et al. [15] developed a systematic approach for tracking cumulative privacy loss across iterative training processes, enabling more precise control over privacy budgets in multi-stage learning scenarios. Pan Ke et al. [16], systematically investigate differentially private deep learning, addressing privacy attacks and preservation with a novel taxonomy. Despite these efforts, optimizing the privacy-utility trade-off continues to pose substantial challenges.

2.2. PRIVACY-UTILITY TRADE-OFFS AND OPTIMIZATION

Transfer learning approaches have shown promising results in medical image diagnosis. Battula and Chandana. [17] demonstrated 99.68% accuracy for cervical cancer classification using an optimized SE-ResNet152 model, highlighting the potential of architecture optimization in healthcare applications. The growing need for privacy preservation, however, necessitates approaches that balance such high performance with robust privacy guarantees. The progress in privacy-preserving machine learning has significantly enhanced our understanding of the privacy-utility trade-off paradigm. Kumar et al. [18] introduced a novel geometric approach using kernel-based methods in Reproducing Kernel Hilbert Spaces (RKHS), effectively reducing accuracy loss while mitigating membership inference risks in sensitive applications. Based on this foundation, Ficiu et al. [19] developed PFairDP, employing Bayesian optimization to identify Pareto-optimal points balancing fairness, privacy, and utility. Significant contributions to federated learning frameworks have emerged, with with Avent et al. [20] presenting a Bayesian optimization methodology to efficiently characterize the privacy-utility trade-off of differentially private algorithms using empirical utility measurements, while Koskela et al. [21] propose a method to enhance differentially private machine learning by tuning hyperparameters on a random data subset and extrapolating optimal values, reducing both privacy leakage and computational cost. Arous et al. [22], demonstrated choice strategies of model parameters (e.g., activation functions) that can significantly impact the privacy utility balance without compromising either aspect.

2.3. HYPER-PARAMETER OPTIMIZATION IN DIFFERENTIALLY PRIVATE DEEP LEARNING (DPDL)

Numerous approaches have been proposed to address the challenge of hyper-parameter optimization in DP. Galli et al. [23] offer foundational insights by dynamically optimizing the clipping threshold in differentially private learning, showing that traditional grid search methods incur excessive privacy costs, while Wang et al. [24] developed DP-HyPO, an adaptive framework leveraging Gaussian process-based optimization.

Significant algorithmic contributions include evolutionary approaches for exploring hyperparameter spaces, Bayesian optimization for probabilistic performance modeling, and enhanced Particle Swarm Optimization (EPSO) as Gao et al. [25] demonstrated for optimizing learning rates while minimizing noise impact. Bu et al. [26] introduced a novel book-keeping technique that improves computational costs while maintaining accuracy, making private optimization comparable to standard training. Tobaben [27] provides foundational insights by analyzing hyperparameter and architectural impacts on the privacy-utility trade-off in DPDL, revealing grid search's inefficiencies with private data.

2.4. META-HEURISTIC APPROACHES AND PARETO OPTIMIZATION

The work by Ramalingam et al. [28] and Banerjee et al. [29] has demonstrated the effectiveness of genetic algorithms, particle swarm optimization, and ant colony optimization in navigating vast solution spaces. These approaches have shown a particular promise in healthcare applications, with Singh et al. [30] successfully applying them to enhance feature selection for disease diagnosis while maintaining privacy constraints. The Pareto optimization aspect of these methods, as explored by Harkare et al. [31],is crucial in balancing multiple competing objectives, such as model accuracy, privacy guarantees, and computational efficiency. Thakur et al. [32] further extended these concepts to resource-constrained environments, demonstrating significant improvements in operational efficiency while maintaining solution diversity.

Table 1. Overview of Key Related Work

Research Area	Key Contributions	Reference
Differential Privacy	Attack-aware noise calibration beyond ε-based approaches	[13]
	Noise addition strategies in SGD	[14]
Privacy-Utility Trade-offs and optimization	SE-ResNet152 optimization using DHO algorithm for medical image classification	[17]
	Subset-based hyperparameter tuning for privacy-utility optimization	[21]
Hyper-parameter Optimization in DPDL	DP-HyPO adaptive framework	[24]
	EPSO for learning rate optimization	[25]
	Systematic optimization strategy comparison	[26]
Meta-heuristic Approaches	Genetic and particle swarm optimization analysis	[28]
	Healthcare feature selection optimization	[29]

3. PROBLEM FORMULATION

Our problem formulation establishes a unified framework for optimizing the privacy-utility trade-off in differentially private deep learning models. We have defined an objective function that balances model accuracy and privacy guarantees and formulated the optimization problem for four distinct approaches: Grid Search, Random Search [33], Bayesian Optimization [34], and Bat Algorithm [35]. Each method navigates the hyperparameter space Θ in its unique way, aiming to find the optimal configuration $\theta*$ that maximizes our objective function $f(\varepsilon(\theta), A(\theta))$.

3.1. DIFFERENTIAL PRIVACY FRAMEWORK

A randomized algorithm $M : D \to R$ with domain D and range R is (ε, δ)- differentially private if for all $S \subseteq R$ and for all adjacent datasets $D, D' \in D$ [36]:

$$P[M(D) \in S] \leq exp(\varepsilon) \cdot P[M(D') \in S] + \delta \quad (1)$$

where: - ε is the privacy budget - δ is the failure probability.

3.2. DP-OPTIMIZATION COMPONENTS

3.2.1. Hyperparameter Space

Let $\theta = (lr, bs, nm, C)$ be the hyperparameter vector where: - lr: learning rate - bs: batch size - nm: noise multiplier - C: gradient clipping threshold The feasible space Θ is defined by:

$$\begin{gathered}\Theta = \{\theta \mid lr_{min} \leq lr \leq lr_{max}, \\ bs_{min} \leq bs \leq bs_{max}, \\ t\varepsilon_{min} \leq t\varepsilon \leq t\varepsilon_{max}, \\ C_{min} \leq C \leq C_{max}\end{gathered} \quad (2)$$

3.2.2. Privacy-Utility Metric

For any configuration θ: - $A(\theta)$: model accuracy - $\varepsilon(\theta)$: achieved privacy budget

3.2.3. Objective Function

The privacy-utility trade-off is quantified by:

$$f(\varepsilon, A) = \frac{\alpha e^{-\varepsilon}.\beta e^{(1-A)}}{\alpha e^{-\varepsilon}+\beta e^{(1-A)}}$$

where α, β are weighting parameters.

3.3. OPTIMIZATION PROBLEM

3.3.1. Primary Objective

Our goal is to find the optimal hyperparameter configuration $\theta*$ that maximizes $f(\varepsilon(\theta), A(\theta))$, the optimization problem can be formally stated as:

3.3.2. Pareto Optimality

To comprehensively analyze the trade-off between privacy and utility, we introduce the concept of Pareto optimality. The Pareto frontier P represents the set of non-dominated solutions where it's impossible to improve either privacy or utility without degrading the other. Formally, we define P as:

$$P=\{(\varepsilon, A) \in S \mid \nexists\ (\varepsilon', A') \in S: (\varepsilon' < \varepsilon \wedge A' \geq A) \vee (\varepsilon' \leq \varepsilon \wedge A' > A)\} \quad (4)$$

3.4. SOLUTION APPROACHES

To solve this optimization problem, and To find $\theta*$, we employ and compare four approaches distinct approaches:

1. Grid Search: Exhaustive search over a predefined hyper-parameter space,

$$\theta^*_{GS} = \arg\max_{\theta \in \Theta_{GS}} \mathrm{f}(\varepsilon(\theta), \mathrm{A}(\theta)) \quad (5)$$

2. Random Search: Randomly sampling configurations from the hyperparameter space [33]

$$\theta^*_{RS} \approx \arg\max_{\theta \in \Theta_{RS}} \mathrm{f}\big(\varepsilon(\theta), \mathrm{A}(\theta)\big) \quad (6)$$

3. Bayesian Optimization: Sequential model-based optimization using Gaussian Processes [34],

$$\theta^*_{BO} \approx \arg\max_{\theta \in \Theta_{BO}} \mathrm{f}\big(\varepsilon(\theta), \mathrm{A}(\theta)\big) \quad (7)$$

4. Bat Algorithm: A nature-inspired meta-heuristic optimization algorithm [35],

$$\theta^*_{BA} \approx \arg\max_{\theta \in \mathrm{BA}} \mathrm{f}\big(\varepsilon(\theta), \mathrm{A}(\theta)\big) \quad (8)$$

Where ΘX represents the search space explored by method X. Each method aims to efficiently navigate the hyperparameter space to find the configuration that maximizes our objective function, thus achieving the best privacy-utility trade-off for our differentially private deep learning model.

3.5. IMPLEMENTATION CONTEXT

The optimization is implemented using the Opacus privacy engine, which: 1. Computes per-sample gradients 2. Clips gradients to bound sensitivity 3. Adds calibrated Gaussian noise 4. Tracks privacy budget consumption The DP optimizer update step is:

$$\theta_{t+1} = \theta_t - \eta . m_t(\tilde{g}_t) \quad (9)$$

Where $\tilde{g}_t$ is the differentially private gradient:

$$\tilde{g}_t = \frac{1}{|B|} \sum_{i \in B} clip(\nabla_\theta L(\theta_t, x_i), C) + \mathcal{N}(0, \sigma_t^2 C^2 \mathbf{I}) \quad (10)$$

where $\tilde{g}_t$ is the privatized gradient, B is the batch size, C is the clipping threshold, and σ_t is the dynamically adjusted noise multiplier at step t.

4. METHODOLOGY

This section presents our comprehensive methodology for optimizing the privacy-utility trade-off in differentially private deep learning. Our framework encompasses model architectures, differential privacy

implementation, hyperparameter optimization techniques, and evaluation procedures. Beyond establishing the foundational approach, we extend our investigation to assess the framework's generalizability by incorporating the PathMNIST dataset — a complex collection of histopathology images that presents more challenging scenarios compared to the BreastMNIST and Breast Cancer Wisconsin datasets. This extension, implemented through a privacy-adapted ResNet-50 architecture with DP-Optimizer, enables us to evaluate our optimization framework's scalability and feasibility on larger, more complex models. Through this comprehensive approach, we aim to provide a clear, reproducible framework for comparing optimization strategies in privacy-preserving deep learning, while demonstrating its applicability across varying levels of task complexity.

4.1. EXPERIMENTAL FRAMEWORK OVERVIEW

Our experimental framework was implemented on the Google Colab Pro+ platform, leveraging TPU v2-8 accelerators that provide 8 cores with up to 180 teraflops of computation power and 64 GB of high-bandwidth memory (HBM). This infrastructure choice was crucial for handling the computational overhead. The summary of the setting parameters is shown in Table 2.

Table 2. Experimental Framework Specifications

Component	Specification
Platform	Google Colab Pro+
Hardware	TPU v2-8 (8 cores, 180 teraflops)
Memory	64 GB HBM
Framework	Python 3 + PyTorch
Cross-Validation	3-fold
Early Stopping	with patience monitoring
Optimizer	ADAM
Loss function	CrossEntropy

4.2. DATASETS AND PREPROCESSING

4.2.1. Breast Cancer Wisconsin Dataset

The Wisconsin Breast Cancer Dataset (UCI Repository) contains 569 samples with 30 features and binary classification (malignant/benign). Data was preprocessed and split into training (301), validation (85), and test (183) sets, then converted to PyTorch tensors for model training.

4.3.1. Model architecture

Multi-Layer Perceptron architecture comprises Input(30) → Linear(20) → Linear(10) → Linear(10) → Linear(10) → Linear(5) → Output(2) with ReLU activations between layers. The architecture employs gradual dimension reduction to prevent overfitting on the breast cancer Wisconsin classification task.

4.3.2. Medical Image Datasets

The BreastMNIST and PathMNIST datasets (MedMNIST v2.2.3) represent distinct medical imaging modalities while sharing standardized preprocessing requirements. Both datasets undergo similar technical preprocessing steps, with images normalized (μ=0.5, σ=0.5) and formatted to 28×28 pixel RGB resolution for deep learning compatibility. However, they differ significantly in their modalities and clinical applications. BreastMNIST focuses on breast imaging diagnostics, containing 780 medical images distributed across training (546), validation (78), and test (156) sets for binary classification tasks. In contrast, PathMNIST encompasses histopathological imaging, presenting a larger collection of 107,180 microscopic tissue images from colon pathology. These are divided into training (89,996), validation (10,004), and test (7,180) sets, supporting a more complex nine-class classification challenge that reflects the diverse cellular patterns and tissue characteristics encountered in pathological analysis.

4.3.3. ResNet Architectures

Our implementation utilizes modified ResNet architectures (ResNet-18 and ResNet-50) with specific privacy-focused adaptations for medical image classification. Both models share fundamental privacy-preserving modifications, replacing BatchNorm layers with GroupNorm (32 groups) to comply with Opacus's privacy requirements, as BatchNorm operations can leak private information across training examples. The architectures maintain pre-trained backbones in frozen evaluation mode while incorporating trainable classification heads. The key distinction lies in their complexity and target tasks: ResNet-18 is configured for binary classification of breast images, and ResNet-50, being deeper and more complex, handles the nine-class histopathology classification task for PathMNIST. Both architectures preserve privacy guarantees through GroupNorm's channel-based normalization approach and ensure efficient feature extraction through their frozen pre trained backbones, demonstrating adaptability to different medical imaging modalities while maintaining privacy-preserving characteristics.

4.3.4. Training Configuration

Our training framework implements model-specific configurations to ensure optimal convergence while managing computational resources effectively. The MLP architecture employs a maximum of 500 epochs with a patience value of 250, while the ResNet-18 training is configured with 20 maximum epochs and a patience threshold of 12. For the PathMNIST experiments using the Bat Algorithm, we established two distinct configurations as detailed in Table 3, where both maintain a population size of 2 bats but differ in their convergence parameters: Configuration 1 uses 14 epochs with early stopping at 10, while Configuration 2 employs 15 epochs with early stopping at 13.

Across all models, we implemented early stopping monitoring validation accuracy to prevent overfitting and ensure optimal convergence. These parameter adjustments, particularly for the PathMNIST experiments, were essential to complete the optimization process while managing computational constraints.

Table 3. Bat Algorithm Optimization Parameters: Two Configurations for PathMNIST classification

	1st Configuration	2nd Configuration
Optimization Parameters		
Population Size	2 bats	2 bats
Max-iterations	14 (convergence at 11)	4 (convergence at 2)
Epochs	15	15
Early Stopping Patience	10	13

4.4. DIFFERENTIAL PRIVACY IMPLEMENTATION

Table 4. Differential Privacy Components

Component	Implementation
Library	Opacus
Engine	PrivacyEngine
Delta (δ)	Fixed at 10^{-4}
Privacy Mechanism	Gradient clipping + Gaussian noise

Differential privacy is implemented via Opacus Privacy Engine with two key components:

1. Privacy Engine Operation:
- Per-sample gradient computation
- Gradient clipping for sensitivity bounds
- Calibrated Gaussian noise addition
- Privacy budget tracking
2. Dynamic Privacy Management:
- Target epsilon (εtarget) specification
- Dynamic noise multiplier (σ) adjustment
- Automated noise calibration
- Privacy parameter reporting

The engine adjusts the noise multiplier during training to balance target epsilon and model utility, following equation (10).

4.5. HYPERPARAMETER SPACE

Our hyperparameter optimization space was carefully defined to accommodate both discrete and continuous optimization methods as shown in Table 5:

4.5.1. Parameter Adjustment Mechanisms

Parameter adjustment mechanisms for continuous optimization include reflection methods that handle out-of-bounds values through boundary reflection, and value adjustment processes that discretize batch sizes, enforce integer constraints, clip boundary values, and prevent negative values.

Table 5. Hyperparameter Space Definition

Parameter	Discrete Values	Continuous Range	Notes
Learning Rate	{0.001, 0.01, 0.1}	[0.001, 0.1]	Three orders of magnitude
Batch Size	{16, 32, 64, 128, 512}	[16, 512]	Powers of 2
Max Gradient Norm	{1.2, 5.6}	[1.2, 5.6]	Conservative range
Privacy Budget (ε)	{0.5, 1.0, 8.0}	[0.2, 8.0]	Strict to relaxed privacy
Privacy Delta (δ)	Fixed at 10^{-4}		Standard failure probability

4.6. OPTIMIZATION FRAMEWORK

The optimization framework is designed to systematically achieve an optimal privacy-accuracy balance in Differentially Private Deep Learning (DPDL) models. It comprises three key phases: initialization, optimization, and Pareto analysis. Fig.1 illustrates the comprehensive workflow, while Fig.2 provides a detailed flowchart of the optimization methods. This framework establishes a robust foundation for achieving optimal privacy-utility trade-offs through systematic parameter tuning and multi-objective optimization.

4.6.1. Objective Function

For the objective function defined in equation (3), we set $\alpha = \beta = 0.5$ to ensure equal importance between privacy preservation (ε) and model accuracy (A), thus achieving a balanced privacy-utility optimization without favoring either aspect.

4.6.2. Hyperparameter Tuning Methods

To identify the optimal hyperparameter configurations, we employ four distinct methods G.S, R.S, B.O, and B.A. Each method is described below:

a) Grid Search (G.S): Deterministic Exploration

G.S operates as an exhaustive search method, systematically evaluating every possible combination of hyperparameters within a predefined discrete space. By exploring the entire search space, G.S identifies the optimal configuration that effectively balances the privacy-utility trade-off. This method ensures a thorough and methodical approach to hyperparameter tuning, albeit at a higher computational cost.

b) Random Search (R.S): Stochastic Exploration

R.S samples hyperparameter configurations from a uniform distribution $U(\theta_{RS})$ over a predefined discrete search space θ_{RS}. It evaluates a fixed number of configurations (N), typically covering approximately 25% of the parameter space. By focusing on a subset of the search space, R.S efficiently approximates the optimal solution while significantly reducing computational overhead compared to exhaustive methods like G.S.

c) Bayesian Optimization (B.O): Sequential Model-Based Optimization

Bayesian Optimization employs a Gaussian Process (GP) as a surrogate model to approximate the objective function $f(\varepsilon(\theta), A(\theta))$. The process begins with 10 warm-up points, randomly sampled to initialize the GP model. At each iteration, the next hyperparameter configuration is selected by maximizing the Expected Improvement (EI):

$$\theta_{t+1} = \arg\max_{\theta \in \Theta} \mathrm{EI}(\theta|D) \tag{11}$$

Where $D=\{(\theta_i, f_i)\}_{i=1}^{t}$ represents the set of observations up to iteration t. This approach sequentially refines the GP model, guiding the search toward regions of the hyperparameter space that are most likely to improve performance.

Through this iterative process, equation (11) guides us toward the optimal solution represented in equation (7). At each step t:

- The GP model is updated using all previous observations D
- $EI(\theta|D)$ estimates where the next evaluation might most improve upon our current best solution
- This sequential refinement helps us approximate the optimal hyper-parameters ${\theta^*}_{BO}^*$ that maximize our objective function $f(\varepsilon(\theta), A(\theta))$

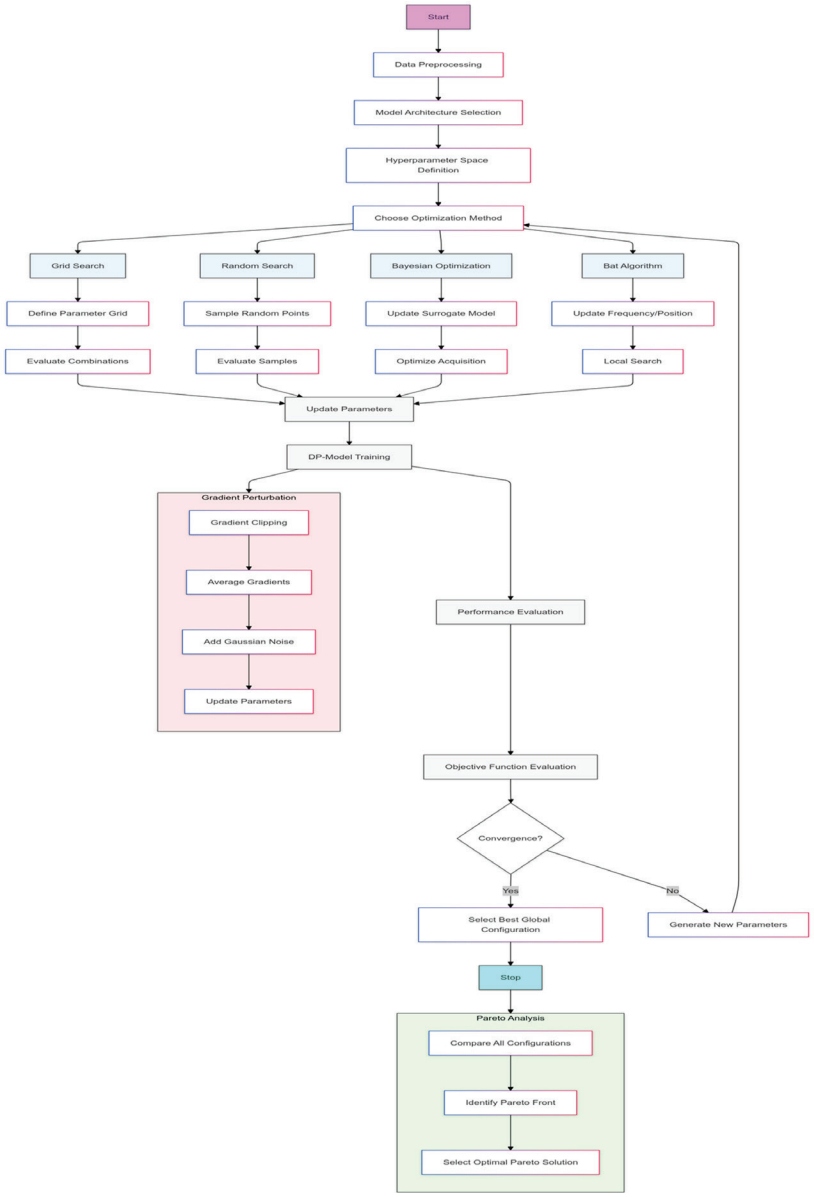


Fig. 1. Flowchart of Hyperparameter Optimization with Gradient Perturbation for Privacy-Accuracy Balance in DPDL

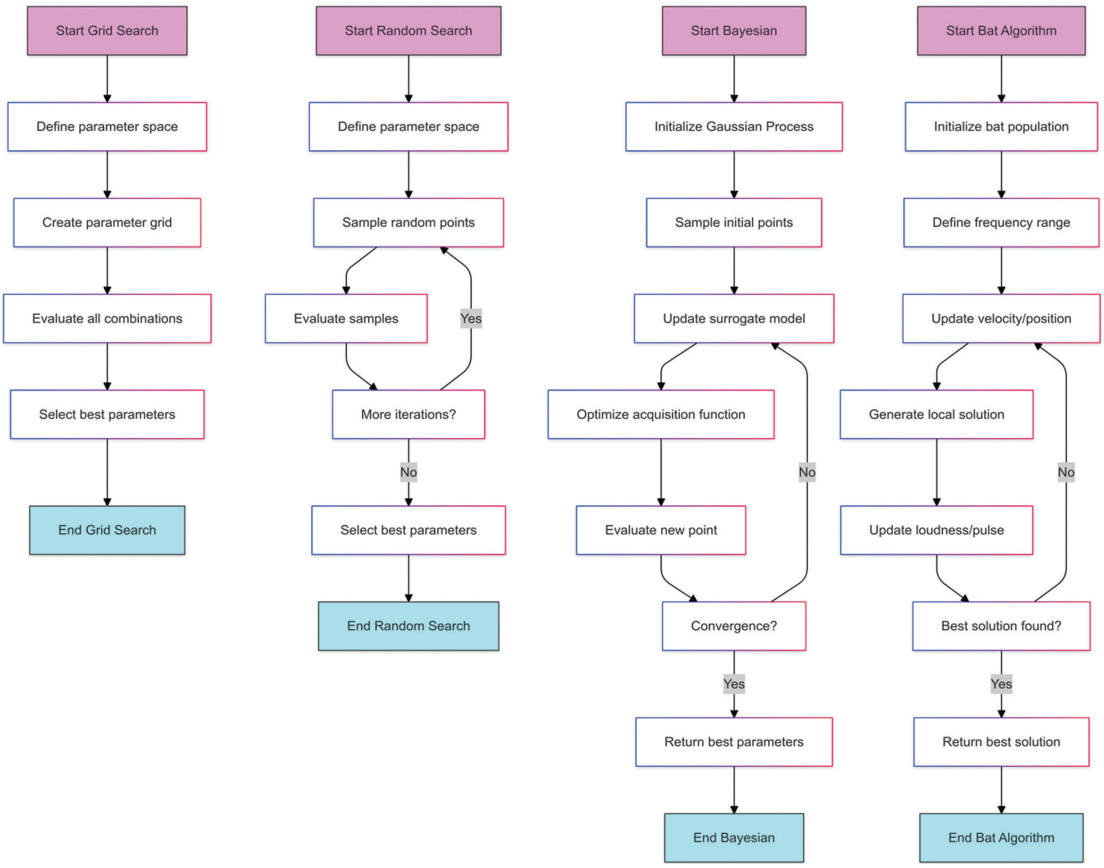


Fig. 2. Flowchart of each Optimization Method

B.O. effectively balances exploration and exploitation, making it highly efficient for high-dimensional spaces.

- Nature-inspired metaheuristic for hyperparameter tuning: The Bat Algorithm is a swarm intelligence-based optimization method that mimics the echolocation behavior of bats to navigate the hyperparameter space. It combines global exploration and local exploitation by iteratively updating the position x_i and velocity v_i of each bat i. The algorithm is guided by frequency f_i, loudness A_i, and pulse emission rate r_i, which are dynamically adjusted to balance exploration and exploitation. Algorithm 1 provides the complete pseudocode for the Bat Algorithm, illustrating its iterative parameter updates for balanced local and global search space exploration.

