Hindawi
Journal of Sensors
Volume 2021, Article ID 5047808, 18 pages
https://doi.org/10.1155/2021/5047808

Research Article

IoT Cloud-Based Framework for Face Spoofing Detection with Deep Multicolor Feature Learning Model

Sajad Einy,[1,2] **Cemil Oz**,[1] **and Yahya Dorostkar Navaei**[3]

[1]*Computer Engineering Department, Sakarya University, Turkey*
[2]*Application and Research Center for Advanced Studies, Istanbul Aydin University, Turkey*
[3]*Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran*

Correspondence should be addressed to Yahya Dorostkar Navaei; y.dorostkar@qiau.ac.ir

Received 12 May 2021; Revised 4 July 2021; Accepted 3 August 2021; Published 31 August 2021

Academic Editor: Yunze He

Copyright © 2021 Sajad Einy et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A face-based authentication system has become an important topic in various fields of IoT applications such as identity validation for social care, crime detection, ATM access, computer security, etc. However, these authentication systems are vulnerable to different attacks. Presentation attacks have become a clear threat for facial biometric-based authentication and security applications. To address this issue, we proposed a deep learning approach for face spoofing detection systems in IoT cloud-based environment. The deep learning approach extracted features from multicolor space to obtain more information from the input face image regarding luminance and chrominance data. These features are combined and selected by the Minimum Redundancy Maximum Relevance (mRMR) algorithm to provide an efficient and discriminate feature set. Finally, the extracted deep color-based features of the face image are used for face spoofing detection in a cloud environment. The proposed method achieves stable results with less training data compared to conventional deep learning methods. This advantage of the proposed approach reduces the time of processing in the training phase and optimizes resource management in storing training data on the cloud. The proposed system was tested and evaluated based on two challenging public access face spoofing databases, namely, Replay-Attack and ROSE-Youtu. The experimental results based on these databases showed that the proposed method achieved satisfactory results compared to the state-of-the-art methods based on an equal error rate (EER) of 0.2% and 3.8%, respectively, for the Replay-Attack and ROSE-Youtu databases.

1. Introduction

Nowadays, the Internet of Things (IoT) affects human lives in a wide range of technology from smart homes to smart cities. An enormous number of IoT devices are utilized for collecting and analyzing information for different reasons, such as healthcare, security, and management. According to the estimation of scientifics, around 90% of storing data would be useless [1]. Therefore, the researchers proposed [1] utilizing the edge devices in the architecture of applications or services for cloud computing. In this way, the data can be analyzed and filtered in edge devices and send more enhanced data for processing in the cloud. For example, the deployed sensors for traffic monitoring can be also utilized for fire detection with low-cost and low-performance devices. However, IoT-based systems are faced with different problems such as security threats from the Internet. For instance, let us consider an IoT-based healthcare application which contains critical information such as blood sugar level and blood pressure. The authentication system for data communication through wireless channels should be secured for protecting critical information of clients. Biometric authentication can be utilized for identifying a person in wireless communication. This authentication requires using personal attributes, such as speech, face, fingerprints, palmprint, gait, and iris [2]. This kind of authentication is based on a comparison between the physical aspect of the client that is collected with the help of different sensors and a copy that was stored. The physiological information of clients is more reliable when compared to knowledge-based or token-based methods because this information is unique and not shareable. For this reason, IoT-based cloud computing systems for authentication of clients applied their biometric information.

For instance, Kumari and Thangaraj [3] proposed a feature selection technique in biometric authentication using a cloud framework. In another similar study, Shakil et al. [4] proposed a biometric authentication system and data management application for security of healthcare data in the cloud. Also, Vidya and Chandra [5] proposed a multimodal biometric authentication system based on entropy-based local binary pattern feature description technique for cloud computing. Additionally, Masud et al. [6] proposed a deep learning-based approach for face recognition in IoT environments. Face recognition systems have achieved significant interest in many applications such as cell phones' and laptops' authentication or registration systems at places such as online exam centers and airports [1]. These kinds of security systems in the Big Data analytics platform are a topic of concern for real-time applications. Consider the scenario when a person is to be recognized in an airport for registration or a student is attending an online exam. In these scenarios and other similar conditions, the camera captures images of the face continuously and sends these data for processing in the cloud environment. Based on meaningful information of face image, a certain person can easily be identified. Nevertheless, these kinds of authentication and registration systems are vulnerable to different types of attacks. For improving the security of biometric authentication systems, various methods and models are proposed.

For example, Ali et al. [1] proposed a multimodal biometric authentication system using an encryption method for protecting the privacy of biometric information in the IoT-based cloud environment. In another study, Gomez-Barrero et al. [2] proposed a framework for the protection of the privacy of multibiometric templates with an encryption method. However, the aforementioned methods are designed for protection based on man-in-the-middle attacks in wireless communication. According to the literature, face spoofing attacks in IoT cloud environments are not discussed and studied yet. The main objective of this study is to present an IoT cloud-based framework for protecting client's information from face spoofing attacks. In a face spoofing attack, the intruder bypasses the authentication system by presenting a fake face of the victim. Due to this threat, robust and stable face Presentation Attack Detection (PAD) methods must be developed and designed. Face spoofing attacks may be classified into four main groups: print, display, replay, and mask attacks [7].

According to the types of sensors for detection of these kinds of attacks, different algorithms are proposed [9–11]. Generally, light field camera sensors are more popular compared to other sensors such as infrared and thermal ones [8] or multibiometric fusion systems [9] because this additional equipment increases the cost of authentication systems. In this case, many researchers investigate feature-based methods. These kinds of spoofing detection methods attempt to extract discriminative features to recognize the genuine user from a fake face. For example, in print, display, and mask attacks, facial liveness features such as lip movement, head movement, and eye blinking can help recognize spoofing attacks. Furthermore, detection of replay attacks is more challenging because they contain this kind of liveness feature [7]. In some cases, the intruder applies liveness features in a mask attack by cropping the lip and eye area from a mask, which shows that liveness features alone cannot detect spoofing attacks properly. Replay display and printed attack images contain some noise and defects because of recapturing of information by a camera. During recapturing of information, the fake face loses the high-frequency information by getting affected in terms of the texture and color information of images, and these features can help distinguish a genuine person and a recaptured face image. Especially in printing and displaying attacks, during recapturing of information, some defects and noises appear in the spoofing face image. These artifacts lead to inadequate color reproduction in comparison to real biometric samples [10]. RGB is the commonly employed color space for sensing and displaying color images on many devices. Nevertheless, this color space in image analysis is inadequate due to the high correlation between the red, green, and blue color components and incomplete separation of the luminance and chrominance information [11]. Therefore, a different color space may help extract discriminative features for extraction of liveness cues of skin tones for detection of live and fake images. Therefore, image texture analysis based on different color spaces has attracted the consideration of research areas in the field of face spoofing attacks [11, 12]. By the success of deep learning algorithms in the field of computer vision and multimedia analysis, deep texture analysis-based algorithms have been employed in face spoofing problems. Nevertheless, deep learning-based face spoofing detection algorithms are faced with some problems such as few numbers of spoofing data and lack of diversity of scenarios which make it difficult to train a deep network [13, 14]. Additionally, IoT-based authentication systems encountered several difficulties such as storing or processing in a real-time manner [6].

To address these problems, we presented a novel approach based on hybrid convolutional neural network (CNN) models on different color spaces for IoT-based cloud computing. The proposed deep learning approach utilized three pretrained models in different color spaces for extracting luminance and chrominance information which are useful in recognition of spoofing face images. Due to extracted robust and discriminative features from a single image, this proposed model can achieve satisfactory results with less training dataset. This advantage of the proposed approach helps to decrease the storing training data in cloud computing which tackles one of the major problems of cloud computing systems. To the best of our knowledge, for the first time, in this paper, an IoT security framework is proposed for face spoofing detection. Extensive experimental analysis was conducted based on two challenging public access spoofing databases with their predefined evaluation protocols for comparison of our proposed approach against state-of-the-art methods. These experimental results show that our proposed approach outperforms all existing deep-based methods among state-of-the-art methods based on benchmark databases. In addition, experimental results show that the proposed approach can achieve stable results with less training dataset compared to benchmark deep learning models.

In light of this information, the main contributions of this paper are presenting an IoT security framework for face spoofing detection which achieved significant results compared to the state of algorithms based on two public databases. Also, the proposed approach achieved stable results with less training dataset compared to benchmark deep learning models.

This paper is briefly organized as follows: In Section 2, short information about types of existing systems and related works on face spoofing methods are available. In Section 3, the methodology of the proposed approach is briefly presented. In Section 4, the experimental results and state-of-the-art algorithms with benchmark databases and protocols are presented. As the final section, conclusion statements are provided in Section 5.

2. Related Work for Face Spoofing Methods

Recently, a lot of face spoofing detection algorithms have been proposed [1–7], based on different cues and attacks. Based on our prior knowledge, the algorithms can be categorized into four different groups: texture analysis, motion analysis, image quality analysis, and hardware-based methods.

