MLbFA: A Machine Learning-Based Face Anti-Spoofing Detection Framework under Replay Attack

Original Scientific Paper

Vijay V. Chakole*
G H Raisoni University,
Amravati, Maharashtra, India
chakole.vijay@gmail.com

Dr. Swati R. Dixit
G H Raisoni College of Engineering and Management,
Nagpur, Maharashtra, India
swati.dixit@raisoni.net

*Corresponding author

Abstract *– The primary aim of the research paper is to deploy an efficient automated face antispoofing system that could consider replay attacks in the presence of partial occlusions. For this purpose, the article introduces a novel machine learning-based face-antispoofing (MLbFA) framework. The system incorporates a modified version of the difference of the Gaussian technique to compute the overall contrast of the input images which is later used to enhance the contrast of the image using contrast correction. On the other hand, the image details, especially the edges are enhanced for significant feature contribution using a Beltrami filter. The contrast-cured and extremity-enhanced images are averaged to obtain a finer image. Face cropping is achieved using the Bounding-Box algorithm to reduce computational complexity and improve classification accuracy for region-bounded feature extraction. Quality conventional or handcrafted features (CF/HF) are extracted through various descriptors from the region of interest (ROI). The features are reduced in dimension using principal component analysis (PCA) and portioned in training and testing sets with a 75%:25% ratio respectively. An experimental study showed that the proposed MLbFA model using a Support Vector Machine (SVM) outperforms other recent existing face anti-spoofing competing techniques with an improvement of 0.11% compared to the best-performing Edge-Net Autoencoder model concerning the classification accuracy.*

Keywords*:* *Face anti-spoofing, machine learning, Difference of Gaussian, Beltrami filter, Bounding-box algorithm, conventional features, PCA, and SVM*

Received: September 30, 2024; Received in revised form: December 23, 2024; Accepted: December 23, 2024

1. INTRODUCTION

Biometric authentication can be primarily classified as possession-based and knowledge-based authentication. Knowledge-based is secretly available with the subject that is only known to the subject himself. However, biometric authentication models are vulnerable to spoof threats where an intruder or fraudster endeavors to steal or compromise it. The type of attack can be differentiated depending on the biometric modality whether the system uses the subject iris pattern, his fingerprints, his facial features, audio, or a keystroke. Despite this, few traits are non-comparable. Therefore, a specifically designed algorithm is required to identify the spoof attacks since each biometric trait has its merits and limitations. In the case of spoof attacks, the fraudster peeps through the vulnerabilities in the biometric system. The fraudster enters the secured biometric-based system attempting to deceive another person. This leads to the significant requirement of an anti-spoofing technique for securing biometric systems and preventing unauthorized trait replicas.

Despite the advancements being made in the field of biometrics, various authentication systems still fail miserably while detecting a spoofing attack. This leads to a positive response being made to an unauthorized individual thereby giving him the overall control of the system unknowingly. Many well-known organizations and industries have already fallen into this trap of breach and spoofing attacks. Hence, a well-developed automated system that can recognize the same in real time has become mandatory. For this purpose, various researchers have used many machine and deep learning classifiers, who tend to extract relevant features from the dataset and later pre-process them for better accuracy. However, due to data imbalance, certain interference peeping in CCD and CMOS sensors, uneven illuminations, distortions, and partial occlusions due to hair, specs, scarfs, organs, etc. the anti-spoofing systems have suffered setbacks in classifying the spoofed and real images.

The work proposed by authors [1] tends to automate a robust system by making use of a Laplacian filter that not only filters the relevant images but also enhances the overall output thus generated. On the other hand authors [2] enlightened the implementation of Schmid as the filter which could uplift the disadvantages of the image and convert them into a readable format. Meanwhile, authors [3] proposed to modify the existing working of the DOG filter by adding an extra layer that could destroy the noise and generate high-frequency edges. A correction algorithm was thus used and further pre-processing of the images was done by eliminating the interference of light. Even though various pre-processing techniques were primarily used, authors [4] executed certain anti-spoofing techniques that could easily differentiate between original and fake images.

In addition to this, the generated 2D-based spoofed images were exposed to various forms of noise distortions that included glossy photo papers and digital screens. Color distortions were also faced in the same process which occurred as a result of resolution from the screen. Apart from noise and color distortions, distortions in the form of facial deformations also occurred which were proposed by authors [5]. Hence, the concept of landmark components was brought into the picture by the author [6-7] who finally segmented the image into equal halves so that optimized accuracy could be obtained.

The techniques [8, 9] incorporating textural properties of the spoofed and authentic images lacked generalization ability even though they showed rapidness in response. Models based on extracting color depth information from a 3-dimensional face using multiple image frames [10, 11] required higher reliability to classify the 2D spoof faces. On the other hand, losing low-frequency details while highlighting the higher-frequency components was associated with the process of recapturing video replays or printed photos. These models were successful and showed better sensitivity when the intra-dataset test was concerned, however, for improved performance and inter-dataset tests, diverse multiscale features were required.

Deciding on a perfect spoof mechanism out of the existing approaches would need studying the features, advantages, and negative remarks of each approach. With this in mind, facial biometrics can be seen as a special case, since multimodality can take advantage of multiple facial properties (e.g., texture, shape, and temperature) to avoid spoof attacks. Spoofing attacks persist be a security challenge for face biometric systems, and there was much effort in the field to find robust methods. However, all these efforts have been following the same recipe, not favouring breakthroughs in the field. Many works of face spoofing detection emphasize 2D attacks by presenting printed photos or replaying recorded videos, and 3D attacks have been recently studied due to the technological advancements in 3D printers and reconstruction. Although perfect results on public data sets have been achieved by many works, there is a considerable gap in moving from academic research to real-world applications in an effective way.

The paper's contributions can be summarized as:

- The ROI is extracted so that the system can eliminate distortions and focus on the facial image by creating a boundary box along the edges of the image. This reduces the overall run-time complexity of the classifier and thereby increases the probability of classification accuracy of the model.
- Sufficient informative handcrafted features are extracted from the ROI that represented the faces more accurately and significantly.
- The feature dimensionality was reduced so that the obtained dataset features are minimized and only the relevant features are executed in the training and testing phase.
- The proposed MLbFA model is simpler and more efficient to discriminate real and unauthentic faces from dataset test samples and real-time test samples.