Algorithm 1
Bat Algorithm for Hyperparameter Tuning

Initialize Parameters:
Population size: $N = 10$ bats
Dimensions: $D = 4$ parameters
Frequency range: $[f_{min}, f_{max}] = [0, 10]$
Loudness: $\boldsymbol{A_i} = 1.0$ (initial)
Pulse emission rate: $\boldsymbol{r_i = r_i^0}$ (initial)
Alpha ($\boldsymbol{\alpha}$): 0.9
Gamma ($\boldsymbol{\gamma}$): 0.9
Maximum iterations: $\boldsymbol{T_{max}}$ (dataset-specific)
Initialize Population:
Randomly initialize positions $\boldsymbol{x_i}$ and velocities $\boldsymbol{v_i}$ for each bat i.
Evaluate Initial Fitness:
Compute fitness $\boldsymbol{f(x_i)}$ for each bat i.
Identify the global best solution $\boldsymbol{x_{best}}$.
Main Loop (**for** $\boldsymbol{t}$=**1 to** $\boldsymbol{T_{max}}$):
for each bat i **do**
    Generate frequency $\boldsymbol{fi}$:
        $\boldsymbol{fi = f_{min} + (f_{max} - f_{min}) \cdot \beta}$, $\boldsymbol{\beta} \in [0,1]$
    Update velocity $\boldsymbol{vi}$:
        $\boldsymbol{v_i^t + 1 = v_i^t + (x_i^t - x_{best}) \cdot fi}$
    Update position $\boldsymbol{x_i}$:
        $\boldsymbol{x_i^{t+1} = x_i^t + v_i^{t+1}}$
    if $\boldsymbol{rand > r_i}$ **then**
    Perform local search:
        $\boldsymbol{x_{new} = x_{old} + \varepsilon \cdot A^t}$, $\varepsilon \in$ [**-1,1**]
    end if
    Evaluate fitness $\boldsymbol{f(x_i^{t+1})}$.
    if $\boldsymbol{rand < A_i}$ and $\boldsymbol{f(x_i^{t+1}) < f\,x_{best}}$) **then**
    Update $\boldsymbol{x_{best} = x_i^{t+1}}$.
    end if
    Update loudness $\boldsymbol{A_i}$ and pulse emission rate $\boldsymbol{r_i}$:
    $\boldsymbol{A_i^{t+1} = \alpha \cdot A_i^t}$
    $\boldsymbol{r_i^{t+1} = r_i^0 \cdot [1 - \exp^{(-\gamma \cdot t)}]}$
end for
Return Optimal Solution:
Output the global best solution $\boldsymbol{x_{best}}$.

4.7. PARETO ANALYSIS

Our Pareto efficiency analysis identifies the optimal trade-off between privacy preservation (ε) and model accuracy (A) across all optimization methods. The analysis involves two key steps:

1. Pareto Front Identification

- We use a non-dominated sorting algorithm to identify the Pareto front, comprising solutions that are not dominated by any other configuration in terms of ε and A.
- Optimization occurs in a two-dimensional space, where each point represents a unique (ε, A) combination.
- Pair-wise dominance comparisons determine dominance: a solution (ε_1, A_1) dominates (ε_2, A_2) if $\varepsilon_1 \leq \varepsilon_2$ and $A_1 \geq A_2$, with at least one strict inequality.

2. Optimal Point Selection

- The optimal Pareto point is selected from the set of non-dominated solutions (P), satisfying the dominance relation in Equation (4).
- If the global best solution matches the Pareto point for a method, it indicates the method's superiority in achieving the best non-dominated trade-off. This highlights the method's robustness, efficiency, and practical applicability for privacy-preserving machine learning tasks while providing a basis for method comparison and future research.

5. COMPREHENSIVE ANALYSIS OF RESULTS

Based on our experimental evaluation across multiple datasets and optimization methods, we present a detailed analysis of the performance metrics, hyperparameter configurations, computational resources utilized, and comparative assessment of optimization strategies in our framework. Our analysis includes a comprehensive performance evaluation across key metrics, a detailed examination of how hyperparameters influence outcomes, a specific analysis of the PathMNIST dataset's performance, an assessment of resource utilization patterns, and a comparative investigation of the strengths and limitations of different methods across various datasets and optimization objectives.

5.1. PERFORMANCE METRICS ANALYSIS

We begin our analysis by examining the performance metrics detailed in Table 6, focusing on accuracy, privacy preservation, and computational efficiency across all optimization methods and datasets. This analysis provides insights into how each method balances these critical performance dimensions.

The Grid Search method demonstrated consistent performance across datasets, achieving an accuracy of 93.40% on the Breast Cancer Wisconsin dataset and 74.18% on BreastMNIST. Maintaining a privacy budget (ε) of 0.500 required substantial computational resources, particularly evident in the 4,000.33 seconds processing time for the Wisconsin dataset.

Random Search exhibited comparable accuracy metrics, reaching 92.53% on the Wisconsin dataset and 74.91% on BreastMNIST. Notably, it achieved these results with varying privacy budgets - 0.500 for Wisconsin and 1.000 for BreastMNIST. The method showed improved time efficiency compared to Grid Search, completing the Wisconsin dataset analysis in 2,749.65 seconds.

Bayesian Optimization achieved the highest accuracy on the Wisconsin dataset at 93.62%, while maintaining a privacy budget of 0.501. The method demonstrated consistent performance on BreastMNIST with 73.99% accuracy. Its computational requirements were significant, requiring 17,828.72 seconds for the Wisconsin dataset, though with reduced memory usage of 787.75 MB.

The Bat Algorithm showed distinct characteristics, achieving Pareto optimality on both datasets. While its accuracy was slightly lower (93.18% for Wisconsin, 73.08% for BreastMNIST), it demonstrated efficient privacy preservation with ε values of 2.258 and 0.293 respectively.

Table 6. Performance Metrics Across Datasets and Optimization Methods

Dataset	Method	Privacy (ε)	Accuracy (%)	Fitness (f)	Time (s)	Memory (MB)	Pareto Optimal?
Breast cancer Wisconsin	Grid Search	0.500	93.40	0.1840	4,000.33	1,064.83	Yes
	Random Search	0.500	92.53	0.1834	2,749.65	845.08	No
	Bayesian Opt.	0.501	93.62	0.1840	17,828.72	787.75	No
	Bat Algorithm	2.258	93.18	0.0470	14,562.20	1,103.16	Yes
BreastMNIST	Grid Search	0.500	74.18	0.1699	12,931.66	1,796.29	Yes
	Random Search	1.000	74.91	0.1697	6,499.74	1,555.48	Yes
	Bayesian Opt.	0.500	73.99	0.1697	38,076.04	1,322.85	Yes
	Bat Algorithm	0.293	73.08	0.1887	10,132.33	4,348.25	Yes

5.2. HYPERPARAMETER CONFIGURATION ANALYSIS

To understand the factors driving performance differences, we examine the optimal hyper-parameter configurations identified by each method across datasets, as presented in Table 7. This analysis reveals key patterns in parameter selection and their impact on optimization outcomes.

The optimal learning rates varied significantly across methods. Grid Search performed best with smaller learning rates (0.01), while the Bat Algorithm required higher rates (0.0965 for Wisconsin). Batch sizes showed a clear pattern, with most optimal configurations favoring larger batches (512) for the Wisconsin dataset and varying sizes for BreastMNIST. Privacy budgets demonstrated method-specific patterns. Random Search and Grid Search maintained consistent budgets (0.500), while the Bat Algorithm and Bayesian Optimization showed more variation. The max gradient norm values remained relatively stable across methods for BreastMNIST but showed greater variation in the Wisconsin dataset. Training epochs exhibited method-specific patterns, with Grid Search requiring 273 epochs for optimal performance on the Wisconsin dataset, while BreastMNIST achieved optimal results with significantly fewer epochs (12-14) across all methods.

Table 7. Best Hyperparameter Configurations for Each Method.

Dataset	Method	Learning Rate (lr)	Batch Size(bs)	Privacy Budget (ε_{terget})	Max Grad Norm (C)	Noise Multiplier (σ)	Training Epochs	Final Accuracy(%)
Breast cancer Wisconsin	Grid Search	0.01	32	0.500	1.2	42.5	273	93.40
	Random Search	0.1	512	0.500	5.6	135.0	252	92.53
	Bayesian Opt.	0.018	128	0.501	3.118	77.5	265	93.62
	Bat Algorithm	0.0965	512	2.258	3.388	95.5	248	93.18
BreastMNIST	Grid Search	0.1	32	0.500	1.2	6.25	12	74.18
	Random Search	0.1	64	1.000	1.2	8.75	12	74.91
	Bayesian Opt.	0.1	128	0.500	4.906	12.19	15	73.99
	Bat Algorithm	0.0429	512	0.293	3.114	18.5	14	73.08

5.3. PATHMNIST-SPECIFIC PERFORMANCE ANALYSIS

As detailed in Table 8, our focused analysis of the PathMNIST dataset provides additional insights into the optimization framework's capability to handle complex medical imaging data and adapt configurations for improved performance. The results demonstrate significant variations between configurations and their impact on multiple performance dimensions.

Table 8. Results on the PathMNIST dataset

	1st Configuration	2nd Configuration
Global Best Solution		
Learning Rate (lr)	0.03106	0.01760
Batch Size	128	512
Max-Grad-Norm (C)	2.179	1.732
TargetEpsilon(ε)	0.508	2.603
Performance Metrics		
Objective Function	0.1435	0.0328
GlobalBest Accuracy	39.97%	44.71%
Global Best Epsilon	0.508	2.603
Time (S)	18,238.194	13,678.6455
Memory (MB)	8,479.0875	7,522.4725
Pareto Optimal?	No	Yes

The PathMNIST dataset results revealed significant improvements between configurations. The global best accuracy increased from 39.97% to 44.71%, accompanied by changes in the learning rate from 0.03106 to 0.01760. The second configuration achieved Pareto optimality while reducing memory requirements from 8,479.0875 MB to 7,522.4725 MB.

The objective function improved from 0.1435 to 0.0328, indicating enhanced optimization performance. The global best epsilon value increased from 0.508 to 2.603, suggesting a different privacy-utility trade-off in the optimal configuration. Execution time increased from 8,238.194 seconds to 13,678.6455 seconds, demonstrating the computational cost of achieving improved performance metrics.

5.4. RESOURCE UTILIZATION ASSESSMENT

Understanding the computational demands of each optimization method is crucial for practical implementation. Our analysis of resource utilization reveals significant variations in memory and time requirements across methods and datasets.

Memory requirements varied significantly across methods and datasets. For the Wisconsin dataset, Grid

Search utilized 1,064.83 MB, while the Bat Algorithm required 1,103.16 MB. BreastMNIST showed higher memory requirements overall, with the Bat Algorithm consuming 4,348.25 MB.

Execution times demonstrated substantial variation, ranging from 2,749.65 seconds for Random Search to 17,828.72 seconds for Bayesian Optimization on the Wisconsin dataset. BreastMNIST generally required longer processing times, with Grid Search taking 12,931.66 seconds and Bayesian Optimization requiring 38,076.04 seconds.

The fitness values across methods remained relatively consistent within each dataset, suggesting that different optimization approaches converged to similarly optimal solutions despite varying computational requirements and privacy-utility trade-offs.

5.5. COMPARATIVE ANALYSIS

To synthesize our findings, we examine the relative strengths and limitations of each optimization method across datasets, highlighting key trade-offs and operational considerations.

Our comparative analysis reveals distinctive patterns across optimization methods and datasets, highlighting the inherent trade-offs between privacy, accuracy, and computational efficiency. On the Wisconsin dataset, Grid Search and Bayesian Optimization achieved comparable accuracy levels (93.40% and 93.62% respectively) while maintaining similar privacy budgets (0.500 and 0.501). However, Bayesian Optimization required approximately 4.5 times more computational time, suggesting a significant efficiency trade-off for marginal accuracy improvement.

The BreastMNIST dataset results demonstrate different optimization dynamics. Random Search emerged as the top performer with 74.91% accuracy, albeit requiring a higher privacy budget (ε = 1.000) compared to other methods. This illustrates the inherent tension between privacy preservation and model performance. Grid Search achieved comparable accuracy (74.18%) with half the privacy budget (ε = 0.500), representing a potentially more balanced solution for privacy-sensitive applications.

The Bat Algorithm's performance presents an interesting case study in multi-objective optimization. Despite achieving lower accuracy scores on both datasets, it consistently achieved Pareto optimality, suggesting superior performance in balancing multiple competing objectives. Its moderate memory requirements (1,103.16 MB for Wisconsin, 4,348.25 MB for BreastMNIST) and execution times position it as a practical choice for resource-constrained environments.

A cross-dataset comparison reveals that optimization methods demonstrate dataset-specific strengths. While Grid Search exhibited stable performance across both datasets, Random Search showed higher variability, performing notably better on the BreastMNIST dataset. This suggests that dataset characteristics significantly influence the effectiveness of different optimization strategies.

5.6. EXTENDED ANALYSIS WITH VISUALIZATION RESULTS

This analysis presents a comprehensive visualization-based examination of our optimization results across three distinct datasets: Wisconsin Breast Cancer, BreastMNIST, and PathMNIST. Our visualization framework employs two key components: privacy-accuracy trade-off plots with fitness value indicators, and convergence plots showing the evolution of fitness values over iterations. This dual visualization approach enables us to understand both the final solution space and the optimization trajectory for each method.

5.6.1. Breast Cancer Wisconsin Dataset

The Breast Cancer Wisconsin Dataset optimization results reveal distinct patterns across the four optimization methods, showcasing various approaches to balancing privacy and accuracy.

Grid Search and Random Search (Fig. 3a, 3b) demonstrate similar exploration patterns, characterized by discrete, well-defined sampling points. The privacy-accuracy trade-off plots show concentrated exploration in specific regions, with Grid Search providing more systematic coverage while Random Search offers more scattered distribution. Both methods achieve relatively quick convergence as shown in their fitness evolution plots, reaching stable fitness values early in the optimization process.

Bayesian Optimization (Fig. 3c) exhibits a more sophisticated exploration strategy, with dense sampling in promising regions of the solution space. The privacy-accuracy plot reveals a continuous distribution of points, suggesting a more thorough exploration of the trade-off space. The convergence plot shows rapid initial improvement followed by consistent refinement, indicating efficient optimization behavior.

The Bat Algorithm (Fig. 3d) demonstrates a unique exploration pattern, with initial broad coverage followed by concentrated sampling in high-performing regions. The privacy-accuracy plot shows clusters of solutions, particularly in areas of favorable trade-offs. The fitness evolution plot reveals a distinctive stepped pattern, suggesting periodic improvements in solution quality as the algorithm explores the search space.

Regarding optimal solutions, all methods successfully identified configurations that balance privacy and accuracy, with the Bayesian Optimization and Bat Algorithm showing a particularly effective exploration of the solution space near the Pareto frontier. The convergence behavior suggests that while Grid Search and Random Search reach stable solutions quickly, Bayesian Optimization and the Bat Algorithm continue to refine their solutions throughout the optimization process, potentially discovering more nuanced trade-offs.

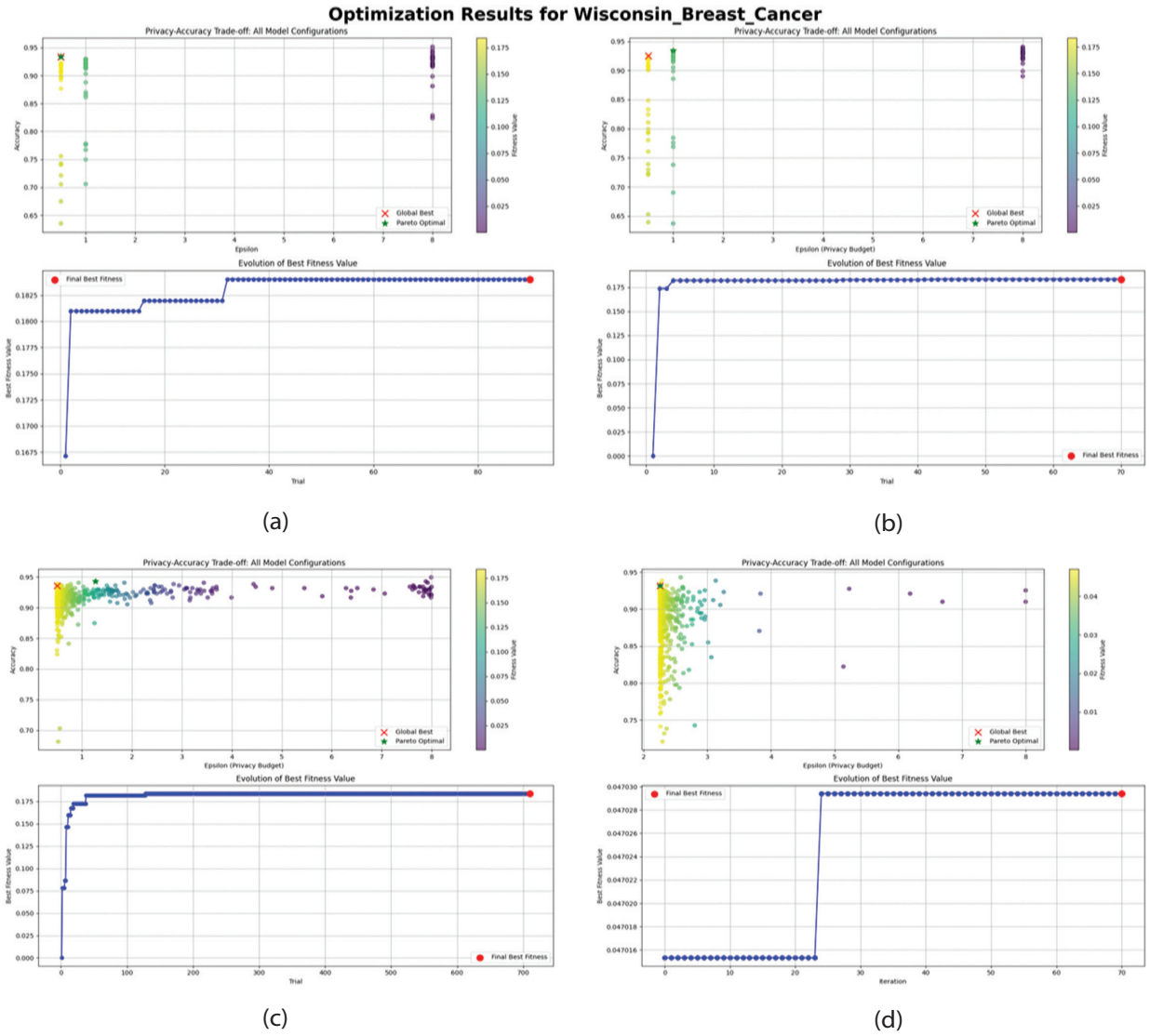


Fig. 3. Optimization Results for Wisconsin Breast Cancer Dataset: (a) Grid Search, (b) Random Search, (c) Bayesian Optimization, and (d) Bat Algorithm. The top row shows privacy-accuracy trade-off plots with color indicating fitness values. The bottom row shows the evolution of the best fitness value over iterations.

5.6.2. BreastMNIST Dataset

The optimization results for the BreastMNIST dataset reveal distinctive characteristics and performance patterns across the four optimization methods, with notable differences from the Wisconsin Breast Cancer dataset analysis.

Grid Search (Fig. 4a) shows a structured exploration pattern with evenly distributed sampling points across the privacy-accuracy space. The convergence plot displays a step-like pattern, indicating discrete improvements in fitness values at specific intervals. This suggests that the method systematically identified better solutions through its predefined search grid.

Random Search (Fig. 4b) demonstrates a more scattered distribution of solutions, yet maintains coverage across the solution space. The fitness evolution plot shows rapid initial improvement followed by sustained performance, suggesting early discovery of promising regions in the search space.

Bayesian Optimization (Fig. 4c) exhibits a more nuanced exploration strategy, with concentrated sampling in regions of higher fitness values. The privacy-accuracy trade-off plot reveals clusters of solutions in promising areas, indicating the algorithm's ability to adapt its search based on previous results. The convergence plot shows progressive improvement, with multiple optimization stages visible in the fitness trajectory.

The Bat Algorithm (Fig. 4d) presents a unique exploration pattern characterized by focused sampling in specific regions of the solution space. The convergence plot demonstrates consistent performance throughout the optimization process, suggesting stable exploration of the search space.

However, the distribution of solutions appears more concentrated compared to other methods, indicating a potentially more focused search strategy. Compared to the Wisconsin dataset results, the BreastMNIST optimization exhibits different convergence patterns

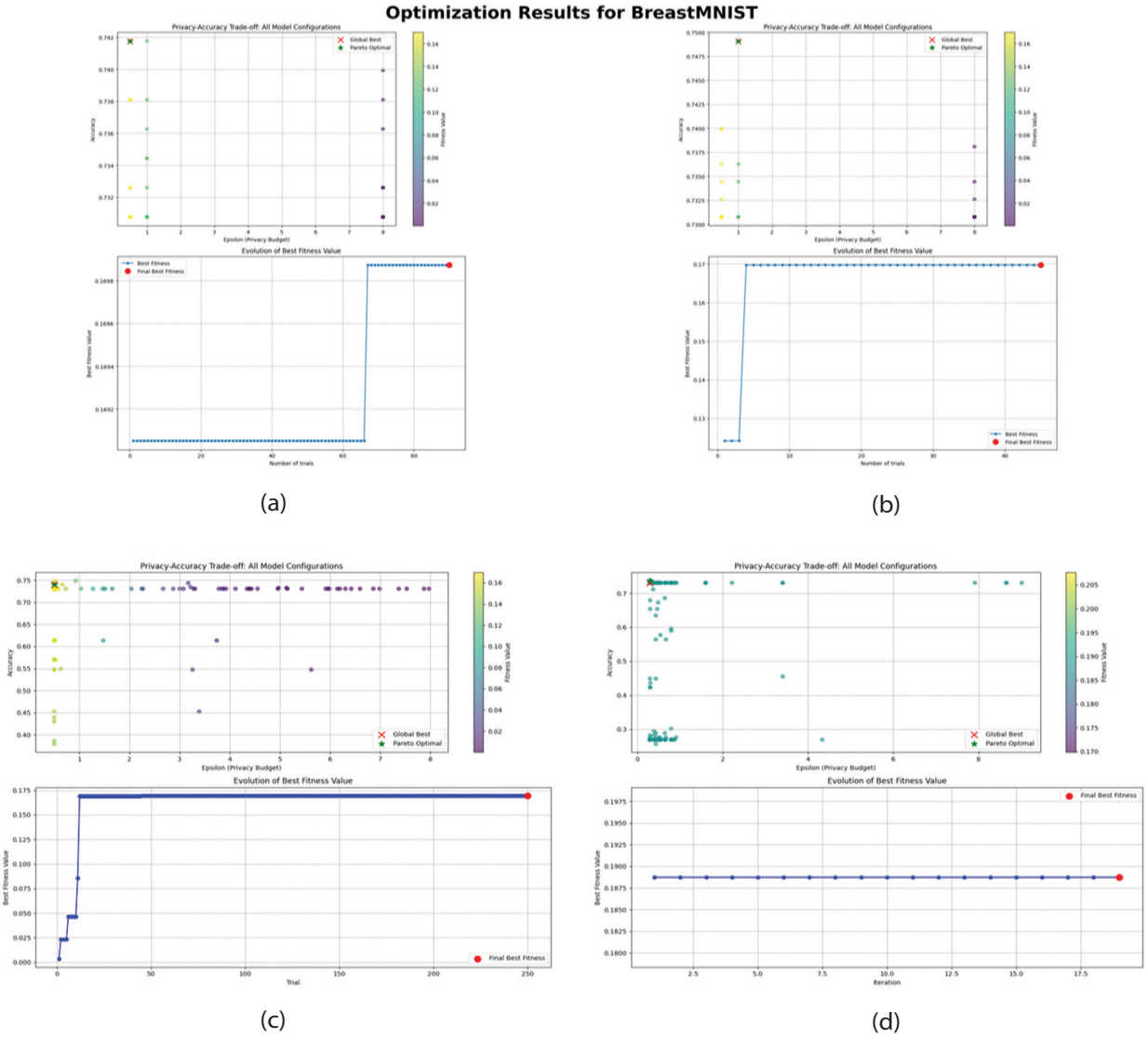


Fig. 4. Optimization Results for BreastMNIST Dataset: (a) Grid Search, (b) Random Search, (c) Bayesian Optimization, and (d) Bat Algorithm. The top row shows privacy-accuracy trade-off plots with color indicating fitness values. The bottom row shows the evolution of the best fitness value over iterations.

and solution distributions, likely due to the increased complexity and distinct characteristics of the dataset. This highlights the importance of algorithm selection based on specific dataset characteristics and optimization objectives.

5.6.3. PathMNIST Dataset

5.6.3.1. Configuration 1

The optimization results for Configuration 1 of the Bat Algorithm on the PathMNIST dataset, as shown in Fig. 5, demonstrate interesting characteristics in both solution distribution and convergence behavior.

The privacy-accuracy trade-off plot reveals two distinct clusters of solutions. The first cluster appears concentrated in the lower epsilon range (around 0.5-1.0) with accuracy values between 0.38 and 0.42. The second, smaller cluster is positioned at a higher epsilon value (approximately 4.0) with improved accuracy values of approximately 0.44-0.45. This bimodal distribution suggests the algorithm identified two potentially promising regions in the solution space.

The fitness evolution plot demonstrates remarkably rapid convergence, reaching near-optimal fitness values within the first two iterations. After this initial sharp improvement, the fitness value stabilizes and maintains consistency throughout the remaining iterations, reaching a final best fitness value of approximately 0.14350. This quick convergence pattern indicates that Configuration 1 efficiently identified a promising solution early in the optimization process.

The iteration markers in the trade-off plot show that later iterations (represented by different colors) focused exploration around these two identified regions, particularly the higher-accuracy cluster. This behavior suggests that the algorithm effectively balanced the exploration of the solution space with the exploitation of promising areas.