2.1. Texture-Based Methods. Face liveness detection algorithms based on texture analysis usually recognize the effects of illumination limitations of a printer or any other device during display, such as printing failures, blurring, and other effects. The RGB color space, as discussed in Section 1, cannot clearly present features regarding illumination and chrominance. In this case, a previous study [12] proposed a deep learning system based on the RGB, HSV, and YCbCr color spaces. In the paper, the CompactNet model was proposed as a layer-by-layer progressively generated color space. Additionally, features of spoofing databases are extracted by a pretrained feature extractor model. Researchers [11] proposed a color feature descriptor method based on different color spaces. In this method, information on the luminance and chrominance channels was extracted by a low-level feature descriptor. Due to the impact of a smaller number of databases in face spoofing detection on training deep learning methods and overfitting problems, researchers investigate the extraction of discriminative and deep features. For instance, a study [15] proposed a perturbation layer (low-level deep features) to extract the deep features of a convolutional neural network (CNN) for classification. Another study [16] presented an adaptive fusion of convolutional feature models to learn the features of face images, and a deep autoencoder was utilized for generating a face image to detect spoofing face images. Some authors [7] proposed a Spatial Pyramid Coding Microtexture (SPMT) feature extractor with a deep learning system for detection of liveness cues and employed the Single Shot Multibox Detector (SSD) as an end-to-end face spoofing detection model. Besides the aforementioned color-based deep learning methods, some methods presented local binary pattern- (LBP-) based feature descriptors for spoofing detection. For instance, a hybrid method was proposed [17] based on the Chromatic Cooccurrence of Local Binary Pattern (CCoLBP) and Ensemble Learning (EL) algorithms. In the case of reducing the parameters of CNN models and extraction of deep features, an end-to-end learnable LBP network was proposed [18]. A previous study [19] proposed an algorithm by integrating the LBP descriptor with a modified convolution neural network that extracted deep texture. For extraction of discriminative features of presentation attacks, the Extended Local Ternary Corelation Pattern (ELTCP) feature extraction method was proposed [20]. This feature descriptor with extraction of spatial information of an image in multiple directions achieved robust results on presentation attacks. In recent years, with increasing attention to 3D face spoofing attacks, several studies have been devoted to recognizing 3D mask attacks. For instance, the 3D wax face attacks [21] approach is proposed with a convolutional neural network based on the Residual Attention Network (RAN) for 3D face spoofing detection. In another similar study, a multichannel CNN [22] approach with a one-class Gaussian mixture model is proposed for the detection of 2D and 3D attacks. Another study [23] presented a shading-based 3D feature description method to extract discriminative and robust 3D features from the face image. In another study, researchers proposed [24] a face spoofing framework with the help of convolutional autoencoders for the detection of 3D mask attacks. Another study [10] investigated various factors of affection of acquisition conditions and devices with different resolutions on the generalization of color texture features for spoofing detection. In this light, another possibility seems to be analyzing image textures based on deep features from multiple color spaces, which is proposed in this paper. The experimental results show that our proposed algorithm is superior in color texture extraction and classification over state-of-the-art methods.

2.2. Motion Analysis Algorithms. Among texture recognition techniques, motion-based analysis also plays an important role in spoofing detection. For instance, a study [25] proposed a motion-based analysis approach based on rigid and nonrigid facial movements. The proposed system extracted motion cues such as face movement, lip movement, and hand shaking and classified them into natural and fake motions. In another study [8], an undirected conditional random field in video processing was proposed for the detection of eye blinking. Other researchers [26] proposed a dynamic mode decomposition pipeline with SVM and LBP. This algorithm extracted facial dynamic information in videos as an image sequence.

2.3. Image Quality Analysis. In spoofing attacks, the image quality is mostly reduced due to the image being reproduced. Based on this inability of devices, some methods have been proposed. For instance, in a previous study [27], an algorithm was proposed where real and fake face images were determined by analysis and comparison of both reflections taken from an LCD screen. In another study [28], it was posited that it is possible to differentiate a fake image from a real one by analyzing the noise signatures with the Fourier spectrum.

2.4. Hardware-Based Analysis. Researchers [29] proposed video-based stereo face antispoofing recognition systems. In this approach, for learning a dynamic disparity map, a CNN classifier with a disparity layer was proposed. In

another study [30], it was proposed to assign a light field to traditional HOG which was utilized for gathering texture information from 2D images and Light Field Histogram Of Gradient (LFHOG).

Apart from the mentioned proposed systems based on single cues, some methods have been proposed based on multicue approaches. For instance, another study [31] proposed a multicue face spoofing detection framework involving image quality analysis by employing the Shearlet method and motion analysis by utilizing the dense optical flow method. In this study, the extracted multicue features were fused and classified with a deep neural network.

3. Proposed IoT-Based Framework Face Spoofing Detection

The smart city framework contains multiple components such as smart devices, high-speed wireless networks, and cloud servers, as presented in Figure 1. The captured face images by IoT devices are analyzed and preprocessed with edges. The preprocessing section with edges and smart devices included Viola and Jone's [32] face detection algorithm for extracting face images and sending more enhanced data to optimize the resource of the cloud. Then, the captured faced images are continually sent to a cloud environment using wireless technology. In the cloud section, several Virtual Machines (VMs) work in a parallel mode. These VMs by employing a deep learning approach recognize spoofing attacks.

Before feeding the face image to the deep model for classification in cloud computing environments, RGB color space is transformed to the HSV and YCbCr color spaces. Three parallel pretrained models are utilized in the proposed deep learning approach. Based on the literature, because of the small number of data and lack of scenarios in controlled environments, it is quite hard to train CNN models from scratch and achieve a stable and high-performance model. In this case, we utilized the VGG-face [33] model in the RGB color space for face spoofing detection [14, 18]. In addition, the transformed images of the HSV and YCbCr color spaces are trained by the VGG16 [34] model individually on the cloud side. After fine-tuning models by a different color space, the features of the last fully connected layer which consists of 4096 features for each deep model are extracted. These features are combined and then selected by employing the Minimum Redundancy Maximum Relevance (mRMR) feature selection algorithm. These selected features are classified with the help of different classification algorithms such as linear regression (LR), Support Vector Machine (SVM), Linear Discriminative Analysis (LDA), and K Nearest Neighborhood (KNN) for detection of the spoof image, as presented in Figure 2.

Suppose a scenario where a student wants to access an online exam. A smart device such as a smart phone or computer captures the student's facial image and sends this image to the cloud using 5G wireless technology. In the cloud server, by employing the face spoofing image database and deep learning method, a deep feature set of face image is extracted in three different color spaces. These combined feature sets contain various aliveness keys from face skin tones which help to detect face spoofing in the online exam scenario. The proposed method is tested and evaluated based on two public access databases, namely, Replay-Attack and ROSE-Youtu. The Replay-Attack database is captured by a MacBook laptop webcam and the ROSE-Youtu database captured by Huawei, iPhone 5s, ZTE, and Hasee smart phone.

3.1. Color Space Transform. RGB is a common color space for many devices and sensors for displaying and sensing color images. Nevertheless, this color space is quite limited for analyzing images because of the high correlation of red, green, and blue colors and incomplete separation of the luminance and chrominance information.

In this case, for the detection of recapping artifacts in spoofing databases, different color spaces are utilized [12]. HSV and YCbCr in addition to RGB provide robust features to detect different liveness cues from face skin tones. Both the HSV and YCbCr color spaces provide color texture information such as the luminance and the chrominance components. In the HSV color space, the H and S define the hue and saturation dimensions for presenting the chrominance information, and V defines the value dimension for presenting the luminance information of images. The YCbCr space separates RGB into luminance (Y), Chrominance Blue (Cb), and Chrominance Red (Cr). The HSV and YCbCr spaces provide discriminative color-based texture from face skin tones in different spoofing attacks [11, 12]. Figure 3 presents different color spaces on the Replay-Attack database for both live and fake face images.

3.2. Convolutional Neural Networks. Convolutional neural networks (CNNs) are designed and developed to automatically learn the spatial hierarchies of features with the help of back propagation algorithms [35]. CNNs are designed based on multiple layers of neurons which mainly include multiple basic structural blocks such as the convolution, pooling, and fully connected (FC) layers. Each convolutional layer contains a set of filters whose sizes can be 3×3, 5×5, or 7×7 pixels. Therefore, each convolutional layer, by applying a filter, creates the input of the next layer [36]. The results of this convolution process are activation maps which contain local distinctive features. Based on Equation (1), the output of $Y_i^{(l-1)}$ of the L layer contains $m_1^{(l)}$ feature maps with sizes of $m_2^{(l)} \times m_3^{(l)}$. In this equation, $B_i^{(l)}$ and $k_{I,j}^{(l)}$ represent, respectively, the basis matrix and the filter size for the ith feature map [37]:

$$Y_i^{(l)} = f\left(B_i^{(l)} + \sum_{j=1}^{m_i^{(l-1)}} k_{i,j}^{(l)} \times Y_j^{(l-1)}\right). \quad (1)$$

The pooling layer reduces the spatial size of the image to reduce the number of parameters and computations in the model. This layer operates on each feature map independently to keep the image features and information intact. Each pooling layer L contains two main parameters as the spatial size of the filter $F^{(l)}$ and $S^{(l)}$ step. The input of the pooling layer is data

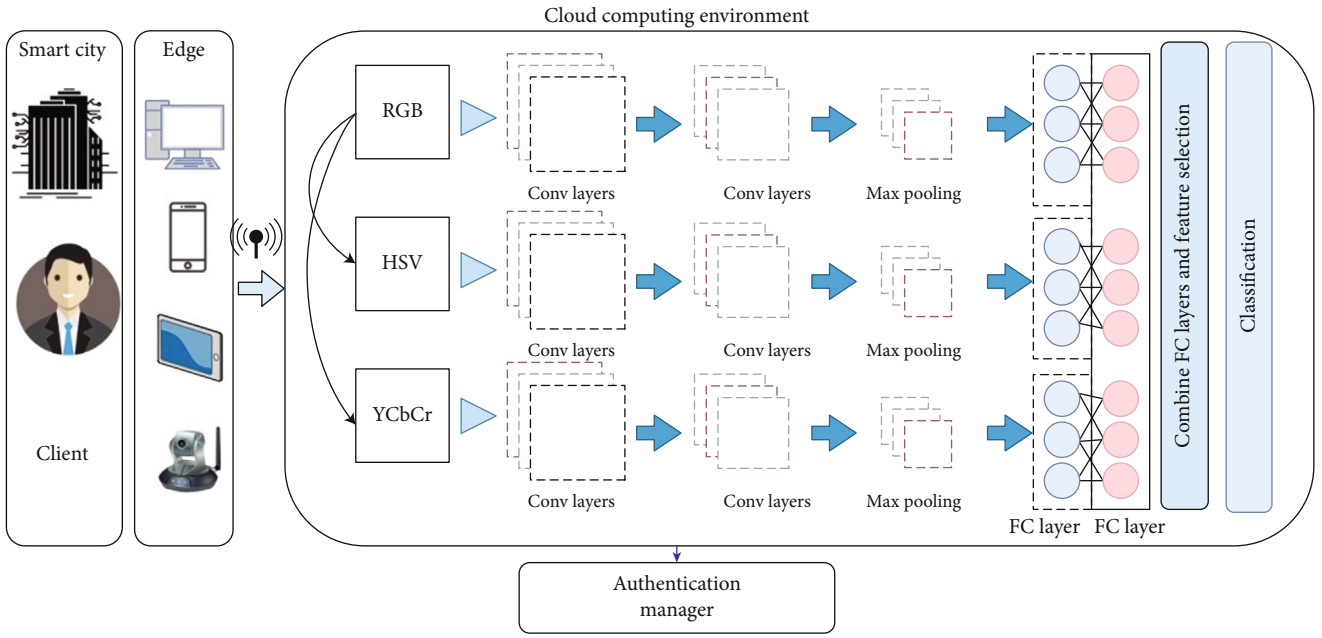


Figure 1: Proposed IoT-based framework for face spoofing detection.