2. RELATED WORK

A generative method of probabilistic voting was introduced by authors [12] wherein they made use of an ensemble classifier along with a discrete wavelength filter. Their study aimed to perform segmentation of the face image and further calibrate it using face alignment so that the overall frequency of the residual image could be reduced. The residue from this image was later converted into a YCbCr model and texture features were extracted from the same by making use of a texture descriptor. The repository used for the implementation of the same comprised four datasets. In a similar work performed by authors [13] they extended their research work by making use of SURF features which led to the conversion of grayscale images to color images. For this purpose, the author used specific color bands for each image and further concatenated them with the SURF features. In the next stage, the overall complexity of the system was reduced by using the PCA algorithm and Fisher as the Vector. All the descriptors used for implementation were obtained using HAAR as the wavelet function comprised of 4x4 blocks. In the final stage, the sub-regions around the image were concatenated by using HSV as the implementation algorithm which eventually resulted in the overall feature vector to reduce the dimension to 64.

In another work authors [14] implemented the techniques of chromatic illumination so that the objective of differentiating between a real and forged image could be made. They proposed to extract inter-channel chromatic occurrence to obtain chromatic texture features. Softmax was used as the classifier along with LBP as the feature extractor. Combinations of 4 datasets were used (MSU, MFSD, CASIA, and FASD) and the overall discriminating factor of the system model was improvised. Similar work was offered by authors [15] wherein they made use of an ensemble learning methodology which was eventually used to analyze the existing chromatic discrepancies. The dataset however consisted of imbalanced images. In another work suggested by authors [16] they made use of SVM as the fuzzy classifier which eventually analyzed the acquired dataset from different angles and perspectives. All the images from the dataset underwent the process of feature extraction by making use of the HOG feature extractor. In addition to this, LPQ features were also simultaneously used which minimized the invariance of images that occurred.

Abhishek Mittal *et al*. [17] obtained a 10% improvement for accurately distinguishing the natural and unnatural faces using features based on the Gray level co-occurrence matrix (GLCM). The texture features-based hybrid approach was evaluated using an integration of 3 different ML classifiers.

Mays Alshaikhl *et al*. [18] suggested an attention module to consider the relevant features only using the spatial features and the color depth features. Their dual module approach was efficient in seeking context-based significant features that helped their framework in improving the classification accuracy. On the other hand, their objective to improve the generalization ability of the deep learned (DL) classifier was also enhanced. The integrated model incorporating the attention module and DL network succeeded in aiding the overall performance. The authors suggested that their framework can be implemented for pixel-based attention mechanisms including quality assessments, segmentation, and face detection jobs.

Spoof strikes were detected by Junwei Zhou *et al*. [19] utilizing an inventive approach that combined LDN-TOP representation and Pro CRC (probabilistic collaborative representation-based classifier) classification pipeline. The LDN and a derivative Gaussian mask were used to learn the texture patterns of the concerned region under disturbances caused by illuminations. While LDN was broadened to spatial-temporal variant to occupy the motion features. The Pro CRC was made to learn from the LDN-TOP represented features extracted from depth images. The experimental evaluation was carried out on 3-different sets of dataset images concerning the EER and HTER (half total error rate). The authors obtained 0.37% EER on the CASIA dataset and 5.7% HTER on the UVAD dataset using a sequence-based protocol. They also carried out a time-window length analysis and demonstrated that improved outcomes can be obtained using larger lengths at the cost of sufficient video frames. Competitive results were also shown for replay attacks in their article.

The authors [20] focused on pre-processing the images before they were submitted to a convolutional neural network constructed with 12 layers. The pre-processing concentrated on proper and precise face alignment, extracting the concerned region, and controlling the uneven illumination from face regions. Cropping was accomplished using a bounding box algorithm while the latter part dealt with three cascaded networks. The P, R, and O-net cascaded network structure controlled the illumination of the bright regions of the face images obtained from the CASIA-FASD dataset. Through experimentation, the author showed that their model attained an HTER of 1.02% and improved the classification accuracy.

The immense success of pre-trained networks in several applications was noticed by authors in [21]. They used 2-different color spaces of the input images and extracted face embeddings which were jointly used for classification using the VGG network. For better face representation, the images were denoised priory and then converted to other color spaces. The joint embeddings from CIE LUV and YCbCr color spaces obtained a false rejection ratio of 0.3% and an acceptance ratio of 0.4% approximately with an accuracy and specificity above 99%.

The LBP and SVM integrated classification approach used by authors in [22] considered a rotation invariant scheme over photo imposter dataset images. The illumination variations introduced in the dataset images were distilled using a modified Difference of Gaussian filter and filtered images were represented employing the LBP operator. Comparison with the normal combination of LBP+SVM and LBPV+SVM showed that the latter outperformed the former, however, the approach failed to perform in the case of cross-dataset samples.

Multi-channel images from the PAD dataset were part of the work introduced in [23] that are prone to various Presentation Attacks. The author utilized all the aspects of multichannel images including the depth, NIR, and color channels, and presented a PAD protocol using a combination of autoencoder and Multi-Layer Perceptron. The challenges due to Presentation Attacks were handled using a dual path framework by the authors who claimed that the individual faces represented higher disparities compared to that of complete faces.

Several researchers over the globe have contributed to developing antispoofing schemes using conventional features by different classifiers. They have used different dataset images with different levels of complexity and analyzed the performance of their scheme on single and cross-domain datasets. However, few of them lacked the generalization ability when cross-dataset or real samples were tested while others failed to classify the samples accurately. Also, in most cases, the classifier models were complex and required a longer time for training.

3. MATERIALS AND METHOD

The system thus proposed in the research paper is divided into five main stages. The implementation begins at the pre-processing stage followed by extraction of relevant features. A process of dimensionality reduction also takes place before the classification. The research finally comes to an end by evaluating and analyzing the results.

Initially, a dataset is obtained which comprises real and fake images. The first stage of pre-processing is done on these images which eventually comprises two primary techniques in parallel used to correct the image thus obtained. The technique involves contrast correction of the image and then using a filter to preserve the edges of the image. The resulting output from these techniques is further given as an input to the next stage which appears to be the feature extraction process. This happens to be the second stage of implementation and is mandatory to enhance the edges of the image and further extract only the relevant features so that the final distinguishing between the real and the fake ones can be made. A total of 2056 image features are extracted and further reduced in terms of dimensions using the PCA in the third stage. The fourth stage of system implementation is characterized using the classification technique. For this purpose, we have used SVM as the classifier, and the dataset is fed to the training and testing phase so that evaluation of the model can be made. A percentage ratio of 75 and 25 is made respectively for training and testing purposes. Finally, the system is evaluated using performance metrics.