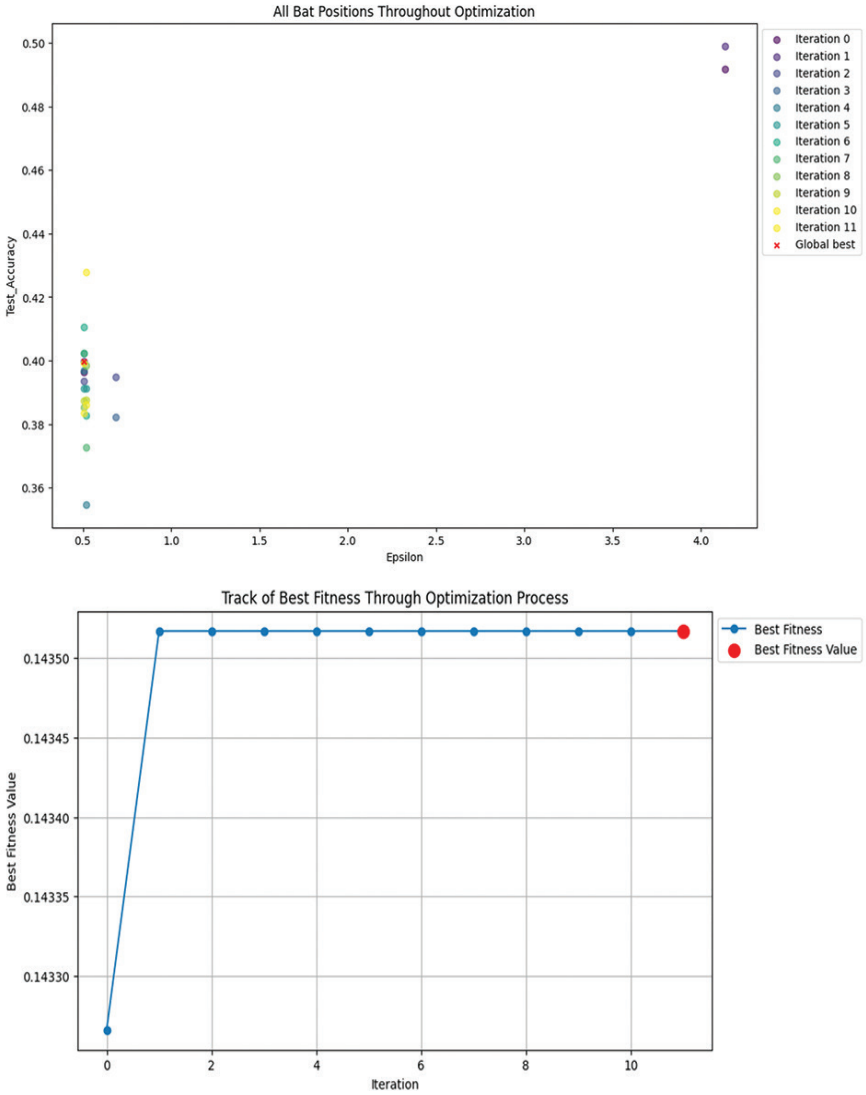


Fig. 5. The optimization results for Configuration 1 of the Bat Algorithm on the PathMNIST dataset

5.6.3.2. Configuration 2

The optimization results for Configuration 1 and 2 illustrated in Fig. 5 and 6, of the Bat Algorithm on the PathMNIST dataset reveal interesting patterns when accounting for their different maximum iteration settings (14 and 2 iterations, respectively).

The privacy-accuracy trade-off plots show distinct exploration patterns. Configuration 1, with its longer optimization period of 14 iterations, demonstrates a more refined clustering of solutions, particularly around two key regions: one at lower epsilon values (0.5-1.0) and another at higher values (approximately 4.0). This extended iteration period allowed for more thorough exploration and refinement of promising areas.

Configuration 2, limited to 2 iterations, shows a more dispersed distribution of solutions across the epsilon range (1-8). While this might initially appear as a broader exploration, it's important to note that this distribution is the result of significantly fewer optimization steps rather than a fundamentally different search strategy.

The fitness evolution plots for both configurations show improvement from their initial values, but the apparent differences in their convergence patterns must be interpreted within the context of their different iteration limits. Configuration 1's longer optimization period provides a more complete picture of the algorithm's convergence behavior, while Configuration 2's shorter run offers only an initial glimpse of the optimization trajectory.

Given the identical starting conditions and population size, the primary differentiating factor between these configurations is the maximum iteration count. This suggests that Configuration 1's more refined solution clusters and stable convergence pattern are primarily the result of having more iterations to optimize, rather than fundamental differences in the algorithm's behavior or efficiency.

This analysis highlights the importance of considering optimization duration when evaluating algorithm performance, as the number of iterations directly impacts the algorithm's ability to refine its solutions and explore the solution space effectively.

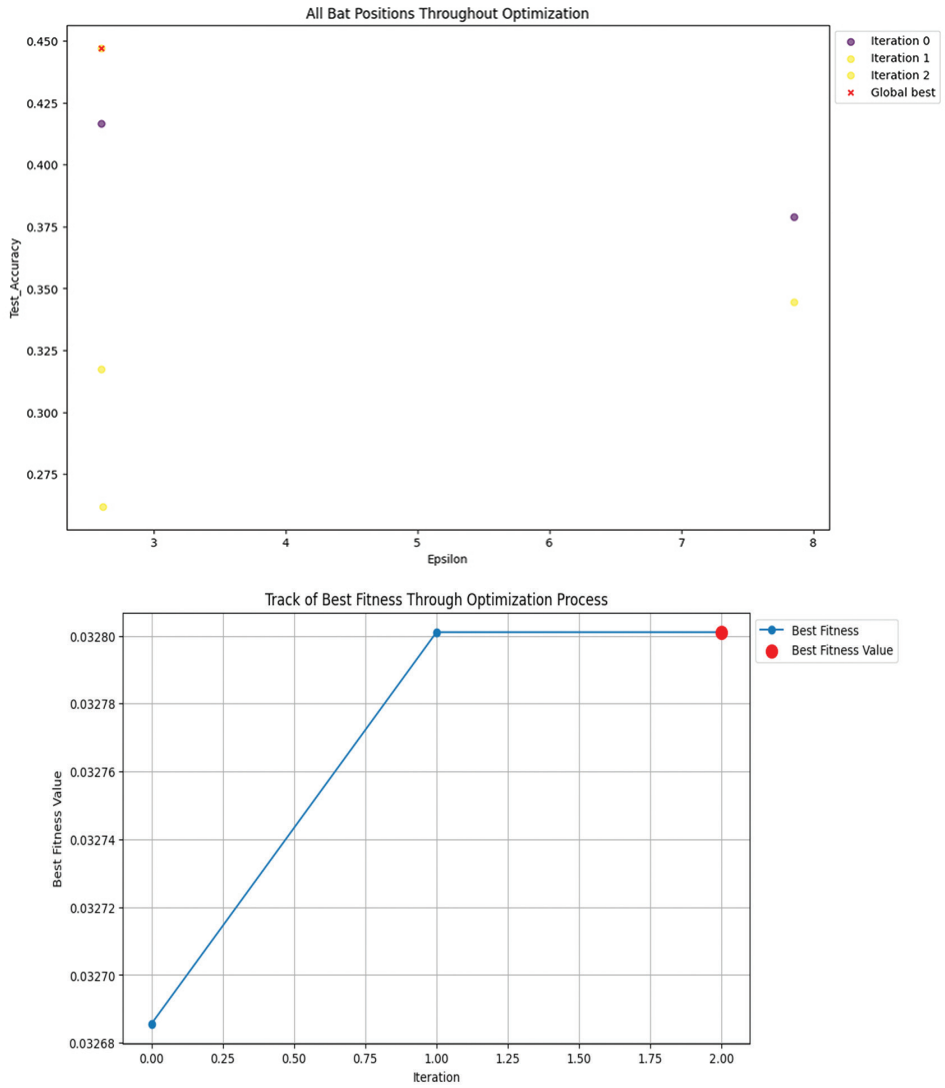


Fig. 6. The optimization results for Configuration 2 of the Bat Algorithm on the PathMNIST dataset

6. DISCUSSION

The comprehensive analysis of our privacy-preserving optimization framework reveals significant insights into the performance, computational feasibility, and practical implications of different optimization approaches in medical image analysis. This discussion examines the critical aspects of our findings while contextualizing them within broader theoretical and practical frameworks.

Critical Analysis of Performance Trade-offs

Our results demonstrate complex interrelationships between privacy preservation, model accuracy, and computational efficiency across optimization methods. The Bayesian Optimization method's superior accuracy (93.62% on the Wisconsin dataset) while maintaining a strict privacy budget ($\varepsilon = 0.501$) constitutes a breakthrough in balancing these competing objectives. However, this performance comes at a substantial computational cost, requiring 17,828.72 seconds of processing time - approximately 6.5 times longer than Random Search. This trade-off exemplifies the fundamental tension between optimization quality and computational efficiency in privacy-preserving machine learning.

The Bat Algorithm's achievement of Pareto optimality across datasets, even with reduced absolute accuracy, suggests a more nuanced approach to multi-objective optimization. Its ability to maintain competitive accuracy (93.18% for Wisconsin) while achieving varying privacy budgets ($\varepsilon = 2.258$ and 0.293) demonstrates adaptive capability in managing privacy-utility trade-offs. This performance characteristic is particularly relevant for applications where balanced optimization across multiple objectives outweighs maximizing individual metrics.

Computational Feasibility and Resource Requirements

The substantial variation in computational requirements across methods necessitates careful consideration of deployment scenarios. Grid Search's consistent but resource-intensive approach (4,000.33 seconds for Wisconsin) contrasts with Random Search's more efficient execution (2,749.65 seconds), suggesting different optimal use cases based on available computation-

al resources. Memory utilization patterns, ranging from 787.75 MB for Bayesian Optimization to 4,348.25 MB for the Bat Algorithm on BreastMNIST, indicate potential scalability challenges for larger datasets.

Our analysis reveals that computational overhead scales non-linearly with dataset complexity, particularly evident in the BreastMNIST results where processing times increased by factors of 2-3 compared to the Wisconsin dataset. This scaling behavior suggests potential limitations for enterprise-scale implementations, particularly in resource-constrained environments.

Method-Specific Performance Analysis

The distinctive performance patterns of each optimization method provide insights into their operational characteristics. Despite higher computational costs, Bayesian Optimization's superior accuracy reflects its sophisticated exploration-exploitation balance, which is particularly effective in complex parameter spaces. The Bat Algorithm's consistent achievement of Pareto optimality demonstrates its effectiveness in navigating multi-objective optimization landscapes, though at the cost of absolute accuracy.

Grid Search's stable performance across datasets (93.40% and 74.18% accuracy) suggests reliability in finding good solutions, albeit with limited ability to adapt to specific dataset characteristics. Random Search's competitive performance (92.53% and 74.91% accuracy) with reduced computational overhead, indicates its viability as a practical alternative under resource-limited conditions.

Scalability and Real-world Applications

The PathMNIST results provide crucial insights into scalability challenges, with accuracy dropping to 44.71% despite increased computational resources. This performance degradation highlights potential limitations in scaling current approaches to more complex medical imaging tasks. The observed increase in memory requirements (7,522.4725 MB) and execution time (13,678.6455 seconds) suggests that practical implementations may require significant computational infrastructure.

Architecture and Implementation Considerations

Hyperparameter sensitivity analysis reveals distinct patterns across methods, with optimal learning rates varying from 0.01 (Grid Search) to 0.0965 (Bat Algorithm). This variation suggests method-specific stability characteristics that must be considered during implementation. The consistent preference for larger batch sizes (512) in the Wisconsin dataset indicates potential optimization opportunities through batch processing strategies.

Comparative Analysis of Solution Quality

Visualization results demonstrate distinct convergence patterns across methods. The Bat Algorithm's stepped convergence pattern suggests periodic improvements in solution quality, while Bayesian Optimization shows more gradual refinement. These patterns offer insights into the exploration-exploitation dynamics of each method, with implications for selecting optimization strategies.

Limitations and Practical Constraints

Current framework limitations include substantial computational requirements for complex datasets and potential scalability challenges. The observed trade-offs between privacy preservation and model performance suggest inherent constraints that may limit applicability in highly privacy-sensitive scenarios. Memory requirements for complex datasets indicate potential deployment challenges in resource-constrained environments.

Future Research Directions and Improvements

Future work should focus on improving computational efficiency through techniques such as parallel processing and adaptive sampling strategies. Investigation of hybrid optimization approaches combining the efficiency of Random Search with the accuracy of Bayesian Optimization could address current limitations. Developing more sophisticated privacy preservation mechanisms while maintaining computational feasibility represents another promising research direction.

Broader Implications and Impact

Our findings have significant implications for privacy-preserving machine learning in medical imaging. The demonstrated feasibility of maintaining privacy while achieving competitive accuracy suggests potential applications across various medical domains. However, the computational requirements and performance trade-offs identified indicate the need for careful consideration of implementation strategies in clinical settings.

The framework's ability to balance privacy preservation with model performance contributes to the broader field of privacy-preserving machine learning while highlighting important considerations for practical deployment. These insights inform future development of privacy-preserving optimization strategies and their application in sensitive medical imaging contexts.

7. CONCLUSION

This study has demonstrated the effectiveness of privacy-preserving deep learning optimization for medical data classification through a comprehensive evaluation of four distinct optimization approaches, achieving significant results across different data modalities (93.62% accuracy for tabular data and 74.91% for image data) while maintaining robust privacy guarantees. Notably, the Bat Algorithm achieved an unprecedented privacy level ($\varepsilon = 0.293$) for medical image analysis while our framework's strength lies in its holistic approach to optimization, simultaneously fine-tuning

both model hyperparameters and privacy parameters through an objective function that effectively balances the privacy-utility trade-off. Our investigation revealed that different medical data modalities require specialized optimization strategies, with Bayesian Optimization excelling in tabular data applications and Random Search providing efficient solutions for image data processing, as demonstrated by the successful application of PathMNIST's complex histopathology images using ResNet-50 architecture.

Looking forward, several promising research directions emerge, including developing distributed learning approaches for improved computational efficiency, integrating federated learning techniques, extending applications to diverse medical data modalities, investigating advanced model architectures, and implementing transfer learning strategies to enhance model generalization across different medical domains. Additionally, future work should address the ethical implications and practical challenges of deploying privacy-preserving models in clinical settings, including developing robust validation frameworks, investigating model interpretability while maintaining privacy guarantees, and assessing the framework's resilience to various privacy attacks. This research establishes a strong foundation for privacy-preserving medical data analysis while highlighting the importance of balanced optimization strategies in healthcare applications, suggesting promising potential for wider adoption in clinical practice, provided that future developments continue to address the challenges of scalability, efficiency, and ethical implementation.

8. REFERENCES

[1] W. N. Price, I. G. Cohen, "Privacy in the Age of Medical Big Data", Nature Medicine, Vol. 25, No. 1, 2019, pp. 37-43.

[2] A. C. Valdez, M. Ziefle, "The Users' Perspective on the Privacy-Utility Trade-offs in Health Recommender Systems", International Journal of Human-Computer Studies, Vol. 121, 2019, pp. 108-121.

[3] C. Dwork, "Differential Privacy", Proceedings of the 33rd International Colloquium on Automata, Languages, and Programming, Venice, Italy, 10-14 July 2006, pp. 1-12.

[4] C. Dwork, A. Roth, "The Algorithmic Foundations of Differential Privacy", Foundations and Trends in Theoretical Computer Science, Vol. 9, No. 3-4, 2014, pp. 211-407.

[5] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, L. Zhang, "Deep Learning with Differential Privacy", Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24-28 October 2016, pp. 308-318.

[6] J. P. Near, D. Darais, C. Abuah, "Duet: An Expressive Higher-Order Language and Linear Type System for Statically Enforcing Differential Privacy", Proceedings of the ACM on Programming Languages, Vol. 3, 2019, pp. 1-30.

[7] D. Yu, H. Zhang, W. Chen, J. Yin, T.-Y. Liu, "Gradient Perturbation is Underrated for Differentially Private Convex Optimization", Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Yokohama, Japan, 7-15 January 2021, pp. 3117-3123.

[8] K. L. van der Veen, R. Seggers, P. Bloem, G. Patrini, "Three Tools for Practical Differential Privacy", Proceedings of PPML18: Privacy-Preserving Machine Learning Workshop at NeurIPS 2018, Montreal, QC, Canada, 7-8 December 2018.

[9] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, N. De Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization", Proceedings of the IEEE, Vol. 104, No. 1, 2015, pp. 148-175.

[10] M. Feurer, F. Hutter, "Hyperparameter Optimization", Automated Machine Learning: Methods, Systems, Challenges, Springer, 2019, pp. 3-33.

[11] B. Bischl et al. "Hyperparameter Optimization: Foundations, Algorithms, Best Practices, and Open Challenges", Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 13, No. 2, 2023, p. e1484.

[12] H. Singh, D. Witarsyah, A. Misra, R. Handayani, "Bat Algorithm: A Review on Theory, Modifications and Applications", Proceedings of the 3rd International Conference on Intelligent Cybernetics Technology & Applications, Bandung, Indonesia, 13-15 December 2023, pp. 78-83..

[13] B. Kulynych, J. Gomez, G. Kaissis, F. Calmon, C. Troncoso, "Attack-Aware Noise Calibration for Differential Privacy", Proceedings of the 37th Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 10-15 December 2024, pp. 134868-134901.

[14] L. Kangjie, "Noise Addition Strategies for Differential Privacy in Stochastic Gradient Descent", Trans-

actions on Computer Science and Intelligent Systems Research, Vol. 5, 2024, pp. 960-967.

[15] P. Thantharate, D. A. Todurkar, T. Anurag, "PRIVML: Analyzing Privacy Loss in Iterative Machine Learning with Differential Privacy", Proceedings of the Cloud Summit, Washington, DC, USA, 27-28 June 2024, pp. 107-112.

[16] K. Pan, Y.-S. Ong, M. Gong, H. Li, A. K. Qin, Y. Gao, "Differential Privacy in Deep Learning: A Literature Survey", Neurocomputing, Vol. 589, 2024, pp. 127663.

[17] K. P. Battula, B. S. Chandana, "Multi-class Cervical Cancer Classification using Transfer Learning-based Optimized SE-ResNet152 model in Pap Smear Whole Slide Images", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 6, 2023, pp. 613-623.

[18] M. Kumar, B. A. Moser, L. Fischer, "On Mitigating the Utility-Loss in Differentially Private Learning: A New Perspective by a Geometrically Inspired Kernel Approach", Journal of Artificial Intelligence Research, Vol. 79, 2024, pp. 515-567.

[19] B. Ficiu, N. D. Lawrence, A. Paleyes, "Automated Discovery of Trade-off Between Utility, Privacy and Fairness in Machine Learning Models", Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Turin, Italy, 18-22 September 2023, pp. 127-144.

[20] B. Avent, J. González, T. Diethe, A. Paleyes, B. Balle, "Automatic Discovery of Privacy-Utility Pareto Fronts", Proceedings on Privacy Enhancing Technologies, Vol. 2020, No. 4, 2020, pp. 5-23.

[21] A. Koskela, T. D. Kulkarni, "Practical Differentially Private Hyperparameter Tuning with Subsampling", Proceedings of the 36th Annual Conference on Neural Information Processing Systems, New Orleans, LA, USA, 10-16 December 2023, pp. 28201-28225.

[22] A. Arous, A. Guesmi, M. A. Hanif, I. Alouani, M. Shafique, "Exploring Machine Learning Privacy/Utility Trade-Off from a Hyperparameters Lens", Proceedings of the International Joint Conference on Neural Networks, Queensland, Australia, 18-23 June 2023, pp. 1-10.

[23] F. Galli, C. Palamidessi, T. Cucinotta, "Online Sensitivity Optimization in Differentially Private Learning", Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 38, No. 11, 2024, pp. 12109-12117.

[24] H. Wang, S. Gao, H. Zhang, W. J. Su, M. Shen, "DP-HyPO: An Adaptive Private Framework for Hyperparameter Optimization", Proceedings of the 36th Annual Conference on Neural Information Processing Systems, New Orleans, LA, USA, 10-16 December 2023.

[25] Q. Gao, H. Sun, Z. Wang, "DP-EPSO: Differential Privacy Protection Algorithm Based on Differential Evolution and Particle Swarm Optimization", Optics & Laser Technology, Vol. 173, 2024, p. 110541.

[26] Z. Bu, Y. X. Wang, S. Zha, G. Karypis, "Differentially Private Optimization on Large Model at Small Cost", Proceedings of the International Conference on Machine Learning, Honolulu, HI, USA, 23-29 July 2023, pp. 3192-3218.

[27] M. Tobaben, "Hyperparameters and Neural Architectures in Differentially Private Deep Learning", University of Helsinki, Helsinki, Finland, PhD thesis, 2022.

[28] R. Ramalingam, J. Shobana, K. Arthi, G. Elangovan, S. Radha, N. Priyanka, "An Extensive Investigation of Meta-Heuristics Algorithms for Optimization Problems", Metaheuristics Algorithm and Optimization of Engineering and Complex Systems, IGI Global, 2024, pp. 223-241.

[29] A. Banerjee, S. Pradhan, B. Misra, S. Chakraborty, "A Guide to Meta-Heuristic Algorithms for Multi-objective Optimization: Concepts and Approaches", Applied Multi-objective Optimization, Springer, 2024, pp. 1-19.

[30] A. P. Singh, Y. Dixit, G. Sharma, "Role of Meta-Heuristic Optimization Approaches in Feature Selection for Disease Diagnosis: A Comprehensive Review", Recent Advances in Computational Intelligence and Cyber Security, CRC Press, 2024, pp. 132-139.

[31] V. Harkare, R. Mangrulkar, O. Thorat, S. R. Jain, "Evolutionary Approaches for Multi-objective Optimization and Pareto-Optimal Solution Selection in Data Analytics", Applied Multi-objective Optimization, Springer, 2024, pp. 67-94.

[32] G. Thakur, A. Pal, N. Mittal, M. S. A. Yajid, F. Gared, "A Significant Exploration on Meta-heuristic Based Approaches for Optimization in the Waste Management Route Problems", Scientific Reports, Vol. 14, No. 1, 2024, p. 14853.

[33] J. Bergstra, Y. Bengio, "Random Search for Hyper-Parameter Optimization", Journal of Machine Learning Research, Vol. 13, No. 2, 2012, pp. 281-305.

[34] A. H. Victoria, G. Maragatham, "Automatic Tuning of Hyperparameters Using Bayesian Optimization", Evolving Systems, Vol. 12, No. 1, 2021, pp. 217-223.

[35] X. S. Yang, "A New Metaheuristic Bat-Inspired Algorithm", Nature Inspired Cooperative Strategies for Optimization, Springer, 2010, pp. 65-74.

[36] C. Dwork, F. McSherry, K. Nissim, A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis", Proceedings of the Third Conference on Theory of Cryptography, New York, NY, USA, 4-7 March 2006, pp. 265-284.

A Semantic Analysis Approach to Extract Personality Traits from Tweets (X)

Original Scientific Paper

Marouane Echhaimi*
Ibn Tofail University,
Science and Engineering, ENSA
(National School of Applied Sciences)
Kenitra, Morocco
Marouane.echhaimi@uit.ac.ma

Khadija Lekdioui
Ibn Tofail University,
Science and Engineering, ENSA
(National School of Applied Sciences)
Kenitra, Morocco
khadija.lekdioui@uit.ac.ma

*Corresponding author

Youness Chaabi
Royal Institute of Amazigh Culture,
CEISIC
Rabat, Morocco
chaabi@ircam.ma

Tarik Boujiha
Ibn Tofail University,
Science and Engineering, ENSA
(National School of Applied Sciences)
Kenitra, Morocco
tarik.boujiha@uit.ac.ma

Abstract *– The utilization of social networks has experienced a substantial surge in the past decade, with individuals routinely exchanging and consuming personal data. This data, subject to analysis and utilization across diverse contexts, has spurred scholarly interest in discerning the personality traits of social network users. Personality, as an intrinsic characteristic, distinguishes individuals in terms of cognition, emotion, and behavior, thereby influencing social relationships and interactions. Among the extensively studied frameworks elucidating personality variance is the Five Factor Model, commonly referred to as the "Big Five," encompassing Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism (OCEAN). Personality assessment holds practical utility across domains such as education, security, marketing, e-learning, healthcare, and personnel management. Prior investigations have demonstrated the feasibility of automatic text analysis in personality discernment. This paper introduces a multi-agent methodology grounded in semantic similarity metrics for personality trait recognition via automatic text analysis of Tweets. Our approach leverages WordNet and information content-based semantic similarity measures to analyze tweet content and classify users' personality traits. Experimental results demonstrate the effectiveness of our method, achieving a remarkable 96.28% accuracy in identifying personality traits from Tweets. This high success rate underscores the potential of our semantic analysis approach in accurately profiling social media users' personalities, offering valuable insights for various applications in behavioral analysis and personalized services.*

Keywords: *Big data, personality traits, big five personality, semantic similarity, Tweets*

Received: March 28, 2024; Received in revised form: September 28, 2024; Accepted: September 23, 2024

1. INTRODUCTION

The utilization of social networking platforms on the internet has experienced a substantial surge over the past decade, with platforms like Facebook and Twitter gaining widespread popularity for information dissemination and social interaction purposes. The online activities of users on these platforms offer valuable insights into their personalities, encompassing individual differences in cognition, behavior, experiences, emotions, opinions, and interests [1]. Understanding personality entails grasping how various aspects of an individual coalesce into a cohesive whole, representing a blend of characteristics and behaviors across diverse situations [2]. Moreover, personality plays a pivotal role in influencing decision-making processes across various domains [3]. It significantly impacts interpersonal interactions, relationships, and one's immediate surroundings, showcasing relevance in diverse contexts such as job satisfaction, professional success, and user preferences [4].

Personality delineation holds significance in numerous processes, including personnel recruitment, digital marketing, psychological interventions, educational mentoring, teaching methodologies, and health ad-

visory services. Hence, several applications stand to benefit from insights into personality traits, prompting organizational interest in profiling individuals' personalities. The existing literature presents a multitude of approaches to personality identification [5-7]. However, a common issue arises when these approaches overlook semantic similarity metrics, which are crucial for achieving concrete results in text-based semantic comparisons. Incorporating semantic similarity measures into personality identification frameworks is imperative for enhancing the accuracy and reliability of personality assessments based on textual data.

This article introduces a multi-agent system designed to analyze messages from social networking platforms and extract personality traits of Internet users utilizing semantic similarity measures. The methodology employed in this approach is grounded in the "Big Five" factor theory [8], which is currently the most widely acknowledged personality model within the scientific community. The Big Five model has gained prominence through numerous independent studies [9], culminating in its widespread acceptance and adoption as a comprehensive model for understanding personality traits [10-11].

