SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
http://saucis.sakarya.edu.tr/
e-ISSN: 2636-8129
Publisher: Sakarya University
Vol. 7, No. 3, 2024, 427-438
DOI: 10.35377/saucis...1503989

RESEARCH ARTICLE

A Comparison of Transfer Learning Models for Face Recognition

Dalhm Al-Shammari[1]*, Devrim Akgun[2]

[1] Computer Engineering Department, Institute of Natural Sciences, Sakarya University, Sakarya, Türkiye
[2] Software Engineering Department, Computer and Information Sciences Faculty, Sakarya University, Sakarya, Türkiye

Corresponding author:
Dalhm Al-Shammari,
Computer Engineering Department,
Institute of Natural Sciences,
Sakarya University, Sakarya, Türkiye
dalhm.ashammari@ogr.sakarya.edu.tr

Article History:
Received: 24.06.2024
Accepted: 04.09.2024
Published Online: 23.12.2024

ABSTRACT
Face recognition (FR) is a method that uses face feature analysis and comparison to identify or verify individuals. Siamese neural networks (SNNs) are an effective method for FR, providing high accuracy and versatility, especially in situations where data is restricted. Unlike standard neural networks, SNNs learn to distinguish between pairs of inputs rather than individual inputs. However, detecting and recognizing faces in unconstrained environments poses a significant challenge due to various factors such as head pose, illumination, and facial expression variations. The aim of this paper is to design and develop an efficient approach based on SNNs and Transfer Learning methods. For this purpose LFW dataset and transfer learning architectures like VGG-16, EfficientNet, RestNet50 and ConvNext have been utilised. Performance of the architectures were measured using 5-Fold cross validation. According to results, EfficientNet, RestNet50 and ConvNext produced 78% accuracy, 95% and 93 % accuracy respectively. SNN with VGG-16 exhibited a low loss and produced the best accuracy in face recognition with 96%.

Keywords: Face recognition, Siamese neural network, VGG-16, ConvNext, EfficientNet, RestNet50

1. Introduction

FRsystems can be tackled either as an identification or verification problem. Face identification, also known as the 1: n matching problem, refers to the process of matching a given face with multiple faces in a database. The unidentified face is compared to all the faces in the known identities database, and a decision is reached based on the outcome of these comparisons. If the person is known to be in the database, the task is referred to as closed-set. Otherwise, it is referred to as open-set. Face verification is commonly referred to as the 1:1 matching problem. The query face's identity is determined by comparing it to the face data of the identity that is claimed in the database. it is either confirmed or rejected based on this comparison [1].

Various FRtechniques have achieved significant success in controlled environments [2], [3]. Due to the nonstable nature of facial images and the fact that it influences how the same person appears in multiple face images in real-life situations, these techniques commonly fail. Therefore, it is necessary to modify these techniques or design new suitable techniques. Due to illumination, pose variation, plastic surgery, expression, ageing, low resolution difficulties, as well as the fact that face acquisition processes may go through a broad range of modifications, the majority of FRalgorithms have not yet achieved the ideal level of recognition accuracy [4]. Therefore, it is important for any proposed solution to focus on enhancing the efficacy of the detection and recognition models while reducing the complexity of resource utilization during the recognition process. Illumination refers to the variations in lighting that can cause a face image to appear differently. The extraction of illumination invariant features continues to be a challenge for robust FR systems. the use of illumination has significant effects on an image, resulting in changes to its location, shadow shape, and contrast gradients [5]. Humans have the ability to recognize faces even when there are changes in lighting. However, FR systems struggle with this task and testing or training these systems is also challenging when lighting conditions vary. Therefore, it is necessary to use image processing techniques to normalize face images that have different lighting conditions.

The general face structure of people may vary significantly as they age, and their distinctive facial features may also alter. As a result, FRtechniques commonly fail in these types of situations [6]. In literature, it has been observed that there is a significant decrease in performance in face recognition-based ageing when there is a large age difference between the target image and the query image [7]. Therefore, the query imagine must be represented using sparse representation classification in order to acquire the residuals for each class, where each pixel is estimated as either occluded or not in the residual for each class.