3.1. THE DATASET

The IDIAP Replay attack dataset is used for the implementation of the proposed research. It comprises 1300 videos including the video and photo attack. The videos lasting for 9 seconds were converted to image frames for all 50 subjects. The color images obtained bear a resolution of 320x240 dimension where the first dimension corresponds to width and the latter the height. The original sampling rate was 25 Hz and the videos were captured in two environments. The controlled environment uses homogeneous backgrounds, and curtain-covered windows in the background with good surrounding illumination.

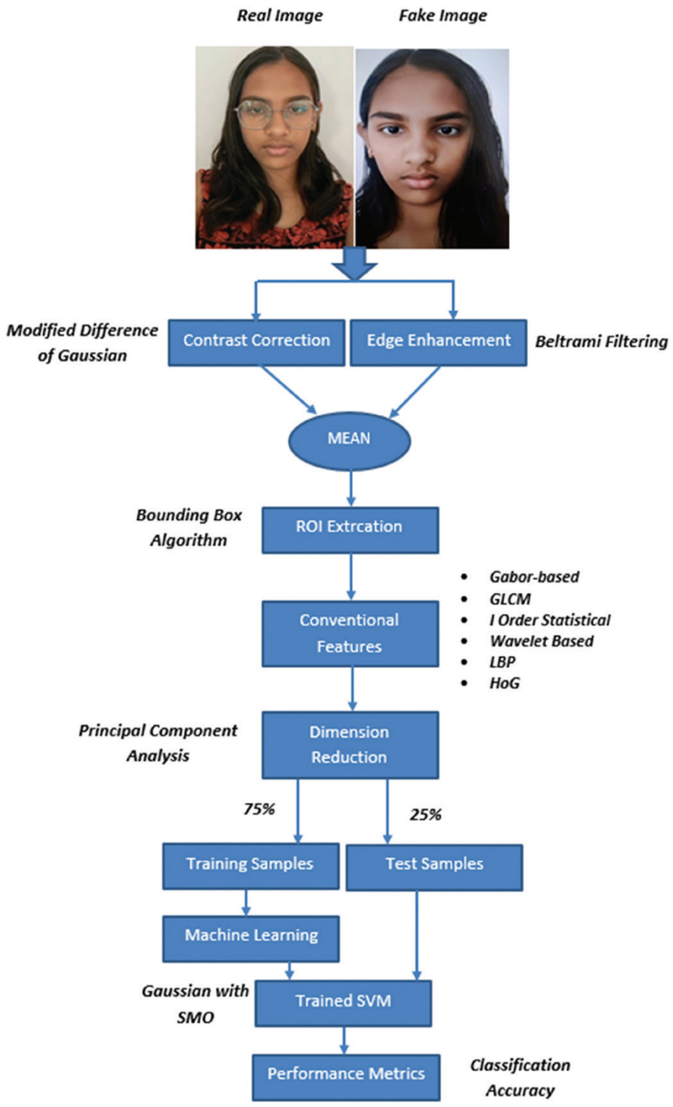


Fig. 1. The proposed MLbFA Model

The adverse environment on the other hand used a complex background with ill-illuminated surroundings and raised curtains over the windows. Ten attack video attacks for each were performed with the subject's fixed biometry and biometry from a device [8, 24].

We constructed our dataset from the above available images. We chose real images belonging to 80 subjects and fake images that included 199 subjects. The number of images for each subject from real and fake categories that were included for experimentation was 50. This was intentionally done to construct an imbalanced dataset. Therefore, images belonging to the real images class were 4000, and the fake images were 9950. Further real and fake images from real-time videos captured using Samsung Galaxy F34 with a 50 MP Camera were used to test the robustness of the proposed system and evaluate cross-test performance. The input images shown in Fig. 1 are real-time images where the left image belongs to the real category while the right belongs to the attack class.

Column 1 and column 3 of Fig. 2 show the adverse condition for real images while column 2 images are acquired in a controlled environment. Similarly, the upper row of Fig. 3 shows images from adverse environments while the bottom row depicts images obtained from the controlled environment under attack conditions.

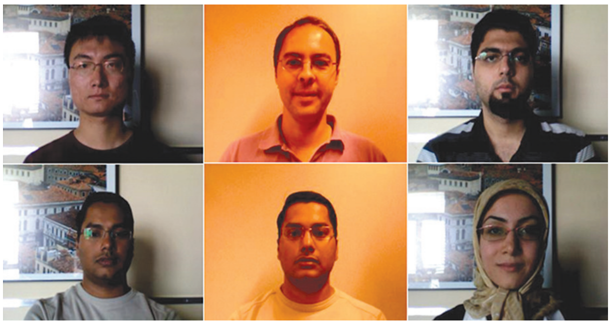


Fig. 2. Sample of Real images from the IDIAP dataset. Column 1 and column 3 show the adverse condition (uneven illumination) for real images while column 2 images are acquired in a controlled environment

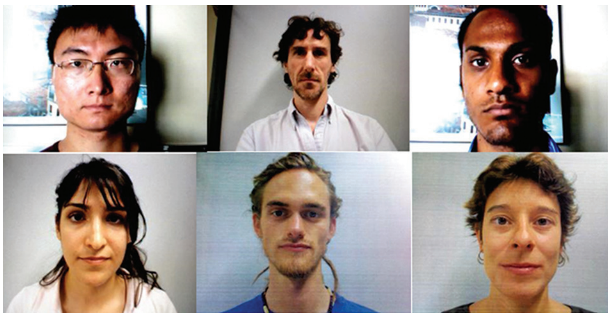


Fig. 3. Sample of spoofed images from the IDIAP dataset. Upper row of Fig. 3 shows images from adverse environments (uneven illumination) while the bottom row depicts images obtained from the controlled environment under attack conditions

3.2. CONTRAST MEASUREMENT AND CORRECTION

Thus the dataset has dual complexity issues: real and fake images are not balanced and they have discerning illuminations and backgrounds. In addition to this, the edges over the image also appear to be blurred which can affect the quality features of the image and makes it less significant during the feature extraction process. Due to this reason, the conventional features will probably fail to absorb latent details from the ROI and thus affect adversely the classification.