2. STATE OF THE ART

The field of personality recognition has experienced a notable increase in research activity over recent years [12-13]. The pervasive presence of social media platforms has incentivized researchers to leverage these platforms for valuable insights that can aid in personality prediction. Numerous studies have highlighted the correlation between personality traits and online behavior [14-15]. Quercia et al. [16] were among the pioneers in investigating the association between personality traits and Twitter usage. They proposed a model capable of accurately inferring user personalities based solely on three publicly available metrics from profiles: followers, following, and listed counts. Similarly, Jusupova et al. [17] utilized demographic and social activity data to predict personalities, particularly focusing on children.

Liu et al. [18] introduced a deep learning approach to construct hierarchical systems for word and sentence representations, enabling the inference of user personalities across three languages: English, Italian, and Spanish. Van de Ven et al. [19] conducted analyses using LinkedIn, a platform primarily used for job-related decision-making, and found notable correlations with personality traits, particularly Extraversion. Furthermore, YouYou et al. [20] demonstrated the potential for computerized assessments to surpass human judgments in accuracy, particularly when sufficient data are available, surpassing judgments made by friends, spouses, and even individuals themselves.

3. BIG FIVE MODEL

The Five Factor Model of personality is a cornerstone in psychological research [21-23]. These factors are not theoretically derived but have been empirically identified through natural language analyses and psychological assessments, aiming to capture personality traits independently while providing a comprehensive description of personality. The five primary traits, known as OCEAN [24], are as follows:

- Openness: Individuals scoring high on Openness exhibit a penchant for learning new things and embracing novel experiences. This trait encompasses qualities like insightfulness, imagination, and diverse interests.
- Conscientiousness: Those with high conscientiousness levels are characterized by reliability and punctuality. Traits associated with Conscientiousness include organization, methodicalness, and rigor.
- Extroversion: Extroverts derive energy from social interactions, contrasting with introverts who draw energy from within. Extroversion involves traits such as dynamism, talkativeness, and assertiveness.
- Agreeableness: Individuals high in Agreeableness display friendliness, cooperativeness, and compassion. Conversely, lower scores in Agreeableness may indicate a more distant demeanor. Traits associated with Agreeableness include kindness, affection, and sympathy.
- Neuroticism: Also known as emotional stability, Neuroticism refers to an individual's emotional steadiness and the presence of negative emotions. High Neuroticism scores are often associated with emotional instability and heightened negative emotions. Traits linked to Neuroticism include moodiness and tenseness.

4. PROPOSED APPROACH

The proposed methodology involves automatic analysis of tweet content to determine the personalities of individual Internet users. The primary challenge lies in identifying the personality traits of Twitter users through automated semantic analysis of tweet content. Each tweet undergoes a sequence of treatments [25].

This section outlines five major personality profiles that have been identified and characterized based on various criteria. Four treatments are executed to establish a profile. Initially, tweets are retrieved, followed by simplification through the removal of unnecessary information as the second treatment. The third treatment involves linguistic analysis for word normalization, while the fourth treatment entails semantic analysis of tweets to ascertain a profile based on the Big Five personality traits for each user within the system. The architectural depiction of this process is illustrated in Fig. 1.

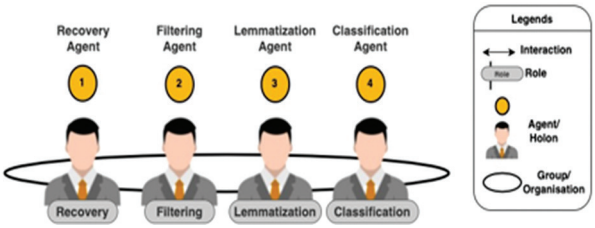


Fig. 1. General architecture of the system

4.1. RECOVERY AGENT

Initially, a retrieval agent employing the Tweepy algorithm [26] is utilized to extract tweets from Twitter and prepare them for subsequent processing steps. In our methodology, the first treatment applied to the tweet corpus involves correcting spelling and grammar errors. Such errors can significantly impact text analysis, both for human comprehension and software algorithms. A single misspelled word or sentence can drastically alter the analysis outcomes. Spelling and grammar correction is achieved using a dictionary corpus integrated with an algorithm that considers language variations, including verbal conjugations, nouns, and adjectives. This process involves comparing words in the tweets with the dictionary corpus, while also considering the context of sentences.

However, it's important to note that while the spelling and grammar checker can be beneficial, it should not replace a thorough manual review for accuracy and precision.

4.2. FILTERING AGENT

Once the retrieval task is completed, the subsequent step involves filter processing to remove words that contribute little to the information conveyed in text messages. These words, termed "empty words," are automatically filtered out for each language [27].

The most commonly occurring words in a corpus typically belong to the category of empty grammatical words, also known as stop words. These include articles, prepositions, linking words, determiners, adverbs, indefinite adjectives, conjunctions, pronouns, and auxiliary verbs, among others [27]. While these words constitute a significant portion of the text, they do not significantly contribute to the overall meaning of the text as they are ubiquitous across all texts. As per Zipf's law [28], removing these empty words during corpus pre-processing streamlines the modeling and analysis process by saving time and reducing computational complexity.

4.3. LEMMATIZATION

After the filtering step, the message undergoes linguistic analysis for word normalization, a process that involves transforming words into their canonical forms through stemming [29]. This normalization process leads to a notable reduction in the lexicon sample size [30]. Lemmatization rules are applied to various words in the corpus to unify morphological variants into a common form, such as converting verbs to their infinitive form and eliminating plural forms. Morphological variants of a word are grouped under the same lemma, allowing them to be treated as a single element (term or concept) during analysis. By reducing the total number of distinct terms, lemmatization contributes to simplifying the complexity of the analyzed text, providing significant advantages to the system.

In many languages, words can exist in multiple forms. For instance, in French, the verb "marcher" may appear as "marche," "marchait," "marchent," or "marchaient." The base form "marcher," typically found in dictionaries, is referred to as the lemma of the word. The combination of the base form with its grammatical properties is often termed the lexeme of the word.

4.4. CLASSIFICATION AGENT

The classification agent evaluates the semantic similarity of a newly acquired tweet and identifies its corresponding personality category (openness, conscientiousness, extroversion, agreeableness, and neuroticism) based on the ratio of training tweets associated with each category.

4.4.1. Semantic similarity measure

In various research domains like psychology, linguistics, cognitive science, and artificial intelligence, assessing semantic similarity among words stands as a critical concern [31]. Semantic similarity, also known as semantic proximity, refers to a measure applied to a set of messages or terms, where the concept of distance between them is predicated on the similarity of their semantic meanings or contents [32]. Conversely, syntactic similarity pertains to a different type of similarity that can be gauged based on the syntactic structures of terms.

Mathematical methodologies are employed to gauge the degree of semantic association between linguistic units, concepts, or instances, through numerical representations. This quantification is achieved by comparing the information that underpins their meanings or describes their essence. Topological similarity can be defined to estimate semantic similarity, utilizing ontologies to determine the distance between terms or concepts [33]. For instance, a basic metric for comparing concepts organized in a partially ordered set and depicted as nodes in a directed acyclic graph (e.g., a taxonomy) could be the shortest path connecting the two concept nodes.

Furthermore, semantic proximity among language units such as words and sentences can also be assessed using statistical techniques like vector space models to correlate words and textual contexts derived from an appropriate text corpus.

4.4.2. Taxonomy

The concept of semantic similarity is more narrowly focused compared to kinship or semantic relationship because the latter encompasses concepts like antonymy and meronymy, whereas similarity does not. However, there is considerable interchangeability in the literature regarding these terms, including semantic distance [34-35]. Fundamentally, semantic similarity, semantic distance, and semantic proximity address the question: "How similar are terms A and B?" The response to this query typically yields a numerical value between -1 and 1, or 0 and 1, where 1 signifies exceedingly high similarity.

4.4.3. Measuring Topological Similarity

There are primarily two approaches for calculating topological similarity between ontological concepts:

- Edge-based approach: This approach utilizes edges and their types as the primary data source [36]. It focuses on the relationships represented by the edges connecting different concepts within the ontology.

- Information content approach: In contrast, the information content approach relies on nodes and their properties as the main data sources [37-38]. It places emphasis on the inherent characteristics and attributes associated with each node or concept in the ontology.

These approaches offer distinct methodologies for evaluating topological similarity within ontological structures, with each approach leveraging different aspects of the ontology's structure and content.

4.4.4. Semantic similarity

Semantic similarity or semantic relationship refers to the measurement of closeness between terms or documents based on their meaning. There are two distinct methods for calculating semantic similarity. One method involves defining topological similarity using ontology to establish a distance metric between words. The other method relies on statistical techniques, such as the vector space model, to correlate words and their textual contexts extracted from a suitable text corpus. In this study, we concentrate on the first approach, utilizing the WordNet ontology for semantic similarity computation [39]. This approach computes similarity by considering the shared and distinct characteristics of objects as the basis for similarity assessment.

4.4.5. WordNet

WordNet is a lexical ontology designed for the English language, serving as a semantic network developed by Princeton University [40]. It structures lexical knowledge in a taxonomic hierarchy, comprising three separate databases: one for nouns, one for verbs, and one for adverbs and adjectives. Within WordNet, terms and concepts are organized into Synsets, which are lists of synonymous terms or concepts. The core component of WordNet is the Synset, which gathers synonyms associated with a specific concept. These Synsets are interconnected through various relationships, such as hypernymy (type of), meronymy (part of), and antonymy (opposite word) [41].

Semantic similarity within WordNet can be computed using two main methods: path length and information content. The path length method calculates the number of nodes or relationships between nodes within the taxonomy. This method offers advantages as it is not reliant on the static distribution of the corpus or word distributions. In our study, we focus on two concepts, "relation" and "name," within the WordNet hierarchy. We utilize WordNet 2.1, which encompasses nine distinct name hierarchies. It's worth noting that in some instances, the path between two concepts may not exist in this version of WordNet (refer to Figure 2). To address this, we introduce a root node labeled "Entity" (refer to Fig. 2), encompassing all nine provided hierarchies within WordNet.

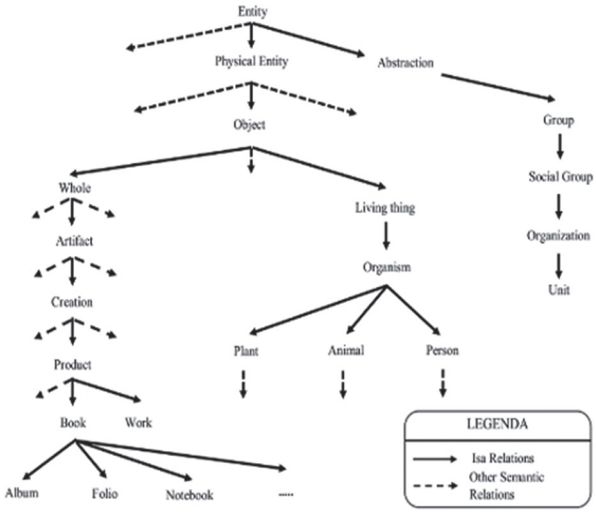


Fig. 2. Extract from the nominal hierarchy in WordNet [40].

4.4.6. Semantic similarity measurement process

The classification agent facilitates a comprehensive processing sequence, as depicted in Figure 3, for semantic similarity computation. This process comprises three distinct phases:

- **Phase 1**: Temporary construction module.
- **Phase 2**: Semantic computation module.
- **Phase 3**: Semantic similarity measurement procedures.

In Phase 1, the temporary construction module sets the groundwork for subsequent semantic computations. Phase 2 involves the actual computation of semantics, while Phase 3 encompasses procedures for measuring semantic similarity between entities.

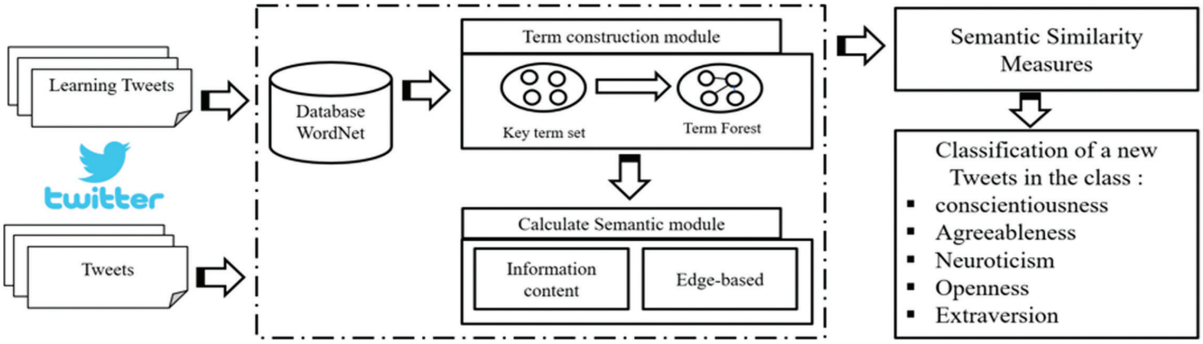


Fig. 3. Semantic Similarity Computation Diagram

- **Phase 1: Term Construction Module**

The primary goal of this module is to identify all words within the tweets that are present in WordNet and to establish the relationships between these words [42]. WordNet is leveraged to enhance the representation of text by incorporating a broader range of semantic information. Specifically, this module utilizes the hypernyms provided by WordNet as valuable features for text analysis. Therefore, the module aims to extract all tweet words found in WordNet and ascertain the relationships between these words based on the hypernym relationships provided by WordNet.

- **Phase 2: Semantic calculation module**

Philip Resnik [42] and Sun Microsystems laboratories propose an alternative to pathfinding in semantic hierarchies by introducing the concept of information content. The information content is a measure of specificity assigned to each concept within a hierarchy based on evidence extracted from a corpus. A concept with high information content is considered highly specific, whereas concepts with low information content are associated with more general ideas. The information content of a concept is calculated by tallying its frequency in a large corpus, as well as the frequency of all subordinate concepts in the hierarchy. The probability of a concept is determined through maximum likelihood estimation, and its information content is derived from the negative logarithm of this probability.

Resnik's similarity measure establishes a semantic relationship between two concepts based on the extent of shared information between them. This shared information is determined by the information content of the least specific concept in the hierarchy that encompasses both concepts.

The similarity between words based on information content:

- Relies on the structure of the thesaurus.
- Improves the path-based approach by normalizing based on hierarchy depth.
- Represents the distance associated with each edge in the hierarchy.
- Integrates probabilistic information derived from a corpus.

The probability that a random word belongs to a concept is calculated as follows (Equation 1) [43]:

$$p(c)=(\sum w \in w_{(c)}\ count\ (w))/\ N \quad (1)$$

Here:

Words (c) represent the set of words subsumed by the concept c.

N is the total number of words in the corpus and the thesaurus.

P(root) = 1 since all words are subsumed by the root concept.

Furthermore, it's worth noting that the probability decreases as the concept descends lower in the hierarchy, reflecting the decreasing specificity and generality as we move down the hierarchy levels.

We need two more definitions:

1) Information Content of a Concept ($IC(c)$) [43]:

$$IC(c)=-log\ P(c) \quad (2)$$

This equation represents the information content of a concept c based on the probability $P(c)$ that a random word from the corpus belongs to the concept c. It quantifies the specificity of the concept within the hierarchy.

2) Lowest Common Subsumer (LCS($c1$, $c2$)) [43]:

The LCS($c1$, $c2$) refers to the lowest node in the hierarchy that serves as a hypernym of both concepts $c1$ and $c2$. It denotes the most specific common ancestor shared by the two concepts.

3) Resnik Similarity Measurement (simResnik($c1$, $c2$)) [43]:

$$simResnik\ (c1,c2) = -log\ P\ (LCS(c1,c2)) \quad (3)$$

This equation calculates the Resnik similarity between concepts $c1$ and $c2$. It estimates the shared amount of information between the concepts by utilizing the information content of their lowest common subsumer (LCS).

- **Phase 3: Semantic similarity measures**

Semantic vectors for $T1$ and $T2$ can be constructed using T statistics and corpus information. The process of deriving semantic vectors for $T1$ (Equation 4) can be described as follows:

Given a word w, let us define [43]:

$$Sim(W_1, W_2) = max_{c1,c2}[smin(c1, c2)]$$
$$Sim(T1, T2) = \sum_{i=1}^{n} \left(\frac{sim(Wi,Wi+1}{n}\right) \quad (4)$$

We obtain measurement values of semantic similarity for Resnik between Tweet 1 and Tweet 2 (5) [43]:

$$Sim\ Resnik\ (T1,T2) = value\ 2 \quad (5)$$

Tweets are comprised of words, hence it is rational to represent a Tweet using the words it contains. Unlike conventional methods that utilize pre-compiled word lists with numerous words, our approach dynamically constructs semantic vectors solely based on the Tweets being compared. Recent advancements in semantic analysis focus on automatically extracting semantic word vectors for sentences [40]. Given two Tweets $T1$ and $T2$, a word set is formed with (Equation 6) [43]:

$$T=T1 \cup T2 = \{W_1, W_2, ..., W_n\} \quad (6)$$

The word set T encompasses all distinct words from $T1$ and $T2$. Inflectional morphology may lead to a word appearing in various forms within a message, each form having a specific meaning in a given context. Therefore, we consider the word form as it appears in the Tweet for our analysis.

5. EXPERIMENTS AND RESULTS

The objective of this research is to automate the process of identifying the personalities of Internet users by conducting a semantic analysis of their Tweets. To achieve this, we conducted a comparative study between human evaluation and the results produced by our model.

We performed experiments involving intuitive analysis of Tweets from Internet users based on notes from a test corpus. Our focus was on both qualitative and quantitative analyses conducted with the input of three experts. We compiled a corpus of Tweets from a sample of 10 Internet users, each contributing 100 Tweets to our dataset. The intuitive analysis of these Tweets included assigning a personality to each user and then identifying the language acts that contribute to determining the personality traits.

For the identification of personality traits, we utilized the MyPersonality database as the learning base for our system. This database served as the foundation for training our model to accurately classify and infer personality traits based on the semantic analysis of Tweets.

5.1. MYPERSONNALITY DATABASE

To test our approach, we utilize a dataset derived from the MyPersonality project. This dataset was curated for research purposes by David Stillwell and Michael Kosinski through a Facebook application designed to administer a personality test and gather diverse personal information and activities from the profiles of consenting Facebook users. The MyPersonality application operated from 2007 to 2012, accumulating a substantial volume of data.

Our study is built upon a subset of the original MyPersonality dataset, which has been made publicly available [12]. This subset comprises 9913 English status updates extracted from 250 users, with their identities anonymized. The dataset is further annotated with scores for personality traits and includes basic statistics describing the users' social networks.

5.2. TEST CORPUS

In order to have a suitable test base, more than 1.5 million tweets were retrieved using Twitter's Tweepy algorithm [25]; then 10 users were selected with at least 100 tweets per person, the base is in CSV format and each line has 6 fields:

0 - the polarity of the Tweet

1 - Tweet id

2 - the date of the Tweet

3 - the request. If there is no query, then this value is NO_QUERY.

4 - the user who tweeted

5 - the text of the Tweet

The text of the Tweets was compared with MyPersonality learning base taking into account the semantic similarity measure to extract personality traits more accurately.

5.3. EXECUTION RESULTS

The first set of results from the analysis on Tweets was monitored by 3 experts. Each expert analysis was done in two steps. The first step consisted of assigning a profile to each internet user according to their personality traits. The second step consisted of analyzing the Tweets exchanged by internet users. The experts were asked to classify the Tweets of internet users into five personality traits (Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness) by analyzing their content, i.e. identifying the speech acts that characterize Tweets (see Table 1).

When the same Tweets are submitted between internet users to the automatic analysis system that is proposed, the results shown in Tables 1 and 2 are obtained, for the same internet users.

The analysis of these results in light of the characteristics of the internet users' profiles allows associating a personality profile to each user. Seen the results of the semantic analysis to calculate the percentage of each personality type for the users JBnVFCLover78, five broad personality traits emerge: extraversion, agreeableness, openness, conscientiousness, and neuroticism. However, by analyzing the resulting percentages of each personality, we find that the percentage of neuroticism personality is important and characterizes the personality of JBnVFCLover78.

Table 1. Results of the intuitive analysis

Twitter users	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
User 1	29.15 %	16%	16.21 %	16.2 %	22.44 %
User 2	42.01 %	8.9 %	8.6 %	20.82 %	19.67 %
User 3	27%	11.4%	14.13 %	31.47 %	16%
User 4	10.54 %	18%	21.56 %	21.06%	28.84 %
User 5	27.17 %	7.05 %	37.22 %	20.8 %	7.76 %
User 6	65.75 %	15.27 %	3.98 %	13.15 %	1.85 %
User 7	4%	45.4 %	10.38 %	32.56 %	7.66 %
User 8	14.46 %	5.54 %	51.58 %	10.69 %	15.73 %
User 9	10.85 %	11.09 %	68.21 %	4.53 %	5.32 %
User 10	5.96%	6.15%	13.65 %	69.06 %	5.18 %

The table below represents the results of running the system on the test corpus:

Table 2. Results of the execution of the system on the test corpus

Twitter users	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
User 1	34.33 %	14.91 %	13 0/0	17.93 %	19,83 %
User 2	46.41 %	4.57 %	10,11 %	22.22 %	16.69 %
User 3	31.62 %	0.68 %	12,42 %	33.51 %	21.77 %
User 4	06.12 %	24.56 %	17.26	16.22 %	35.84 %
User 5	33.08 %	09.00 %	46.62 %	20.00 %	6.57 %
User 6	60.15 %	12.14 %	05.98 %	17.73 %	04.00 %
User 7	07.56 %	51.12 %	07.00 %	28.00 %	06.32 %
User 8	11.00%	08.00 %	56.27 %	14.00 %	10.73 %
User 9	08.00 %	13.16 %	71.06 %	02.78 %	05.00 %
User 10	07.13 %	04.22 %	08.65 %	80.00 %	02.56 %

Table 3. Comparison between analysis system and intuitive analysis

Twitter users	Agreeableness	Conscientiousness	Extraversion	Neuroticism	Openness
User 1	94,82%	98,91%	96,79%	98,27%	97,39%
User 2	95,6%	95,67%	98,49%	98,6%	97,02%
User 3	95,38%	89,28%	98,29%	97,96%	94,23%
User 4	95,58%	93,44%	95,7%	95,16%	93%
User 5	94,09%	98,05%	90,6%	99,2%	98,81%
User 6	94,4%	96,87%	98%	95,42%	97,85%
User 7	96,44%	94,28%	96,62%	95,44%	98,66%
User 8	96,54%	97,54%	95,31%	96,69%	95%
User 9	97,15%	97,93%	97,15%	98,25%	99,68%
User 10	98,83%	98,07%	95%	89,06%	97,38%
Average	95,88%	96,00%	96,19%	96,40%	96,90%
Final result			**96,28 %**		

Table 3 shows the margin of error between the intuitive analysis of the expert and the system analysis. This error margin gives a confidence degree for results validation. A result of 100% means that the system is perfectly aligned with the human expert. In the example below, from the intuitive analysis the internaut « Bigeny » emerges as an Agreeableness personality with 29,15% (see table 2) and 34,33 % according to the results of the system (see table 1). We have considered the result of the intuitive analysis expert as a reference; we can see that the error rate is 3.72 %.

Table 4. A comparison of the proposed approach with alternative methods.

Methode	Accuracy				
	Agreeableness	Conscientiousness	Extraversión	Neuroticism	Openness
Wang et al [44]	76.8 %	75%	85%	70%	79%
Zheng andWu [45]	65%	62%	71%	68%	70%
Xue et al [46]	78.57 %	77.20 %	82.35 %	83.08 %	79.55 %
BOW [47]	88.75 %	87.86 %	88.15 %	89.52 %	87.94 %
Skip-Vec [47]	88.81 %	88.19 %	88.39 %	89.71 %	88.27 %
Our method	**95.88 %**	**96.00 %**	**96.19 %**	**96.40 %**	**96.90 %**

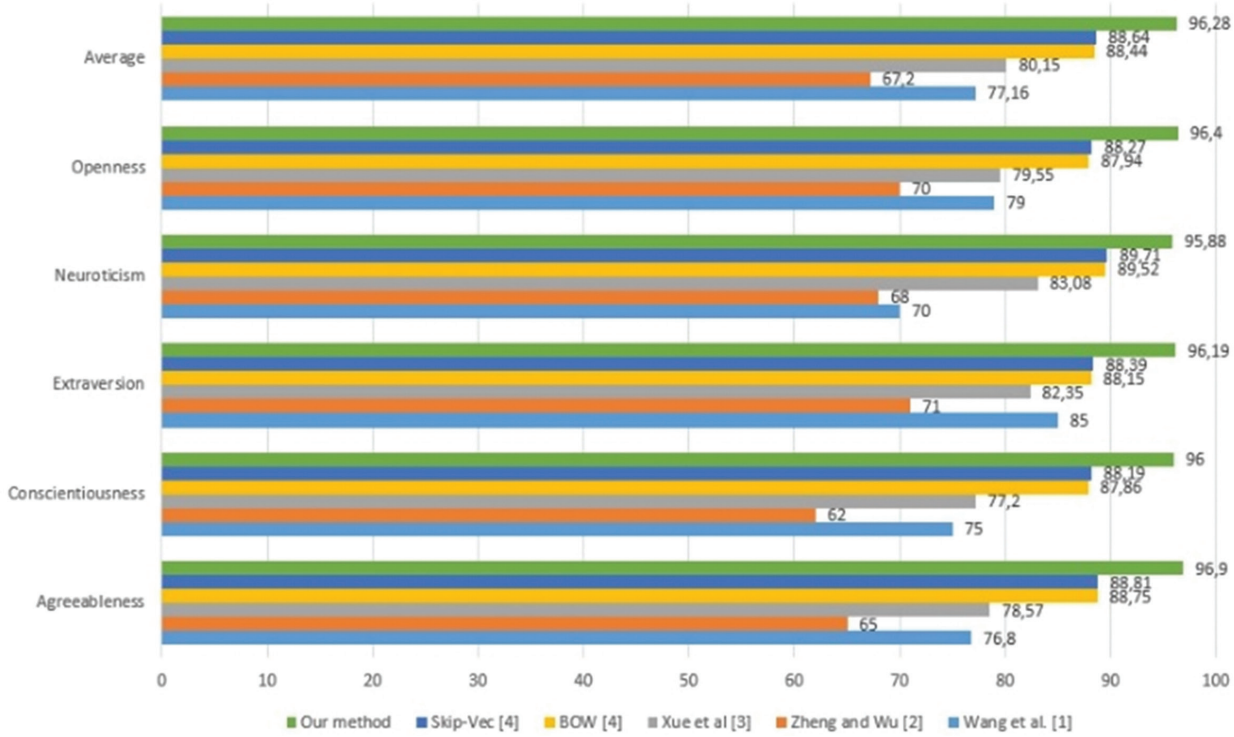


Fig. 4. Graphical representation comparing the proposed approach with alternative methods

The comparison results demonstrate that our method significantly outperforms other evaluated approaches, achieving an average precision of 96.28%. This exceptional performance is particularly evident in traits such as Openness and Neuroticism, showcasing its superior ability to capture nuanced textual cues associated with these personality dimensions.

