Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.031561
Article

Tech Science Press

Challenge-Response Emotion Authentication Algorithm Using Modified Horizontal Deep Learning

Mohamed Ezz[1], Ayman Mohamed Mostafa[1,*] and Ayman Elshenawy[2,3]

[1]College of Computer and Information Sciences, Jouf University, Sakaka, 72314, Saudi Arabia
[2]Systems and Computers Engineering DEPT, Faculty of Engineering, Al-Azhar University, Cairo, 11651, Egypt
[3]Software Engineering & Information Technology, Faculty of Engineering & Technology, Egyptian Chinese University, Egypt
*Corresponding Author: Ayman Mohamed Mostafa. Email: amhassane@ju.edu.sa

Received: 21 April 2022; Accepted: 09 June 2022

Abstract: Face authentication is an important biometric authentication method commonly used in security applications. It is vulnerable to different types of attacks that use authorized users' facial images and videos captured from social media to perform spoofing attacks and dynamic movements for penetrating security applications. This paper presents an innovative challenge-response emotions authentication model based on the horizontal ensemble technique. The proposed model provides high accurate face authentication process by challenging the authorized user using a random sequence of emotions to provide a specific response for every authentication trial with a different sequence of emotions. The proposed model is applied to the KDEF dataset using 10-fold cross-validations. Several improvements are made to the proposed model. First, the VGG16 model is applied to the seven common emotions. Second, the system usability is enhanced by analyzing and selecting only the four common and easy-to-use emotions. Third, the horizontal ensemble technique is applied to enhance the emotion recognition accuracy and minimize the error during authentication processes. Finally, the Horizontal Ensemble Best N-Losses (HEBNL) is applied using challenge-response emotion to improve the authentication efficiency and minimize the computational power. The successive improvements implemented on the proposed model led to an improvement in the accuracy from 92.1% to 99.27%.

Keywords: Face authentication; challenge-response authentication; transfer learning and horizontal ensemble technique

1 Introduction

Recently, mobile devices and applications are used for accessing confidential files, bank accounts, and business applications. Due to the importance of mobile applications in daily use, authentication mechanisms are considered crucial in verifying users. Traditional authentication systems in mobile applications are considered weak as the attacker can access confidential information by unlocking the secret patterns or using authorized users' images or videos. Biometric Authentication Systems (BAS) are suggested to solve the problems of traditional authentication methods. It utilizes the username/password, secret key or token, and one or more biometric features of the authorized users that strongly distinguish persons from each other. Nowadays, BAS is the most common authentication method for mobile user authentication [1]. These biometric features may be a person's face [2], voice [3], iris [4], and fingerprint [5], while the face and fingerprint are the most involved features in mobile user authentication. These features are not enough for authenticating the user as it is considered a single model that uses only one biometric feature. Although BAS makes the authentication process more robust, it still has significant challenges due to the easiness of reproducing the same biometric feature of humans, such as spoofing attacks using available images and videos of the user on social media or fake pictures and videos using recent technologies [1]. As presented in [6], a multi-model mechanism for user authentication using lip analysis, while the authors of [7] provided a mechanism for facial expression recognition. Multi-model BASs outperforms the single model BAS as they can increase the system's strength and accuracy subjected to hacking and vulnerabilities [1]. Attackers can use any video or facial movements containing the user's face to gain access to the application. These attacks can be prevented by applying a challenge-response authentication that verifies the user with high accuracy and robustness. Another implementation of the challenge-response authentication was proposed in [6], where a multi-model based on a silent password recognition framework for lip analysis is proposed. It identifies the user by detecting his face and extracting the silent password from the input video to provide the same level of security by adding challenge-response to it. Facial expression recognition (FER) has attracted the attention of many researchers in recent years to improve FER accuracy. Facial expressions can be used in many applications such as acting based on emotions, capturing images by smiling or receiving customer feedback during shopping [6], patients' reactions [7], and other applications that benefit from recognizing the user's emotions.

Different research methodologies are applied for improving the recognition of facial expressions. As presented in [8], the authors proposed an automatic FER system of three stages. First, the face is localized using the Voila-Jones algorithm. Second, the face features are extracted using Gabor filters; in the third stage, the facial expressions are classified using a three-state Hidden Markov Model (HMM). The proposed system gives a suitable increase in classification accuracy compared with some other methods. The authors of [9] applied Support Vector Machine (SVM) to classify facial expressions based on HOG feature extraction of facial images. The proposed methods are tested using both JAFFFE and KDEF datasets. The obtained results provide a surprise increase in the recognition rates. As shown in [10], the FER model is presented based on a region-based convolution fusion network. They have applied a BAS based on a muscle movement model to extract the basic crucial face regions (eye, forehead, mouth). A neural network is used to extract region features, and the combination of these features is used to classify the FER. The proposed system is characterized by its adaptability to variant samples of frontal faces. As presented in [11], a static image-based classification method is proposed using deep learning Convolution Neural Network (CNN) for FER classification. In [12], an Auxiliary Classifier Generative Adversarial Network (AC-GAN) is presented to recognize ten facial expressions. The proposed system uses U-net as a generator and capsule-Nets as a discriminator.

All the previous researchers have applied their proposed models to frontal image faces. The authors of [13] proposed a GAN for multi-view FER. They train the model on a frontal face then the GAN is applied to synthesis faces with different deflections. Finally, CNN is used to fuse the synthesis images and the deflected images, which is considered a multi-view FER solution. Facial Expression can present a lot in the trust of security systems because it preserves a lot of information about facial components and their spatial relations [14], which is difficult to be spoofed by attackers and make security systems more trusted based on this feature. Although many researchers have studied the FER subject, they have not investigated the point of applying FER in authentication systems. In [15], a hybrid BAS is presented that verifies the person based on one facial expression recorded earlier for the user in a database. The selected feature plays the role of a password to distinguish a user if he is a real user or fraudulent. In [16], the authors proposed an authentication model to combine people's physiological and behavior stamps. The proposed model applies face recognition and FER to legitimate users. In [17], the authors presented a secure login system based on FER.

In this paper, an innovative challenge-response emotion authentication technique is proposed to authenticate users based on a random sequence of emotions that can verify whether the user is authorized or not. An enhanced authentication model based on the horizontal ensemble technique is used to recognize users independently. Different experiments are applied to improve the model's accuracy and increase its robustness and efficiency. A novel mechanism is applied using Horizontal Ensemble Best N-Losses (HEBNL) on four common and easy-to-express emotions to improve authentication efficiency using challenge-response emotion and minimize the runtime computational power required.

As a result, impersonal and spoofing attacks can be prevented due to the randomization of emotion sequences. The contribution of the paper is as follows:

i) A challenge-response BAS based on verifying a random sequence of user emotions during the authentication process is proposed to authenticate authorized users with high efficiency and accuracy.

ii) Tune the pre-trained VGG16 model on the seven primary user emotions to select the best re-trained points for detecting emotions with high accuracy.

iii) Enhance the authentication usability by selecting the four most common and easy-to-express emotions selected by the end-users via survey.

iv) Improve the accuracy of the selected four common emotion recognition and minimize recognition error by applying the traditional horizontal ensemble technique.

v) Provide a novel Horizontal Ensemble Best N-Losses (HEBNL) algorithm for improving the accuracy of authentication using challenge-response emotion.

The remainder of the paper is organized as follows: Section 2 presented a literature review about commonly used BASs. Section 3 describes the methodology of the paper that contains transfer learning, the Horizontal Ensemble Technique, and the proposed Horizontal Ensemble Authentication framework. Section 4 provides the experiments on the proposed model. Discussion of the proposed experimental results is provided in Section 5. Finally, the conclusion and future works are presented in Section 6.