Cite as: D. Al-Shammari, D. Akgun, "A Comparison of Transfer Learning Models for Face Recognition," *Sakarya University Journal of Computer and Information Sciences*, vol.7, no. 3, pp. 427-438, 2024. Doi: 10.35377/saucis…1503989

Humans are capable of displaying a variety of expressions on their faces at all times, unless they are frozen in a static position. These expressions are utilized for communicating diverse feelings and mental states to others. Facial emotion alters the geometry and appearance of the face, which decreases the accuracy of facial recognition [8]. Therefore, it is necessary to apply customized pre-processing techniques that can only extract features specific for certain expressions-from the face image.

It has been observed that FRtechniques are unable to identify individuals' faces accurately after they undergo plastic surgery [2]. In this scenario, the facial image of a person is completely altered, resulting in a completely different individual. the process of cosmetic surgery on the human face results in a change in skin texture between photographs of the same person, making the job of face identification challenging [9]. Therefore, it is necessary to apply the local binary pattern to important regions of the face image, rather than the entire face image. The concept is grounded in the notion that only that local binary patterns (lbp) containing essential information, such as corners and edges, will be valuable for facial recognition following plastic surgery.

The goal of this research is to develop and implement an efficient methodology based on SNN and transfer learning techniques. This is done using the LFW dataset with transfer learning architectures such as VGG-16, EfficientNet, RestNet50, and ConvNext. The performance of the architectures was evaluated using 5-fold cross validation on LFW dataset. The rest of this paper is organized as follows. Section 2 presents the proposed methodology that utilized to achieve the goals as well this section illustrates the used datasets. The results discussion and analysis of the proposed system is explained is section 3. Section 4 concludes this study.

2. Materials and Methods

This section consists of three actions: designing the SNN with pre-trained networks VGG-16, EfficientNetB0, ConvNeXtBase and ResNet50 as a backbone based FRmodel and providing an explanation of the LFW dataset. Finally, the SNN with VGG-16, EfficientNetB0, ResNet50 and ConvNextTiny model was trained using the same dataset.

2.1. Transfer Learning

A Transfer Learning [10] a particular techniques used to transfer previously learned knowledge gained from a related domain to a target field. Since our data set is compact with around 13,000 samples, to avoid training issues, we perform data transfer. There are specific steps to transfer the particular knowledge. First, we load the model we want to transfer in this study use VGG-16, EfficientNetB0, ConvNeXtBase and ResNet50, and then determine how many layers the previous training requires to keep. Next, we use the new model to be the backbone of the SNN model. The training set used in this second stage is the LFW in the next stage, the pre-trained network model. It is used for LFW testing. Finally, accuracy is calculated based on the task Transfer learning. Architectural flow plan for transfer of learning.

While selected to train the final two layers of the model without freezing it. This approach was selected to optimize the model's adaptability and responsiveness to the dataset, leading to consistent performance and learning In this research, the pre-trained models of VGG16, EfficientNetB0, ConvNeXtBase and Wide ResNet-50 were used. The VGG16, EfficientNetB0, ConvNeXtBase and Wide ResNet-50 pre-trained models expect input images normalized in mini-batches of 3-channel RGB images of shape (3 × H × W), where H and W are expected to be 62 x 47. The final classification layers were replaced with fully connected layers. ReLU activation functions and with a binary cross-entropy loss were also used. The initial layers from training were frozen, and the modified layer was fine-tuned with the LFW dataset. The model was trained for 120 epochs with a batch size of 32. Adam [11]. was used as the model optimizer with training rate (1e-4). To enhance the model's efficiency in terms of computation time and performance, model enhancement techniques were used.

2.1. Siamese Neural Network Algorithm

A Siamese neural network [12] refers to a type of neural network architecture that consists of two or more subnetworks that are identical to each other. In this context, 'identical' refers to having the same configuration, parameters, and weights. The parameter updating is mirrored across both sub-networks. Artificial neural networks are utilized to determine the similarity between inputs by comparing their feature vectors. As a result, these networks find application in numerous domains [13].