In contrast, alternative methods exhibit varying levels of performance. Wang et al.'s use of graph convolutional networks for text encoding achieves an average precision of 77.16%, indicating moderate effectiveness [44]. Zheng and Wu's approach, employing semi-supervised learning on Facebook status data, shows a lower precision of 67.2%, suggesting limitations in leveraging social media for precise personality trait recognition [45].

Xue et al.'s method, which employs semantically-enhanced sequential modeling, improves upon these results with an average precision of 80.15% [46]. This method excels particularly in traits like Agreeableness and Extraversion, highlighting its ability to capture contextual relationships within texts.

The Bag of Words (BOW) and Skip-Vec methods achieve average precisions of 88.44% and 88.64%, respectively, demonstrating solid performance but still trailing behind our approach [47]. Skip-Vec slightly outperforms BOW, likely due to its superior incorporation of contextual relationships [47].

Our method clearly surpasses others, demonstrating superior efficacy in personality trait recognition from texts. These findings underscore the robustness and accuracy of our approach, even outperforming newer and more sophisticated techniques in the field.

6. CONCLUSION AND PERSPECTIVES:

Personality traits significantly influence decision-making processes, interpersonal interactions, and individual success. Understanding people's personalities is essential for various applications, such as job candidate selection, targeted marketing, and security measures. Our study focused on detecting personality traits by analyzing Tweets using semantic similarity measures and a learning base grounded in the Big Five model. The experimental results demonstrated a high accuracy rate of 96.28% in identifying personality traits, underscoring the potential of our approach in accurately profiling social media users' personalities.

We have incorporated recent literature to contextualize our findings, highlighting the alignment and relevance of our methodology with current research trends in automated personality detection. The integration of semantic similarity measures, particularly using WordNet and information content-based similarity measures, played a crucial role in enhancing the accuracy and reliability of personality assessments based on textual data.

Future work will focus on improving the system's execution time and expanding the test base to include more comprehensive user information. Additionally, we plan to explore the use of deep learning methods and generative AI to further optimize the accuracy of personality trait detection. By incorporating advanced new techniques, we aim to enhance the robustness and applicability of automated personality profiling in various domains.

7. REFERENCES

[1] E. Kazdin, Encyclopedia of Psychology: 8 Volume Set, American Psychological Association and Oxford University Press, 2000.

[2] B. Y. Pratama, R. Sarno, "Personality classification based on Twitter text using Naïve Bayes, KNN and SVM", Proceedings of the International Conference on Data and Software Engineering, Yogyakarta, Indonesia, 25-26 November 2015, pp. 170–174.

[3] I. Cantador, I. Fernández-Tobías, A. Bellogín, "Relating personality types with user preferences in multiple entertainment domains", Project Papers and Workshop Proceedings of the 21st Conference on User Modeling, Adaptation, and Personalization, Vol. 997, 2013.

[4] J. Golbeck, C. Robles, M. Edmondson, K. Turner, "Predicting personality from Twitter", Proceedings of the IEEE Third International Conference on Privacy, Security, Risk and Trust and the IEEE Third International Conference On Social Computing, Boston, MA, USA, 9-11 October 2011, pp. 149-156.

[5] E. Kafeza, A. Kanavos, C. Makris, P. Vikatos, "T-PICE: Twitter personality based influential communities extraction system", Proceedings of the IEEE International Congress on Big Data, Anchorage, AK, USA, 27 June - 2 July 2014, pp. 212-219.

[6] Y. Zheng, "Identifying Dominators and Followers in Group Decision Making based on The Personality Traits", Proceedings of the IUI Workshops, 2018.

[7] N. Hawi, M. Samaha, "Identifying commonalities and differences in personality characteristics of Internet and social media addiction profiles: traits, self-esteem, and self-construal", Behaviour & Information Technology, Vol. 38 No. 2, 2019, pp. 110-119.

[8] S. V. Paunonen, M. C. Ashton, "Big five factors and facets and the prediction of behavior", Journal of Personality and Social Psychology, Vol. 81, No. 3, 2001, p. 524.

[9] M. B. Donnellan, F. L. Oswald, B. M. Baird, R. E. Lucas, "The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality", Psychological Assessment, Vol. 18, No. 2, 2006, p. 192.

[10] J. M. Digman, "Higher-order factors of the Big Five", Journal of Personality and Social Psychology, Vol. 73 No. 6, 1997, p. 1246.

[11] B. De Raad, "The Big Five Personality Factors: The psycholexical approach to personality", Hogrefe & Huber Publishers, 2000.

[12] F. Celli, B. Lepri, J. Biel, D. Gatica-Perez, G. Riccardi, "The workshop on computational personality recognition 2014", Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 2014, pp. 1245–1246.

[13] M. Tkalcic, B. D. Carolis, M. D. Gemmis, A. Odi´c, A. Košir, "Preface: EMPIRE 2014-2nd Workshop Emotions and Personality in Personalized Services", Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization Aalborg, Denmark, 7-11 July 2014.

[14] D. Hughes, M. Rowe, M. Batey, A. Lee, "A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage", Computers in Human Behavior, Vol. 28, 2011, pp. 561–569.

[15] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, D. Stillwell, "Personality and patterns of Facebook usage", Proceedings of the 4th Annual ACM Web Science Conference, Evanston, IL, USA, 22-24 June 2012.

[16] D. Quercia, M. Kosinski, D. Stillwell, J. Crowcroft, "Our Twitter profiles, our selves: Predicting personality with Twitter", Proceedings of the IEEE Third International Conference on Privacy, Security, Risk and Trust and the IEEE Third International Conference on Social Computing, Boston, MA, USA, 9-11 October 2011, pp. 180-185.

[17] A. Jusupova, F. Batista, R. Ribeiro, "Characterizing the Personality of Twitter Users based on their Timeline Information", Proceedings of the Atas da 16 Conferência da Associacao Portuguesa de Sistemas de Informação, Porto, Portugal, Vol. 16, 22-24 December 2016, pp. 292-299.

[18] F. Liu, J. Perez, S. Nowson, "A Language-independent and Compositional Model for Personality Trait Recognition from Short Texts", arXiv:1610.04345, 2016.

[19] N. Van de Ven, A. Bogaert, A. Serlie, M. J. Brandt, J.J. Denissen, "Personality perception based on LinkedIn profiles", Journal of Managerial Psychology, Vol. 32, 2017, pp. 419-429.

[20] W. YouYou, M. Kosinski, D. Stillwell, "Computer-based personality judgments are more accurate

than those made by humans", Proceedings of the National Academy of Sciences, USA, Vol. 112, No. 4, 2014, pp. 1036-1040.

[21] S. D. Gosling, P. J. Rentfrow, W. B. Swann Jr, "A very brief measure of the Big-Five personality domains", Journal of Research in personality, Vol. 37 No. 6, 2003, pp. 504-528.

[22] M. R. Barrick, M. K. Mount, "The big five personality dimensions and job performance: a meta-analysis", Personnel Psychology, Vol. 44 No. 1, 1991, pp. 1-26.

[23] L. M. Hough, "The 'Big Five' personality variables-construct confusion: Description versus prediction", Human performance, Vol. 5, No. 1-2, 1992, pp. 139-155.

[24] O. Plaisant, J. Guertault, R. Courtois, C. Réveillère, G. A. Mendelsohn, O. P. John "Big Five History: OCEAN of personality factors. Introduction of the French Big Five Inventory or BFI-Fr", Annales Médico-psychologiques, revue psychiatrique, Elsevier Masson, Vol. 168, No. 7, 2010, pp. 481-486.

[25] Y. Chaabi, L. Khadija, B. Mounia, "Semantic Analysis of Conversations and Fuzzy Logic for the Identification of Behavioral Profiles on Facebook Social Network", International Journal of Emerging Technologies in Learning, Vol. 14, No. 7, 2019.

[26] J. Roesslein, "tweepy Documentation", http://tweepy.readthedocs.io/en/v3, 5. (accessed: 2009)

[27] Y. Chaabi, L. Khadija, F. Jebbor, R. Messoussi, "Determination of Distant Learner's Sociological Profile Based on Fuzzy Logic and Naïve Bayes Techniques", International Journal of Emerging Technologies in Learning, Vol. 12, No. 10, 2017, pp. 56-75.

[28] H. Guiter, M. V. Arapov, "Studies on Zipf's law", Brockmeyer, 1982.

[29] J. B. Lovins, "Development of a stemming algorithm", Mechanical Translation and Computational Linguistics, Vol. 11, No. 1-2, 1968, pp. 22-31.

[30] Y. Chaabi, R. Messoussi, V. Hilaire, Y. Ruichek, K. Lekdioui, R. Touahni, "Design of an Intelligent System to Support Tutors in Learning Communities Using Multi-Agent Systems and Fuzzy Logic", International Review on Computers and Software, Vol. 10, No. 8, 2015, pp. 845-855.

[31] Z. Wu, M. Palmer, "Verbs semantics and lexical selection", Proceedings of the 32nd annual meeting on Association for Computational Linguistics, 1994, pp. 133-138.

[32] G. Zhu, C. A. Iglesias "Computing Semantic Similarity of Concepts in Knowledge Graphs", IEEE Transactions on Knowledge and Data Engineering, Vol. 29, No. 1, 2017, pp. 72-85.

[33] W. Hua, Z. Wang, H. Wang, K. Zheng, X. Zhou, "Understand Short Texts by Harvesting and Analyzing Semantic Knowledge", IEEE transactions on Knowledge and Data Engineering, Vol. 29, No. 3, 2017, pp. 499-512.

[34] J. J. Jiang, D. W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy", arXiv:cmp-lg/9709008, 1997.

[35] T. Pedersen, S. V. Pakhomov, S. Patwardhan, C. G. Chute, "Measures of semantic similarity and relatedness in the biomedical domain", Journal of Biomedical Informatics, Vol. 40, No. 3, 2007, pp. 288-299.

[36] D. Girardi, S. Wartner, G. Halmerbauer, M. Ehrenmüller, H. Kosorus, S. Dreiseitl, "Using concept hierarchies to improve calculation of patient similarity", Journal of Biomedical Informatics, Vol. 63, 2016, pp. 66-73.

[37] H. Y. Wang, W. Y. Ma, "Integrating Semantic Knowledge into Lexical Embeddings Based on Information Content Measurement", Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Vol. 2, 2017, pp. 509-515.

[38] W. U. Hao, H. Huang, "Efficient Algorithm for Sentence Information Content Computing in Semantic Hierarchical Network", IEICE Transactions on Information and Systems, 2017.

[39] A. Meštrović, A. Calì, "An ontology-based approach to information retrieval", Semantic Keyword-based Search on Structured Data Sources, 2016, pp. 150-156.

[40] C. Fellbaum, G. A. Miller, "WordNet: An electronic lexical database, language, speech, and communication", MIT Press, 1998.

[41] S. Sudha, "WordNet-An On-line Lexical Database", 2016.

[42] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy", Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Canada, 1995, pp. 448-453.

[43] L. Meng, R. Huang, J. Gu, "A review of semantic similarity measures in Wordnet", International Journal of Hybrid Information Technology, Vol. 6, No. 1, 2013, pp. 1-12.

[44] Z. Wang, C. H. Wu, Q. B. Li, B. Yan, K. F. Zheng, "Encoding text information with graph convolutional networks for personality recognition", Applied Sciences, Vol. 10, No. 12, 2020, p. 4081.

[45] H. Zheng, C. Wu, "Predicting personality using Facebook status based on semi-supervised learning", Proceedings of the 11th International Conference on Machine Learning and Computing, Zhuhai, China, 22-24 February 2019, pp. 59-64.

[46] X. Xue, J. Feng, X. Sun, "Semantic-enhanced sequential modeling for personality trait recognition from texts", Applied Intelligence, Vol. 51, 2021, pp. 7705-7717.

[47] H. J. Escalante, H. Kaya, A. A. Salah, S. Escalera, Y. Gucluturk, U. Guclu, "Modeling, recognizing, and explaining apparent personality from videos", IEEE Transactions on Affective Computing, Vol. 13, No. 2, 2022, pp. 894-911.

FusionNet- A Hybrid Deep Learning Approach for Accurate Drug-Target Binding Prediction

Original Scientific Paper

Tintu Vijayan*
Department of CSE, Presidency University,
Bangalore, India
tintuidhaya@gmail.com

Pamela Vinitha Eric
Department of CSE, Presidency University,
Bangalore, India
pamelavinitha.eric@presidencyuniversity.in

*Corresponding author

Abstract *– Identifying drug-target binding affinities (DTBA) is crucial in drug discovery, to understand how effectively drugs interact with their targets. However, traditional methods often struggle to accurately capture the complex relationships in biological data, leading to limitations in their predictive power. This paper introduces FusionNet, an advanced deep-learning model designed to improve DTBA prediction. FusionNet combines the strengths of Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Transformers, to better understand both short-range and long-range interactions in biological sequences and employs the Layer-wise Adaptive Moments (LAMB) optimizer, which ensures the model is more efficient and stable, especially when working with large datasets. FusionNet achieved an MSE of 0.20 and an* rm^2 *of 0.681 on the Davis dataset and an MSE of 0.18 and an* rm^2 *of 0.71 on the KIBA dataset, significantly outperforming existing models like SimBoost, GANsDTA, DeepCDA, and DeepDTA, making it a powerful tool for drug discovery and bioinformatics. This work not only enhances the accuracy of DTBA prediction but also sets new performance standards by integrating advanced neural network architectures and optimizing their training process. FusionNet effectively addresses the limitations of previous approaches, offering a more reliable and efficient method for predicting drug-target interactions.*

Keywords*: Drug Target Binding Affinity (DTBA), Convolutional neural network (CNN), Layer-wise Adaptive Moments (LAMB)*

Received: August 22, 2024; Received in revised form: November 14, 2024; Accepted: November 15, 2024

1. INTRODUCTION

Drug discovery and development is a complex, lengthy, and expensive process, often requiring over a decade and billions of dollars to bring a new drug to market. A critical aspect of this process is understanding the interaction between drug molecules and their target proteins, quantified by the drug-target binding affinity (DTBA). Accurate prediction of DTBA is essential for determining the efficacy and safety of potential therapeutics. Traditional experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, surface plasmon resonance (SPR), and isothermal titration calorimetry (ITC) are critical for understanding drug-target interactions at the molecular level [1]. However, these methods are labor-intensive, time-consuming, and expensive, often limiting their scalability and practical application in early-stage drug discovery. For example, X-ray crystallography, while highly precise, can take several months to years for data collection and structural determination, making it impractical for high-throughput screening [2]. Similarly, NMR spectroscopy requires a high concentration of samples and extensive computational resources, limiting its efficiency [3]. Consequently, there is a significant demand for computational methods that can predict DTBA accurately, efficiently, and at scale.

Recent advances in computational power and the availability of large-scale biochemical datasets have catalyzed the development of various computational approaches for DTBA prediction. These methods range from classical machine learning techniques to more sophisticated deep learning models. Traditional machine learning models, such as random forests and support vector machines (SVMs), have been widely used in drug-target binding affinity prediction due to their simplicity and interpretability [4]. However, these models typically rely on manually engineered features,

such as molecular fingerprints and physicochemical descriptors, which fail to capture the complex, nonlinear relationships inherent in biological data. This reliance on handcrafted features restricts the models' generalization ability to unseen data, reducing prediction accuracy and reliability [5]. Additionally, SVMs and other classical models are computationally inefficient when dealing with high-dimensional, large-scale datasets, limiting their effectiveness in practical, large-scale drug discovery applications [6]. In contrast, deep learning models can learn feature representations directly from raw data, showing promise in overcoming these limitations. Convolutional Neural Networks (CNNs) and Long Short-Term Memory networks (LSTMs) have been particularly effective due to their ability to model spatial and sequential data, respectively. [7]

Although CNNs and LSTMs have achieved great success in DIBA prediction, there are still some significant challenges. CNNs excel at capturing local spatial patterns but struggle to model long-distance dependencies, which are critical for understanding interactions between distant residues or a drug and its binding site on a protein. LSTMs, targeted at sequential data processing, can partially capture sequential dependencies. However, they may run into difficulties such as vanishing gradients and computational inefficiency when dealing with lengthy sequences, resulting in their model for the full complexity of drug-target interactions. [8]

Moreover, both CNNs and LSTMs fall short of effectively leveraging contextual information, which is essential for accurately modeling biological sequences. For instance, the interaction between a drug and a protein is influenced by the broader context of the protein's structure and the physicochemical properties of the drug, which these models struggle to capture comprehensively. This limitation hinders their ability to provide accurate and reliable DTBA predictions Another limitation lies in the optimization of deep learning models. Standard optimizers, such as stochastic gradient descent (SGD) and Adam, may not be efficient enough to train large models on complex datasets, leading to suboptimal performance and longer training times, particularly when processing large bioinformatics datasets [9].

To address these challenges, we propose a hybrid deep learning model that integrates CNNs, LSTMs, and Transformers to improve DTBA prediction. The model leverages CNNs for local feature extraction, LSTMs for capturing sequential dependencies, and Transformers for modeling long-range dependencies and contextual information via self-attention [10]. This combination provides a comprehensive representation of sequences, enhancing accuracy and reliability. CNNs extract key patterns, LSTMs maintain temporal order, and Transformers capture broader interactions between drug and protein sequences.

To further enhance the training efficiency and performance of the hybrid model, the Layer-wise Adaptive Moments (LAMB) optimizer is employed. LAMB is specifically designed for large-batch training, making it well-suited for deep learning models on large datasets. It adapts the learning rate for each layer individually, ensuring stable and efficient convergence, and addressing the shortcomings of traditional optimizers in training large models efficiently, which is particularly beneficial in the context of our complex hybrid model.

The contributions are:

- A novel hybrid model combining CNNs, LSTMs, and Transformers to improve DTBA prediction by addressing key limitations of existing methods
- Implementation of the LAMB optimizer to enhance training efficiency and performance.
- Extensive experiments demonstrating the superiority of the model over traditional approaches.

2. RELATED WORKS

Deep learning has become a powerful tool for predicting drug-target binding affinity (DTBA). Öztürk et al. introduced DeepDTA, a model that uses CNNs to analyze sequence information from both targets and drugs, achieving high accuracy in predicting binding affinities [11]. Feng et al. developed PADME, which consistently outperformed baseline methods across multiple datasets using a deep learning-based framework [12]. Furthering this, Öztürk et al. introduced WideDTA, combining chemical and biological textual sequence information to enhance binding affinity predictions [13]. Zeng et al. improved upon these approaches by integrating multiple attention blocks, effectively encoding correlations between atoms and modeling drug-target interactions [14]. In other applications, Elansary et al. developed a bat-inspired optimizer using RNNs for predicting anti-viral cure drugs, highlighting the versatility of deep learning in drug discovery [15]. Makowski et al. explored machine learning models for co-optimizing therapeutic antibody affinity and specificity, emphasizing their importance in therapeutic development [16].

Hybrid models that integrate CNNs and LSTMs have shown significant improvements in drug discovery and drug repurposing. These models leverage CNNs' ability to capture local features and LSTMs' strengths in modeling sequential dependencies. For example, Yoon et al. outperformed traditional models in predicting DNA-protein binding sites using this hybrid approach [17]. In DTBA, hybrid CNN-LSTM models have demonstrated remarkable success by extracting key features from SMILES representations and protein sequences before processing them with LSTMs for accurate affinity predictions. DeepBind, which also integrates CNNs and LSTMs, has been used to predict protein-DNA binding affinity with high accuracy.

Transformers have revolutionized sequence modeling by effectively capturing long-range dependencies

with self-attention mechanisms. In bioinformatics, transformers have been applied to tasks like protein structure prediction, achieving state-of-the-art results. The LAMB optimizer, developed by You et al., enhances training efficiency and performance in large-batch scenarios by adapting learning rates layer-wise, which is especially beneficial for models with extensive parameter spaces [18]. LAMB builds on Adam [19], offering better generalization and faster convergence. The proposed hybrid model, combining CNNs, LSTMs, and Transformers with the LAMB optimizer, enhances DTBA prediction by leveraging their strengths for improved accuracy and generalization.

3. METHODOLOGY

3.1. DATASETS

The study utilizes two well-established datasets for predicting drug-target binding affinities: the KIBA dataset and the Davis dataset [20]. These datasets provide comprehensive and experimentally validated information on drug-target interactions, making them suitable benchmarks for evaluating the performance of the proposed hybrid model. These datasets provide comprehensive and experimentally validated information on drug-target interactions, making them suitable benchmarks for evaluating the performance of the proposed hybrid model.

Table 1. The statistics of datasets

Dataset	No of drugs	No of proteins	Known DTI
Davis	68	442	30,056
Kiba	2,116	229	118,254

The KIBA dataset integrates information from multiple sources, including Kinase Inhibitor Bioactivity data, to provide a unified measure of drug-target binding affinities. It combines data from Ki, Kd, and IC50 measurements, offering a robust and comprehensive resource for kinase inhibitor bioactivity.

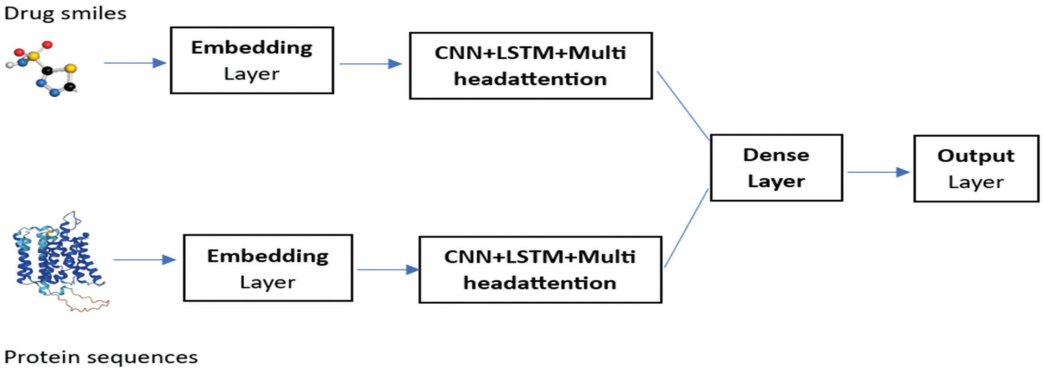


Fig.1. Proposed Model Structure

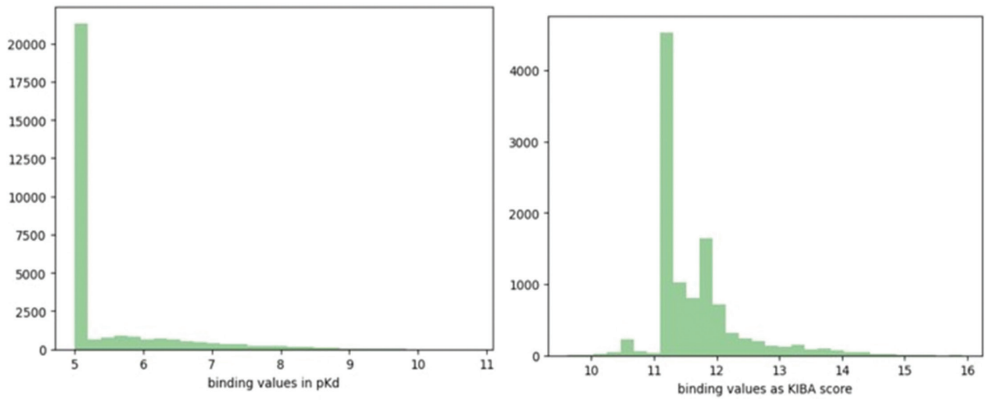


Fig. 2. Distribution of binding affinity values in Davis and Kiba dataset

3.2. PROPOSED METHODOLOGY

Predicting drug-target binding affinity (DTBA) is crucial in drug discovery, where the goal is to quantify the interaction strength between a drug molecule and a target protein. The problem can be mathematically stated as, given a drug represented by its SMILES (Simplified Molecular Input Line Entry System) string, S, and a target protein represented by its amino acid sequence, P, the task is to predict the binding affinity y, which can be mathematically represented as

$$y = f(S, P) \quad (1)$$

where f is the predictive model.

The ultimate objective is to optimize the function. $f(S, P)$ such that the predicted binding affinity y' close-

ly approximates the true binding affinity y minimizing the mean squared error (MSE) across the dataset.

In the proposed model as in Fig.1, Convolutional Neural Networks (CNNs) to extract local features from SMILES strings and protein sequences, utilizing multiple 1D convolutional layers with filter sizes of 4, 6, and 8, chosen based on empirical testing to capture various sequence patterns. Bidirectional Long Short-Term Memory (LSTM) networks handle sequential dependencies, with 64 units selected to balance complexity and efficiency. To model long-range dependencies, Transformer layers with multi-head self-attention (4 heads and key dimension of 64) are incorporated, chosen through preliminary experiments to optimize performance while managing computational demands. Hyperparameters were tuned using cross-validation and grid search, including filter sizes, LSTM dimensions, and dropout rates (set at 0.5 to mitigate overfitting). The LAMB optimizer was selected for its ability to adjust learning rates for each layer adaptively, improving training efficiency, particularly with large batch sizes (128) and a learning rate chosen to ensure stable and rapid convergence. The model was trained for 100 epochs, with Mean Squared Error (MSE) as the loss function and evaluation metric to measure the alignment between predicted and actual binding affinities.