2 Literature Review

BAS depends on various technologies to identify the user by exploiting their unique, measurable physiological and behavioral features. Traditional BAS are subjected to attacks using biometric tools such as photo attacks, video attacks, masks, lenses, or fingerprint images. In [18], the author presented a multimodal BAS that uses both Electroencephalography (EEG) and gait signals to overcome the attacks of BAS. The proposed system consists of an invalid ID filter model and an attention-based identification model using RNN. The proposed system is achieving a false acceptance rate (FAR) and a false rejection rate (FRR) of 0% and 1.0% respectively. Although this model provides good performance, it has a problem that the user must have wearable inertial measurement units (IMU) and EEG headset which requires large-scale deployment in practice.

The author of [19] presented a BAS based on the eye movement of the user to extract reliable biometric features in a low time. The proposed system can prevent replay attacks and have a good performance related to other systems. It achieves equal error rates and significantly lower authentication times (5 s). It was able to detect the replay of the recorded eye with a probability of 99.94. In [20], the author proposes a multimodal BAS model to improve the security of facial recognition systems. It can detect both photo-attack and video attacks using challenge-response based on both user speech and user mouth move during the speaking of random words. The proposed model analyzes both the recognized speech and the mouth motion traits during the speaking and applies it as input to supervised machine learning to check if the challenged user is authenticated or not. In [21], the author presented a biometric Challenge-Response Authentication (CRA) model based on the user's hand movement while writing a specific word in the air. A hardware device such as a leap motion controller is required to capture the user's hand movement and extract his right style. If the user success to pass the challenge the system authenticates the user. In [22], the authors propose a novel BAS based on electrical muscle forearm muscles stimulation of a user. The proposed model measures the user's involuntary finger movements as a response to the challenge. The proposed model ensures that the challenge is never requested twice.

In [23], the authors presented a multi-modal BAS resilient to presentation attacks. The proposed model provided a solution for both liveliness detection and hybrid deep feature-based recognition. It involves the face and iris biometrics with a challenge in form of arbitrary emotion invoking images at a random position on the screen. In [24], the authors proposed a behavioral BAS based on how the user holds the phone and how it moves his finger when he writes or signs on the touchscreen. It analyses the number of points pressed on the touch screen and pressing strength. It achieves a 95% true acceptance rate and a 3.1% false acceptance rate on a dataset of 30 samples. In [25], the authors develop a BAS based on analyzing the micro-movement of both eyes to identify the user view. The model is learned on a dataset of 150 participants. It also achieves lower error rates compared to other researchers at this time. The authors of [26] presented two BAS techniques for face anti-spoofing that detect both eye movement and liveness. The authors of [27] offered a challenge-response BAS approach that records a user's gaze in reaction to a moving stimulus. To identify spoofing attacks, it extracts the center of the user's eyes and computes landmark features to determine whether the gaze corresponds with the challenging trajectory. A database of 70 people is used to train the suggested model. The authors of [28] described a BAS that uses pupillary movements to match it with a moving visual challenge to capture some pupillary motion. In [29], the authors developed a deep CNN that analyzes eye-tracking signals and verifies the identity of the viewer. By analyzing both the randomized stimulus and the ocular reaction to this stimulus, the method detects replay attempts with far lower mistake rates than previous studies. The authors of [30] described Y-eyes, a challenge-response BAS that uses lightweight machine learning approaches to construct numerous screen picture patterns as a challenge and detect relevant corneal specular reflection responses from human eyes. In [31], the authors presented a BAS based on detecting the most frequently used region in five-and ten-second data of mouse movements. The authors of [32] proposed a BAS that employs facial recognition as well as the unique gestures of that particular face when pronouncing a password.

The previously presented research studies in detail more figures on exploiting human biometric features in the authentication process. Biometric authentication is applied by performing a match between the user's biometric features and the stored features of the user in the database. Therefore, biometric authentication is considered a solid method for authenticating users to determine whether they are authorized or not but, it has some subtle issues. One of the problems of the previously implemented BAS's is that it requires hardware that must be installed and configured for reading user features. In addition, the sensitivity level of the user feature samples must be applied to determine the matching degree between the physical user features and the recorded features. This degree of sensitivity may vary from one feature to another. Another major concern is that the biometric authentication method is based on the presented image during the authentication process. This method can be breached by a forged image. Therefore, different solution requirements must be applied to increase the accuracy of the authentication process. This paper presents an innovative challenge-response emotions authentication model based on the horizontal ensemble technique for providing a highly accurate face authentication process by challenging the authorized user using a random sequence of emotions. This model can provide a specific response for every authentication trial with a different sequence of emotions.

3 Methodology

3.1 Transfer Learning

Transfer learning is a flexible method of machine learning that uses a pre-trained models based on the preprocessing of feature extraction and then integrates the trained models into new models for solving problems. Based on the learning experience of the solved problem, new related problems can be adapted and applied for solving. Traditional machine learning algorithms are used to predict unknown data based on the learned patterns of the training datasets. In contrast, transfer learning starts from the learned patterns of other related tasks or problems for solving a specified problem. The conducted experimental results can be incentivized to solve major machine learning problems by applying transfer learning. In machine learning, the models can take months to train, so the best-recommended method is to solve a specified problem based on pre-trained models for decreasing the neural network model's training time, leading to a low generalization error. ImageNet is considered one of the publicly available computer vision models used for detecting generic features from images. The ImageNet is trained for more than 1 million images for 1000 categories or classes. The implementation of transfer learning models is based on three main parts. (i) The convolutional layers closer to the input layer represent the low-level features such as lines. (ii) The middle layer learns complex features to combine the low-level features extracted from the input layers, such as corners and simple shapes. (iii) The output layer that interprets the extracted features in the context of a classification task. The model output is used as an input to a new classifier model. The pre-trained model is integrated into a new neural network model where the trained or non-trained layers can be controlled. The non-trained layers are considered frozen layers where the weights of the pre-trained models will not be updated to a newly trained model. To use deep learning algorithms for predicting tasks and problems on labeled images, there must be massive amounts of datasets to train parameters in the neural network to verify and apply the labeled dataset [33]. The key issue in using transfer learning is that it is highly recommended with deep learning algorithms where it can be applied to a small/medium dataset size while maintaining the accuracy and performance of the conducted results high [34]. Applying transfer learning with deep learning algorithms in authentication, especially in face recognition is considered one of the recently applied mechanisms in security applications. As presented in [35], a face recognition framework using transfer learning is presented for training data in detecting and recognizing face patterns based on a labeled dataset of 1000 images. The results achieved high accuracy due to the training of data using transfer learning.

Different researches are conducted for increasing image classification accuracy that can enhance the learning of images and features. As proposed in [36], a CNN is applied to analyze images with fewer layers and high efficiency. A framework is used to reduce the training time of images and improves the model performance for detecting images. As presented in [37], a deep learning model is applied for improving low-level image features. The main goal is to extract input images with spatial features that can solve the edit propagation problem. As shown in [38], another deep neural network is applied that increases the classification accuracy of images where mobile net models are used as dense blocks. A network structure is generated to create a large number of image features with low convolution cores.

3.2 System Model

3.2.1 Mathematical Presentation

Previous biometric authentication methods suffer from different subtle issues that can affect the performance and accuracy of the authentication process such as the need for special hardware for verifying users and the need for special software that must store user features in a secured database server. In addition, there is a sensitivity degree between the authenticated user and the recorded user features. This degree of sensitivity may vary from one feature to another. Another major concern is that the biometric authentication method is based on the presented image during the authentication process. In this paper, an authentication model is applied for verifying a random sequence of user emotions using transfer learning of emotion classes.