Neural networks are nearly perfect at every task in the deep learning era of today, but they need more data to do so. However, in certain cases such as FRobtaining more data may not always be feasible. To address these challenges, a novel neural network architecture known as Siamese neural networks has been developed [14]. It utilizes a limited number of images to improve its predictions. Moreover, SNN have gained popularity in recent years due to their remarkable ability to learn from minimal amounts of data.

In this study, the Siamese network consists of 2 identical neural networks, each with VGG-16 as the backbone, that receive the input images. The last layers of these networks are subsequently connected to a binary cross-entropy loss function [15], responsible for measuring the similarity between the two inputs. To ensure simultaneous learning of identical features for both input images, the twin networks share the same configuration and parameters [16].Two scenarios are presented in this study; first scenario is having 2 images of the same person "positive pair". Second scenario is feeding 2 images of different individuals "negative pair". Through training, the network becomes capable of differentiating between positive and negative pairs, which is also used to compare a new, unseen image with a stored image for authentication purposes. As an output, the network gives out a metric that quantifies how much the two input images are different. This metric is then compared with a threshold to decide whether the two images are of the same person or not. The threshold is determined based on the distribution of distances observed during the training phase. To summarize, the SNN with VGG-16 as the backbone presents a reliable and effective approach for FR. Its ability to handle diverse factors like pose, lighting, and expression renders it well-suited for practical applications, including the specific scenario described in this study. As seen in Figure 1 explain the architecture of SNN with VGG-16, EfficientNetB0, ConvNeXtBase and ResNet50 model as a backbone.

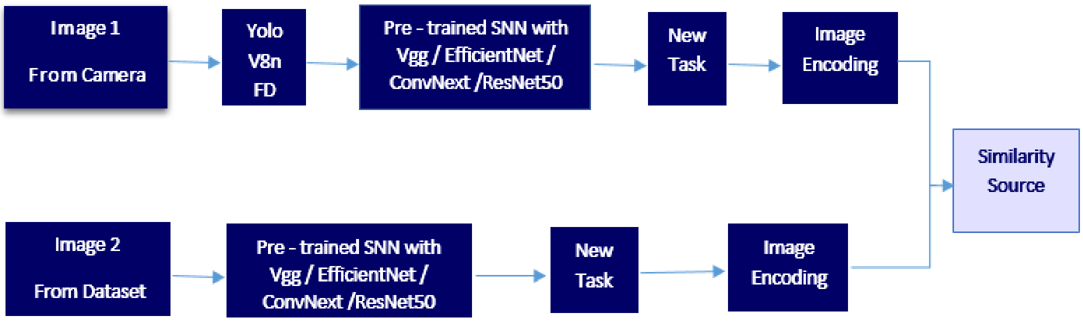


Figure 1. The Architecture of SNN with VGG-16, EfficientNetB0, ConvNeXtBase and ResNet50 Model as a Backbone

2.2. VGG-16 Network

VGG-16 [17], named after the visual geometry group (VGG) at oxford, is a convolutional neural network (CNN) architecture that has proven to be one of the most effective vision model architectures to date. It consists of 16 layers with weights, making it a relatively extensive network with a total of 138 million parameters. The architecture of VGG16 is simple and incorporates the most important features of convolutional neural networks. It uses small 3×3 convolution filters and a stride of 1, which are in the same padding, and a max pool layer of 2×2 filters for stride 2. This arrangement of convolution and max pool layers is consistent throughout the whole architecture. The network also includes three fully connected layers. the simplicity of the VGG-16 architecture is its main attraction, and its ability to handle diverse factors like pose, lighting, and expression makes it well-suited for practical applications [18]. As seen in Figure 2 illustrates the general architecture of the VGG-16 network before transfer learning.