The image thus obtained from the dataset is heavily dependent on the spatial arrangements, edge measurements, and light illuminations in which the image is captured. Apart from this, the color of the image, resolution, and pixel size of the image also have an overall impact. To overcome this issue, we used contrast correction and edge preservation both independently on the input image. The perceived contrast is measured by computing the difference between the extremely low and extremely high-intensity pixels in the image [1]. Initially, the overall image contrast is measured using the Modified-DoG method suggested by Tadmor and Tolhurst [2], and based on the measured contrast value (equation 1) the image contrast is corrected (equation 1-8) to improve the overall quality of the image. In parallel, we also carried out distilling the image using the Beltrami Filter for edge enhancement. This filter is responsible for preserving the image edges which are distorted due to illuminations and/or background noise. Both techniques are however carried out to enhance the quality of the overall image and obtain good classification results.

Fig. 4 shows the output of modified DoG filtering. The contrast 'Cm' of an image is measured using equation (1).

$$C_m(xx, yy) = \frac{R_c(xx,yy) - R_s(xx,yy)}{R_c(xx,yy) + R_s(xx,yy)} \tag{1}$$

Where the output of the central component is,

$$R_c(xx, yy) = \sum_{i=xx-3r_c}^{i=xx+3r_c} * \sum_{j=yy-3r_c}^{j=yy+3r_c} C(i-xx, j-yy) I(i,j) \tag{2}$$

While the output of the surround component is,

$$R_s(xx, yy) = \sum_{i=xx-3r_c}^{i=xx+3r_c} * \sum_{j=yy-3r_c}^{j=yy+3r_c} S(i-xx, j-yy) I(i,j) \tag{3}$$

The center and surround components of the receptive field are given by, *Center component*,

$$C(xx, yy) = exp\left[-\left(\frac{xx}{r_c}\right)\left(\frac{xx}{r_c}\right) - \left(\frac{yy}{r_c}\right)\left(\frac{yy}{r_c}\right)\right] \tag{4}$$

(xx, yy) is the spatial coordinates of the receptive field, and rc is the radius at which the sensitivity decreases to 1/e w. r. t. the peak level.

$$Surround\ component,\quad S(xx,yy) = 0.85\left(\frac{r_c}{r_s}\right) exp\left[-\left(\frac{xx}{r_s}\right)^2 - \left(\frac{yy}{r_s}\right)^2\right] \tag{5}$$

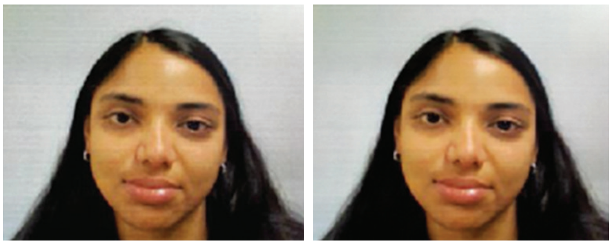


Fig. 4. Contrast measurement. (**a**) The original image and (**b**) Image obtained from the patch level contrast values

3.3. CONTRAST CORRECTION

Once the contrast measure is estimated over the image by averaging the patch level contrast values, the overall contrast value is used to enhance the image quality using expressions 6 to 8. This henceforth diminishes the processing time for feature extraction as well. 'g' represents the contrast-corrected image.

$$mm = 255^* C_m \tag{6}$$

$$f = 259 * \frac{(mm+255)}{(255*(259-mm))} \tag{7}$$

$$g = (f \ * (\text{Image} - 128)) + 128 \tag{8}$$

Where 'mm' and 'f' are intermediate results obtained using the contrast Cm computed using equation (1). 'Image' here represents the original image which is considered for contrast correction.

3.4. EDGE PRESERVING AND ENHANCING FILTER - BELTRAMI FILTER

The conceptual working theory of the Beltrami filter was introduced by [25]. This was majorly done to preserve the image edges from losing data due to the presence of noise. The technique makes use of a denoising filter and applies the same to the 2D images. The Beltrami filter is however capable of discarding any aliases present in the process and enhancing all the weak textures of the image while preserving the edges. The filter also makes use of various color channels which tends to separate the input image. In the proposed research we have used 20 iterations of the filter with a time step of 0.5.

The original input image and the peak signal-to-noise ratio between the filtered image, contrast-corrected image, and the mean image are shown in Figs. 5 and 6.

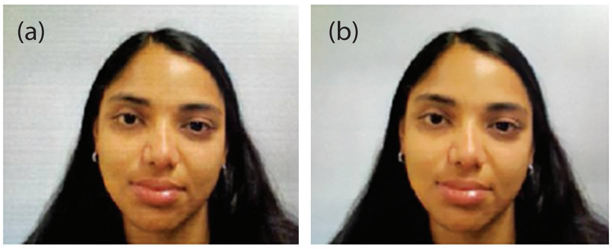


Fig. 5. (**a**) Original color image and (**b**) output of Beltrami filter with PSNR between them

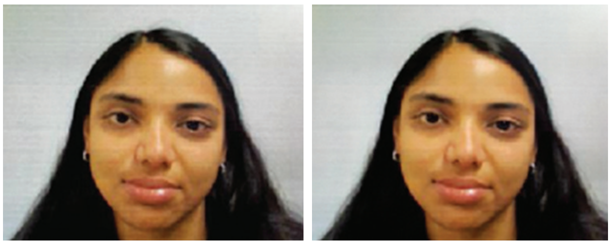
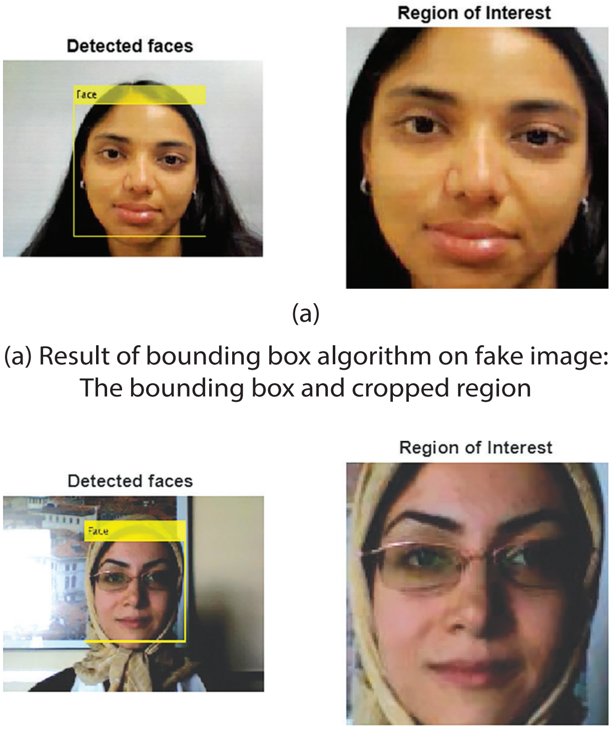