Given a specific problem domain D that contains the features f and the probability distribution $P(F)$, where $F = \{f_1, f_{2,}, \ldots, f_n\} \in f$. For a given problem domain $D = \{f, P(F)\}$, the emotion classes consist of label space l_s and the prediction function $Pf : f \rightarrow l_s$, where the prediction function Pf is used to predict the corresponding label of the emotion features $Pf(f)$. This process or task is known as $T = \{l_s, Pf(f)\}$ is used in transfer learning of the pairs $\{f_i, ls_i\}$ where $f_i \in F \textit{ and } ls_i \in l_s$. Two major issues must be identified in transfer learning to enhance the learning mechanism. These issues are source domain D_S and target domain D_T while the learning task in both source and target domains should not be equal such that $D_S \neq D_T \textit{ or } T_S \neq T_T$ where T_S is the task in the source domain of the problem and T_T is the task in the target domain of the proposed solution. Transfer learning is considered a primary part of our proposed model for training the image dataset to classify user emotions. In transfer learning, only one model is trained using machine learning, and then it is used as a guide for future related problems.

3.2.2 Horizontal Ensemble Technique

The training of a neural network on a small or medium dataset always has the issue of there being no guarantee that the final training model will be generalized to the validation dataset. The reason for this issue is that when the training curve starts overfitting before the end of training epochs, the curve tends to zeros in case of loss in the measurement of the problem, and the validation curve starts to oscillate in most of the training processes. Therefore, there is no guarantee that the final model is the best-generalized model on the validation data. The horizontal ensemble technique (HET) [39] is proposed to solve this issue through the ensemble of the last N models generated in the training epochs and the average predictions of the probability of each label. This mechanism will be able to generalize and improve the validation accuracy. In this paper, the horizontal ensemble technique (HET) is modified to apply the selection of the N best models during the training epochs.

3.2.3 Proposed Horizontal Ensemble Authentication Framework

The proposed model in this paper presents an innovative idea for providing a challenge-response methodology for authenticating users with high efficiency and accuracy. It can recognize the user emotions independently with no need to store the authorized user emotions in the database. The face authentication process is enhanced by involving the challenge-response of four random emotions to verify whether they are authentic. Fig. 1 presented a schematic diagram for the proposed model. It works as the following:

i) The classifier layer of the pre-trained VGG16 model on the ImageNet dataset that contains 1000 classes is replaced by 7-emotion classes only and re-trained for the emotion classification problem.

ii) The new model is tuned for selecting the number of the re-trained layers from the VGG16 model.

iii) The model is retrained to classify the best easy-to-use 4-emotions classes that can be more expressive by the user during the authentication process.

iv) The horizontal ensemble technique is applied to the 4-emotion model to provide final predictions.

v) The enhanced horizontal ensemble technique is applied to measure the best N losses during the training process of validation data.

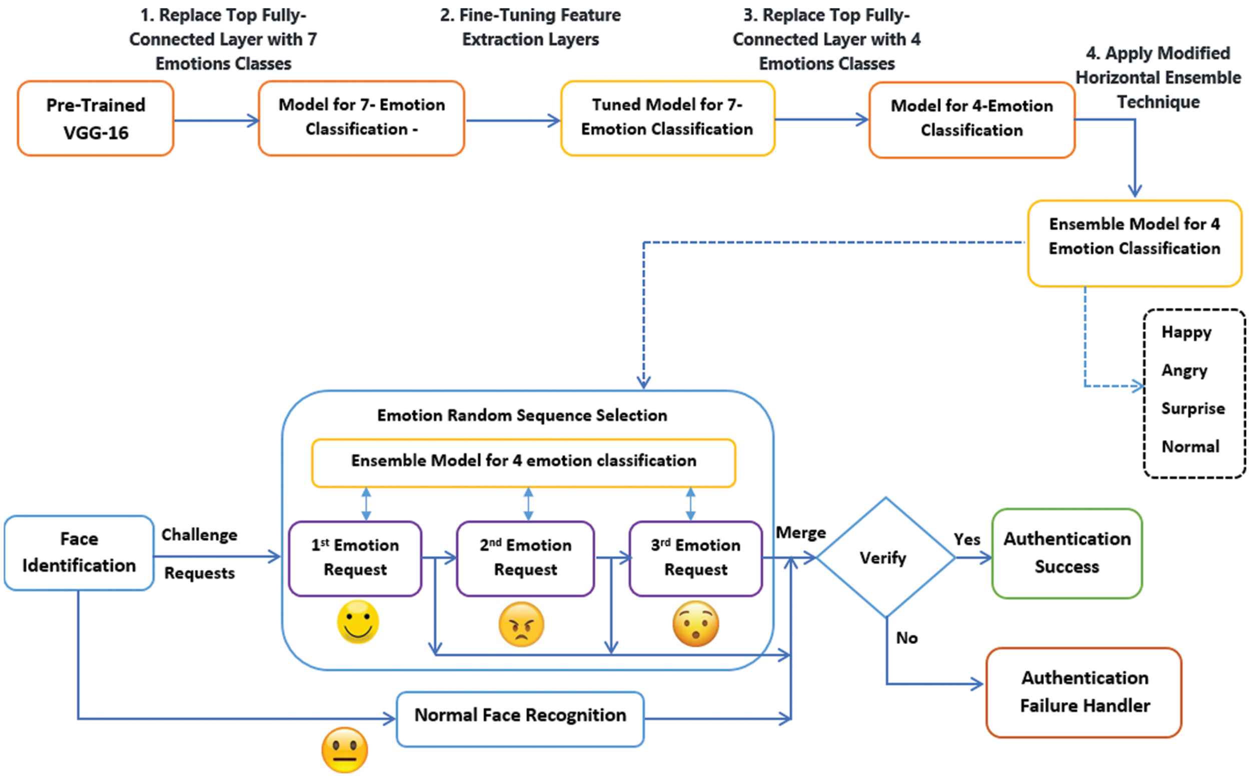

1. Replace Top Fully-Connected Layer with 7 Emotions Classes
2. Fine-Tuning Feature Extraction Layers
3. Replace Top Fully-Connected Layer with 4 Emotions Classes
4. Apply Modified Horizontal Ensemble Technique
Pre-Trained VGG-16
Model for 7- Emotion Classification -
Tuned Model for 7-Emotion Classification
Model for 4-Emotion Classification
Ensemble Model for 4 Emotion Classification
Happy
Angry
Surprise
Normal
Emotion Random Sequence Selection
Ensemble Model for 4 emotion classification
Face Identification
Challenge
Requests
1st Emotion Request
2nd Emotion Request
3rd Emotion Request
Merge
Verify
Yes
No
Authentication Success
Authentication Failure Handler
Normal Face Recognition

Figure 1: Challenge-response authentication framework

4 Results

The evaluation strategy on the results is applied using 10-fold cross-validations without elaboration. More elaboration will be added for the 10-fold cross-validations. The 10-fold cross-validation strategy divides the data into 10 parts, and then builds the model using the first 9 parts and uses the last part (hold out) to test the model. The process is repeated 10 times and each time different 9 parts are selected for training and test using the holdout part. This means that all samples are given the opportunity to be included in the holdout part one time as a test and used to train the model 9 times. The cross-validation splitting was modified to consider the data stratified on the target emotions. After repeating the process 10 times the results of the 10-fold calculation represent the model performance. The experimental results are conducted on four subsequent experiments for classifying emotions during authentication based on different transfer learning and deep learning methods as follows:

4.1 First Experiment

The VGG16 model is loaded without its classifier layer. Then a flatten layer is added after the last pooling layer. A new classifier model is applied with a fully connected dense layer where the output layer will be able to predict the 7-emotion classes of our proposed model. The 7-emotion classes are happy, surprise, sad, disgust, angry, fear, and natural emotions. The VGG16 model is used to train the new layers without updating their weights so that the new output layers will be able to learn and predict the learned features. As a result, there will be an ability to select which layers can be trainable for the new model. The main objective of this experiment is to modify the VGG16 model to classify the 7-available emotions and to tune the features extraction layers. This modification can help select which layers to be retrained again on the emotions dataset. In the first experiment, the pre-trained VGG16 on ImageNet dataset with 1000 classes is used as a baseline for transfer learning after replacing the last layer (i.e., classification layer) of VGG16 that contains 1000 classes with 7-emotion classes on the dense layer as shown in the following equation:

$$VGG16_{1000}\ \langle L_0,\ L_1,\ L_2,\ \ldots,\ L_n[c = 1000]\rangle \rightarrow$$

$$\mathit{Modified\ Neural\ Network}\ \rightarrow VGG16_7\ \langle L_0,\ L_1,\ L_2,\ \ldots,\ L_n[c = 7]\rangle \tag{1}$$

where: L is the VGG16 layer, n is the layer's index, and c is the number of output classes.