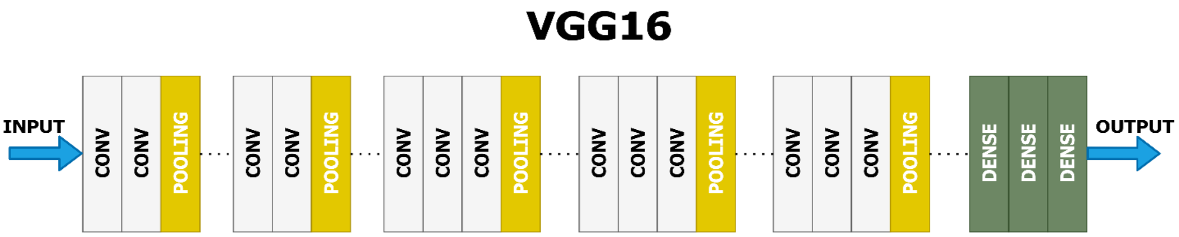


Figure 2. The General Architecture of the VGG-16 Network

To perform the transfer learning process, we freeze the layers to preserve the weights that were previously trained on ImageNet dataset, then separate the Vgg16 head, which consists of three fully connected layers and the final CNN, and replace them with SNN layers to perform the recognition process by adding relu as activation function As seen in Figure 3.

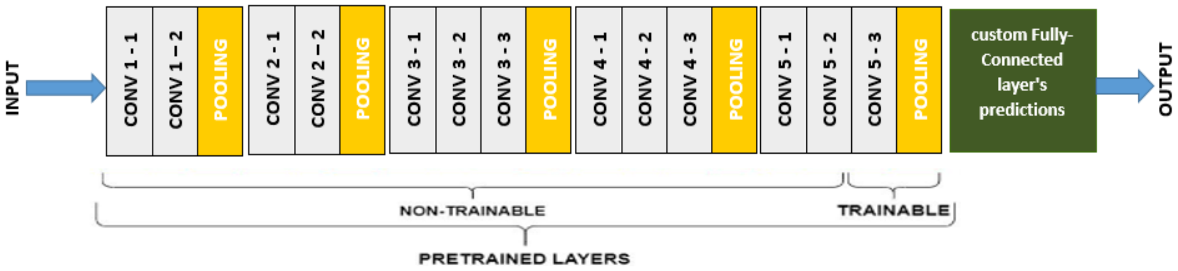


Figure 3. Adding New Layers on the Top of the VGG-16 Model [19]

2.3. EfficientNetB0

EfficientNetB0 [20], is a convolutional neural network architecture, It solves the challenging task of reconciling accurate modeling with speed of computation, with a target of reaching state-of-the-art performance with significantly fewer the parameters and utilized resources than previous versions. The design makes certain that all network parameters modify in a consistent manner, improving performance while decreasing computation weight. The primary feature of EfficientNet is its composite scaling mechanism, which constantly changes the network's depth, width, and accuracy according to a set of predefined scaling factors. EfficientNet includes a hierarchy of layers formed comprised of repeating blocks that gradually improve the network's depth, width, and resolution. EfficientNet-B0 that used in this study has 7 layers in each block. EfficientNet-B0 has fewer layers compared to EfficientNet-B7 has 19 layers in each block. Once the architecture lengths, a total number of layers increases significantly. After construct the SNN, EfficientNetB0's layers should be frozen except for the last layers. Train the SNNs using the LFW dataset while updating the unfrozen layers' weights. Figure.4 shows the general architecture of the EfficientNetB0. As well Figure 5 shows the transfer learning.

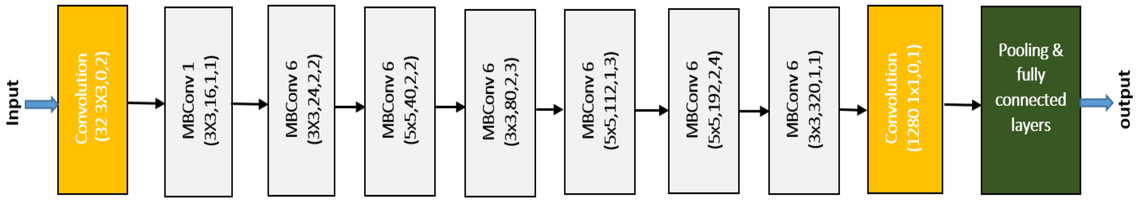


Figure 4 The general architecture of the EfficientNetB0 model [20].