Fig. 6. (**a**) Output of Contrast Correction and (**b**) Final mean image with PSNR related to the original image

The region of the face was further extracted using the bounding box algorithm available with MATLAB 2019b. In the next stage, the extracted region of the face (ROI) is further resized to a fixed dimension [120 120 3] such that there is a minimum loss of features. This is done to ensure that different faces when localized using the bounding box algorithm carry different widths and heights. Extensive experiments were carried out on images from the dataset to set suitable dimensions for all cropped faces. The dimension chosen was relative to the smallest and the largest face detected using the bounding box algorithm. Setting smaller dimensions would cause a significant loss of features from larger cropped faces. While setting larger dimensions would probably distort the features of smaller faces. Fig. 7 below depicts the face regions automatically detected and then resized to a common dimension [120 120 3].



(a)

(a) Result of bounding box algorithm on fake image: The bounding box and cropped region

(b)

(b) Result of bounding box algorithm on real image: The bounding box and cropped region

Fig. 7. Result of Bounding Box Algorithm on samples from both classes. (**a**) Fake and (**b**) Real

3.5. FACE REPRESENTATION USING THE CONVENTIONAL FEATURES

- Gabor Filter-Based Features

The concept of Gabor features was developed by [26] and eventually used to set the obtained image representation with default filter parameters (scales = 5, orientations = 9). The number of rows and columns to be set to 39. The rows and columns were reduced by a factor of 39 through down-sampling. The final resized image was converted to a grayscale image wherein a total of 640 features were functionally extracted for representation.

- Gray Level Co-occurrence Matrix (GLCM) Based Features

The GLCM-based feature extraction technique is primarily used to extract features from the grayscale images by taking into consideration the associated properties such as contrast, homogeneity, energy, and correlation. The representation due to each of the GLCM factors was obtained in all 8 directions. Though features are obtained considering 3600 orientation (8 directions), they were averaged and the GLCM properties were represented by a single value. Thus, overall 4 values corresponding to four factors were considered for evaluation.

- Statistical Features of First-Order

A total of five features were extracted from the cropped and resized grayscale image (I_{face}). The converted images were labeled as I_{face}. The five features however included the calculation of mean, variance, standard deviation, skewness, and kurtosis which were based on probabilities obtained over Iface pixel intensities. The pixel intensities 'L' are assumed to be in the range of [0 255]. The probability 'P' is initially computed using expression (11) and eventually, all the five parameters are computed using expressions from 12 to 16.

$$L = 0{:}255 \tag{10}$$

Probability $Pb(x)$ of each pixel 'x' in the image:

$$P_b(x = 1 \text{ to } 255) = \frac{1}{M*N}\sum_{i=1,j=1}^{i=M,j=N}(I_{face}(i,j) == x) \tag{11}$$

$$\text{Mean, } M = \sum L .* P_b \tag{12}$$

$$\text{Variance, } V = \sum(\,[(L-M)^2].* P_b) \tag{13}$$

$$\text{Standard deviation, } Std = \sqrt{V} \tag{14}$$

$$\text{Skewness, } S = \frac{\sum(\,[(L-M)^3].*P_b)}{Std^3} \tag{15}$$

$$\text{Kurtosis, } K = \frac{\sum(\,[(L-M)^4].*P_b)}{Std^4} \tag{16}$$

- Features through Wavelet transform

Wavelet-based features are constructed using 1-level decomposition over six different mother wavelets. Experiments showed that vertical and diagonal component at level 1 shows discriminative features as compared to other components. We computed the energy and magnitude of the components for all six mother wavelets and enhanced the feature set. The six mother wavelets used for decomposition for the I_{face} image include haar, bior, debauchees and symlet (bior 3.1, bior 3.5, and bior 3.7), debauchees 3 (db3), symlet 3 (sym3), and haar).

The magnitude M_w and energy value E_w over both components is computed using the following expressions (11) and (12).

$$M_w = \frac{1}{m*n}\left(\sum_{r=1}^{m}\sum_{c=1}^{n} W_{xx}\right) \tag{17}$$

$$E_w = \frac{1}{m*n}\left(\sum_{r=1}^{m}\sum_{c=1}^{n} abs({W_{xx}}^2)\right) \tag{18}$$

Where m and n are the wavelet components' row size and column size. W_{xx} is either W_{vert} or W_{diag} and represents the vertical and diagonal components.

- Histogram-Based Color Depth Features

The I_{face} color image is converted to Lab color space for color-based depth features. Histograms using 16 intensity levels are obtained over each channel from both color spaces. All six histograms are normalized using the size of the I_{face} image. The histogram values of independent channels are averaged to obtain a single histogram for both color spaces. Finally, the 16-level histograms are concatenated to obtain a 32-level feature vector to contribute to the color description. The following expressions (19) and (20) are used to compute the histograms:

In the case of RGB color space,

$$h(x \in \text{R, G, B}) = \frac{1}{M*N}\, h(I_{face}(x)) \text{ (16 bins)} \tag{19}$$

In the case of Lab color space,

$$h(y \in L, a, b) = \frac{1}{M*N}\, h(I_{face}(y)) \text{ (16 bins)} \tag{20}$$

Here 'h' denotes the histogram.

- Haar-Wavelet-Based LBP Features

The usage of LBP as the feature extraction method enables the extractor to capture all the details of the image such as the edges, the illumination on the image, and the textural patterns. LBP is initially applied on all color channels the original image Iface and LBP feature are extracted. The LBP-featured channels are then used to compute the overall mean. Further, using the haar mother wavelet, the Iface image is wavelet decomposed to 4 levels. Similar LBP features are obtained on all four wavelet tributaries and averaged at the end.

The final haar-wavelet-based LBP features are obtained by averaging the mean LBP of the color image and mean LBP features obtained at different levels as shown in Fig. 8. Implementing this step ensures that image details are not lost and the system can eventually differentiate between a real image and a fake one. The figure below illustrates the architectural mechanism for the process of feature extraction using LBP.