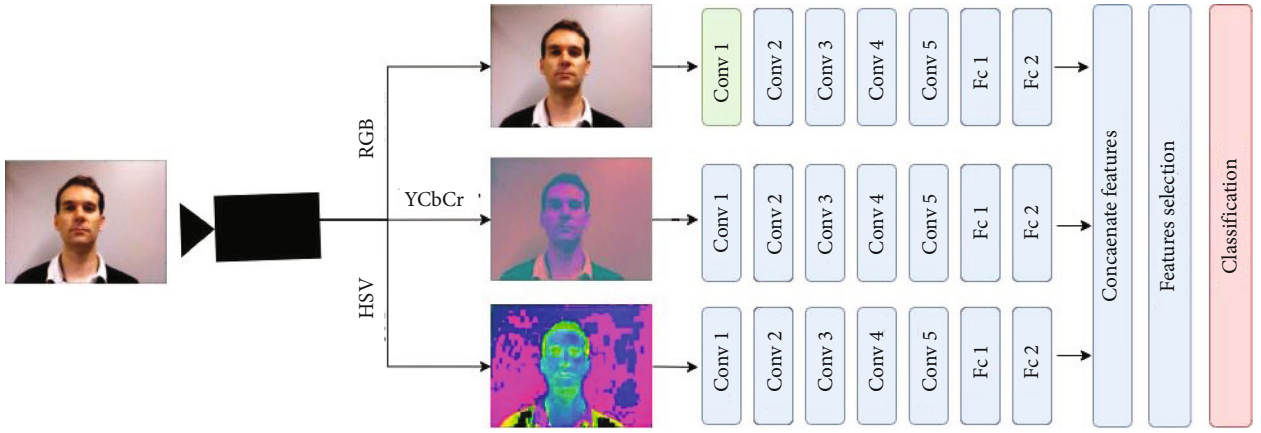


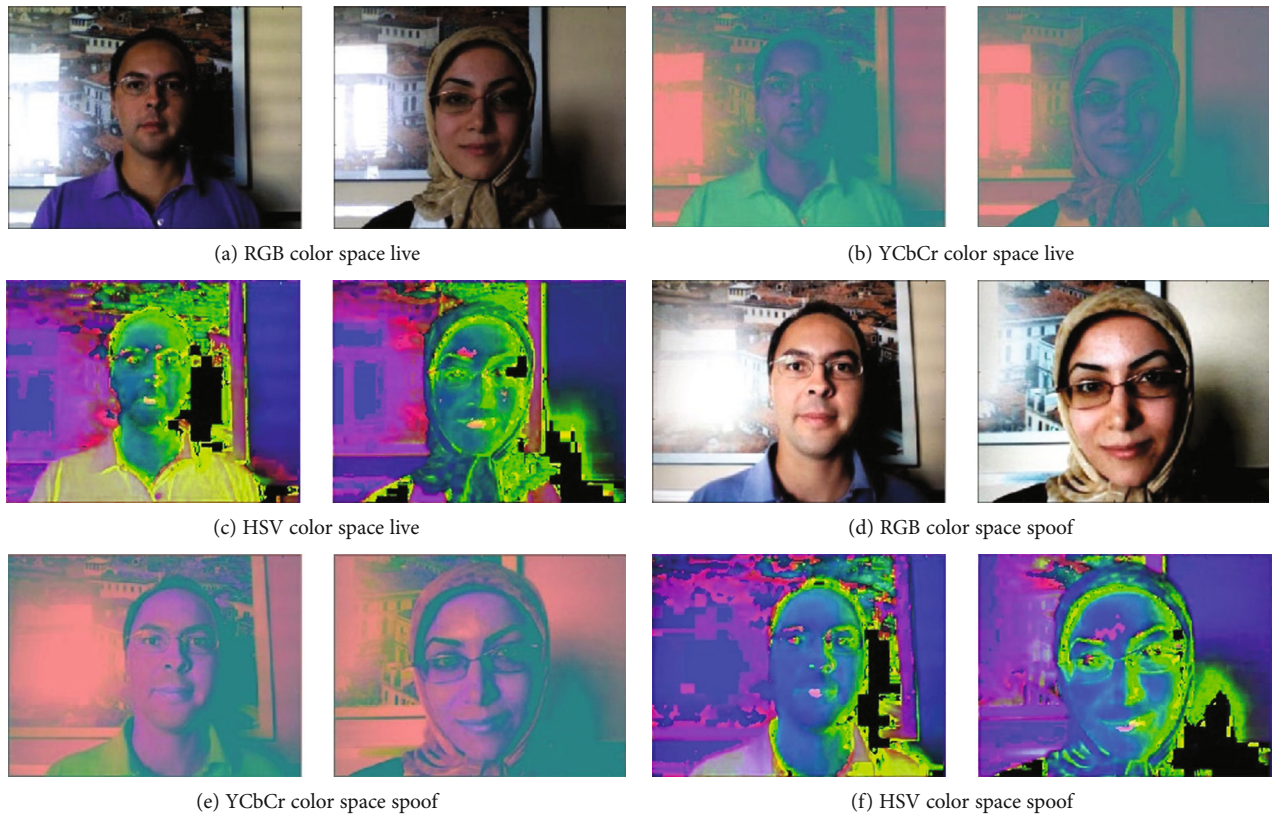
Figure 2: The architecture of deep learning approach.

with the size of $m_1^{(l-1)} \times m_2^{(l-1)} \times m_3^{(l-1)}$, and the output volume of this layer is $m_1^{(l)} \times m_2^{(l)} \times m_3^{(l)}$. Equation (2) briefly presents the operation of the pooling layer:

$$\begin{cases} m_1^{(l)} = m_1^{(l-1)}, \\ m_2^{(l)} = \dfrac{m_2^{(l-1)} - F^{(l)}}{S^{(l)}} + 1, \\ m_3^{(l)} = \dfrac{m_3^{(l-1)} - F^{(l)}}{S^{(l)}} + 1. \end{cases} \tag{2}$$

The output of feature maps of the last convolutional or pooling layer is flattened in the layer named the fully connected layer. The FC layer transforms the output of previous layers into a one-dimensional feature vector, updates the weights, and provides the latest possible values for each label [37]. These layers may be connected to a more fully connected layer which is also known as the dense layer. By employing a learning rate, every input is connected to every output. The features are extracted by the convolution layers, downsampled by the pooling layers, and mapped by the FC layer to the final output of the model. The last FC layer contains a number of nodes equal to the number of classes of classification images. Each FC layer is supported by a nonlinear function such as the ReLU function. Equation (3) presents the FC layer's processing steps by weights (W) and the $f(Z_i^{(l)})$ nonlinear function:

$$Y_i^{(l)} = f\left(Z_i^{(l)}\right) \text{ with} Z_i^{(l)} = \sum_{j=1}^{m_i^{(l-1)}} w_{i,j}^{(l)} \times y_j^{(l-1)}. \tag{3}$$



(a) RGB color space live (b) YCbCr color space live

(c) HSV color space live (d) RGB color space spoof

(e) YCbCr color space spoof (f) HSV color space spoof

Figure 3: Different color spaces based on Replay-Attack databases.

3.2.1. Pretrained Models. To modify the pretrained experiment models for face spoofing recognition, the models were fine-tuned by spoofing databases. The binary classification was utilized for spoofing detection problems and changing the output of the classification layer to two classes of spoof and real face.

After modifying the SoftMax classification layer based on the spoofing database in the training phase, the VGG16 and VGG-face models were fine-tuned based on the spoofing database. The VGG-face model is one of the popular pretrained models for face recognition systems. This model was developed by the Oxford Visual Geometry Group [33]. The model was trained by 2.6 M to face images in the RGB color space, and the default size of an input image is 224 × 224 [18]. This model contains five max pooling, thirteen convolutional layers with the rectified linear unit (ReLU) function, and three fully connected layers, namely, FC6, FC7, and FC8. The last fully connected layer (FC8) modifies from 2622 (face image classes) to 2 classes of spoof and real. The architecture of the VGG-face model is a variant of VGG16, which is trained by face images, as presented in Table 1. In this approach, the fine-tuned VGG-face and VGG-16 models based on the face spoofing database are utilized as a deep feature extractor. The deep features are taken from FC7 (seventh fully connected layer), the last layer before the output layer. The activation values of this FC layer for all models are set as default values equal to 4096 (dimensional feature vectors) for the input images.

Table 1: VGG16 architecture.

Layer	Patch size/stride	Input size
Conv × 2	3 × 3/1	64 × 224 × 224
Pool	2 × 2	64 × 224 × 224
Conv × 2	3 × 3/1	128 × 112 × 112
Pool	2 × 2	128 × 112 × 112
Conv × 3	3 × 3/1	256 × 56 × 56
Pool	2 × 2	256 × 56 × 56
Conv × 3	3 × 3/1	512 × 28 × 28
Pool	2 × 2	512 × 28 × 28
Conv × 3	3 × 3/1	512 × 14 × 14
Pool	2 × 2	512 × 14 × 14
FC	25088 × 4096	25088
FC	4096 × 4096	4096

3.3. Feature Selection. The main purpose of the mRMR method is to select the subset of features which has the most correlation with the class and reduce irrelevant and redundancy features based on mutual information [38, 39]. Measurement of the mutual information of I between two x and y attributes is defined based on

$$I(x,y) = \sum_{i,j} p\left(x_i, y_j\right) \log \frac{p\left(x_i, y_j\right)}{p(x_i)p\left(y_j\right)}, \tag{4}$$

where $p(x_i)$ and $p(y_j)$ represent the marginal probabilities and $p(x_i, y_j)$ represents the joint probabilistic distribution. Let us define each property of the equation as F_i in a K-size vector ($F_i = [F_{1i}, F_{2i}, F_{3i}, \cdots, F_{Ki}]$). In this case, the mutual information of the variables (i, j) is defined as $I(F_i, F_j)$. In order to find the best features of the selected subset, Equations (5) and (6) must be satisfied. The minimum redundancy feature is presented in Equation (8), and the maximum relevance condition is presented in Equation (6):

$$\min W, W = \frac{1}{|s|^2} \sum_{F_i F_j} I(F_i, F_j), \tag{5}$$

$$\max V, V = \frac{1}{|s|} \sum_{F_i F_j} I(H, F_i), \tag{6}$$

where H represents the class label and s shows the number of features selected. The mRMR feature set is obtained by optimizing the combination of feature selection criteria, namely, Mutual Information Difference (MID) and Mutual Information Quotient (MIQ), which are presented in

$$\begin{cases} \text{MID} = \max\ (v - w), \\ \text{MIQ} = \max\ \left(\dfrac{v}{w}\right). \end{cases} \tag{7}$$

For optimizing the MID and MIQ conditions, it is required to combine them into a single criterion function [40], as shown in the following equation:

$$f_{\text{mRMR}}(X_i) = I(H, F_i) - \frac{1}{|s|} \sum_{F_i F_j} I(F_i, F_j), \tag{8}$$

where $I(H, F_i)$ measures the relevance feature to be added for the class and $1/|s| \sum_{F_i F_j} I(F_i, F_j)$ estimates the redundancy of features with respect to previously selected s features. These selected features are classified with a linear regression classification algorithm for detection of face presentation attacks.