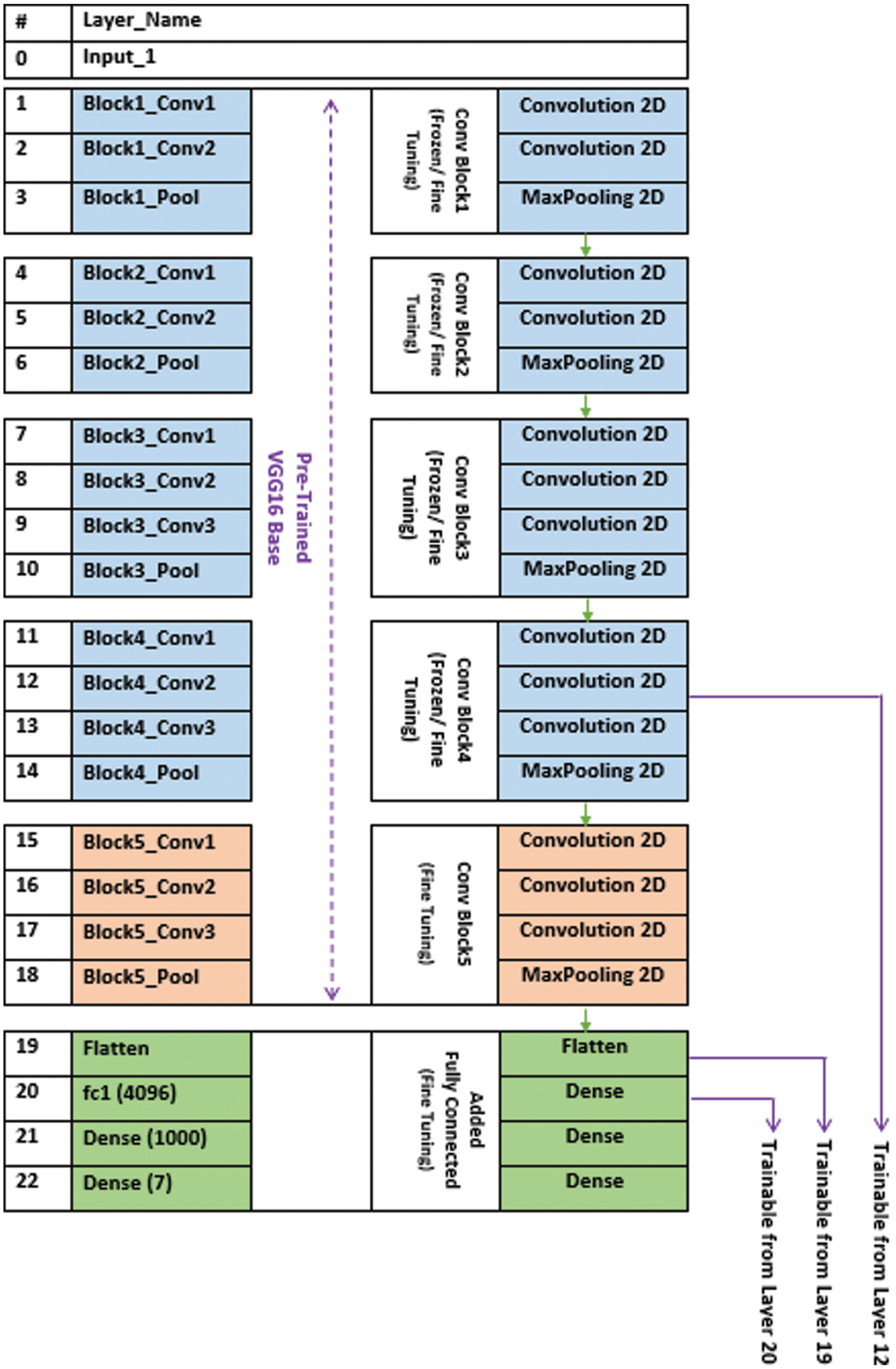
As shown in Fig. 2, the modified $VGG16_7$ was retrained many times to select the best-retrained points. The retrained point of the network is considered the index of the layer that separates the trainable and non-trainable (freezing) layers. This point starts from the last layer (classification layer index 21) and moves backward until there are no improvements in the classification accuracy of the result. The best-retrained point is the point that achieved the best classification accuracy. The algorithm for selecting the best-retrained point is shown below:

Algorithm 1: VGG16 Tuning for 7 Classes

```
1 ret_point = n ( the last layer of the network)
2 net_freez = freez_layers → $VGG16_7$ [$L_0$ : $L_{ret\_point}$]
3 net_train = train_layers → $VGG16_7$ [$L_{ret\_point}$ : $L_n$]
4 SET net = net_freez + net_train
5 $net_{tr}$ = training (net, cross_validation)
6 SET accuracy = evaluate ($net_{tr}$, cross_validation)
7 while accuracy > 50% Do
8     If accuracy > best_ accuracy then
9         best_ accuracy = accuracy
10        best_ret_point = ret_point
11      ret_point = ret_point-1
12      Go to step 2
```

The tuning algorithm is started by initializing the retraining point (ret_point) with the last layer index n. This point marks the VGG16 layers into two types: net freeze and net training. The net_freez is applied by freezing the layers from L_0 to $L_{ret\_point,}$ while the net_train is applied by marking the trainable layers from L_{ret_point} to L_n. The network training is started using the cross-validation mechanism, and then the average accuracy is measured from the net training. The tuning process will not be repeated until the average accuracy drops below 50% (50% is below the random guessing). The tuning process is repeated by moving the retrained point (ret_point) one-step backward until the loop is finished. The repetition process will record the best-achieved average accuracy and the best-retrained point (best_ret_point).

As presented in Fig. 2, the fine-tuning process of VGG16 is retrained starting from the last layer, namely layer index 21 (Dense 1000), for training the 7-emotion classes, and the accuracy of the corresponding layer is measured. A sequential, cumulative process is performed to determine the VGG16 best layer accuracy to train the last *n* layers. As shown in Fig. 3, the best accuracy is found at the retrained point of layer index 15 (Block5_Conv1) with an accuracy of 92.1%, while the retrained point at layer index 13 (Block4_Conv3) achieved 91.6%. The retraining process is continued to layer 12 (Block4_Conv2), where the accuracy drops below 50%. Therefore, the experiments are stopped. The average accuracy retrained of point 15 is distributed on the 7-emotion classes for measuring the best accuracy of each detected emotion where "Happy", "Normal", "Fear", "Surprise", "Sad", "Disgust", and "Angry" emotions recorded 98.2%, 96.6%, 94.4%, 93.1%, 91.1%, 90.8%, and 90.4% respectively.