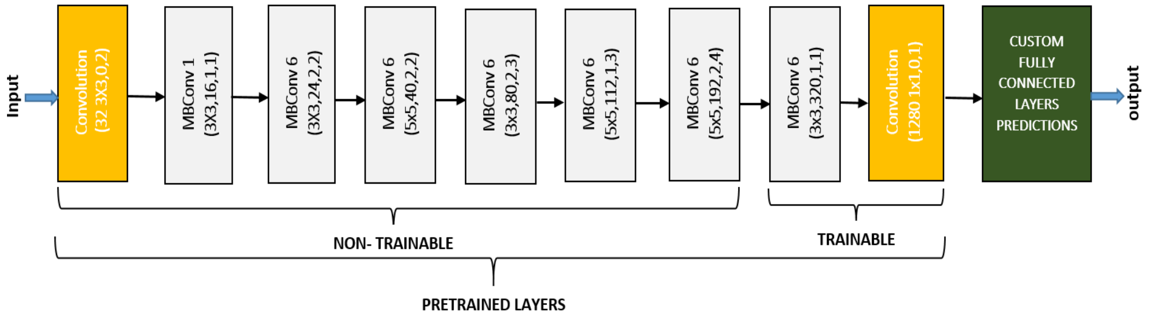


Figure 5. Adding New Layers on the top of the EfficientNetB0 Model [20]

2.4. ResNet50

ResNet50 is an architecture for convolutional neural networks that was first presented by [21]. It belonged to the ResNet (Residual Network) family, which has been developed to use residual learning to solve the degradation issue that extremely deep networks face. Furthermore, ResNet50 was developed using a fundamental component architecture and has 50 layers As seen in Figure 6 and Figure 7 show the trainable model.