Mathematically, the convolution operation for each filter size i can be expressed as

$$conv(x)=ReLU(Conv1D(x, filters_i, Kernel\_size_i)) \quad (2)$$

where x represents the input sequence, and $filters_i$, $Kernel\_size_i$ correspond to the filter and kernel size for the i-th convolutional layer. Following convolution, max-pooling layers are applied to reduce the dimensionality and retain the most significant features, expressed as

$$MaxPool_i=Maxpooling1D\,(Conv_i\,(x)) \quad (3)$$

The pooled features are then concatenated to form a comprehensive feature representation. Subsequently, Long Short-Term Memory (LSTM) networks are employed to capture sequential dependencies in the data. The bidirectional LSTM processes the CNN-extracted features f by considering both forward and backward contexts, which can be mathematically represented as

$$LSTMbi(f)= Concat(LSTMforward\,(f), LSTMbackward\,(f)) \quad (4)$$

Where LSTMforward(f),LSTMbackward(f) are the LSTM operations in the forward and backward directions, respectively. Finally, Transformer layers are incorporated to model long-range dependencies more effectively. The Transformer utilizes a multi-head self-attention mechanism, which allows the model to assign different weights to various parts of the sequences during the interaction between SMILES and protein features.

The multi-head self-attention mechanism applies multiple attention layers in parallel (in our model, 4 heads), each learning distinct patterns from the input sequences. The attention weights are calculated by taking the dot product of queries (Q), keys (K), and values (V), where Q, K, and V are derived from the input. This can be expressed as:

$$Attention(Q,K,V)=Softmax(\frac{QK^T}{\sqrt{d_k}})V \quad (5)$$

where d_k is the dimension of the keys (set to 64 in our model). This mechanism allows the model to focus on different parts of the sequences, making it well-suited for tasks like DTBA, where interactions between distant parts of the sequences are important.

Following the attention mechanism, the outputs are passed through a feed-forward neural network, and layer normalization is applied.

The multi-head self-attention applied to the LSTM-processed features H can be expressed as

$$H_{attn}=Multi\_Head\_Self\_Attention\,(H) \quad (6)$$

This is followed by a feed-forward network to process the attended features.

$$H_{ffn}=Feed\_Forward\_Network(H_{attn}) \quad (7)$$

The processed features are then combined and normalized to produce the final feature representation for the prediction task.

To optimize the model, we employed the LAMB optimizer (Layer-wise Adaptive Moments based on Batch size), which is particularly effective in handling large batch sizes. Unlike traditional optimizers, LAMB adjusts the learning rate for each layer individually, taking into account both the gradient magnitude and the layer's weight norm. This enables the model to maintain stable training dynamics, especially in deep networks, while benefiting from faster convergence. The LAMB optimizer's advantages over traditional optimizers like Adam include improved scalability with large batch sizes (128 in our model) and more efficient training in deep architectures. The learning rate was carefully selected to ensure stable and rapid convergence, while a dropout rate of 0.5 was used to mitigate overfitting.

The model was trained for 100 epochs, with Mean Squared Error (MSE) as the loss function and evaluation metric to measure the alignment between predicted and actual binding affinities. The entire workflow is depicted in Algorithm 1 and Fig 3.

Algorithm 1

1. Input: SMILES strings, S Protein sequences P, Labels Y

2. Output: Predicted interaction scores Y'

3. Data Preparation

3.1. Shuffle the dataset, D containing triplets S, P, Y

3.2. Split the data into training and test sets:

$$train\_smiles,test\_smiles \leftarrow S$$
$$train\_proteins,test\_proteins \leftarrow P$$

4. Tokenization and Padding

4.1. Initialize a character-level tokenizer, $\boldsymbol{Tokenizer_{SMILES}}$

4.2. Fit the tokenizer on the training SMILES strings $\{Si\}_{i\in D}$

4.3. Convert the SMILES strings to sequences of tokens, $Seq_{Smiles} = \boldsymbol{Tokenizer_{SMILES}}(S)$

4.4. Pad the sequences to a maximum length, $\boldsymbol{L_{SMILES}}$

$$Padded_{Smiles} = Pad(Seq_{Smiles}, \boldsymbol{L_{SMILES}})$$

4.5. similarly padding for proteins

$$Padded_{Protein} = Pad(Seq_{Protein}, \boldsymbol{L_{Protein}})$$

5. Define two input layers S_{input} for SMILES and P_{input} for protein sequences.

6. Encode SMILES and proteins using convolutional layers followed by LSTM layers.

$$E_{Smiles} = \boldsymbol{Embedding}(\boldsymbol{S_{input}})$$
$$C_{Smiles} = \boldsymbol{Conv_{1D}}(\boldsymbol{E_{Smiles}})$$
$$H_{Smiles}^{LSTM} = \boldsymbol{LSTM}(\boldsymbol{C_{Smiles}})$$

7. Apply transformer blocks with multi-head attention and LSTM layers.

$$H_{Smiles}^{Trans} = \boldsymbol{MultiHeadAttention}(H_{Smiles}^{LSTM})$$
$$H_{Smiles} = LayerNormalization(H_{Smiles}^{LSTM}, H_{Smiles}^{Trans})$$
$$\boldsymbol{F_{Smiles}} = \boldsymbol{LSTM}(\boldsymbol{H_{smiles}})$$

8. Similar steps 6 and 7 to be followed for the protein encoder.

$$\boldsymbol{F_{Proteins}} = \boldsymbol{LSTM}(\boldsymbol{H_{Proteins}})$$

9. Concatenate the outputs of the SMILES and protein encoders.

$$\boldsymbol{F_{combined}} = Concatenate(\boldsymbol{F_{Smiles}}, \boldsymbol{F_{Protein}})$$

10. Output the final interaction score, y'.

4. RESULTS AND DISCUSSION

4.1. EVALUATION METRICS

The performance of the model is evaluated using metrics such as accuracy, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R-squared (R2), Area Under the Precision-Recall Curve (AUPR), and Concordance Index (CI). These metrics provide a comprehensive assessment of the model's ability to predict drug-target binding affinities accurately.

Mean Squared Error (MSE) measures the average squared difference between the actual and predicted values. It is a fundamental metric for regression problems, providing a clear indication of the model's prediction error.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \quad (8)$$

A lower MSE indicates better model performance, with an ideal value of 0, which would mean no error. High MSE values suggest that the model is not capturing the underlying trend in the data. The Concordance Index (CI) measures the agreement between predicted and actual rankings, commonly used in survival analysis and similar tasks. CI provides an evaluation of the ranking accuracy of predictions. CI values range from 0.5 (random chance) to 1 (perfect prediction). Values above 0.7 are generally considered good. Higher CI values indicate better model performance in terms of ranking predictions correctly.

4.2. PERFORMANCE OF THE PROPOSED FUSIONNET MODEL

The FusionNet model demonstrates strong performance in predicting drug-target interactions on both the Kiba and Davis datasets. The Kiba dataset, achieved a training loss and MSE of 0.1874, with validation loss and MSE of 0.1852, indicating good generalization without overfitting. On the Davis dataset, the model had a training loss and MSE of 0.2073, and a validation loss and MSE of 0.2296, showing slight deviations but maintaining strong performance.

The system architecture shown in Fig. 3. displays MSE over 100 epochs, where the training MSE drops sharply in the first 10 epochs, and the validation MSE stabilizes, demonstrating effective learning and good generalization. Fig. 4. depicts "Actual vs. Predicted Values" and "Residual Plots." For the Davis dataset, predicted values show more spread, and residuals indicate areas of inaccuracy, whereas, for Kiba, predictions are tightly clustered with residuals randomly distributed around zero, reflecting better accuracy. Overall, the model performs more accurately on the Kiba dataset.

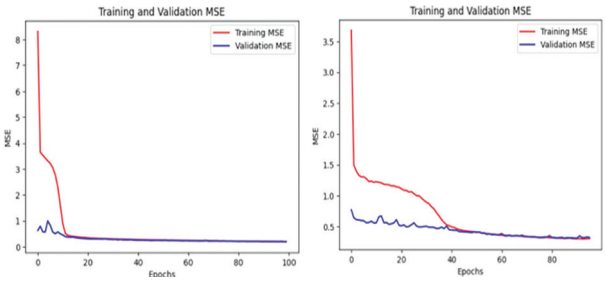


Fig. 3. Training and Validation MSE for Kiba and Davis Dataset

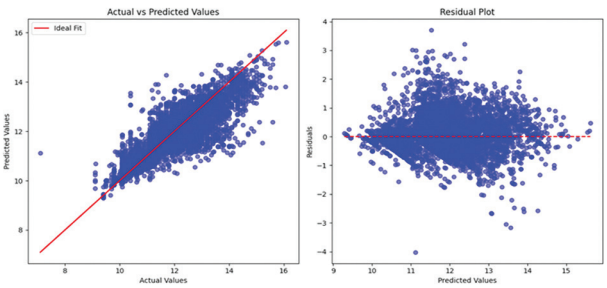


Fig. 4. Actual vs. Predicted Values and Residual Plots for Kiba dataset

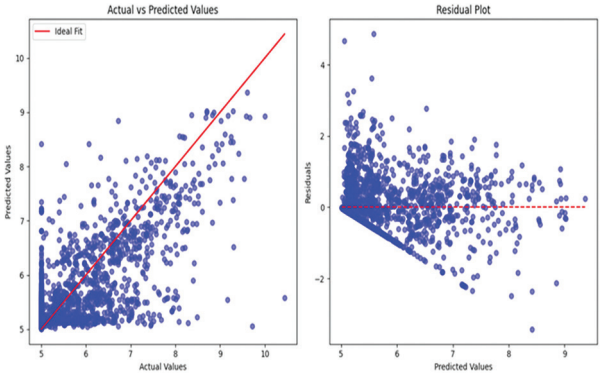


Fig. 5. Actual vs. Predicted Values and Residual Plots for Davis dataset for Davis dataset

Fig. 5 and Fig. 6 show two scatter plots of predicted vs. actual values for two models. The Kiba plot, with an R-squared of 0.710, indicates that 71% of the variance in actual values is explained by the model, showing a strong correlation and good predictive accuracy. The Davis plot, with an R-squared of 0.681, explains 68.1% of the variance, indicating a weaker correlation and lower predictive accuracy.

Fig. 7 displays two precision-recall curves for two models. The precision-recall curves for kiba, have an average precision (AP) of 0.80, showing a model that maintains high precision across various recall levels but drops off as recall approaches 1.0. The precision-recall curves for the Davis curve, have an AP of 0.86, indicating better overall performance, with higher precision maintained over a broader range of recall values.

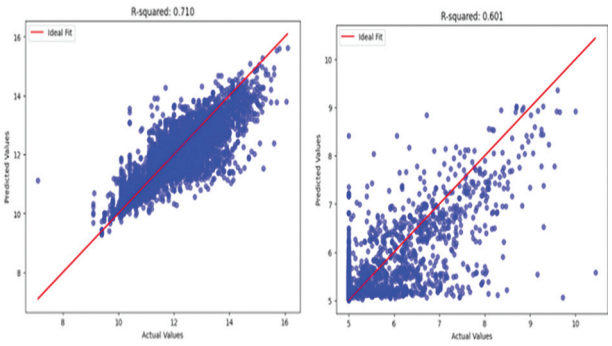


Fig. 6. Actual vs. Predicted Values for Kiba and Davis dataset

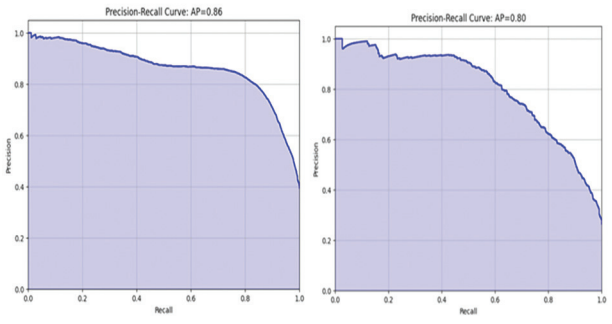


Fig. 7. Precision-recall curves for the Kiba and Davis dataset

Fig. 8 shows an AUC (Area Under the Curve) of 0.9 for the Davis dataset, indicating a high level of model performance with a good balance between sensitivity and specificity. The Fig. 13 shows an AUC of 0.91, which is slightly better, suggesting even better performance. Both models significantly outperform random guessing (represented by the red dashed line).

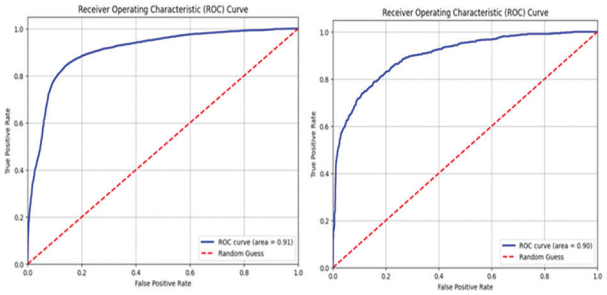


Fig. 8. ROC curve for Kiba and Davis dataset

The Protein Attention Map and SMILES Attention Map as in Fig. 9 provide valuable insights into the model's decision-making process by visualizing how attention is distributed across protein sequences and molecular structures, respectively. In the Protein Attention Map, each vertical strip represents specific positions in the protein sequence, with varying color intensities indicating the attention values assigned by the model. This helps capture important local and long-range dependencies essential for accurate drug-target binding predictions.

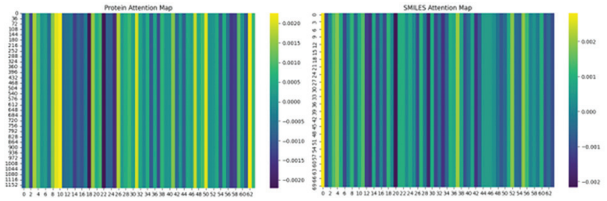


Fig. 9. Attention map for protein and smile sequences

Similarly, the SMILES Attention Map highlights attention distribution across molecular sequences represented in SMILES notation, where each vertical line corresponds to a token (atom or bond) in the molecule. The intensity reflects the importance given by the model to different parts of the molecular structure. These visualizations enhance interpretability, showing how the model focuses on crucial areas in both protein and molecular data, thereby improving transparency in the prediction process.

4.2. COMPARATIVE ANALYSIS OF VARIOUS APPROACHES

In this paper, the proposed FusionNet model has been compared to previous state-of-the-art approaches like SimBoost[21], GANsDTA[22], DeepCDA[23], and DeepDTA[11] using mean squared error (MSE), root mean squared error (RMSE), concordance index (CI),

area under the precision-recall curve (AUPR), and rm2 evaluation metrics, as shown in Table 2. In the case of the Davis dataset, the proposed model has obtained an MSE of 0.20 and an rm2 value of 0.681. This result showcases a clear improvement over existing methods by offering the lowest MSE and the highest rm2, indicating its ability to make more accurate predictions. In more detail, compared to the SimBoost, GANsDTA, DeepCDA, and DeepDTA models with MSE values of 0.28, 0.27, 0.24, and 0.26, respectively, FusionNet demonstrates a higher prediction accuracy due to its ability to capture both local and long-range dependencies within drug-target sequences. Additionally, the rm2 values of these methods lag, further highlighting the efficacy of our approach.

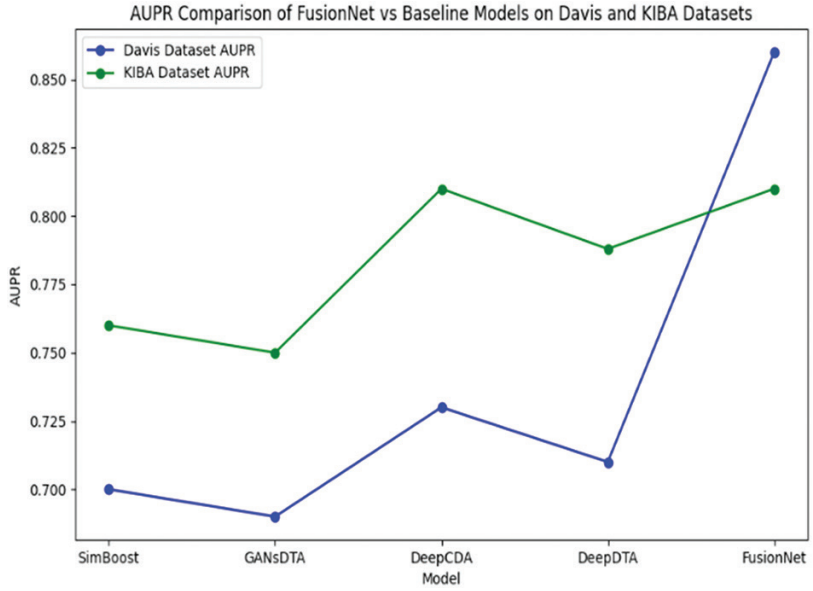


Fig 10. AUPR comparison graph of FusionNet's Vs baseline models on the Davis and KIBA datasets

Similarly, on the KIBA dataset, FusionNet outperforms existing methods, achieving an MSE of 0.18 and an rm2 of 0.71. The Transformer component allows FusionNet to model complex drug-target interactions, leading to more biologically meaningful predictions. For instance, while SimBoost, GANsDTA, DeepCDA, and DeepDTA returned MSE values of 0.22, 0.22, 0.17, and 0.19 respectively, FusionNet's multi-head attention mechanism captures intricate patterns across SMILES and protein sequences, improving generalization. Other metrics, such as root mean squared error and concordance index, reflect this model's robust performance across datasets. Figure 10 provides a side-by-side AUPR comparison graph, highlighting FusionNet's performance against baseline models on Davis and KIBA datasets.

The biological implication is substantial, as more accurate drug-target affinity predictions enhance early-stage drug discovery by efficiently identifying potential drug candidates. In summary, FusionNet's superior accuracy and generalizability make it highly promising for accelerating therapeutic discovery.

Table 2. Comparative analysis of various approaches

Datasets	Methods	MSE	RMSE	CI	AUPR	rm2
Davis	SimBoost	0.28	-	0.83	0.70	0.644
	GANsDTA	0.27	-	0.70	0.69	0.653
	DeepCDA	0.24	-	0.89	0.s73	0.649
	DeepDTA	0.26	-	0.87	0.71	0.67
	Proposed (FusionNet)	**0.20**	**0.55**	**0.89**	**0.86**	**0.681**
KIBA	SimBoost	0.22	-	0.83	0.76	0.629
	GANsDTA	0.22	-	0.86	0.75	0.675
	DeepCDA	0.17	-	0.88	0.81	0.682
	DeepDTA	0.19	-	0.86	0.788	0.673
	Proposed (FusionNet)	**0.18**	**0.40**	**0.88**	**0.81**	**0.71**

5. CONCLUSION

This work presents a new deep-learning framework, called FusionNet, which substantially improves the prediction of drug-target binding affinity. By combining Convolutional Neural Networks, Long Short-Term Memory networks, and transformers, it effectively represents local patterns, sequential dependencies, and long-range interactions in biological sequences. The model's novelty lies in its hybrid architecture, which leverages the strengths of each component to achieve enhanced robustness and accuracy in prediction. It also integrates the Layer-wise Adaptive Moments optimizer to improve training efficiency and performance, making the model scalable and reliable even for large datasets. Extensive tests were conducted with FusionNet on both the Davis and KIBA datasets, yielding state-of-the-art results. FusionNet achieved an MSE of 0.20 and an rm2 of 0.681 on the Davis dataset, and an MSE of 0.18 and an rm2 of 0.71 on the KIBA dataset. These results highlight FusionNet's superior performance, surpassing existing methods like SimBoost, GANsDTA, and DeepDTA. Unlike traditional methods that often focus on either local or global context, FusionNet addresses both simultaneously.

Biologically, this enhanced accuracy could streamline drug discovery pipelines by enabling more precise identification of potential drug-target pairs earlier in the development process, potentially reducing costs and timeframes. Additionally, the LAMB optimizer not only accelerates the training process but also enhances stability and performance on large datasets, underscoring FusionNet's scalability and robustness. Key takeaways include the hybrid model's capability to capture both local and global features, leading to improved predictive accuracy, and the model's scalability for larger datasets, which is essential for real-world applications. However, a notable limitation is FusionNet's reliance on sequence-based features without considering structural or genomic data, which might offer further insights into drug-target interactions. Future work could address this by expanding the dataset to include more diverse biological sequences and incorporating additional data such as 3D protein structures or genomic information. Scaling the model to larger datasets or alternative binding data types could present challenges that need further exploration. Another avenue could involve the use of transfer learning to leverage pre-trained models for related tasks. Finally, enhancing interpretability, such as linking attention to biologically relevant protein or molecular regions, could provide deeper insights into the biological mechanisms underlying drug-target interactions, increasing the model's value in biomedical research. This will not only establish FusionNet as a robust predictor but also enrich our understanding of molecular biology.

ACKNOWLEDGMENTS

The authors would like to thank all anonymous reviewers for their constructive advice.

FUNDING

This research received no external funding

6. REFERENCES

[1] B. Rupp, "Biomolecular crystallography: principles, practice, and application to structural biology", Garland Science, 2009.

[2] R. L. Rich, D. G. Myszka, "Survey of the year 2005 commercial optical biosensor literature", Journal of Molecular Recognition: An Interdisciplinary Journal, Vol. 19, No. 6, 2006, pp. 478-534.

[3] A. Velazquez-Campoy, E. Freire, "Isothermal titration calorimetry to determine association constants for high-affinity ligands", Nature Protocols, Vol. 1, No. 1, 2006, pp. 186-191.

[4] L. Breiman, "Random forests", Machine Learning, Vol. 45, 2001, pp. 5-32.

[5] D. Rogers, M. Hahn, "Extended-connectivity fingerprints", Journal of Chemical Information and Modeling, Vol. 50, No. 5, 2010, pp. 742-754.

[6] H. Luo, Y. Xiang, X. Fang, W. Lin, F. Wang, H. Wu, H. Wang, "BatchDTA: implicit batch alignment enhances deep learning-based drug–target affinity estimation", Briefings in Bioinformatics, Vol. 23, No. 4, 2022.

[7] T. Cheng, Q. Li, Z. Zhou, Y. Wang, S. H. Bryant, "Structure-based virtual screening for drug discovery: a problem-centric review", The AAPS Journal, Vol. 14, 2012, pp. 133-141.

[8] M. Ragoza, J. Hochuli, E. Idrobo, J. Sunseri, D. R. Koes, "Protein-ligand scoring with convolutional neural networks", Journal of Chemical Information and Modeling, Vol. 57, No. 4, pp. 942-957, 2017.

[9] Y. You, I. Gitman, B. Ginsburg, "Large batch training of convolutional networks", arXiv:1708.03888, 2017.

[10] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention is all you need", Advances in Neural Information Processing Systems, Vol. 30, 2017.

[11] H. Öztürk, A. Özgür, E. Ozkirimli, "Deepdta: deep drug-target binding affinity prediction", Bioinformatics, Vol. 34, No. 17, 2018, pp. i821-i829.

[12] Q. Feng, E. Dueva, A. Cherkasov, M. Ester, "Padme: A deep learning-based framework for drug-target interaction prediction", arXiv:1807.09741, 2018.

[13] H. Öztürk, E. Ozkirimli, A. Özgür, "Widedta: prediction of drug-target binding affinity", arXiv:1902.04166, 2019.

[14] Y. Zeng, X. Chen, Y. Luo, X. Li, D. Peng, "Deep drug-target binding affinity prediction with multiple attention blocks", Briefings in Bioinformatics, Vol. 22, No. 5, 2021.

[15] I. Elansary, W. Hamdy, A. Darwish, H. Aboul, "Bat-inspired optimizer for prediction of anti-viral cure drug of Sars-Cov-2 based on recurrent neural network", Journal of System and Management Sciences, Vol. 10, No. 2020, pp. 20-34.

[16] E. K. Makowski et al. "Co-optimization of therapeutic antibody affinity and specificity using machine learning models that generalize to novel mutational space", Nature Communications, Vol. 13, No. 1, 2022, p. 3788.

[17] S. Min, B. Lee, S. Yoon, "Deep learning in bioinformatics", Briefings in Bioinformatics, Vol. 18, No. 5, 2017, pp. 851-869.

[18] M. I. Davis, J. P. Hunt, S. Herrgard Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, C.-J. Hsieh, "Large batch optimization for deep learning: Training Bert in 76 minutes", arXiv:1904.00962, 2019.

[19] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization", arXiv:1412.6980, 2014.

[20] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, P. P. Zarrinkar, "Comprehensive analysis of kinase inhibitor selectivity", Nature Biotechnology, Vol. 29, No. 11, 2011, pp. 1046-1051.

[21] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, M. Ester, "SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines", Journal of Cheminformatics, Vol. 9, No. 1, 2017, pp. 1-14.

[22] T. Song, X. Zhang, M. Ding, A. Rodriguez-Paton, S. Wang, G. Wang, "DeepFusion: A deep learning based multi-scale feature fusion method for predicting drug-target interactions", Methods, Vol. 204, 2022, pp. 269-277.

[23] K. Abbasi, P. Razzaghi, A. Poso, M. Amanlou, J. B. Ghasemi, A. Masoudi-Nejad. "DeepCDA: deep cross-domain compound-protein affinity prediction through LSTM and convolutional neural networks", Bioinformatics, Vol. 36, No. 17, 2020, pp. 4633-4642.