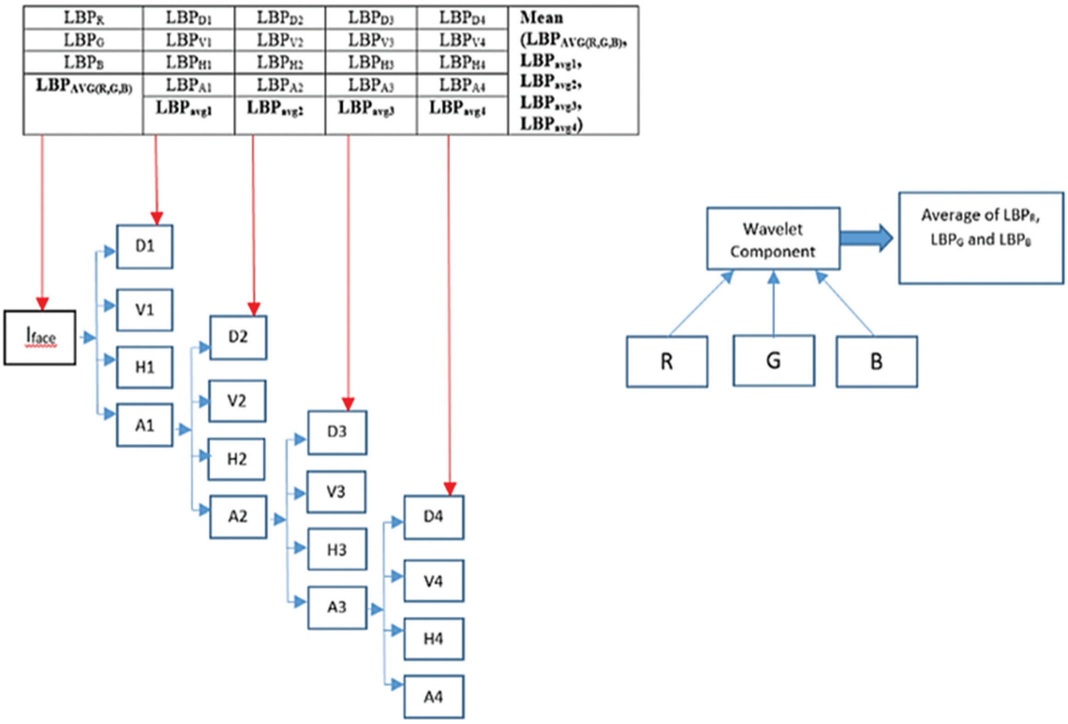


Fig. 8. Haar-wavelet-based LBP feature extraction

- Histogram of Gradient (HoG)-Based Features

HoG-based features are extracted from the Iface image and all its channels (R, G, and B) using a cell size of [16 16]. Later the HoG features were averaged to obtain the final set of features with dimension 1296. The description of conventional features with their respective dimensions is depicted in Table 1 below.

Table 1. Conventional features and respective dimension

Descriptors	Type of I_{face} image	Dimension
HoG	color	1296
Gabor	grayscale	640
Haar-Wavelet-Based LBP	color	59
RGB & Lab color space-based Histograms	color	32
Wavelet-based	grayscale	24
I order Statistical	grayscale	5
GLCM	grayscale	4

4. RESULTS AND DISCUSSIONS

The performance metric that is used to evaluate the performance of the proposed Spoof detection framework is classification accuracy. It is simply computed by calculating the ratio of samples correctly classified to total samples. It is expressed by the following expression (21):

$$\text{Accuracy (\%)} = \frac{\text{Number of samples correctly classified}}{\textit{Total test samples}} \times 100 \quad (21)$$

Experiments conducted on 4000 (80x50) authentic set images and 9950 (199x50) spoofed images showed that the machine learning-based SVM classifier using the conventional features achieved 99.89% training accuracy. We selected 75% of samples from each subject from both categories for training the SVM with a Gaussian kernel and 'SMO' solver. The remaining 25% of samples from each category were used for assessment which were selected randomly each time. The SVM was trained and tested 50 times and the mean accuracy was considered. The targets assigned to both classes were 0 and 1 respectively for fake and authentic samples. As seen from Table 1, more than 2000 features with different descriptors were used to represent a single image from the dataset.

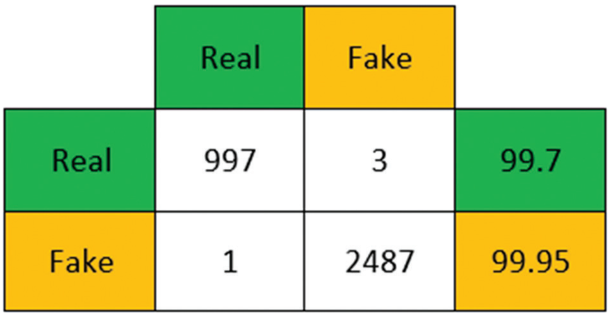
The efficacy of the conventional features can be seen from the performance of the proposed MLbFA model which is shown in Table 2. We compared the proposed MLbFA model performance with other state-of-the-art competing models utilizing a similar type of dataset involving a Replay and presentation attack. As seen from Table 2, several feature-based, machine learning-based approaches and deep learning-based techniques are used by researchers to differentiate the real and spoofed classes. Though they have used different datasets and volumes of samples, the objective is to devise a solution that can classify real images from unauthentic faces.

Agarwal A. et al. [27] used a simple approach to verify the two-class samples based on the weighted sum over SVM fusion obtained from the Haralick features. The weighted sum rule fusion model was used over the conventional features obtained on the color channels of the original and the face-detected image. The authors subjected the color channels to redundant discrete wavelet transform and extracted Haralick features. The features were combined using SVM fusion and based on the weighted sum the decision was considered.

The author obtained a remarkable accuracy of 99.08% (error=0.92%). The authors in [36] utilized an RGCS-ConvNeXt using a convolutional neural network and obtained a significant accuracy of 99.25% whereas the hybrid network CNN-VGG16 introduced in [38] outperformed all previous techniques improving the classification to 99.50%. The author used three different approaches and obtained the best results using the CNN-VGG16-based features and classifying them with machine learning. The features were obtained using the HSV and YUV color space formats. The best performance for this model is limited to the scarcity of datasets required for deep-learned models.

The authors claimed that handcrafted features are unable to obtain sufficient representation for the images owing to satisfactory performance. The superior result using our proposed MLbFA model thus verifies that the features representing the dataset images for spoof detection are efficient and informative. We outperform all other anti-spoofing techniques with a simpler but efficient spoof detection mechanism. The simplicity is regarding the use of conventional features and SVM instead of deep-learned models with complex architectures.