4. Experimental Results

The proposed method as shown in Figure 2 was compiled with an NVIDIA GeForce 4 GB graphics card (GPU). Other hardware details were Intel Core i5 3.6 GHz processor and 16 GB RAM. As presented in Table 2, these parameters were used with their default values. Additionally, the minibatch size was set as 32.

4.1. Experimental Databases

4.1.1. The Replay-Attack Database [41]. The Replay-Attack database consists of 1300 videos of 2D face attacks under different conditions. This database contains three main subgroups for training, validation, and testing folders with names of training data, development data, and test data. Two main different lighting conditions in this database were named as controlled and adverse. The controlled scenario data were collected under homogeneous backgrounds and with office lights turned on, and the adverse data were collected with more complex backgrounds and without office lights as presented in Figure 4.

4.1.2. ROSE-Youtu Face Liveness Detection Dataset [42]. This database contains a large variety of illumination conditions, cameras with different resolutions, and types of attacks such as display, print, and mask attacks. The ROSE-Youtu database contains 4225 videos with 25 subjects, and each video duration average is around 10 seconds. The ROSE-Youtu database is divided into two subsets of training and testing. The first 10 indexed units are separate for training, and the rest of the videos belong to testing. The numbers of samples from this database are presented in Figure 5.

4.2. Evaluation Metric. To measure the performance of the models, accuracy (Acc), sensitivity (Se), specificity (Sp), precision (Pr), and F-score metrics derived from the confusion matrix were used, and the formulations of the metrics were as follows [43]:

$$\begin{cases} \text{Acc} = \dfrac{(\text{TP} + \text{TN})}{(\text{TF} + \text{FN}) + (\text{FP} + \text{TN})}, \\ \text{Se} = \dfrac{(\text{TP})}{(\text{TP} + \text{FN})}, \\ \text{Pr} = \dfrac{(\text{TP})}{(\text{TP} + \text{FP})}, \\ F\text{-score} = \dfrac{(2 \times \text{TP})}{(2 \times \text{TP} + \text{FP} + \text{FN})}. \end{cases} \tag{9}$$

To evaluate our new approach against state-of-the-art methods, we applied the formula of the Half Total Error Rate (HTER) in

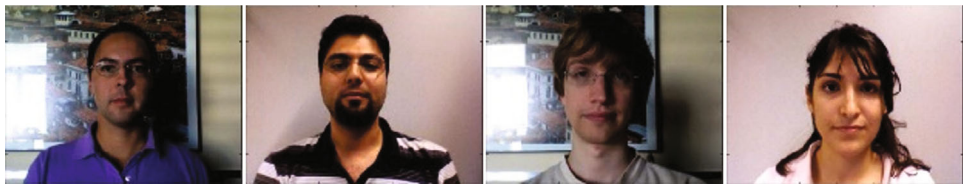
$$\text{HTER} = \frac{\text{FRR}(\mathscr{K}, \mathscr{D}) + \text{FAR}(\mathscr{K}, \mathscr{D})}{2}, \tag{10}$$

where $\text{FRR}(\mathscr{K}, \mathscr{D})$ is a false rejection rate, $\mathscr{D}$ denotes the used database, and $\mathscr{K}$ is estimated on the equal error rate (EER). In this context $\text{FAR}(\mathscr{K}, \mathscr{D})$ stands for the False Acceptance Rate.

4.3. Fine-Tuning VGG-Face Model for Face Spoofing Detection. Our face spoofing recognition approach in the first steps was based on the VGG-face model. The VGG-face model is trained by a large database of face images. As presented in Figure 6, each convolution block contains the rectified linear unit (ReLU) function and a 3×3 kernel size. Also, each convolution block contains a max pooling layer with a kernel size of 2×2. Two FC layers are set with 4096 channels with the ReLU function and batch normalization. The last FC layer contains the ReLU function, batch normalization, and the SoftMax activation function where the output of this layer presents categorical distribution over face spoofing recognition labels.

Table 2: Parameter values of the proposed approach used in this study.

Software	Optimization	Activation function	Momentum	Decay	Minibatch	Learning rate
Keras	Adam	ReLU	0.9	$1e-6$	32	0.01



(a) Live face

(b) Spoof face

Figure 4: Replay-Attack database samples for live and spoof images.

(a) Live face

(b) Spoof face

Figure 5: ROSE-Youtu face liveness detection samples for live and spoof face images.

The performance of the VGG-face model for face spoofing detection databases depends on the level of fine-tuning of the convolutional blocks. For this reason, in this test, we evaluated the effects of each pretrained convolutional block on the accuracy of the model [14]. Different models arranged based on the retrained and frozen levels of the parameters of the network with names of the A, B, C, and D models are presented in Figure 7. Five convolution blocks with the names of Conv1, Conv2, Conv3, Conv4, and Conv5 and two FC layers were trained based on the level of fine-tuning. For example, the first model (A) consisted of the Conv2-5 and FC layers, which means that the convolutional blocks from 2 up to 5 were trained based on new datasets, and the rest of the parameters of the model were frozen. In the same way, the models B, C, and D were, respectively, trained from the third, fourth, and fifth convolutional blocks with the fully connected layer.

Based on the experimental results presented in Figure 8, the best accuracy was for model A (Conv2-5 and FC layers) with 97.99% and 82%, respectively, for the Replay-Attack and ROSE-Youtu databases which were highlighted with gray shading. All models (A, B, C, and D) were trained based on the parameters presented in Table 2 and 1000 epochs. Additionally, for the classification of the images, the SoftMax classifier was utilized with two channels of live and spoof labels. As a result, for the Replay-Attack and ROSE-Youtu databases, model A stayed on the best accuracy, respectively, with (97%, 82%) compared to B (96%, 66%), C (96%, 76%), and D (92%, 66%). Based on these experimental results, it may be proven that, for spoofing detection based on the RGB color space, the optimum level of fine-tuning of the VGG-face model was the trained convolutional blocks numbered 2 up to 5 with two fully connected layers and by freezing the first convolutional block parameters.

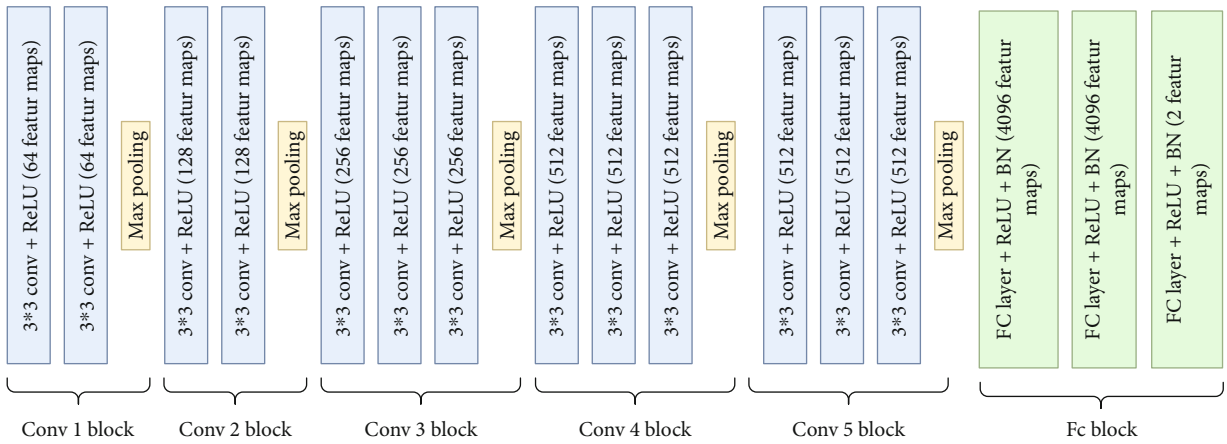


Figure 6: Structure of VGG-face model [33].

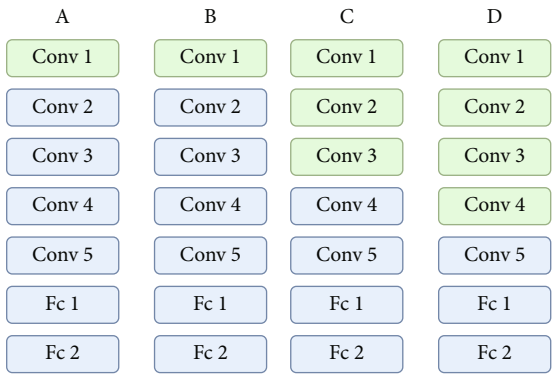


Figure 7: Green shaded blocks are frozen and pretrained, and blue shaded blocks are retrained during the training process.

In this case, in the rest of the experimental results for the RGB color space, we utilized the same level of fine-tuning (model A) which stayed on the best accuracy rate for the VGG-face model for face spoofing detection. For training the deep models with the ROSE-Youtu database, we selected 70% of the data from the first 10 indexed samples of data for training, and the rest of these were used for validation. In this case, the training and validation data were totally separated. Because the ROSE-Youtu database contains data with different rotations such as 90 degrees clockwise and counterclockwise, the image data augmentation technique in the Keras library was utilized.