An Improved MPPT Scheme for Photovoltaic Systems Using a Novel MRAC-FUZZY Controller

Original Scientific Paper

Mahbouba Brahmi*
Laboratory of Industrial Systems Engineering and Renewable Energy (LISIER), Higher National Engineering School of Tunis (ENSIT), University of Tunis, 5 Av.Taha Hussein,5 BP 56, 1008 Tunis, Tunisia.
mahboubabrahmi@gmail.com

Afef Marai Ghanmi
Laboratory of Industrial Systems Engineering and Renewable Energy (LISIER), Higher National Engineering School of Tunis (ENSIT), University of Tunis, 5 Av.Taha Hussein,5 BP 56, 1008 Tunis, Tunisia.
afef.marai@yahoo.com

Hichem Hamdi
Laboratory of Industrial Systems Engineering and Renewable Energy (LISIER), Higher National Engineering School of Tunis (ENSIT), University of Tunis, 5 Av.Taha Hussein,5 BP 56, 1008 Tunis, Tunisia.
hichem.hamdi@gmail.com

Chiheb Ben Regaya
Laboratory of Industrial Systems Engineering and Renewable Energy (LISIER), Higher National Engineering School of Tunis (ENSIT), University of Tunis, 5 Av.Taha Hussein,5 BP 56, 1008 Tunis, Tunisia.
chihebbenregaya@issatkr.u-kairouan.tn

Abderrahmen Zaafouri
Laboratory of Industrial Systems Engineering and Renewable Energy (LISIER), Higher National Engineering School of Tunis (ENSIT), University of Tunis, 5 Av.Taha Hussein,5 BP 56, 1008 Tunis, Tunisia.
rhaeim.zaafouri@gmail.com

*Corresponding author

Abstract *– This paper presents a novel and highly effective fuzzy model reference adaptive control for MPPT based on a boost converter. The design of Model-Referenced Adaptive Control (MRAC) and the adaptive gain selection are discussed. The adjustment of the adaptation gains by a fuzzy logic subsystem and a simplified fuzzy MRAC procedure are presented. The suggested algorithm is assessed through a comprehensive simulation in MATLAB/Simulink. Various scenarios and environmental conditions are considered to assess its robustness and adaptability. The results indicate that the suggested MRAC-Fuzzy MPPT control is extremely robust, with tracking efficiency that can reach 99.97%. Furthermore, it consistently operates the photovoltaic system at or around the MPP, effectively reducing oscillations, improving energy efficiency, and enhancing power production. Under real operating conditions, this new controller can be used for photovoltaic pumping applications.*

Keywords: *Photovoltaic System, MPPT, Fuzzy Model Reference Adaptive Control, MIT Rule*

Received: August 31, 2024; Received in revised form: December 11, 2024; Accepted: February 5, 2025

1. INTRODUCTION

Energy production represents a significant challenge for the years ahead. Furthermore, developing countries will increasingly require energy to enhance their development. Energy sources can be divided roughly into two groups: renewable and non-renewable energies. The first category includes sources such as wind, hydro, waste, biomass, geothermal energy and solar, while the second category comprises uranium, gas, oil, coal, and similar resources. Consumption of fossil fuels from the second category results in greenhouse gas emissions, leading to an increase in atmospheric pollution. Additionally, these fossil fuel resources are exhaustible [1, 2]. Consequently, numerous countries are dedicated to exploring alternative, sustainable, and profitable renewable energy sources. In contrast to fossil fuels, these new energy sources are non-polluting, emitting no greenhouse gases, and pose no inherent danger [3]. The use and advancement of photovoltaic energy are on the rise globally. One of the most promising applications of this renewable energy source is photovoltaic pumping,

which is particularly beneficial in rural areas with high levels of sunlight and no access to an electric grid [4]. Solar energy refers to the portion of electromagnetic energy extracted by photosensitive cells from the radiation emitted by the sun. It involves converting electromagnetic radiation into electricity through the photovoltaic effect. Based on the photovoltaic cells' electrical properties and their configuration, the efficiency of PV systems can be enhanced through techniques known as Maximum Power Point Tracking (MPPT) [5].

Indeed, numerous research studies on MPPT algorithms have been conducted and documented in the literature. Traditional methods like the Incremental Conductance (INC) algorithm [6] and the Perturb and Observe (P&O) algorithm [7] are among the most commonly used techniques. In spite of their simple and functional design, conventional approaches were only adept to follow the maximum power point (MPP) when weather conditions stayed constant. Moreover, conventional MPPT algorithms often demonstrate ripples in the vicinity of the MPP and may not be optimal for extensive solar power installations. Given these limitations mentioned above, researchers worldwide are actively devising innovative approaches to MPPT control in solar systems. Advanced MPPT techniques, including heuristic approaches like genetic algorithm, Fuzzy Logic Control (FL) [8], Particle Swarm Optimization (PSO) [9, 10], and Artificial Neural Network (ANN) [11], are widely used as some of the most prevalent enhanced MPPT control techniques, which guarantee a remarkable capability to track the MPP. MPPT methodology based on soft computing is widely regarded as one of the most powerful approaches for addressing nonlinear problems. The realm of literature abounds with research endeavors focused on enhancing existing methodologies and surmounting their inherent constraints. [12] have presented a novel method called perturbation and observation approach, which has been optimized using Neural Network (NN) technology to achieve MPPT. To validate the effectiveness of this system, simulation tests were conducted, considering various solar radiation levels. The findings of this research suggest that the approach excels in varying light intensities and temperature, the P&O approach optimized by NN is more efficient than traditional INC approaches. This controller demonstrates the ability to generate approximately 99% of the real maximum power. In contrast to the Incremental Conductance approach, which requires approximately 0.3 seconds to attain the reference value, the NN method requires approximately 0.025 seconds to execute, exhibiting minimal overshoot. An efficient and rapid method for MPPT has been devised by employing FL without the need for an expert to construct the membership functions. Using MATLAB, the methodology is put into practice and its effectiveness is assessed by analyzing the results obtained, fuzzy logic significantly outperforms ANN optimized with PSO, ANN-GA (Genetic Algorithm), and ANN-ICA (Imperialist Competitive Algorithm) in terms of stability, precision, rapidity and simplicity of installation in the face of environmental fluctuations, as reported by [13]. In order to tackle the issue of chattering, a novel super-twist sliding-mode controller was proposed and integrated into the system. Additionally, a Type 2 fuzzy set (STSMC-T2FC) was employed to further enhance the performance of the system. The algorithm proposed has been developed using MATLAB and then assessed against both STSMC and traditional SMC methodologies across different radiation scenarios. The efficiency of the STSMC-T2FC MPPT stands at 99.59%, surpassing both STSMC with 99.33% and SMC with 99.20%. Despite the closely matched efficiency performances, STSMC-T2FC emerged as the superior choice, as indicated by [14]. A two-stage global MPPT control mechanism has been proposed to guarantee the utilization of all power generated by the PV for the load [15]. The initial stage employs global perturbation-based extremum seeking control (GPESC) to pinpoint the global Maximum Power Point (MPP). The second stage involves Model Reference Adaptive Control (MRAC), which is utilized to regulate the dynamics of the DC-DC converter. The simulation evaluates the effectiveness of the suggested controller in terms of tracking speed, efficiency, and accuracy under different radiation conditions. The GPESE and GPESC-PID controllers are utilized for comparative analysis. A newly developed high-frequency learning-based adjustable gain Model Reference Adaptive Control (HFLAG-MRAC) system, as proposed by [16], designed for a two-level MPPT control structure in PV systems. This approach aims to optimize power distribution to the load, particularly in the face of rapidly changing environmental conditions. The adaptive principle for the HFLAG-MRAC is formulated through Lyapunov theory, ensuring that the control system is theoretically robust and stable. However, there are several efforts still to be resolved in order to enhance the effectiveness of MPPT control. These efforts involve reducing response times, monitoring MPP, optimizing design parameters, attenuating steady-state oscillations, minimizing the sensor costs involved, and simplifying complexity.

Another issue is the aleatory behavior of optimization approaches in one-shot design methods, with MPPT MRAC control, system performance is affected by the adaptation gain of the adjustment mechanism: a high value of adaptation gain can cause system instability. This implies that the adaptation gain should be selected optimally to minimize this problem. In this context, a novel adaptive MPPT controller have been proposed, whose adaptation gain is defined heuristically using an adequate heuristic method for setting the adaptation gain based on fuzzy logic.

The main contributions of the current study are described as follow:

- A new Fuzzy Model Reference Adaptive controller (MRAC-FUZZY) is suggested for photovoltaic systems in order to obtain an optimal MPP.
- The proposed algorithm reduces complexity by minimizing the adaptation equations mechanism and subsequently the controller.

- MRAC-FUZZY MPPT is intended to provide an adaptive control strategy that optimizes the power output of photovoltaic systems by dynamically adjusting control parameters to track the MPP under varying environmental conditions while maintaining stability and eliminating ripples. This leads to increased energy efficiency and improved performance of PV systems.
- The proposed algorithm reduces response time: by approximately 11 times, 5 times and 2 times, faster than P&O, FL and PSO respectively.
- A comparative study involving simulation is conducted to assess the efficiency of the suggested MPPT controller.

This research is structured as outlined below: In Section 2, we present the photovoltaic system's mathematical model. Section 3 outlines the procedure design of the proposed MRAC-Fuzzy MPPT control algorithm. Section 4 illustrates the simulation outcomes regarding the performance of the photovoltaic systems. These results are obtained by implementing the MRAC-Fuzzy MPPT control algorithm, which is proposed in this study. Furthermore, a comparison is made between the performance of this algorithm and three conventional controllers namely, "MPPT PSO," "MPPT Fuzzy Logic," and "MPPT P&O." Additionally, we evaluate the performance of each MPP controller in comparison to the proposed MRAC-Fuzzy MPPT algorithm. Finally, we conclude with some remarks and a summary of our findings.

2. MODELING OF THE PHOTOVOLTAIC SYSTEM

2.1. PHOTOVOLTAIC CELL MODELING

The typical composition of a photovoltaic cell consists of components depicted in Fig. 1, comprising a current generator, a diode, and a combination of resistors connected in series and parallel [17].

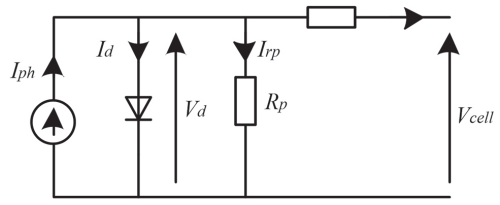


Fig. 1. PV circuit equivalent model

The equation below can be used to calculate the current delivered by a solar panel.

$$I_{cell} = I_{ph} - I_d[exp(\mu) - 1] - \frac{V_{pv} + R_s I_{pv}}{R_p} \quad (1)$$

With

$$\mu = \frac{V_{PV} + R_s I_{pv}}{N_s \frac{\mathrm{Ak_b}}{\mathrm{e}} T} \quad (2)$$

$$I_{ph} = \left(I_{sc} + \mathrm{k_i}(T - 270)\right)\frac{G}{1000} \quad (3)$$

$$I_d = I_0\left(\frac{T}{298}\right)^3 \exp\left(\frac{\mathrm{qE_q}}{\mathrm{k_b} V_t}\left(\frac{1}{298} - \frac{1}{T}\right)\right) \quad (4)$$

Icell (I_{pv}) and V_{pv} represent the output current and voltage of the PV system, respectively, I_{ph} is the photocurrent, Id designates the diode current, I_0 is the inverse saturation current. The short-circuit current is designated by I_{sc}, the series resistance is indicated by R_s, the parallel resistance is noted as R_p. k_i is the temperature coefficient of the short circuit current. T, q, kb, and A correspond to the temperature, the electronic charge, the Boltzmann constant, and the diode factor respectively [18]. T is the temperature of Solar Cells and G is the irradiance and the irradiance reference (kW/m^2).

2.2. BOOST CONVERTER

Fig. 2 depicts the DC-DC converter responsible for optimizing the power transfer from the PV array to the load. This converter plays a crucial role in MPPT by dynamically adjusting the voltage and current levels between the PV source and the load. This allows the system to operate at the Maximum Power Point (MPP) of the PV array, maximizing power extraction under varying environmental conditions.

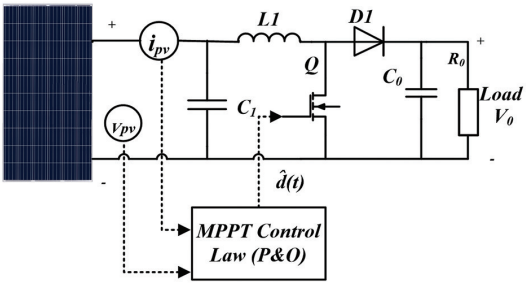


Fig. 2. Photovoltaic system schematic diagram

The fundamental connection between converter duty cycle and mains voltage is given by:

$$V_{pv} = i_{pv} R_0 (1 - d)^2$$

It is essential to consider the interaction among the duty cycle and the grid voltage in MPPT control to improve the transitional response [19]. To facilitate the analysis of the system's transient behavior, we examine a small equivalent signal similar to the one depicted in Fig. 2, as illustrated in Fig. 3.

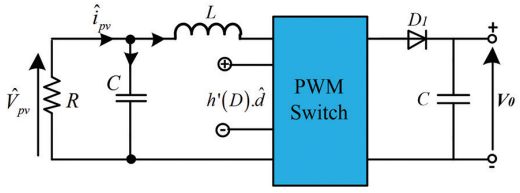


Fig. 3. Small-signal model of the photovoltaic output converter circuit

The duty cycle transfer function at mains voltage in small signal mode is calculated using an operating

point [20]. As shown in Fig. 4, the equation between the mains voltage $\hat{V}_{pv}$ and the variation of the small pulse around the duty cycle $\hat{d}$ of the inverter can be determined in the Laplace domain as below:

$$\frac{\hat{V}_{pv}(s)}{R} + \hat{V}_{pv}(s)Cs = \frac{h'(D)\hat{d}(s) - \hat{V}_{pv}(s)}{sL} \quad (6)$$

with $\hat{d}(s)$ denoting the smaller signal variation near the duty cycle D, $h(d)$ represents the relationship among V_{pv} and D. $h'(D)$ the derivative of $h(D)$. According to Eq. (6), we get:

$$\frac{\hat{V}_{pv}(s)}{\hat{d}(s)} = \frac{h'(D)}{LCs^2 + \frac{L}{R}s + 1} \quad (7)$$

As defined above, $h(D)$ can be written as the following equation:

$$h(D) = V_{PV} = (1 - D)V_0 \quad (8)$$

V_0 represents the output of the boost converter. We simply derive $h(d)$ with regard to the duty cycle D, we obtain:

$$h'(D) = -V_0 \quad (9)$$

The output of the boost converter in a steady-state condition, designated as V_0, is represented by Eq. (8) under the assumption that transient switching behavior does not influence $h(d)$ and V_0. This leads to the derivation of $h'(D) = -V_0$. Consequently, Eq. (7) will be formulated as follows:

$$\frac{\hat{V}_{pv}(s)}{\hat{d}(s)} = \frac{\frac{-V_0}{LC}}{s^2 + \frac{1}{RC}s + \frac{1}{LC}} \quad (10)$$

3. MAXIMUM POWER POINT TRACKER DESIGN

This section introduces the concept of the MRAC-FUZZY, which aims to optimize the power generated by a photovoltaic array. The complete framework of the suggested control methodology is depicted in Fig. 4.

The proposed algorithm consists of two levels. First level presents an MPPT control law based on P&O technique, as shown in Fig. 5. A voltage reference (vref) is set by this control block for any given MPP voltage. In the second level, a proposed MRAC-Fuzzy MPPT controller is developed which is illustrated in Fig. 6.

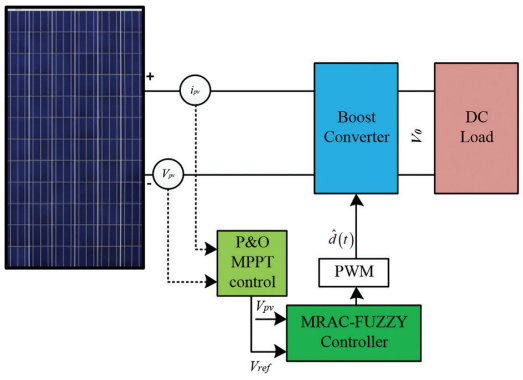


Fig. 4. Photovoltaic system with the suggested MPPT control configuration

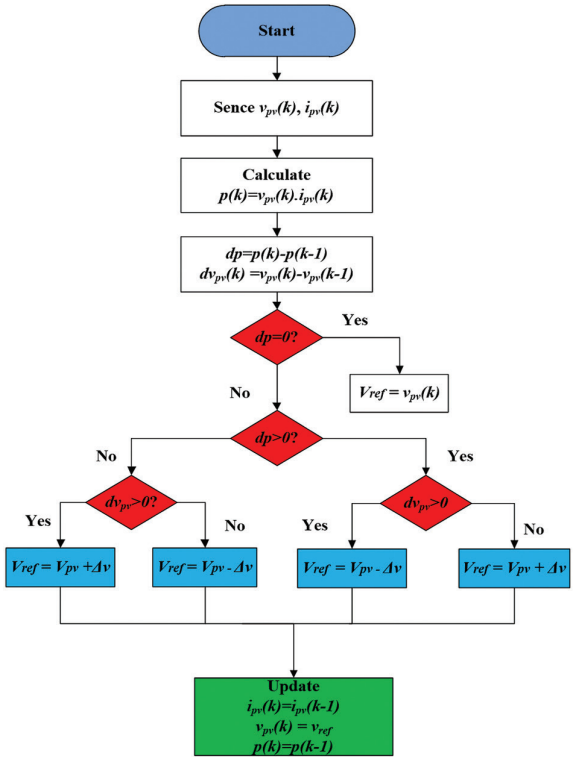


Fig. 5. Voltage Setpoint calculation

The new adaptive MPPT controller has only two inputs: voltage array and reference voltage. Its architecture is based on a reference model, a plant, and an adaptation gain (γ) as shown in Fig. 6.

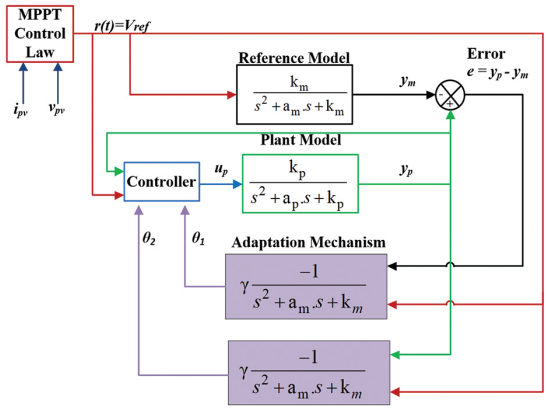


Fig. 6. Proposed MRAC controller architecture

The aim of the MRAC is to ensure that the output of the plant produces the output of the reference model by using γ. To achieve optimal output measurement, it is crucial to select a reference model as the initial step in implementing an MRAC, Additionally, a controller should be formulated to minimize the error (e) among reference and plant value. The Massachusetts Institute of Technology (MIT) law, which employs a gradient strategy, is considered one of the most fundamental adaptive approaches. Developed at MIT in the 1960s for aerospace uses, the MRAC controller enhances this technique by adjusting the adaptation laws to minimize the disparity among the reference and system output.

Conventional MRAC feedback is not sufficient for second-order systems. The second-order system control law and the first to second-order extension of MRAC-FUZZY are described in this section.

The plant model is defined with the following equation:

$$\frac{d^2y_p(t)}{dt^2} = -\mathrm{a_p}\frac{dy_p(t)}{dt} - \mathrm{b_p}y_p(t) + \mathrm{k_p}u_p(t) \tag{11}$$

$$G_p(s) = \frac{y_p(s)}{u_p(s)} = \frac{\mathrm{k_p}}{s^2 + \mathrm{a_p}s + \mathrm{b_p}} \tag{12}$$

a_p, b_p and k_p are plant coefficients which can be determined using equation (9). The reference model has been specially adapted to define the required output $y_m(t)$ for the input $r(t)$ in the next equation.

$$\frac{d^2y_m(t)}{dt^2} = -\mathrm{a_m}\frac{dy_m(t)}{dt} - \mathrm{b_m}y_m(t) + \mathrm{k_m}r(t) \tag{13}$$

$$G_m(s) = \frac{y_m(s)}{r(s)} = \frac{\mathrm{k_m}}{s^2 + \mathrm{a_m}s + \mathrm{b_m}} \tag{14}$$

k_m exhibits a positive gain, a_m and b_m are determined to ensure that the reference model produces a step response that is critically attenuated. The purpose of the control system is to create $y_p(t)$ in a manner that the later proceeds $y_m(t)$.

The MIT rule (The MIT law employs a gradient approach in its implementation) is applied to build the adaptation law of the controller parameters for the MRAC. Using the MIT rule, the cost function is given by:

$$J(\theta) = \frac{e^2}{2} \tag{15}$$

$$e = y_p - y_m \tag{16}$$

Where e represents the error between the plant and the reference model, θ is an adjustable control parameter. Based on the MIT rule, we can write:

$$\frac{d\theta}{dt} = -\gamma\frac{\delta J}{\delta\theta} = -\gamma e\frac{\delta e}{\delta\theta} \tag{17}$$

Where γ is an adaptation gain.

In the proposed algorithm, the equation (17) is used for the controller. In contrast to the MRAC developed by [21], we have streamlined the equations governing the adaptation mechanism and subsequently optimized the controller.

$$u_p = \theta_1 r - \theta_2 y_p = \theta^T\varphi \tag{18}$$

With φ expressed as $[r, y_p]\,T$ and $[\theta_1, \theta_2]\,T$ denotes the estimation vector of the controller variables. Replacing Eq. (18) with Eq. (11), we obtain:

$$\frac{d^2y_p(t)}{dt^2} = -\mathrm{a_p}\frac{dy_p(t)}{dt} - (\mathrm{b_p} + \mathrm{k_p}\theta_2)y_p(t) + \mathrm{k_p}\theta_1 r(t) \tag{19}$$

Based on Eqs.13 and 19, we can get:

$$\mathrm{k_p}\theta_1 = \mathrm{k_m} \tag{20}$$

$$\mathrm{b_p} + \mathrm{k_p}\theta_2 = \mathrm{b_m} \tag{21}$$

$$\mathrm{a_p} = \mathrm{a_m} \tag{22}$$

θ_1, θ_2 converge as follows:

$$\theta_1 = \frac{\mathrm{k_m}}{\mathrm{k_p}} \tag{23}$$

$$\theta_2 = \frac{\mathrm{b_m} - \mathrm{b_p}}{\mathrm{k_p}} \tag{24}$$

Using Laplace, the Eq. 23 become:

$$\frac{y_p(s)}{r(s)} = \frac{\mathrm{k_p}\theta_1}{s^2 + \mathrm{a_p}s + (\mathrm{b_p} + \mathrm{k_p}\theta_2)} \tag{25}$$

According to the error in Eq. (16), we can define:

$$e = \left(\frac{\mathrm{k_p}\theta_1}{s^2 + \mathrm{a_p}s + (\mathrm{b_p} + \mathrm{k_p}\theta_1)} - \frac{\mathrm{k_m}}{s^2 + \mathrm{a_m}s + \mathrm{b_m}}\right)r(s) \tag{26}$$

In order to determine the derivatives of sensitivity ($\delta e/\delta\theta_1$ and $\delta e/\delta\theta_2$), and with the use of Eqs.17 and 26, we obtain:

$$\frac{\delta e}{\delta\theta_1} = \frac{\mathrm{k_p}r}{s^2 + \mathrm{a_p}s + \mathrm{b_p} + \mathrm{k_p}\theta_2} \tag{27}$$

$$\frac{\delta e}{\delta\theta_2} = \frac{-\mathrm{k_p}y_p}{s^2 + \mathrm{a_p}s + \mathrm{b_p} + \mathrm{k_p}\theta_2} \tag{28}$$

As mentioned previously that $s^2 + a_m.s + b_m = s^2 + a_p.s + b_p + k_p\theta_2$, and according to the MIT law and utilizing equations (27) and (28), we can conclude the expressions of θ_1 and θ_2.

$$\frac{d\theta_1(t)}{dt} = -\gamma\left(\frac{1}{s^2 + \mathrm{a_m}s + \mathrm{b_m}}r(t)\right)e(t) \tag{29}$$

$$\frac{d\theta_2(t)}{dt} = \gamma\left(\frac{1}{s^2 + \mathrm{a_m}s + \mathrm{b_m}}y_p(t)\right)e(t) \tag{30}$$

The adaptation gain γ dictates the system's performance and is typically determined heuristically. In order to guarantee an optimum performance, the gain γ in the proposed algorithm is determined using a fuzzy logic controller. As defined previously, the fuzzy controller must contain two inputs and a single output as illustrated in Fig. 7.

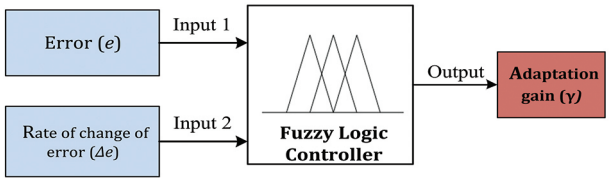


Fig. 7. Fuzzy control input and output variables

We have assigned five triangular membership functions to each fuzzy controller variable as indicated in Fig. 8, resulting in a total of 25 inference rules. The fuzzification method employed is the Max-Min method (Mamdani). These distributions are depicted in the following figures.

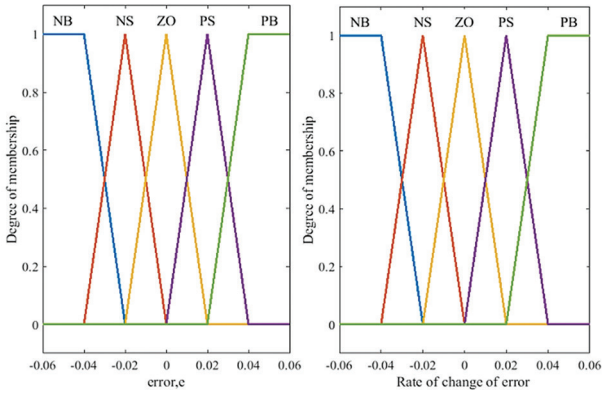


Fig. 8. Membership functions of two inputs

With NB: Negative Big; NS: Negative Small; ZO: Approximately Zero; PS: Positive Small; PB: Positive Big.

Inputs and output are linked by rules called inference rules, which enable conclusions to be executed. The typical form of a fuzzy rule is:

If <Conditions linked by fuzzy operators> Then <Action >

The following table illustrates the representation of inference rules (used to determine the adaptation gain) in matrix form, commonly referred to as the "Inference Matrix".