The confusion matrix corresponding to one of the iterations is shown in Fig. 9. The number of test samples (feature vector) for real and fake images correspondingly were 1000 and 2488 (25%). The SVM achieved 100% accuracy while training and classified the test samples with an accuracy of 99.82%. The individual accuracies are 99.7% and 99.95% respectively.



	Real	Fake	
Real	997	3	99.7
Fake	1	2487	99.95

Fig. 9. Confusion Matrix for an iteration

The reason why more real samples are affected is associated with the class imbalance. The number of samples in the fake category are higher than the real samples. Therefore, even though the SVM trained accurately using the training set, few samples which are ambiguous due to illuminations, backgrounds etc. are mostly aligned towards the higher class. In a similar way, accuracies are obtained using random test samples and the mean accuracy is computed.

The significant part of the proposed work is the preprocessing stage which involves contrast measurement and correction along with edge-preserving filter operation using the Beltrami filter and the conventional coarse (GLCM, Color, First order Statistical, and wavelet features) and fine (Gabor, HoG and LBP features) quality features. The parallel process of contrast correction and filtering and then averaging resultant images mitigated the effect of uneven illumination and uplifted the edges in the original image without much loss. The feature extraction operators were able to perform their duties independently and each of them contributed positively to obtain a robust feature set. Reducing the dimension of the features and the samples improved the performance in terms of time complexity and computational complexity of the classifier.

Table 2. Comparative analysis of the proposed MLbFA model with other competing anti-spoofing techniques

Method	Year	Dataset	Accuracy
Score fusion of Partition images [27]	2017	TIFADB	99.08%
GFA-CNN [28]	2020	Siw	95.02%
NAS-FAS [29]	2021	MSU-MFSD	95.85%
Morphological SVM [30]	2021	FPAD	97.21%
Edge-Net Autoencoder [31]	2021	IDIAP	99.87%
Deep CNN [32]	2022	IDIAP	98.21
EBDG [33]	2022	MSU-MFSD	97.17
CNN [34]	2023	IDIAP	98.36
IADG [35]	2023	MSU-MFSD	98.19
RGCS ConvNeXt [36]	2024	Siw	99.25
UCDCN [37]	2024	Replay Attack	99.18
CNN-VGG16 HSV LUV [38]	2024	NUAA Imposter	99.5
Proposed MLbFA	2024	IDIAP	99.98

In the second part, we acquired real-time videos from two real subjects and their photos. The videos were converted to frames and further partitioned into two sets. A set containing 100 frames sampled at 10 Hz from the videos was added to the training and the remaining 100 frames were added to the test set. The code was modified and all the IDIAP dataset images (4000 (authentic) +9950 (unauthentic) =13950) were provided for training the SVM. The dimension of the features was reduced using the PCA algorithm. The test set samples of subjects were projected using the coefficients and mean of the training samples obtained using the PCA algorithm after their features were extracted. It was observed that all the test set samples were accurately classified by the SVM.

5. CONCLUSIONS

Although tremendous advancements are being carried out to enhance the capability of deep networks and the success stories of deep-learned networks in various recent applications are known, the work proposed in this article uses a simpler handcrafted-based approach for face-antispoofing using a lightweight machine learning classifier. The selected descriptors to extract significant features efficiently possess the capability to represent the face images more informatively and thus can be classified more accurately. These features exhibit dynamic biometrical traits and can used for low dataset images and unbalanced dataset samples. Thus the proposed MLbFA model is more proficient which consolidates the advantages of handcrafted features and supervised learning with lower complexity. The result showed that the proposed MLbFA model achieved remarkable performance with a simpler feature extraction mechanism and classification.

The work will be extended for cross-dataset test samples. It can be tested for other types of attacks incorporating a fusion of conventional and blind features.

6. REFERENCES:

[1] B. Chen, X. Qi, Y. Zhou, G. Yang, Y. Zheng, B. Xiao, "Image splicing localization using residual image and residual-based fully convolutional network", Journal of Visual Communication and Image Representation, Vol. 73, 2020, p. 102967.

[2] Y. Tadmor, D. Tolhurst, "Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes", Vision Research, Vol. 40, No. 22, 2020, pp. 3145-3157.

[3] Md R. Hasan, S. M. H. Mahmud, X. Y. Li., "Face Antispoofing using texture-based techniques and filtering methods", Journal of Physics: Conference Series, Vol. 1229, 2019.

[4] C. Xin, W. Hongfei, Z. Jingmei, C. Hui, Z. Xiangmo, J. Yilin, "DTFA-Net: Dynamic and Texture Features Fusion Attention Network for Face Antispoofing", Complexity, Vol. 2020, 2020, p. 5836596.

[5] P. Keyurkumar, H. Hu, J. A. Kumar, "Secure Face Unlock: Spoof Detection on Smartphones", IEEE Transactions on Information Forensics and Security, Vol. 11, No. 10, 2016, pp. 2268-2283.

[6] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, R. Singh, "Computationally efficient face spoofing detection with motion magnification", Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23-28 June 2013, pp. 105-110.

[7] G. Pan, L. Sun, Z. Wu, S. Lao, "Eye blink-based anti-spoofing in face recognition from a generic web camera", Proceedings of the 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14-21 October 2007, pp. 1-8.

[8] I. Chingovska, A. Anjos, S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing", Proceedings of the IEEE International Conference of Biometric Special Interest Group, Darmstadt, Germany, 6-7 September 2012, pp. 1-7.

[9] J. Maatta, A. Hadid, M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis", Proceedings of the International Joint Conference on Biometrics, Washington, DC, USA, 11-13 October 2011, pp. 1-7.

[10] W. Bao, H, Li, N. Li, W. Jiang, "A liveness detection method for face recognition based on optical flow field", Proceedings of the International Conference on Image Analysis and Signal Processing, Linhai, China, 11-12 April 2009, pp. 233-236.

[11] M. De Marsico, M. Nappi, D. Riccio, J. L. Dugelay, "Moving face spoofing detection via 3D projective invariants", Proceedings of the International Conference on Biometrics, New Delhi, India, 29 March - 1 April 2012, pp. 73-78.

[12] D. Yuting, Q. Tong, X. Ming, Z. Ning, "Towards face presentation attack detection based on residual color texture representation", Security and Communication Networks, 2021, p. 6652727.