#	Layer_Name		
0	Input_1		
1	Block1_Conv1	Conv Block1 (Frozen/ Fine Tuning)	Convolution 2D
2	Block1_Conv2		Convolution 2D
3	Block1_Pool		MaxPooling 2D
4	Block2_Conv1	Conv Block2 (Frozen/ Fine Tuning)	Convolution 2D
5	Block2_Conv2		Convolution 2D
6	Block2_Pool		MaxPooling 2D
7	Block3_Conv1	Conv Block3 (Frozen/ Fine Tuning)	Convolution 2D
8	Block3_Conv2		Convolution 2D
9	Block3_Conv3		Convolution 2D
10	Block3_Pool		MaxPooling 2D
11	Block4_Conv1	Conv Block4 (Frozen/ Fine Tuning)	Convolution 2D
12	Block4_Conv2		Convolution 2D
13	Block4_Conv3		Convolution 2D
14	Block4_Pool		MaxPooling 2D
15	Block5_Conv1	Conv Block5 (Fine Tuning)	Convolution 2D
16	Block5_Conv2		Convolution 2D
17	Block5_Conv3		Convolution 2D
18	Block5_Pool		MaxPooling 2D
19	Flatten	Added Fully Connected (Fine Tuning)	Flatten
20	fc1 (4096)		Dense
21	Dense (1000)		Dense
22	Dense (7)		Dense

Figure 2: Retraining process of the modified VGG16

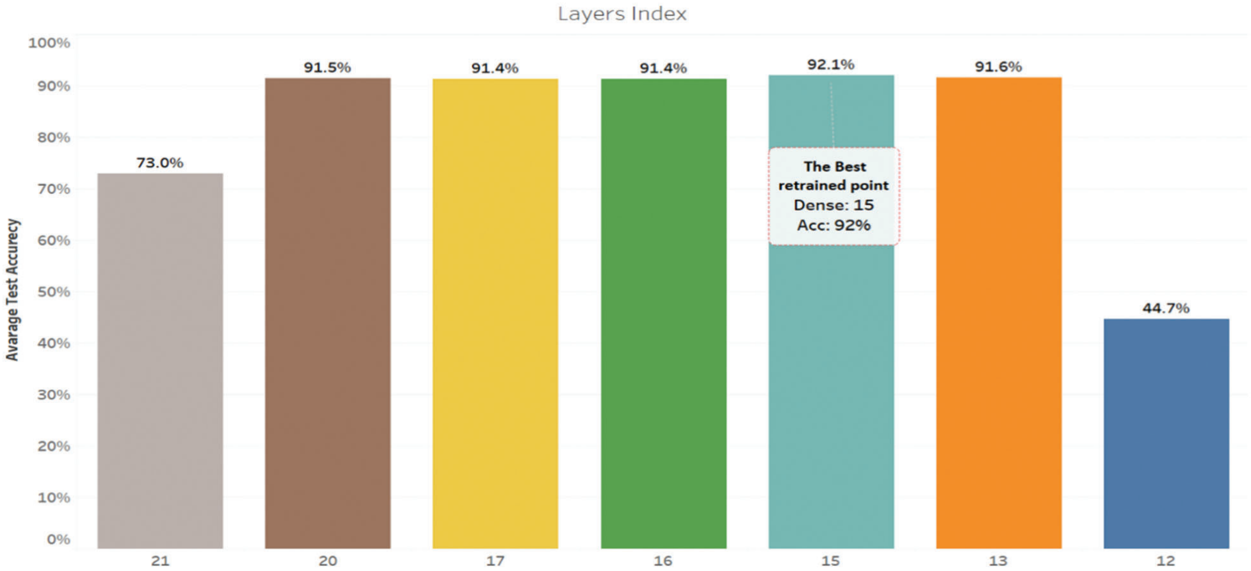


Figure 3: Base model layer accuracy

4.2 Second Experiment

The objective of this experiment is to select the best easy-to-express emotions by the user in the authentication process and improve the recognition accuracy. As stated in the previous experiment, the highest recorded accuracy is 92.1% at the best-retrained point layer of index 15. This accuracy is relatively low when deploying this model for the authentication process, where the average percentage of misuse of the model is 7.9%. The obtained result is an indication of the potential breaches that could take place in the security measures. In addition, the 7-emotions are not common and are not easy to express during the authentication process in a short time by the user. A survey was executed on 85 users to select the best 4-emotions that are common and easy to express quickly from the overall 7-emotions to increase the authentication efficiency of user emotions. The survey objective concentrates on how the users can express their emotions easier in a short time. The survey is processed by selecting the best three emotions from 6 emotions. The "Normal" emotion as normal emotion is the most common and easy, so it is not included in the survey. The results showed that the best expressive emotions are "Happy", "Angry", and "Surprise", with 90.6%, 62.4%, and 51.8% percentage of votes, respectively. The last layer (i.e., classification layer) of VGG16$_7$ that classifies the 7-emotions classes is replaced by four selected emotions, as shown in the following equation.

$$VGG16_7 \langle L_0,\ L_1,\ \ldots,\ L_n[c=7]\rangle \rightarrow \textit{Modified Network} \rightarrow VGG16_4 \langle L_0,\ L_1,\ ,\ \ldots,\ L_n[c=4]\rangle \qquad (2)$$

By utilizing what we achieved in the previous experiments, the modified Neural Network VGG16$_4$ will be retrained using the best-retrained point at layer index 15. The experimental results achieved an average accuracy of 98.7% for recognition of the 4-emotions classes, where the "Happy" emotion achieved 100% accuracy, and the "Surprise" emotion achieved 99.6%. The "Natural" and "Angry" emotions achieved 97.5% and 97.8%, respectively.

4.3 Third Experiment

This experiment involves the horizontal ensemble technique (HET) [39] to improve the authentication reliability using challenge-response emotion, improve the emotion recognition accuracy, and minimize the error during the authentication process. The average accuracy for detecting 4-emotions is 98.7%. In this experiment, the authentication accuracy for emotions is improved using the horizontal ensemble technique for the deep learning model. The horizontal ensemble technique proposes an ensemble of a last number of models during the training of the deep learning model to produce the final predictions of the model.

As presented in Tab. 1, the horizontal ensemble technique involved an ensemble of the last n models during the training process. This experiment turned the N models and tried to ensemble the last 2, 3, 4, 5, 6, and 7 models to evaluate the average recognition accuracy of the 4-emotions classes. Applying the last 2-models result in an average accuracy of 98.51% for detecting the emotion classes. The experiments are continued for the next ensemble models until the average accuracy reached 98.9% by using the last 7-models. In this model, the "Happy" emotion achieved 99.63%, while the "Surprise" emotion achieved 99.60%. The "Normal" and "Angry" emotions achieved 98.54% and 97.82%, respectively.

Table 1: Horizontal ensemble process of 4-emotions

No. of ensemble models	Average accuracy	Happy emotion accuracy	Surprise emotion accuracy	Normal emotion accuracy	Angry emotion accuracy
Last 2	98.51%	99.66%	99.60%	97.74%	97.04%
Last 3	98.15%	98.49%	99.29%	97.76%	97.07%
Last 4	98.72%	98.57%	98.89%	98.97%	98.46%
Last 5	98.53%	99.26%	99.29%	97.74%	97.82%
Last 6	98.33%	99.20%	98.94%	97.05%	98.12%
Last 7	**98.90%**	**99.63%**	**99.60%**	**98.54%**	**97.82%**

4.4 Fourth Experiment

This experiment aims to improve the efficiency of authentication using challenge-response emotion and minimize the computational power required at runtime to deploy the proposed authentication solution on limited resources devices such as IoT and mobile devices. A modified horizontal ensemble technique called Horizontal Ensemble Best N-Loss technique is applied in Algorithm 2.