4.4. Color-Based Approach Model. In this section, we explain the process of converting the color space from RGB to HSV and YCbCr. Furthermore, we evaluated three benchmark VGG models for finding the effects of each color space on the accuracy of classification. In this test, we utilized two VGG16 models and trained the entirety of each network with HSV and YCbCr color space images from spoofing datasets with the default window size. All models were trained based on the parameters presented in Table 2 and 1000 epochs.

Table 3 presents the experimental results on the HSV and YCbCr color spaces and the evaluation of fine-tuning of the entirety of the networks with these color spaces. According to the results obtained, the HSV color space-based image in the Replay-Attack database achieved significant results compared to the YCbCr color space by improving 0.71% in accuracy. Nevertheless, in the ROSE-Youtu database, the YCbCr space provided better results compared to HSV by improving 7.59%. According to these results, it may be concluded that, for face spoofing recognition under different conditions such as illumination changes and displaying a high-resolution camera, both color spaces contain discriminative features which can help distinguish a live image from a fake face in different scenarios.

4.5. Deep Feature Extraction. In the second step of our experimental procedure, the features of the fully connected layer (FC7) of the pretrained VGG-face model based on the RGB color space were extracted, which included 4096 channels. The features extracted from this layer were classified with different typical classifiers such as SVM, LDA, and KNN. Moreover, these results were compared to the SoftMax classifier to evaluate the performance of the extracted deep features with other classification algorithms. Based on the experimental results shown in Table 4, the best results were for SVM and KNN in the Replay-Attack database with 98.93 (Acc), 98.50 (Se), 100 (Sp), 98.97 (Pr), and 98.93% (F-score) for both classification algorithms.

In the Replay-Attack database, the SoftMax classifier was placed on the fourth stage among the other classifiers based on the results. However, in the ROSE-Youtu database, the SoftMax classifier achieved significant results compared to the other classifiers with 82.84 (Acc), 97.42 (Se), 72.41 (Sp), 89.52 (Pr), and 88.00% (F-score).

4.6. Feature Selection and Classification. In this step, we utilized mRMR to reduce the size of the extracted features from three different models and select robust and discriminative feature sets. The size of the extracted features for each model was 4096, and by combining these three VGG models, the size increased to 12288 features. For finding the optimum dimension of feature sets, we analyzed different sizes of features with the help of mRMR feature selection as presented in Figures 9(a) and 9(b). Based on the results, the best feature

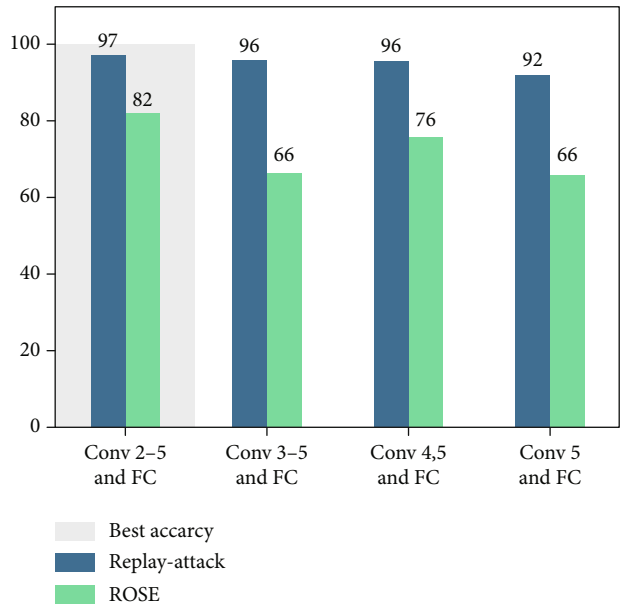


Figure 8: Accuracy of VGG-face model based on level of fine-tuning of the networks.

Table 3: Experimental results of fine-tuning pretrained VGG16 models with the HSV and YCbCr color spaces.

Metrics (%)	HSV		YCbCr	
	Replay-Attack	ROSE-Youtu	Replay-Attack	ROSE-Youtu
Acc	99.46	71.94	98.75	79.53
Se	99.25	77.42	99.25	88.61
Sp	100	66.67	97.50	45.77
Pr	99.47	71.87	98.75	83.75
F-score	99.47	71.87	98.75	71.03

sizes for Replay-Attack were 400, 500, and 700, and those for the ROSE-Youtu database were 300, 500, and 700, respectively, for RGB, HSV, and YCbCr based on the LR classifier. In this case, the optimum feature size for covering both databases and all color spaces may be set to 1600 features. In continuation of this test, we analyzed the effects of the deep features of HSV color spaces on the improvement of accuracy rates. In this case, we combined extracted features from the FC7 layer of the pretrained VGG-face model (RGB) with the VGG16 model (HSV). The experimental results presented in Table 5 show that the accuracy of the face spoofing detection approach was improved drastically in the Replay-Attack database.

In this database, all evaluation metrics with the LR, SVM, and KNN classifiers stayed on significant rates with 99.82 (Acc), 99.75 (Se), 100 (Sp), 99.82 (Pr), and 99.82 (*F*-score) %. In the ROSE-Youtu database, also, all evaluation metrics were improved with four different classifiers, and the best results were obtained for the linear regression classifier by 95.98 (Acc), 99.00 (Se), 93.24 (Sp), 95.98 (Pr), and 95.98 (*F*-score) %. The experimental results in this table compared to Table 4 showed that HSV deep features improved the effectiveness of detection of spoofing data. The comparison of two experimental results of Tables 4 and 5 showed that all evaluation metrics were improved by combining HSV deep features with VGG-face deep features, and these results were improved by 13.14 (Acc), 1.58 (Se), 20.83 (Sp), 6.46 (Pr), and 7.98 (*F*-score) based on the LR classifier in the ROSE-Youtu database.

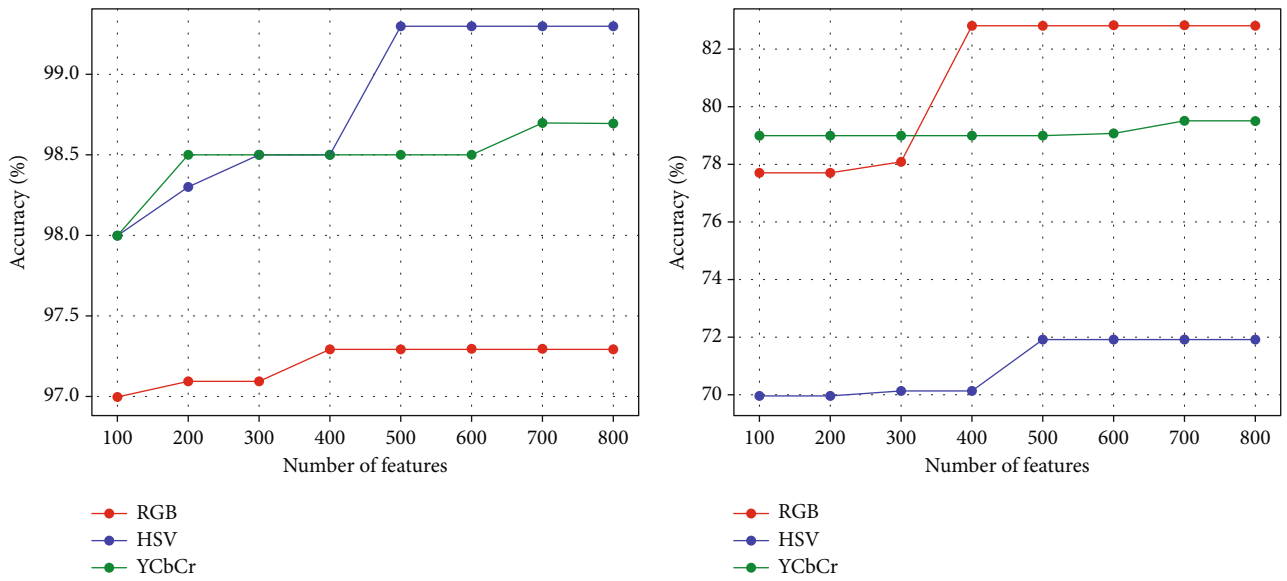
In Table 6, the experimental results of the proposed deep model by applying the feature selection method are presented. After concatenation of three extracted features from different color spaces from the VGG models, mRMR feature selection was applied. As discussed in Section 3.3, the main reason for applying the mRMR algorithm was to reduce the irrelevant features and select robust and discriminative features. Figure 10 presents a visualization of the first four feature maps of each five convolutional blocks with the RGB, HSV, and YCbCr color spaces. According to the extracted features from each convolutional block and specifically the fifth convolutional block, it was obtained that combining features from each model with different color spaces includes redundant and irrelevant features which decrease the effectiveness of our proposed approach. Based on these results in Table 6, the extracted YCbCr features cannot improve the evaluation metrics in the replay-attack database. However, on the other hand, these features improved the effectiveness of recognition of spoofing data in the ROSE-Youtu database and increased the results by 1.18 (Acc), 2.91 (Sp), 2.25 (Pr), and 1.19 (*F*-score) based on the LR classifier. Additionally, the linear regression classifier stayed on the best results compared to SVM, KNN, and LDA.

To better present the results, we utilized ROC curve analysis for both experiment databases as shown in Figure 11. The ROC curve analysis showed that the proposed approach with the help of well-known pretrained models in the RGB, HSV, and YCbCr color spaces extracted discriminative features for the detection of spoofing face images. Based on these results, the LR classifiers stayed on the best AUC compared to the other mentioned classification algorithms by 0.995 and 1.00 for the ROSE-Youtu (Figure 11(b)) and Replay-Attack (Figure 11(a)) databases, respectively. In this case, we selected the LR classifier as the base classification algorithm for our proposed approach and employed this classification algorithm in the rest of the paper.

4.7. Evaluation of Different Attacks. For evaluation of our proposed approach in different scenarios of spoofing attacks and for finding the advantages and disadvantages of our proposed approach, we tested our deep learning approach on different attacks individually. Based on the experimental results on Replay-Attack (Table 6), it may be concluded that our proposed approach had satisfactory results in the replay, display, and print attacks which are presented in the Replay-Attack database. Furthermore, this approach achieved 97.16% accuracy in the ROSE-Youtu database, in which, for finding misclassification reasons, in this test, the spoofing scenarios were individually analyzed. We categorized the ROSE-Youtu database into five different groups such as the real, display and print, mask with cropping, and mask

Table 4: The classification results based on different classifiers and deep features of the VGG-face model on the Replay-Attack and ROSE-Youtu databases.