[13] B. Zinelabidine, K. Jukka, H. Abdenour, "Face Antispoofing using Speeded-Up Robust Features and Fisher Vector Encoding", IEEE Signal Processing Letters, Vol. 24, No. 2, 2017, pp. 141-145.

[14] F. Peng, L. Qin, M. Long, "CCoLBP: Chromatic Co-Occurrence of Local Binary Pattern for Face Presentation Attack Detection", Proceedings of the 27th International Conference on Computer Communication and Networks, Hangzhou, China, 30 July - 2 August 2018, pp. 1-9.

[15] F. Peng, L. Qin, M. Long, "Face presentation attack detection based on chromatic co-occurrence of local binary pattern and ensemble learning", Journal of Visual Communication and Image Recognition, Vol. 66, 2020, p. 102746.

[16] K. Mohan, P. Chandrashekhar, K. V. Ramanaiah, "Object-specific face authentication system for

liveness detection using combined feature descriptors with fuzzy-based SVM classifier", International Journal of Computer Aided Engineering and Technology, Vol. 12, No. 3, 2020, pp. 287-300.

[17] A. Mittal, P. Kaur, A. Oberoi, "Hybrid Algorithm for Face Spoof Detection", International Journal for Research in Applied Science and Engineering Technology, Vol. 10, No. 2, 2022.

[18] M. Alshaikhli, O. Elharrouss, S. Al-Maadeed, A. Bouridane, "Face-fake-net: the deep learning method for image face AS detection", Proceedings of the IEEE 9th European Workshop on Visual Information Processing, Paris, France, 23-25 June 2021.

[19] J. Zhou, K. Shu, P. Liu, J. Xiang, S. Xiong, "Face AS Based on Dynamic Color Texture Analysis Using Local Directional Number Pattern", Proceedings of the 25th International Conference on Pattern Recognition, Milan, Italy, 10-15 January 2021, pp. 4221-4228.

[20] C. Pei, Q. Hui-min, "Face AS algorithm combined with CNN and brightness equalization", Journal of Central South University, Vol. 28, 2021, pp. 194-204.

[21] Balamurali K, Chandru S, Muhammed Sohail Razvi and V. Sathiesh Kumar, "Face Spoof Detection Using VGG-Face Architecture", Journal of Physics: Conference Series, Vol. 1917, 2021.

[22] Z. Ming, M. Visani, M. M. Luqman, J.-C. Burie, "A Survey on AS Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices", Journal of Imaging, Vol. 6, No. 12, 2020, p. 139.

[23] Md R. Hasan, S. M. H. Mahmud, X. Y. Li, "Face AS Using Texture-Based Techniques and Filtering Methods", Journal of Physics: Conferences Series, Vol. 1229, 2019.

[24] P. T. de Freitas, K. Jukka, A. Andre, M. Jose De Mario, H. Abdenour, P. Matti, M. Sebastien, "Face liveness detection using dynamic texture", EURASIP Journal on Image and Video Processing, Vol. 2, 2014.

[25] A. Wetzler, R. Kimmel, "Efficient Beltrami Flow in Patch-Space. In Scale Space and Variational Methods in Computer Vision. SSVM 2011", Lecture Notes in Computer Science, Vol. 6667, Springer, 2012, pp. 134-143.

[26] M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification", Expert Systems with Applications, Vol. 42, No. 21, 2015, pp. 7905-7916.

[27] A. Agarwal, R. Singh, M. Vatsa, A. Noore, "Boosting Face Presentation Attack Detection in Multi-Spectral Videos through Score Fusion of Wavelet Partition Images", Frontiers in Big Data, Vol. 5, No. 22, 2022, p. 836749.

[28] X. Tu, Z. Ma, J. Zhao, G. Du, M. Xie, J. Feng, "Learning generalizable and identity-discriminative representations for face anti-spoofing", ACM Transactions on Intelligent Systems and Technology, Vol. 11, No. 5, 2020, pp. 1-19.

[29] Z. Yu, J. Wan, Y. Qin, X. Li, S. Z. Li, G. Zhao, "NASFAS: Static-dynamic central difference network search for face anti-spoofing", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, No. 9, 2021, pp. 3005-3023.

[30] T.-H. Lai, C.-Y. Peng, C.-L. Chou, "Fast Face Presentation Attack Detection in Thermal Infrared Images Based on Morphological Filtering", International Journal of Network Security, Vol. 25, No. 2, 2023, pp. 185-193.

[31] S. D. Thepade, M. R. Dindorkar, P. R. Chaudhari, S. V. Bang, "Enhanced Face Presentation Attack Prevention Employing Feature Fusion of Pretrained Deep CNN Model and Thepade's Sorted Block Truncation Coding", International Journal of Engineering, Transactions A: Basics, Vol. 36, No. 04, 2023, pp. 807-816.

[32] A. H. Alharbi, S. Karthick, K. Venkatachalam, M. Abouhawwash, D. S. Khafaga, "Spoofing Face Detection Using Novel Edge-Net Autoencoder for Security", Intelligent Automation & Soft Computing, Vol. 35, No. 3, 2023, pp. 2773-2787.

[33] Z. Du, J. Li, L. Zuo, L. Zhu, K. Lu, "Energy-based domain generalization for face anti-spoofing", Proceedings of the 30th ACM International Conference on Multimedia, Lisboa, Portugal, 10-14 October 2022, pp. 1749-1757.

[34] X. Chen, J. Zhou, X. Zhao, H. Wang, Y. Li, "A presentation attack detection network based on dynamic convolution and multilevel feature fusion with security and reliability", Future Generation Computer Systems, Vol. 146, 2023, pp. 114-121.

[35] Q. Zhou, K.-Y. Zhang, T. Yao, X. Lu, R. Yi, S. Ding, L. Ma, "Instance-aware domain generalization for face anti-spoofing", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17-24 June 2023, pp. 20453-20463.

[36] H. Qi, R. Han, Y. Shi, X. Qi, "A Novel High-Performance Face Anti-Spoofing Detection Method", IEEE Access, Vol. 12, 2024, pp. 67379-67391.

[37] J. Zhang et al. "UCDCN: a nested architecture based on central difference convolution for face-antispoofing", Complex and Intelligent Systems, Vol. 10, 2024, pp. 4817-4833.

[38] M. Prasad, S. Jain, P. Bhanodia, A. Priya, "Influence of Standalone and Ensemble Classifiers in Face Spoofing Detection using LBP and CNN Models", European Journal of Electrical Engineering and Computer Science, Vol. 8, No. 2, 2024, pp. 17-30.