Table 1. Inference matrix

		e				
	γ	NB	NS	ZO	PS	PB
Δe	NB	Z	Z	B	B	B
	NS	Z	Z	S	S	S
	ZO	S	Z	Z	Z	S
	PS	S	S	S	Z	Z
	PB	B	B	B	Z	Z

For instance, the rules corresponding to the red cell in the table is interpreted as follows: if error is NB and rate of change of error is ZO, then the adaptation gain is small (S).

4. RESULTS AND DISCUSSION

The MATLAB-Simulink software is utilized to perform various simulations. Table 2, 3 and 4 presents the PV panel parameters, the boost converter data, and the coefficients of the suggested methodology. To assess the performance of our approach, we compare it with other classical MPPT control methods such as Perturb and Observe, fuzzy logic, and metaheuristic PSO controllers under dynamic temperature and irradiation profiles.

To simulate a photovoltaic (PV) system with MPPT control in MATLAB/Simulink, several settings need to be adjusted in the blocks to ensure correct simulation and accurate results. Below are the key block parameters and model settings for running a PV system with a boost converter and MPPT controller.

- **PV Generator (Photovoltaic Array) Settings**

The photovoltaic array block is typically found in the Simscape Electrical library. Key parameters adjusted:

- Irradiation (Ir): We use a standard value for solar irradiance, such as 1000 W/m² (under standard test conditions), or modify it based on real-time weather data or dynamic inputs. Example: Irradiance = 1000 W/m².
- Temperature (T): We set the ambient temperature to a typical value, like 25°C, or adjust it according to simulation parameters (e.g., dynamic temperature changes).
- Number of Series and Parallel Modules: These affect the voltage and current characteristics of the PV system. Adjust based on our desired output. In our case Series = 2 modules, Parallel = 2 string (these values vary depending on the panel's specifications).

- **Boost Converter Settings. Key Parameters to Adjust are given in Table 3.**
  - Duty Cycle (D): The duty cycle is controlled by the MPPT algorithm and adjusts based on the MPPT feedback.
- **Simulation Settings**
  - Solver: We used ode45 (Variable step size).
  - Simulation Time: Simulation time = 1 seconds.
  - Time Step (Sampling Time): The time step should be small enough to capture system dynamics accurately. A typical value might be 0.01 to 0.1 seconds. In our case: Time step = 10^{-6} seconds.
- **Test Cases for Different Conditions. To test various conditions of your system, we modify the following parameters:**
  - We vary the Irradiation: we test our system under different light conditions [800 W/m2 700w/m2 600W/m2 800W/m2 900W/m2].
  - We vary the Temperature: we test for different ambient temperatures (as shown in figure 13) to observe the system's response to temperature changes.

Table 2. PV Model Parameters

DC-DC Boost Parameters	Value
C1	100 µF
VIN	56.6-60.3 V
L	2 mH
R0	20 Ω
C0	100 µF
V0	112.5-129.1 V

Table 3. DC-DC boost converter Parameters

PV Model Parameters	Value	PV Model Parameters	Value
Maximum power	213.15W	PV cell Rpe	100 µF
Maximum current	35 A	PV cell Rse	56.6-60.3 V
Maximum voltage	29 V	Cells per module	2 mH
Short-circuit current (Isc)	7.84 A	R	20 Ω
Open-circuit voltage (Voc)	36.3 V		

Table 4. MRAC-FUZZY control parameters

MRAC_FUZZY Parameters	Value
$a_p = a_m$	8.17 × 103 (rad/s)
$b_p = b_m = 1/L \times C1$	1.67 × 107 (rad/s)²
$k_p = V0/L \times C1$	6.45 × 108V (rad/s)²
Simulation step time (Ts)	1 µs
Switching frequency (fs)	20 kHz
k_m	5.75 ×108V (rad/s)²

The proposed MRAC-FUZZY based MPPT control has been verified by simulations under MATLAB/Simulink using the control scheme illustrated in Fig. 9.

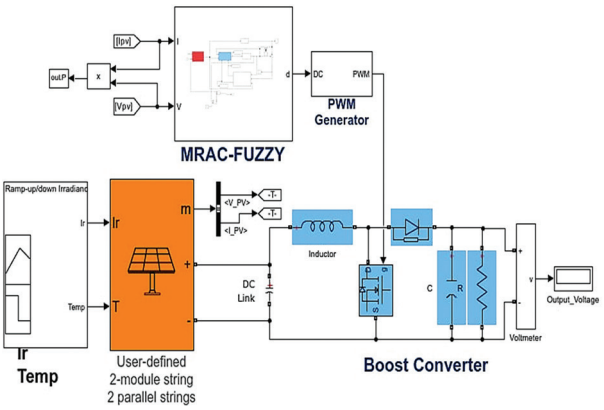


Fig. 9. Block diagram of the photovoltaic system

4.1. CONTROLLER BEHAVIOR IN CASE OF IRRADIATION VARIATION

The efficacy of the proposed MRAC-FUZZY MPPT has been validated through simulations under diverse operating conditions using MATLAB/Simulink software. Initially, we conducted an irradiation test as illustrated in Fig. 10. The irradiance profile begins with an initial irradiance of 800W/m^2. At t=0.2s, the irradiance gradually decreases to 700W/m^2 over a duration of 0.2 seconds. Following this, starting at t=0.6s, the irradiance rises to 600W/m^2, then increases further to 800W/m^2, and finally to 900W/m^2. The temperature remains constant at 25°C throughout the used profile.

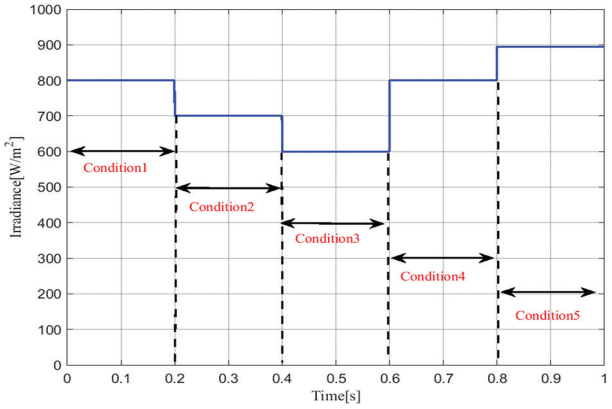


Fig. 10. Variable irradiation profile

Fig. 11 illustrate the photovoltaic output power using the proposed MRAC-FUZZY, PSO, P&O, and Fuzzy Logic controllers. Zooms are applied: one focusing on the transient mode to enhance response time and another on the steady-state mode to illustrate ripples in the MPP. Fig. 10 shows that the P&O technique has a maximum time to reach MPP which about 0.046s, followed by the FL technique at 0.016s, and the PSO at 0.0095s, while the proposed technique takes only 0.0034s to reach the MPP.

The FL and P&O methods show a significant ripple around the MPP but do not reach it. The PSO algorithm presents less ripples, compared with the MRAC-FUZZY method, which has practically no ripple and follows the MPP easily over all five irradiation conditions.

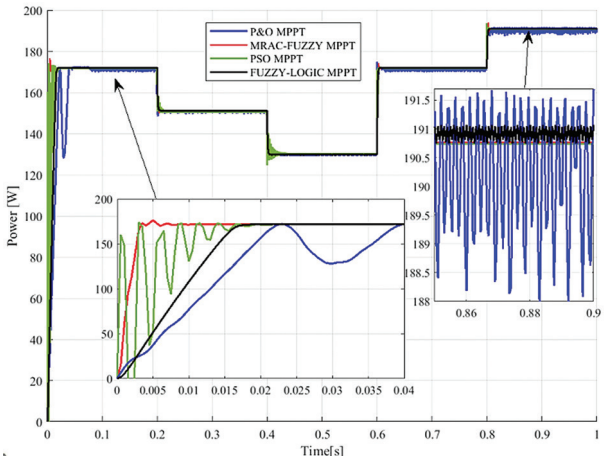


Fig. 11. Simulation results of four MPPT controllers with variable irradiation and constant temperature

To highlight the effectiveness of the proposed approach, in addition to response time and ripple, other performance criteria are calculated like energy losses and efficiency during five irradiation conditions, as shown in Table 5. As a result, the novel control algorithm can ensure the superior effectiveness compared to the other conventional MPPT control, FL and metaheuristic controller under sudden change in irradiance conditions. It can be noted that the response time is minimal, ripple and energy losses are the lowest, and efficiency is the very highest in the event of the proposed MRAC-FUZZY MPPT.

Table 5. Performance comparison for the 4 algorithms

MPPT Techniques	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
Response time (s)					
P&O	0.046	0.043	0.049	0.046	0.044
FUZZY LOGIC	0.016	0.018	0.022	0.016	0.02
PSO	0.0095	0.0096	0.019	0.0095	0.01
MRAC_ FUZZY	0.0034	0.0036	0.004	0.0034	0.0038
Ripples (W)					
P&O	5.2	4.1	3	3.2	3.5
FUZZY LOGIC	2.6	3	2.3	2.6	2
PSO	1.4	1.6	1	1.4	1.3
MRAC_ FUZZY	0.05	0.01	0.01	0.01	0.02
Energy losses (%)					
P&O	3	3.7	2.30	1.86	1.81
FUZZY LOGIC	1.5	1.98	1.76	1.51	1.03
PSO	0.8	1.05	0.76	0.81	0.67
MRAC_ FUZZY	0.02	0.006	0.007	0.005	0.01
Efficiency (%)					
P&O	95.2	94.6	95.11	95.94	95.83
FUZZY LOGIC	96.41	95.73	96.42	96.58	96.94
PSO	97.92	97.25	98.2	97.6	98.1
MRAC_ FUZZY	99.88	99.97	99.86	99.98	98.87

4.2 CONTROLLER BEHAVIOR IN CASE OF TEMPERATURE VARIATION

The four MPPT controls are simulated under two variable profiles of temperature and constant irradiation as illustrated in Fig. 12. Firstly, it's evident from Fig. 13 and Fig. 14 that power varies inversely with temperature. The performance criteria are summarized in Table 6. It can be noted that the MPPT MRAC-FUZZY control has better PPM tracking performance with minimal response time, almost zero ripple around the MPP and very high efficiency with little energy loss compared to P&O, Fuzzy Logic and PSO MPPTs controllers.

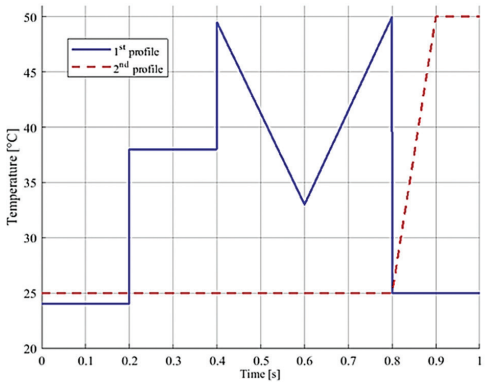


Fig. 12. Variable Temperature profiles

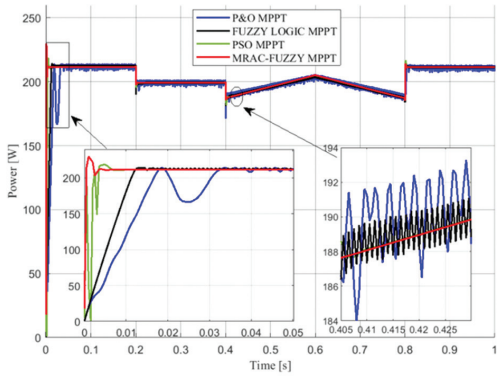


Fig. 13. Simulation results of four MPPT controllers with variable temperature and constant irradiation (1st profile)

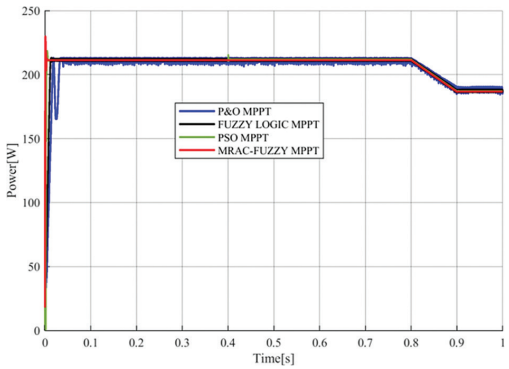


Fig. 14. Simulation results of four MPPT controllers with variable temperature and constant irradiation (2nd profile).

Table 6. Performance comparison for a variable temperature profile

MPPT techniques	P&O	FUZZY-LOGIC	PSO	MRAC-FUZZY
Response time (s)	0.034	0.015	0.005	0.0032
Ripple (W)	7.8	2.4	1.2	0.01
Energy loss (%)	4.53	1.93	0.69	0.005
Efficiency (%)	95.32	96.67	97.98	99.96

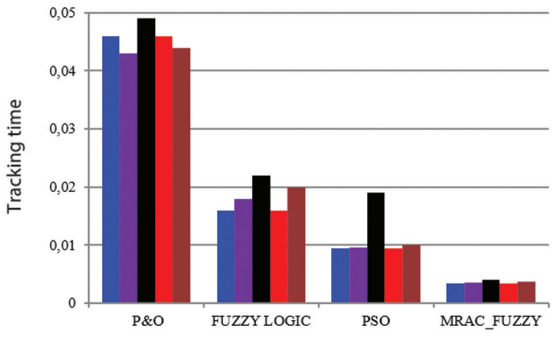
The results presented in Table 6 show that the P&O method has the highest response time to reach the Maximum Power Point (MPP), with a value of 0.034 s. This is followed by the FUZZY_LOGIC technique at 0.015 s, and the PSO method at 0.005 s. In contrast, the proposed technique captures the MPP in just

0.0032 s. Although the P&O, FUZZY_LOGIC, and PSO methods exhibit ripple content near the MPP, they fail to precisely achieve it. On the other hand, the MRAC-FUZZY method demonstrates nearly zero ripple content and tracks the MPP with high accuracy. Regarding energy losses, the P&O, FUZZY_LOGIC, and PSO methods experience losses of 4.53%, 1.93%, and 0.69%, respectively, while the proposed technique incurs virtually no energy losses. In terms of tracking efficiency, the P&O method achieves 95.32%, FUZZY_LOGIC achieves 96.67%, and PSO achieves 97.98%. In comparison, the proposed technique achieves an impressive tracking efficiency of 99.96%.

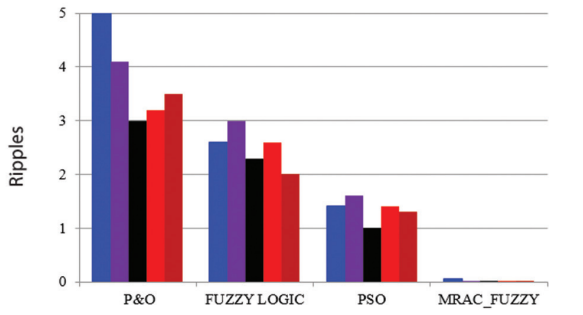
The results highlight the exceptional MPP tracking capability of the MRAC-FUZZY MPPT controller, excellent tracking of maximum power point with total elimination of ripples. In contrast, other MPPT controllers experience delays in reaching the MPP. Moreover, the time required to reach the MPP is 0.0034s, approximately 11 times faster than P&O, 5 times faster than FL, and 2 times faster than PSO. Figure 15 show the comparison between tracking time, ripples, and efficiency of proposed MRAC-FUZZY, PSO, Fuzzy Logic and P&O MPPT under variable condition as shown in Figure 9.

In comparison to other recent studies, the PV system's tracking efficiency has ultimately been improved, as outlined in Table 7.

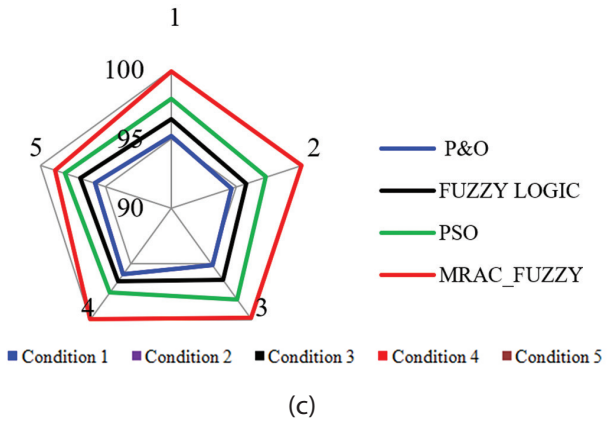
It can be seen that MRAC-FUZZY algorithm enhances the system's average efficiency from 95.32%, 96.67%, and 97.98 in comparison to the P&O, Fuzzy Logic and conventional PSO algorithms respectively, achieving a rate of 99.96%.



(a)



(b)



(c)

Fig. 15. Comparison evaluation. (**a**) Tracking time, (**b**) Ripples, (**c**) Efficiency.

Moreover, the MRAC-FUZZY algorithm, as proposed, exhibits superior accuracy in tracking maximum power following fluctuations in climatic conditions, with minimal observable oscillation around the MPP, distinguishing itself from other control methods, As observed in the zoomed-in views of Figs. 11 and 13, notable improvements are evident when comparing the results obtained with the PSO, Fuzzy Logic, and P&O MPPT algorithms. These enhancements are delineated below:

- The tracking time is notably shorter compared to other algorithms
- The power ripple has been significantly reduced

Table 7. An analysis comparing the suggested approach to other methods

Performance Parameters	Adaptive MPPT Controller [21]	ANFIS-TRSMC [19]	(PSO) [22]	(GWO)-PID [23]	Proposed MPPT
Tracking time	0.0036	0.04	0.012	0.018	0.0034
Oscillations at MPP	Low	Medium	Medium	No	No
Complexity	Medium	Medium	Medium	Medium	low
Efficiency	99.69%	98.9%	96.96%	99.50%	99.97%

5. CONCLUSION

In order to improve the photovoltaic system's efficiency, a new fuzzy model reference adaptive control based MPPT has been proposed. The proposed algorithm combines the strengths of the MRAC concept, which can handle non-linear systems adequately, and the advantages of fuzzy logic, which can determine the adaptation gain heuristically. In order to simulate the system behavior for MPP tracking, simulations were carried out in the MATLAB/Simulink environment. A comparative analysis was conducted again to the classical algorithms like (P&O), fuzzy logic, metaheuristic (PSO), according to different criteria (dynamic response time, low ripples, and efficiency). The simulation outcomes have confirmed the exceptional performance of the innovative controller, showcasing a reduction in response time (3.4ms to reach the MPP) it is around 11 times, 5 times and 2 times faster than P&O, FL and PSO respectively with high efficiency up to 99.97%. Also, the proposed MPPT algorithm based MRAC-Fuzzy can ensure good tracking of MPP without ripple.

The proposed MRAC-Fuzzy-based MPPT controller offers higher tracking efficiency, reduced ripple content, and lower energy losses compared to traditional MPPT techniques. However, successful implementation in a real PV system requires addressing challenges related to sensor accuracy, hardware limitations, environmental factors, and system stability. In the upcoming study, the proposed MPPT algorithm will be deployed in a real testing environment. This will include the setup of temperature and irradiation sensors, as well as instrumentation to measure the power and voltage generated by the photovoltaic panel. The MPPT controller will be integrated with an inverter or a battery charger to validate the effectiveness of the MPPT algorithm based MRAC-FUZZY controller. During the implementation of the developed MPPT controller, several challenges and potential issues may arise, including:

- Problems related to model parameter estimation: Photovoltaic system models may not be perfectly accurate. Real-world conditions can present unexpected variations compared to theoretical assumptions (e.g., shadows on the panels, dust, or panel degradation). This could lead to errors in adjusting the MRAC-FUZZY controller.
- Sensitivity to irradiation and temperature conditions: MPPT algorithms are sensitive to rapid variations in irradiation and temperature. In particular, sudden changes in irradiation due to clouds or shadows could disrupt the tracking efficiency. The fuzzy algorithm must be adjusted to handle these variations smoothly and stably.
- Control latency: In a real system, sensor and controller latency may introduce delays in adjusting the control, which could slow down the system's response. Our proposed algorithm needs to be designed to minimize this latency and ensure an appropriate real-time response.

6. REFERENCES

[1] H. T. Pham, L. Dai, "A Feasible MPPT Algorithm for the DC/DC Boost Converter: An Applied Case for Stand-Alone Solar Photovoltaic Systems", Inter-

national Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 6, 2023, pp. 713-724.

[2] M. Brahmi, C. B. Regaya, A. Zaafouri, "Comparative Study Of P&O and PSO Particle Swarm Optimization MPPT Controllers for Photovoltaic Systems", Proceedings of the 8th International Conference on Control, Decision and Information Technologies, Istanbul, Turkey, 17-20 May 2022, pp. 1608-1613.

[3] C. B. Regaya, H. Hamdi, F. Farhani, A. Marai, A. Zaafouri, A. Chaari, "Real-time implementation of a novel MPPT control based on the improved PSO algorithm using an adaptive factor selection strategy for photovoltaic systems", ISA Transactions, Vol. 146, 2022, pp. 496-510.

[4] A. Elgharbi, D. Mezghani, A. Mami, "Intelligent Control of a Photovoltaic Pumping System", Engineering, Technology & Applied Science Research, Vol. 9, No. 5, 2019, pp. 4689-4694.

[5] A. R. Danila, R. Roshini, E. Priyanka, M. Sindhuja, A. S. Jones, "Design and Implementation of Photovoltaic Powered SEPIC DC-DC Converter Using Particle Swarm Optimization (PSO) Technique", Computing and Social Networking. Lecture Notes in Networks and Systems, Vol. 317, Springer, 2022.

[6] M. Jagadeshwar, D. K. Das, "A novel adaptive model predictive control scheme with incremental conductance for extracting maximum power from a solar panel", Iranian Journal of Science and Technology, Transactions of Electrical Engineering, Vol. 46, 2022, pp. 653-664.

[7] M. Jagadeshwar, D. K. Das, "A novel MRAC-MPPT scheme to enhance speed and accuracy in PV systems", Iranian Journal of Science and Technology, Transactions of Electrical Engineering, Vol. 47, 2023, pp. 233-254.

[8] A. K. Mohaisen, "Artificial Intelligent Maximum Power Point Controller based Hybrid Photovoltaic/ Battery System", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 10, 2023, pp. 1183-1191.

[9] S. Saha, M. Cintury, C. Roy, "A PSO based MPPT controller for solar PV system at variable atmospheric conditions", Advanced Computational Paradigms and Hybrid Intelligent Computing, Advances in Intelligent Systems and Computing, Vol. 1373, Springer, 2022.

[10] B. Mahbouba, B. R. Chiheb, H. Hichem, Z. Abderrahmen, "Comparative Study Of P&O and PSO Particle Swarm Optimization MPPT Controllers Under Partial Shading", International Journal of Electrical Engineering and Computer Science, Vol. 4, 2022, pp. 45-50.

[11] Md. T. Hussain, A. Sarwar, M. Tariq, S. Urooj, A. BaQais, Md. A. Hossain, "An Evaluation of ANN Algorithm Performance for MPPT Energy Harvesting in Solar PV Systems", Sustainability, Vol. 15, No. 14, 2023, p. 11144.

[12] Y. El Aidi Idrissi, K. Assalaou, L. Elmahni, E. Aitiaz, "New improved MPPT based on artificial neural network and PI controller for photovoltaic applications", International Journal of Power Electronics and Drive Systems, Vol. 13, No. 3, 2022, pp. 1791-1801.

[13] M. Fathi, J. A. Parian, "Intelligent MPPT for photovoltaic panels using a novel fuzzy logic and artificial neural networks based on evolutionary algorithms", Energy Report, Vol. 7, 2021, pp. 1338-1348.

[14] K. Kayisli, "Super twisting sliding mode-type 2 fuzzy MPPT control of solar PV system with parameter optimization under variable irradiance conditions", Ain Shams Engineering Journal, Vol. 14, No. 1, 2023.

[15] R. D. Tehrani, F. Shabaninia, "Two-level control of photovoltaic systems using global perturbation-based extremum seeking control and model reference adaptive control", Transactions of the Institute of Measurement and Control, Vol. 40, No. 13, 2018, pp. 1-12.

[16] P. Sahu, R. Dey, "An improved 2-level MPPT scheme for photovoltaic systems using a novel high-frequency learning based adjustable gain-MRAC controller", Scientific Reports, Vol. 11, 2021.

[17] F. Z. Kebbab, L. Sabah, H. Nouri, "A Comparative Analysis of MPPT Techniques for Grid Connected PVs", Engineering, Technology & Applied Science Research, Vol. 12 No. 2, 2022, pp. 8228-8235.

[18] M. Brahmi, C. B. Regaya, H. Hamdi, A. Zaafouri, "MPPT Using PSO Technique Comparing to Fuzzy

Logic and P&O Algorithms for Photovoltaic System", Proceedings of the 9th International Conference on Control, Decision and Information Technologies, Rome, Italy, 3-6 July 2023, pp. 2457-2461.

[19] B. Mbarki, F. Fethi, J. Chrouta, A. Zaafouri, "Adaptive neuro-fuzzy inference system algorithm-based robust terminal sliding mode control MPPT for a photovoltaic system", Transactions of the Institute of Measurement and Control, Vol. 46, No. 2, 2023, pp. 316-325.

[20] C. B. Regaya, F. Farhani, H. Hamdi, A. Zaafouri, A. Chaari, "A new MPPT controller based on a modified multiswarm PSO algorithm using an adaptive factor selection strategy for partially shaded PV systems", Transactions of the Institute of Measurement and Control, Vol. 46, No. 10, 2023, pp. 1991-2000.

[21] S. Manna et al. "Design and implementation of a new adaptive MPPT controller for solar PV systems", Energy Reports, Vol. 9, 2023, pp.1818-1829.

[22] W. M. Ersalina, R. Agus, K. Ardath, D. Rudi, "Particle Swarm Optimization (PSO) based Photovoltaic MPPT Algorithm under the Partial Shading Condition", Journal Electronika Dan Telekominikasi, Vol. 23, No. 2, 2023, pp. 99-107.

[23] J. Aguila-Leon, C. Vargas-Salgado, C. Chinas-Palacios, D. Diaz-Bello, "Solar photovoltaic Maximum Power Point Tracking controller optimization using Grey Wolf Optimizer: A performance comparison between bio-inspired and traditional algorithms", Expert Systems with Applications, Vol. 211, 2023, p. 118700.

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