Model	Database	Classification	Acc (%)	Se (%)	Sp (%)	Pr (%)	*F*-score (%)
VGG-face (RGB color space)	Replay-Attack database	SoftMax	97.32	99.25	99.50	97.34	97.30
		SVM	98.93	98.50	100	98.97	98.93
		LDA	98.91	98.91	100	99.78	99.78
		KNN ($K = 1$)	98.93	98.50	100	98.97	98.93
	ROSE-Youtu	SoftMax	82.84	97.42	72.41	89.52	88.00
		SVM	78.38	59.75	90.03	78.46	77.65
		LDA	70.30	50.13	82.91	69.61	69.39
		KNN ($K = 1$)	78.38	59.75	90.03	78.46	77.65



(a) Replay-Attack database (b) ROSE-Youtu database

Figure 9: Accuracy of LR classification based on different sizes of features.

Table 5: The classification results of the extracted features from RGB and HSV on the Replay-Attack and ROSE-Youtu databases.

Model	Databases	Classification	Acc (%)	Se (%)	Sp (%)	Pr (%)	*F*-score (%)
VGG-face (RGB)+VGG16 (HSV)	Replay-Attack database	LR	99.82	99.75	100	99.82	99.82
		SVM	99.82	99.75	100	99.82	99.82
		LDA	98.75	99.50	96.88	98.75	98.75
		KNN($K = 1$)	99.82	99.75	100	99.82	99.82
	ROSE-Youtu	LR	95.98	99.00	93.24	95.98	95.98
		SVM	95.98	97.51	94.59	96.04	95.98
		LDA	83.34	77.11	92.79	85.97	85.22
		KNN($K = 1$)	94.79	97.51	92.34	94.96	94.80

without cropping groups containing videos from persons as presented in Figure 12. Display and replay attacks are already tested in different conditions such as light change and shaking hands in experimental databases, namely, Replay-Attack. We set the displayed attack and print attack categories together and labeled them as display. However, the main difference of the ROSE-Youtu database is mask attack in different conditions and scenarios which are not available in other experimental databases. Mask attack in the ROSE-Youtu database contains scenarios such as a mask with two

Table 6: The classification results of the extracted features from RGB and HSV and YCbCr.

Model	Databases	Classification	Acc (%)	Se (%)	Sp (%)	Pr (%)	F-score (%)
VGG-face (RGB)+VGG16 (HSV)+VGG16 (YCbCr)	Replay-Attack database	LR	99.82	99.75	100	99.82	99.82
		SVM	99.82	99.75	100	99.82	99.82
		LDA	98.75	99.50	96.88	98.75	98.75
		KNN($K = 1$)	99.82	99.75	100	99.82	99.82
	ROSE-Youtu	LR	97.16	98.41	96.15	97.21	97.17
		SVM	95.98	93.12	98.29	96.05	95.97
		LDA	96.45	97.73	95.73	96.49	96.46
		KNN ($K = 1$)	88.17	86.77	89.32	88.18	88.18

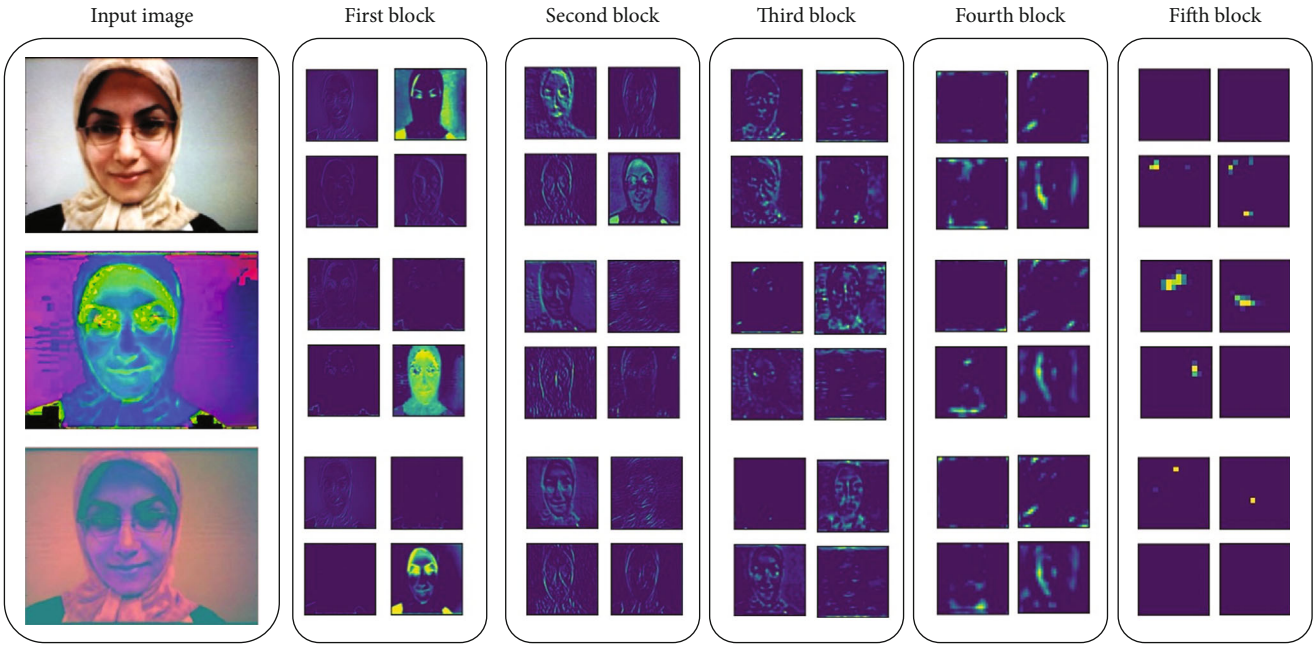


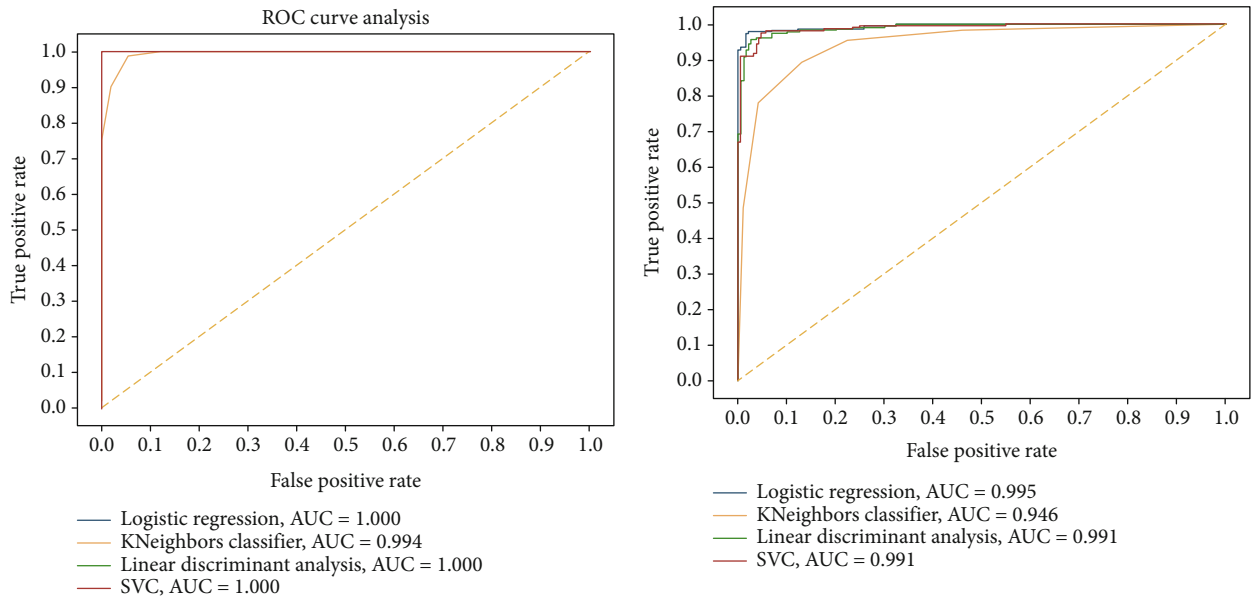
Figure 10: Extracted feature maps from each convolutional block.

eyes and mouth cropped out, mask without cropping, mask with the upper part cut in the middle, and mask with the lower part cut in the middle.

In this test, we categorized these mask attack scenarios into two main groups as a mask without cropping and mask with cropping. Based on the experimental results which are presented in Table 7, it appeared that the main advantage of the proposed approach was the detection of spoofing attacks such as display and print attacks. The accuracy of recognition of display and print attacks was 98.00%, which stayed on the highest value compared to other spoofing data. The second highest value of accuracy was for the mask with cropping attacks with 97.82% accuracy. The results for replay attacks were also compatible with 94.64% accuracy. On the other hand, the lowest results were for a mask without cropping with 92.59 (Acc), 96.81 (Se), 98.93 (Sp), 92.70 (Pr), and 92.33 (F-score) %. These results proved that the proposed approach has a significant accuracy in recognition of display and printed attack and compatible accuracy in a mask without cropping scenarios.

In continuation of this test, we utilized the scatter plot of the extracted features based on the attack groups and real videos. In this part, we selected one frame from each video from the test set and reduced the dimensions of the features with the help of Principal Component Analysis (PCA) from 1600 to 3 to obtain the X, Y, and Z values for each image and present them in 3D scatter plots. As presented in Figure 13, it appeared that the mask without cropping and replay attack features were overlapped with real video frames. Furthermore, other spoofing attacks such as display and mask with cropping were clearly separated from real videos.

4.8. Evaluation Efficiency of the Proposed Method in Cloud System. As presented before, one of the main problems of cloud computing systems is the management of storing data and optimizing resources. For this reason, we proposed a deep learning approach that trains with fewer data and achieved significant results based on accuracy compared to existing models. For evaluating our approach, we train the model in four different types. First, the models are trained



(a) Replay-Attack database

(b) ROSE-Youtu database

Figure 11: ROC curve analysis based on different classifiers.