Algorithm 2: Horizontal Ensemble Best N-Loss (HEBNL)

```
N = 2 ( number of ensemble models)
SET best_N = 2 ( best ensemble models)
best_accuracy = 0
epoch = 50
models = []
loss = []
For e = 1 to epoch Loop
    models[e] = train ( $VGG16_4$, train_data)
    loss[e] = evaluate (models[e], validation_data)
best_models = rank_and_select(models, loss, N)
For m = 1 to N Loop
    m_prediction[m] = best_models[m].predict(X_test_data)
```

(Continued)

Algorithm 2: (continued)

```
13 test_prediction = average (m_prediction[1:N])
14 accuracy = evaluate (target_test_data , test_prediction )
15 If accuracy > best_accuracy then
16     best_accuracy = accuracy
17     Best_N = N
18     N = N + 1
19     Go to step 4
```

The results in this experiment are conducted on Intel ® Core i9 @ 2.81 GHz and 32.0 GB RAM. The average time required for prediction of one emotion is 24 ms in case using one model while it takes 26, 28, 30, 32, 34, and 36 ms when ensemble 2, 3, 4, 5, 6, 7 models. As shown in the time analysis the consumed time & computational power of experiment 4 is better than experiment 3 by 2 ms. The algorithm started by initializing the number of ensemble models (N = 2). Two arrays are defined during the experiment. The first array is used to store the generated models during the training process, while the second array is used to store the archived loss in each epoch. The training epoch is executed one by one for training the $VGG16_4$ on the training data (train_data) and storing the generated model and its corresponding losses. The developed models are ranked using the related losses and the best N losses based on the lowest. The best-selected models are evaluated on the testing data by measuring the average value of the test prediction from each model and then comparing it with the actual target of the test data (target_test_data) to calculate the average accuracy of the selected N best models. This process is repeated until the best accuracy is improved. At this point, the experiments are stopped, and the average accuracy is measured for the best N losses. The modified technique ensembles the N models that give the best losses during the training process on the validation data. In this experiment, the N models are tuned for the models from 2 to 7 to achieve the best losses on validation data. As explained in Fig. 4, the best five ensemble models are selected during the training process. The objective is to tune the N models to achieve the best average recognition accuracy of emotions. As explained, five models are detected that represent the lowest losses in the training models.

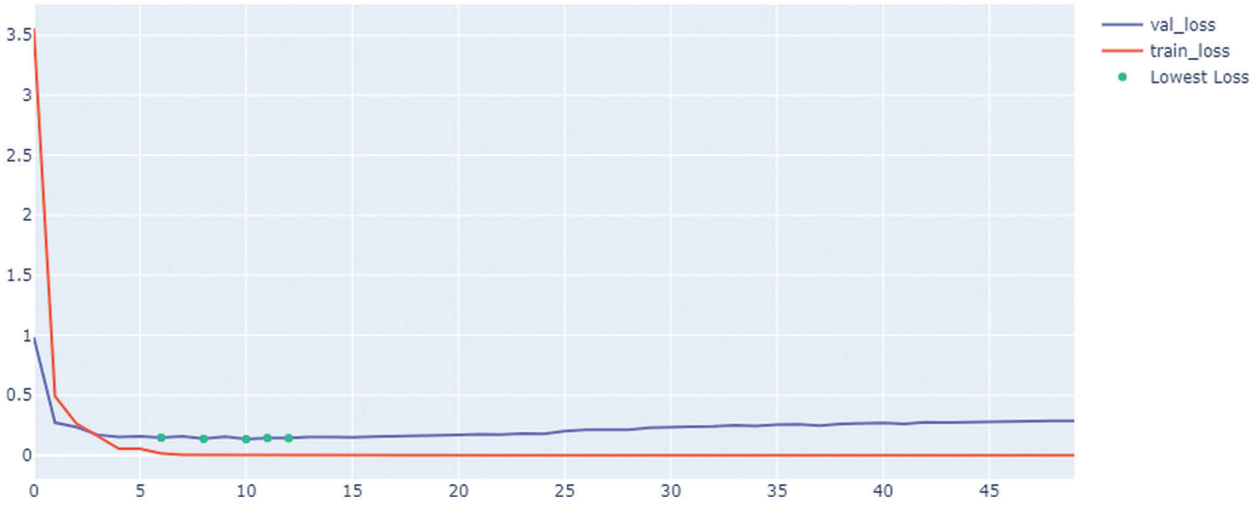


Figure 4: Horizontal ensemble best N losses

The results of tuning the horizontal ensemble Best N Losses technique are presented in Fig. 5, where the best number of ensemble models is five. The five models achieved the highest accuracy with minimum N losses of 99.27%, while model 6 achieved a close accuracy of 99.26%. Models 7 and 2 achieved 99.09% and 99.08% accuracy, respectively, while models 3 and 4 achieved 98.91% and 98.17%, respectively.

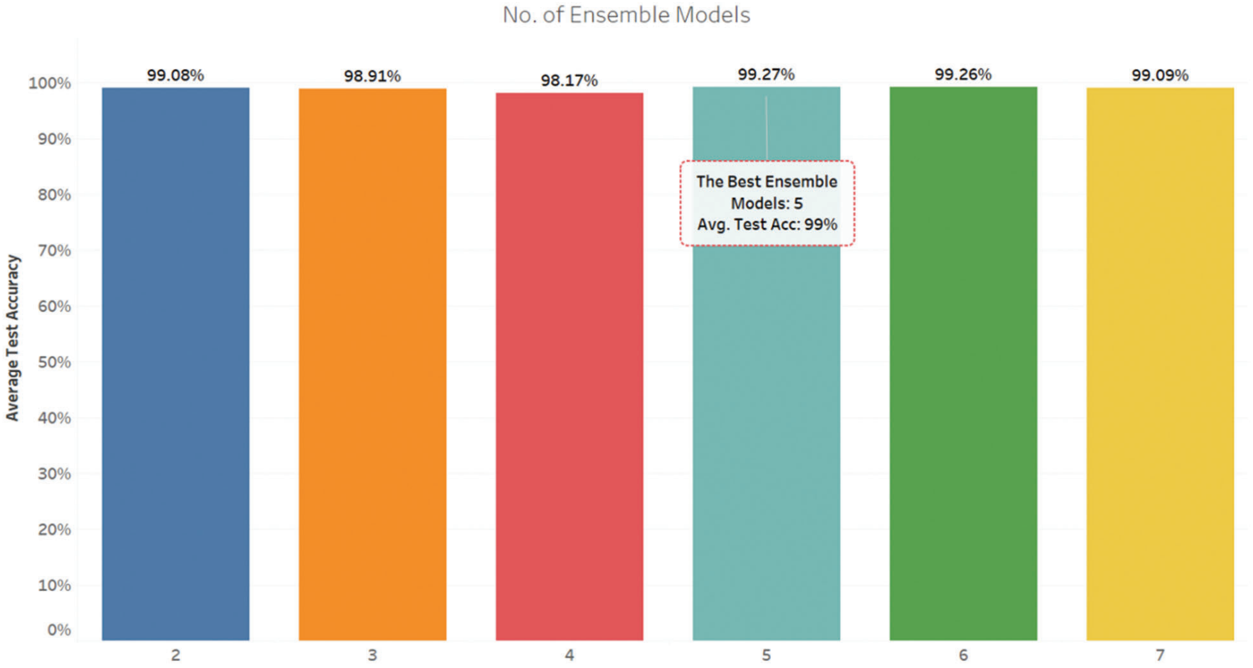


Figure 5: Accuracy of enhanced horizontal ensemble models

5 Discussion

As presented in Tab. 2, the Horizontal Ensemble Best N-Loss technique achieved the best average accuracy of 99.27% on five models with a minimum standard deviation of 0.009, while the average accuracy on two models achieved 99.08% with a standard deviation of 0.013.