(a) Display and print attacks

(b) Mask with cropping

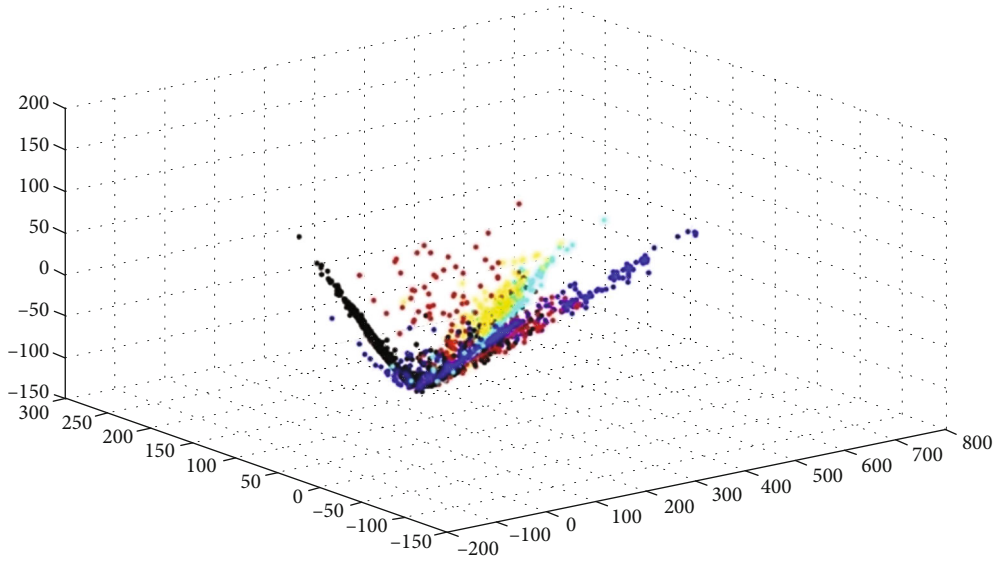
(c) Mask without cropping

(d) Replay attack

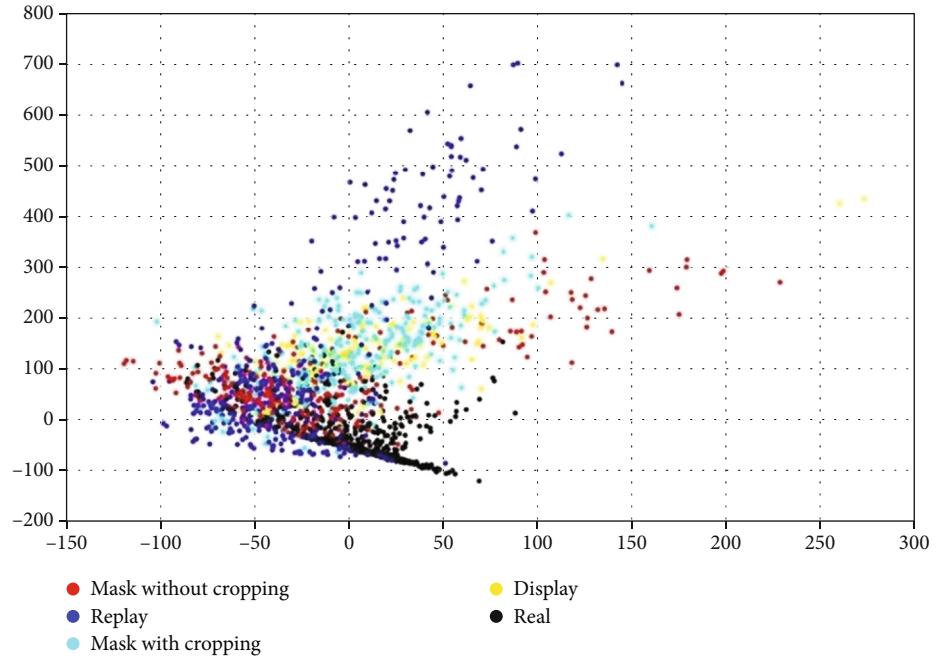
Figure 12: Categorization of spoofing attacks of the ROSE-Youtu database.

Table 7: Evaluation of different types of attacks on the ROSE-Youtu database.

Database	Types of attacks	Acc (%)	Se (%)	Sp (%)	Pr (%)	*F*-score (%)
ROSE-Youtu	Mask without cropping	92.59	96.81	98.93	92.70	92.33
	Replay attack	94.64	90.99	96.81	94.64	94.63
	Mask with cropping	97.82	95.83	98.89	97.83	97.82
	Display and print attack	98.00	96.46	98.93	98.00	98.00



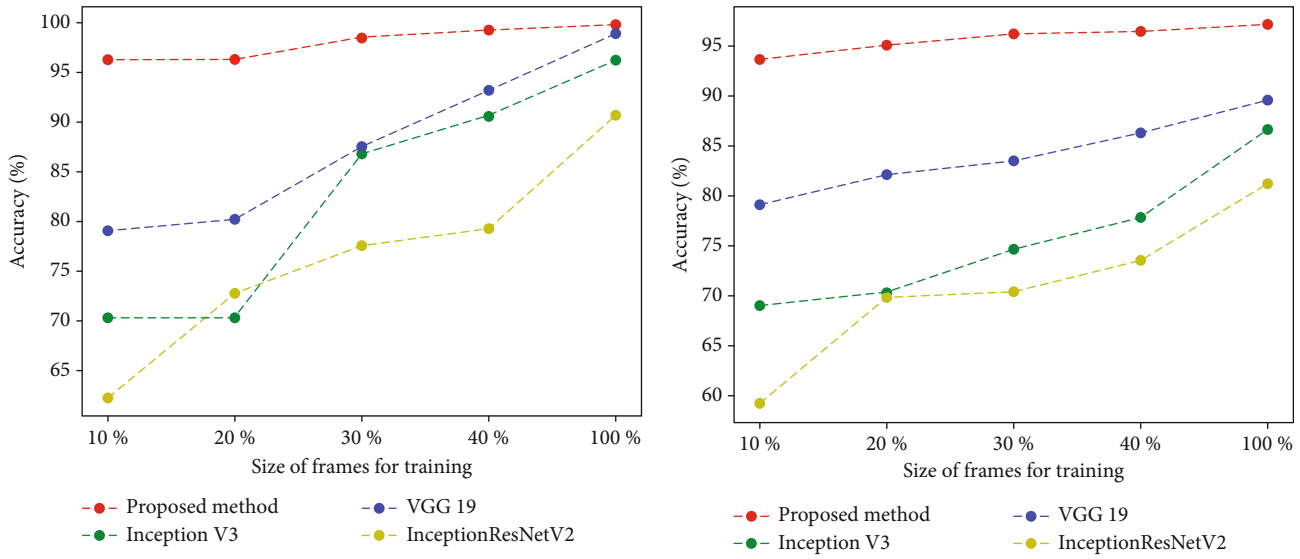
(a) Replay-Attack database



(b) ROSE-Youtu database

Figure 13: 3D and 2D scatter plots of features based on attacks.

on 10% of frames of each video and test on all frames. The second, third, and fourth modes of evaluation are in the same condition, such as 20, 30, and 40% of the frames for training and evaluating on all frames of test sets. These scenarios are tested on well-known deep learning models in RGB color space such as Inception V3 [44], InceptionResNetV2 [45], and VGG 19 [34]. These pretrained models on the ImageNet database are employed as a deep feature extractor. For fine-



(a) Replay-Attack database

(b) ROSE-Youtu database

Figure 14: Evaluation of different sizes of training data on the accuracy of classification.

Table 8: Comparison of the proposed approach against state-of-the-art algorithms based on the Replay-Attack database.

Method	EER (%)	HTER (%)
Motion+LBP [45]	4.5	5.1
DMD [26]	3.8	5.3
SURF color texture [10]	1.2	4.2
Color texture [11]	0.4	2.8
LBP net [18]	0.6	1.3
Color LBP [46]	0.9	4.9
Partial CNN [14]	2.9	4.3
CompactNet [12]	0.8	0.7
Dense optical flow+Shearlet [31]	0.83	0.0
Proposed method	0.2	0.4

Table 9: Comparison of the proposed approach against state-of-the-art algorithms based on the ROSE-Youtu database.

Method	EER (%)
Deep color-based feature [42]	8.0
SE-ResNet 18 [48]	7.2
3D CNN [49]	7.0
Two-stage deep model [47]	4.56
Proposed method	3.8

tuning of parameters in these models with the face spoofing database, we changed the SoftMax classification layer to two classes of spoof and real face. In addition, the small number of learning rates with 0.0001 is set for all models; besides, we employed Adam optimization, batch size 16, and 10000 epochs. Suppose we capture a one-minute video with 720 p resolution at 30 fps containing 1800 frames which are around 60 MB. Therefore, training the model with 10% of frames of each video not only reduced the size of data for training (around 6 MB) but also decreased the computation cost in the training phase.

Based on experimental results presented in Figure 14, it appears that the proposed method achieved significant results in the detection of spoofing attacks with less training data compared to benchmark deep learning methods. The proposed method achieved the accuracy of classification with 96.3% in ten percent of frames of each video for training and testing on entire videos which this score is better than the results achieved by Inception V3, InceptionResNetV2, and VGG 19 with 70.32, 62.3, and 79.1, respectively, in the Replay-Attack database. The results of the proposed method are 96.3, 96.3, 98.5, 99.2, and 99.8% which are better than other experimented deep learning methods, respectively, for 10, 20, 30, 40, and 100% of frames of each video in the Replay-Attack database. In the same condition, in the ROSE-Youtu database, also, our proposed method stayed on the best results with 93.7, 95.1, 96.2, and 96.5 in 10, 20, 30, and 40 percent of frames of each video for training.

4.9. Comparison of the Proposed Approach against State-of-the-Art Algorithms. Table 8 provides a comparison between the proposed approach and state-of-the-art methods. The experimental results shown in Table 8 demonstrated the effectiveness of our extracted deep features in the Replay-Attack database.

We may observe that, among the state-of-the-art methods presented in this table, the best results were for deep learning-based methods like the LBP net [18] with 0.6 (EER) and 1.3 (HTER). The best HTER was for dense optical flow +Shearlet [31] with 0.0. Furthermore, our proposed method achieved 0.2 (EER), which was better than the multicue deep method proposed in a previous study [31] with a single cue (color texture analysis).