Table 2: Horizontal ensemble best N-loss of 4-emotions

N-Best ensemble models	Average accuracy	Standard deviation	Angry accuracy	Happy accuracy	Normal accuracy	Surprise accuracy
2	99.08%	0.013	98.88%	99.60%	98.57%	99.29%
3	98.91%	0.018	98.06%	99.66%	98.23%	99.63%
4	98.17%	0.015	97.71%	98.54%	97.76%	98.59%
5	**99.27%**	**0.009**	**98.88%**	**99.60%**	**99.28%**	**99.31%**
6	99.26%	0.010	98.91%	99.23%	99.31%	99.60%
7	99.09%	0.018	98.06%	99.66%	98.57%	100.00%

As presented in Tab. 3, a comprehensive analysis of different classifiers for emotion detection is proposed. These classifiers are applied to the KDEF dataset using 10-fold cross-validations. A different sequence of user emotions is deployed on the classifiers to evaluate the accuracy of detecting the user emotion. The table shows that each detected user emotion achieved different accuracy according to the classifier type and emotional expressiveness. The obtained results showed the highest accuracy of the conducted experiments, especially the proposed HEBNL that recorded an accuracy of 99.27%.

Table 3: Comparative analysis of recent emotion classifiers

Ref	Classifier	Data set	Surprise %	Angry %	Fear %	Disgust %	Happy %	Normal %	Sad %	Average accuracy
[40]	Virtual variation training	KDEF	92.86	73.57	52.14	81.43	95.71	NA	72.14	77.98
[41]	CNN + SVM	KDEF	100	95.24	90.48	95.24	97.62	100	95.24	96.26
[42]	POEM	KDEF	81.7	59.2	50.0	79.2	97.5	97.5	50.0	73.6
[9]	HOG + SVM	KDEF	97.62	83.0	52.38	76.19	96.0	NA	62.0	NA
[43]	Transfer leaning using deep CNN	KDEF	NA	NA	NA	NA	NA	NA	NA	98.78
[44]	FER-net convolution neural network.	KDEF	94	100	37	76	100	84	71	83.0
[45]	LA-Net lightweight attention CNN	KDEF	NA	NA	NA	NA	NA	NA	NA	93
[46]	ShuffleNet + node_163 + SVM	KDEF	NA	NA	NA	NA	NA	NA	NA	87.8
[47]	Pre-trained VGG-Face	KDEF	77.14	72.14	80.71	90.00	99.29	91.43	91.43	83.98
	Random patches	KDEF	80.71	60.71	74.29	94.29	100.00	95.00	84.29	84.18
	Triplet loss	KDEF	85.00	70.71	72.86	91.43	97.14	93.57	68.57	82.76
Paper results	VGG16$_7$	KDEF	93.1	90.4	94.4	90.8	98.2	96.7	91.1	92.1
	VGG16$_4$	KDEF	99.6	97.8	NA	NA	100	97.5	NA	98.7
	HET	KDEF	99.6	97.82	NA	NA	99.63	98.54	NA	98.9
	HEBNL	KDEF	99.31	98.88	NA	NA	99.60	99.28	NA	**99.27**

6 Conclusion

In this paper, an innovative challenge-response emotions authentication model based on the horizontal ensemble technique is proposed and applied to authenticate a user based on a random sequence of emotions with high efficiency and accuracy. The applied model depends mainly on challenging the authorized user using a random sequence of emotions to provide a user response for every authentication trial with a different sequence of emotions. Many improvements are applied to the system to get the best accuracy and performance. First, the VGG16 is applied to seven user emotions to tune the best re-train point to enhance the measure of overall detection accuracy. Second, the VGG16 is applied to four common and expressive emotions from the seven emotions. These emotions are selected based on a survey of different users to choose the most common emotions that are easy and can be expressed in a short time. Third, a horizontal ensemble technique (HET) is applied to minimize recognition error and enhance emotion detection accuracy. Finally, a novel Horizontal Ensemble Best N-Losses (HEBNL) mechanism is applied to improve the efficiency of user authentication during the challenge-response process. The proposed HEBNL achieved an accuracy face of 99.27% which is considered the highest accuracy compared to other emotion detection classifiers. The proposed method has a limitation regarding the user response to the challenge emotion sequence. Two procedures will be applied where the user account will be locked or a different authentication technique will be proposed to the user such as a PIN code or password that is not recommended as it reduces the security performance. In feature work, the research can be extended to use different challenge-response techniques such as eye blinking or lip movement. These techniques can be added as a high-level challenge-response so the user is challenged by either emotion, eye blanking, or lip movement. Another future work is to use these techniques cascaded one after another. Therefore, if the user failed to authenticate using one technique, the next technique will be applied.

Funding Statement: This work is partially supported by the Deanship of Scientific Research at Jouf University under Grant No (DSR-2021–02–0369).

Conflicts of Interest: The authors state that they have no conflicts of interest to report regarding the present study.

References

[1] M. Said, K. Mohamed, A. Elshenawy and M. Ezz, "A survey on smartphone protecting identification against attacks using biometric authentication systems," *Journal of Al-Azhar University Engineering Sector*, vol. 16, no. 59, pp. 288–299, 2021.

[2] A. Ali, T. El-Hafeez and Y. Mohany, "An accurate system for face detection and recognition," *Journal of Advances in Mathematics and Computer Science*, vol. 33, pp. 1–19, 2019.

[3] J. Galka, M. Masior and M. Salasa, "Voice authentication embedded solution for secured access control," *IEEE Transactions on Consumer Electronics*, vol. 60, no. 4, pp. 653–661, 2014.

[4] G. Singh, R. Singh, R. Saha and N. Agarwal, "IWT based iris recognition for image authentication," *Procedia Computer Science*, vol. 171, pp. 1868–1876, 2020.

[5] W. Yang, S. Wang, J. Hu, G. Zheng and C. Valli, "A fingerprint and finger-vein based cancelable multibiometric system.," *Pattern Recognition*, vol. 78, pp. 242–251, 2018.

[6] M. Bouzakraoui, A. Sadiq and N. Enneya, "A customer emotion recognition through facial expression using POEM descriptor and SVM classifier," in *Int. Conf. on Big Data, Cloud and Applications (BDCA'17)*, Morocco, pp. 1–6, 2017.

[7] M. Leo, P. Carcagnì, P. Mazzeo, P. Spagnolo, D. Cazzatoet *et al.,* "Analysis of facial information for healthcare applications: A survey on computer vision-based approaches," *MDPI: Information*, vol. 11, no. 123, pp. 1–26, 2020.

[8] E. Zineb, E. Othmane, K. Mustapha and A. Hakim, "Robust facial expression recognition system based on hidden markov models," *International Journal of Multimedia Information Retrieval*, vol. 5, pp. 229–236, 2016.

[9] S. Eng, H. Ali, A. Cheah and Y. Chong, "Facial expression recognition in JAFFE and KDEF datasets using histogram of oriented gradients and support vector machine," in *Int. Conf. on Man Machine Systems*, Malaysia, vol. 705, pp. 1–8, 2019.

[10] Y. Ye, X. Zhang, Y. Lin and H. Wang, "Facial expression recognition via region-based convolutional fusion network," *Journal of Visual Communication and Image Repesentation*, vol. 62, pp. 1–11, 2019.

[11] T. Ozcan and A. Basturk, "Static facial expression recognition using convolutional neural networks based on transfer learning and hyperparameter optimization," *Multimedia Tools and Applications*, vol. 79, pp. 26587–26604, 2020.

[12] V. Dharanya, A. Raj and V. Gopi, “Facial expression recognition through person-wise regeneration of expressions using auxiliary classifier generative adversarial network (AC-GAN) based model,” *Journal of Visual Communication and Image Representation*, vol. 77, pp. 1–12, 2021.

[13] Z. Han and H. Huang, “GAN based three-stage-training algorithm for multi-view facial expression recognition,” *Neural Processing Letters*, vol. 53, pp. 4189–4205, 2021.

[14] A. Butalia, M. Ingle and P. Kulkarni, “Facial expression recognition for security,” *International Journal of Modern Engineering Research (IJMER)*, vol. 2, no. 4, pp. 1449–1453, 2012.

[15] D. Yin, S. Omar, B. Talip, A. Muklas, N. Norain *et al.,* “Fusion of face recognition and facial expression detection for authentication: A proposed model,” in *Int. Conf. on Ubiquitous Information Management and Communication*, Japan, pp. 1–8, 2017.

[16] D. Yin, A. Muklas, R. Chik, A. Othman and S. Omar, “A proposed approach for biometric-based authentication using of face and facial expression recognition,” in *IEEE Int. Conf. on Communication and Information Systems (ICCIS)*, Singapore, pp. 28–33, 2018.

[17] J. K. B. N. D. U. ArtiPadole, “Facial biometric for secure login using facial feature extraction,” *International Journal of Future Generation Communication and Networking*, vol. 13, no. 2, pp. 1246–1255, 2020.

[18] X. Zhang, L. Yao, C. Huang, T. Gu, Z. Yang *et al.,* “A multimodal biometric authentication system via deep decoding gaits and brainwaves,” *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 4, pp. 1–24, 2020.

[19] I. Sluganovic, M. Roeschlin, K. Rasmussen and I. Martinovic, “Using reflexive eye movements for fast challenge-response authentication,” in *ACM Conf. on Computer and Communications Security*, USA, pp. 1056–1067, 2016.

[20] C. Chou, “Presentation attack detection based on score level fusion and challenge-response technique,” *Journal of Supercomputing*, vol. 77, pp. 4681–4697, 2021.

[21] W. Xu, J. Tian, Y. Cao and S. Wang, “Challenge-response authentication using in-air handwriting style verification,” *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 1, pp. 51–64, 2020.

[22] Y. Chen, Z. Yang, R. Abbou, P. Lopes, B. Zhao *et al.,* “User authentication via electrical muscle stimulation,” in *CHI Conf. on Human Factors in Computing Systems*, USA, pp. 1–15, 2021.

[23] A. Hanumanthaiah and M. Eraiah, “Challenge responsive multi modal biometric authentication resilient to presentation attacks,” *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 2, pp. 494–507, 2022.

[24] A. Buriro, B. Crispo, F. Delfrari and K. Wrona, “Hold and sign: A novel behavioral biometrics for smartphone user authentication,” in *IEEE Security and Privacy Workshops*, USA, pp. 276–285, 2016.

[25] S. Makowski, L. Jäger, P. Prasse and T. Scheffer, “Biometric identification and presentation-attack detection using micro-and macro-movements of the eyes,” in *IEEE Int. Joint Conf. on Biometrics (IJCB)*, USA, pp. 1–10, 2020.

[26] P. Linn and E. Htoon, “Face anti-spoofing using eyes movement and CNN-based liveness detection,” in *IEEE Int. Conf. on Advanced Information Technologies (ICAIT)*, Myanmar, pp. 149–154, 2019.

[27] N. Alsufyani, A. Ali, S. Hoque and F. Deravi, “Biometric presentation attack detection using gaze alignment,” in *IEEE 4th Int. Conf. on Identity, Security, and Behavior Analysis (ISBA)*, Singapore, pp. 1–8, 2018.

[28] A. Ali, S. Hoque and F. Deravi, “Biometric presentation attack detection using stimulated pupillary movements,” in *IEEE Int. Conf. on Imaging for Crime Detection and Prevention*, UK, pp. 80–85, 2019.

[29] S. Makowski, P. Prasse, D. Reich, D. Krakowczyk, L. Jäger *et al.,* “Deepeyedentificationlive: Oculomotoric biometric identification and presentation-attack detection using deep neural networks,” *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 3, no. 4, pp. 506–518, 2021.

[30] M. Mohzary, K. Almalki, B. Choi and S. Song, “Your eyes show what your eyes see (Y-EYES): Challenge-response anti-spoofing method for mobile security using corneal specular reflections,” in *ACM Workshop on Security and Privacy for Mobile AI*, USA, pp. 25–30, 2021.

[31] M. Rahman and S. Basak, “Identifying user authentication and most frequently used region based on mouse movement data: A machine learning approach,” in *IEEE Annual Computing and Communication Workshop and Conf. (CCWC)*, USA, pp. 1245–1250, 2021.

[32] M. Raghavendra, P. Omprakash and R. Mukesh, "AuthNet: A deep learning based authentication mechanism using temporal facial feature movements," in *AAAI Conf. on Artificial Intelligence*, India, vol. 35, no. 18, pp. 15873–15874, 2021.

[33] J. Fan, J. Lee and Y. Lee, "Application of transfer learning for image classification on dataset with not mutually exclusive classes," in *IEEE Int. Technical Conf. on Circuits/Systems, Computers and Communications*, Singapore, pp. 1–4, 2021.

[34] M. Soumik, A. Aziz and A. Hossain, "Improved transfer learning based deep learning model for breast cancer histopathological image classification," in *IEEE Int. Conf. on Automation, Control and Mechatronics for Industry*, Bangladesh, pp. 1–4, 2021.

[35] S. Khan, E. Ahmed. M. Javed, S. Shah and S. Ali, "Transfer learning of a neural network using deep learning to perform face recognition," in *IEEE Int. Conf. on Electrical, Communication and Computer Engineering*, Pakistan, pp. 1–5, 2019.

[36] W. Wang, X. Huang, J. Li, P. Zhang and X. Wang, "Detecting COVID-19 patients in X-ray images based on MAI-nets," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 1607–1616, 2021.

[37] Y. Gui and G. Zeng, "Joint learning of visual and spatial features for edit propagation from a single image," *The Visual Computer*, vol. 36, no. 3, pp. 469–482, 2020.

[38] W. Wang, Y. T. Li, T. Zou, X. Wang, J. Y. You *et al.,* "A novel image classification approach via Dense-MobileNet models," *Mobile Information Systems*, vol. 14, no. 1, pp. 1607–1616, 2020.

[39] J. Xie, B. Xu and Z. Chuang, "Horizontal and vertical ensemble with deep representation for classification," in *ICML Workshop on Representation Learning*, USA, pp. 1–6, 2013.

[40] E. Cruz, C. Jung and C. Franco, "Facial expression recognition using temporal POEM features," *Pattern Recognition Letters*, Elsevier, vol. 13, pp. 13–21, 2020.

[41] Z. Sun, Z. Hu, M. Wang and S. Zhao, "Individual-free representation-based classification for facial expression recognition," *Signal, Image and Video Processing*, vol. 11, pp. 597–604, 2016.

[42] A. Garcia, M. Elshaw, A. Altahhan and V. Palade, "A hybrid deep learning neural approach for emotion recognition from facial expressions for socially assistive robots," in *Neural Computing and Applications*, Springer, Germany, vol. 29, pp. 359–373, 2018.

[43] M. Akhand, S. Roy, N. Siddique, A. Kamal and T. Shimamura, "Facial emotion recognition using transfer learning in the deep CNN," *MDPI: Electronics*, vol. 10, no. 9, pp. 1–19, 2021.

[44] K. Mohan, A. Seal, O. Krejcar and A. Yazidi, "FER-net: Facial expression recognition using deep neural net," in *Neural Computing and Applications*, Springer, Germany, vol. 33, pp. 9125–9136, 2021.

[45] H. Ma, T. Celik and H. Li, "Lightweight attention convolutional neural network through network slimming for robust facial expression recognition," *Signal, Image and Video Processing*, vol. 15, pp. 1507–1515, 2021.

[46] Z. Fei, E. Yang, L. Yu, X. Li, H. Zhou *et al.,* "A novel deep neural network-based emotion analysis system for automatic detection of mild cognitive impairment in the elderly," *Neurocomputing*, Elsevier, vol. 468, pp. 306–316, 2021.

[47] W. Dias, F. Andaló, R. Padilha, G. Bertocco, W. Almeida *et al.,* "Cross-dataset emotion recognition from facial expressions through convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 82, pp. 1–18, 2021.