Table 9 provides a comparison between the proposed approach and state-of-the-art methods in the aspect of EER. According to these experimental results, it may be argued that our proposed approach is more applicable and stayed on the best EER (%) values in comparison to state-of-the-art methods in the Replay-Attack database. In the other benchmark public access database (ROSE-Youtu), our proposed approach also stayed on the best ERR (%) values. In this database, the best EER in state-of-the-art algorithms was for the two-stage deep model [47] approach over which our proposed approach improved the EER value by 0.76%. Based on these experimental results and comparison with state-of-the-art algorithms, it may be concluded that our proposed approach achieved robust and significant results for distinguishing fake faces from live faces with 0.2 and 3.8 for EER (%) in the replay-attack and ROSE-Youtu databases, respectively.

5. Conclusion

The IoT cloud-based framework for face spoofing detection is proposed and implemented in this study. The proposed system detects face spoofing attacks by applying the new deep learning framework. This approach can be used reliably in the cloud-based environment by storing less data which decreased both processing cost and size of data in the training phase. Moreover, the proposed multicolor deep feature-based approach outperformed the baseline methods on the Replay-Attack database, while achieving competitive results on the ROSE-Youtu database. The results obtained for the Replay-Attack and ROSE-Youtu databases proved that environmental factors and scenarios such as background changes, shaking hands, high-resolution camera, and illumination did not limit the effectiveness of our proposed approach. Furthermore, our proposed approach achieved satisfactory results in scenarios such as print, display, and replay attacks. In the case of mask attacks in different scenarios such as without cropping, with cropping, upper part cut, lower part cut, and mask with two eyes and mouth cropped out, the proposed approach presented compatible results. Furthermore, in mask without cropping attacks, the proposed approach achieved the lowest rate of accuracy (92.59%) compared to different attacks such as replay or print attacks. This inefficiency of the proposed approach in mask attack types makes us eager to solve this problem in future work. In future work, we will investigate adding depth information to our color-based deep features to improve the effectiveness of recognition of spoofing attacks in different mask scenarios in IoT cloud environments.

Data Availability

The data used to support the findings of this study are available from the authors upon reasonable request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Z. Ali, M. S. Hossain, G. Muhammad, I. Ullah, H. Abachi, and A. Alamri, "Edge-centric multimodal authentication system using encrypted biometric templates," *Future Generation Computer Systems*, vol. 85, pp. 76–87, 2018.

[2] M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, and J. Fierrez, "Multi-biometric template protection based on homomorphic encryption," *Pattern Recognition*, vol. 67, pp. 149–163, 2017.

[3] P. Kumari and P. Thangaraj, "A fast feature selection technique in multi modal biometrics using cloud framework," *Microprocessors and Microsystems*, vol. 79, p. 103277, 2020.

[4] K. A. Shakil, F. J. Zareen, M. Alam, and S. Jabin, "BAMHealthCloud: a biometric authentication and data management system for healthcare data in cloud," *Journal of King Saud University - Computer and Information Sciences*, vol. 32, no. 1, pp. 57–64, 2020.

[5] B. Sree Vidya and E. Chandra, "Entropy based local binary pattern (ELBP) feature extraction technique of multimodal biometrics as defence mechanism for cloud storage," *Alexandria Engineering Journal*, vol. 58, no. 1, pp. 103–114, 2019.

[6] M. Masud, G. Muhammad, H. Alhumyani et al., "Deep learning-based intelligent face recognition in IoT-cloud environment," *Computer Communications*, vol. 152, pp. 215–222, 2020.

[7] X. Song, X. Zhao, L. Fang, and T. Lin, "Discriminative representation combinations for accurate face spoofing detection," *Pattern Recognition*, vol. 85, pp. 220–231, 2019.

[8] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *2007 IEEE 11th International Conference on Computer Vision*, Rio de Janeiro, Brazil, 2007.

[9] A. Gumaei, R. Sammouda, A. M. S. Al-Salman, and A. Alsanad, "Anti-spoofing cloud-based multi-spectral biometric identification system for enterprise security and privacy-preservation," *Journal of Parallel and Distributed Computing*, vol. 124, pp. 27–40, 2019.

[10] Z. Boulkenafet, J. Komulainen, and A. Hadid, "On the generalization of color texture-based face anti-spoofing," *Image and Vision Computing*, vol. 77, pp. 1–9, 2018.

[11] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818–1830, 2016.

[12] L. Li, Z. Xia, X. Jiang, F. Roli, and X. Feng, "CompactNet: learning a compact space for face presentation attack detection," *Neurocomputing*, vol. 409, pp. 191–207, 2020.

[13] D. T. Nguyen, T. D. Pham, N. R. Baek, and K. R. Park, "Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors," *Sensors (Switzerland)*, vol. 18, no. 3, p. 699, 2018.

[14] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 16–21, Oulu, Finland, 2017.

[15] Y. A. U. Rehman, L. M. Po, and J. Komulainen, "Enhancing deep discriminative feature maps via perturbation for face presentation attack detection," *Image and Vision Computing*, vol. 94, p. 103858, 2020.

[16] Y. A. U. Rehman, L. M. Po, M. Liu, Z. Zou, W. Ou, and Y. Zhao, "Face liveness detection using convolutional-features fusion of real and deep network generated face images," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 574–582, 2019.

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

[18] L. Li, X. Feng, Z. Xia, X. Jiang, and A. Hadid, "Face spoofing detection with local binary pattern network," *Journal of Visual Communication and Image Representation*, vol. 54, pp. 182–192, 2018.

[19] G. B. De Souza, S. Member, D. Felipe, R. G. Pires, A. N. Marana, and J. P. Papa, "Deep texture features for robust face spoofing detection," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 64, no. 12, pp. 1397–1401, 2017.

[20] R. J. Raghavendra and R. S. Kunte, "Extended local ternary corelation pattern: a novel feature descriptor for face anti-spoofing," *J. Inf. Secur. Appl.*, vol. 52, p. 102482, 2020.

[21] S. Jia, C. Hu, X. Li, and Z. Xu, "Face spoofing detection under super-realistic 3D wax face attacks," *Pattern Recognition Letters*, vol. 145, pp. 103–109, 2021.

[22] A. George and S. Marcel, "Learning one class representations for face presentation attack detection using multi-channel convolutional neural networks," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 361–375, 2021.

[23] J. M. Di Martino, Q. Qiu, and G. Sapiro, "Rethinking shape from shading for spoofing detection," *IEEE Transactions on Image Processing*, vol. 30, pp. 1086–1099, 2021.

[24] S. Arora, M. P. S. Bhatia, and V. Mittal, "A robust framework for spoofing detection in faces using deep learning," *The Visual Computer*, vol. 13, 2021.

[25] T. Edmunds and A. Caplier, "Motion-based countermeasure against photo and video spoofing attacks in face recognition," *Journal of Visual Communication and Image Representation*, vol. 50, pp. 314–332, 2018.

[26] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 762–777, 2015.

[27] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015.

[28] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4726–4740, 2015.

[29] Y. A. U. Rehman, L. M. Po, and M. Liu, "SLNet: stereo face liveness detection via dynamic disparity-maps and convolutional neural network," *Expert Systems with Applications*, vol. 142, p. 113002, 2020.

[30] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Arlington, VA, USA, 2013.

[31] L. Feng, L.-M. Po, Y. Li et al., "Integration of image quality and motion cues for face anti-spoofing: a neural network approach," *Journal of Visual Communication and Image Representation*, vol. 38, pp. 451–460, 2016.

[32] Y.-Q. Wang, "An analysis of the Viola-Jones face detection algorithm," *Image Processing On Linec*, vol. 4, pp. 128–148, 2014.

[33] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Procedings of the British Machine Vision Conference 2015*, Swansea, UK, 2015.

[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 1–14, Kuala Lumpur, Malaysia, 2015.

[35] N. Aloysius and M. Geetha, "A review on deep convolutional neural networks," in *2017 International Conference on Communication and Signal Processing (ICCSP)*, pp. 588–592, Chennai, India, 2018.

[36] Z. Wang, "Deep convolutional neural networks for image classification: a comprehensive review," *Neural Comput*, vol. 2733, pp. 2709–2733, 2017.

[37] M. Toğaçar, B. Ergen, and Z. Cömert, "A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models," *Irbm*, vol. 1, pp. 1–11, 2020.

[38] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1–6, 2005.

[39] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Proc. 2003 IEEE Bioinforma. Conf. CSB 2003*, vol. 3, no. 2, pp. 523–528, 2003.

[40] C. Perez, J. Tapia, P. Estévez, and C. Held, "Gender classification from face images using mutual information and feature fusion," *International Journal of Optomechatronics*, vol. 6, no. 1, pp. 92–119, 2012.

[41] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Spec. Interes. Group, BIOSIG 2012*, pp. 1–7, 2012.

[42] Z. Yang, W. Chen, F. Wang, and B. Xu, "Unsupervised domain adaptation for face anti-spoofin," in *2018 24th International Conference on Pattern Recognition (ICPR)*, pp. 338–343, Beijing, China, 2018.

[43] D M W Powers and Ailab, "Evaluation: from precision, recall and F-factor to ROC, informedness, markedness & correlation," *J. Mach. Learn. Technol.*, vol. 2, pp. 37–63, 2007.

[44] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *31st AAAI Conf. Artif. Intell. AAAI 2017*, pp. 4278–4284, 2017.

[45] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *2013 International Conference on Biometrics (ICB)*, Madrid, Spain, 2013.

[46] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 2636–2640, Quebec City, QC, Canada, 2015.

[47] M. M. Hasan, M. S. U. Yusuf, T. I. Rohan, and S. Roy, "Efficient two stage approach to detect face liveness : motion based and deep learning based," in *2019 4th International Conference on Electrical Information and Communication Technology (EICT)*, pp. 20–22, Khulna, Bangladesh, 2019.

[48] G. Wang, H. Han, S. Shan, and X. Chen, "Unsupervised adversarial domain adaptation for cross-domain face presentation attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. , 202156–69, 2021.

[49] H. Li, P. He, S. Wang, A. Rocha, X. Jiang, and A. C. Kot, "Learning generalized deep feature representation for face anti-spoofing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 10, pp. 2639–2652, 2018.