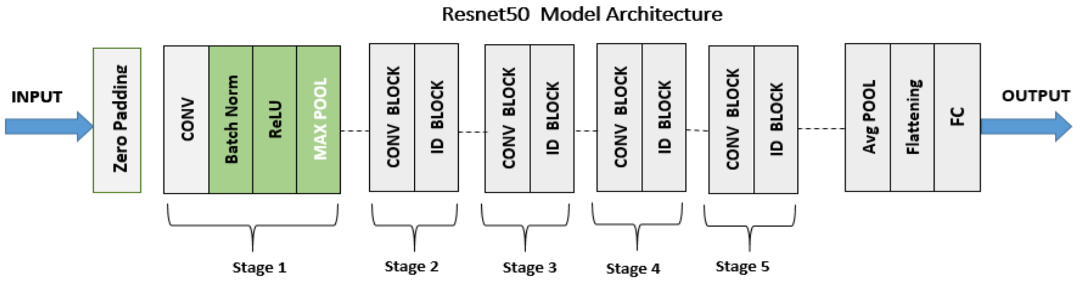


Figure 6. The General Architecture of Baseline ResNet50 Model

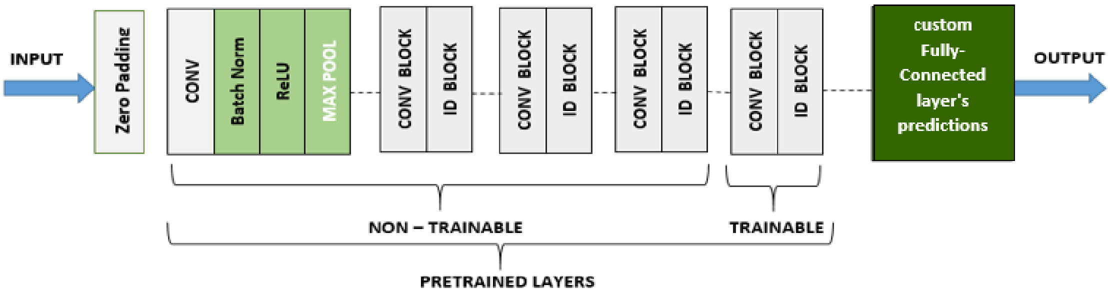


Figure 7. Adding New Layers on the Top of the. Baseline ResNet50 Model

2.5. ConvNeXtBase

ConvNets [22] another developed categorized image method. The ConvNeXtBase architecture is a convolutional neural network (CNN) design that improves traditional CNNs by using structured set convolutions. It is proposed as part of the "Aggregated Residual Transformations for Deep Neural Networks" (ResNeXt) [21], framework, which seeks to increase model performance and efficiency. The number of layers in a ConvNeXt network varies based on the version and architecture configuration, with a preference for residual block composition and pooling convolutions over a fixed overall number of layers. Consider using a moderately deep architecture, such as ResNeXt-50 or ResNeXt-101 [21], for FRtasks that use faces tagged in the Wild (LFW) dataset, which comprises a very limited number of face photographs per subject (typically approximately 10). These architectures find a mix between model complexity and computational efficiency, making them ideal for workloads that require medium-sized datasets, such as LFW. After loading the pre-trained ConvNeXt model, drop its last fully connected layer. This layer is typically used for classification tasks and is not required for feature extraction in the SNN weights As seen in Figure 8 [22].

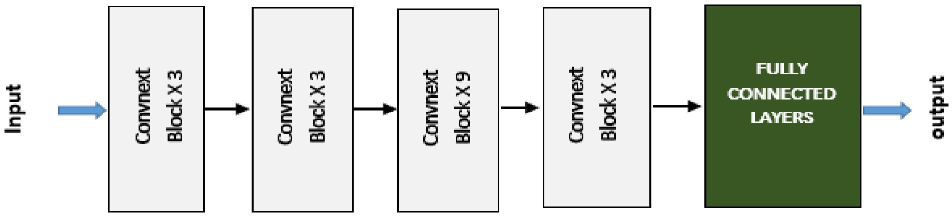


Figure 8. The General Architecture of ConvNeXtBase Model

2.6. Labelled Faces in the Wild (LFW) Dataset

The labelled faces in the wild (LFW) dataset [23] is a collection of face photographs specifically gathered for the purpose of studying unconstrained FR. The dataset comprises more than 13,000 facial images that have been gathered from various sources on the internet. The name of each person in the image was labelled on their respective face in this data set. The LFW dataset consists of four distinct sets of images. These sets include the original images as well as three types of aligned images. These aligned images are specifically designed to evaluate algorithms in various conditions. the dataset utilizes funnelled images [24], LFW-a, and deep funnelled images [25] for alignment purposes. Most face verification algorithms tend to yield better results when using deep funnelled and LFW-a images compared to funnelled images and original images. As seen in Figure 9 illustrates the LFW image dataset.

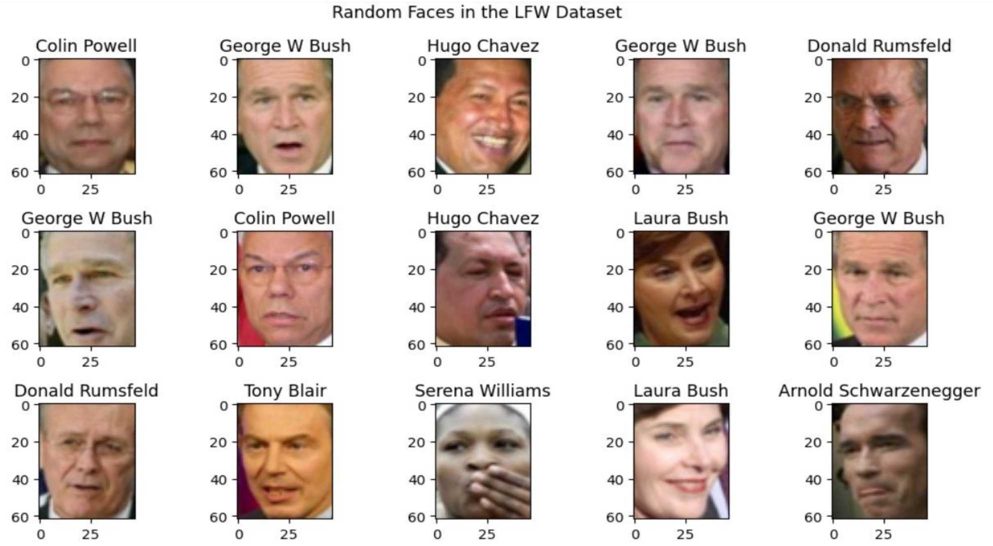


Figure 9. An Example of the LFW Image Dataset

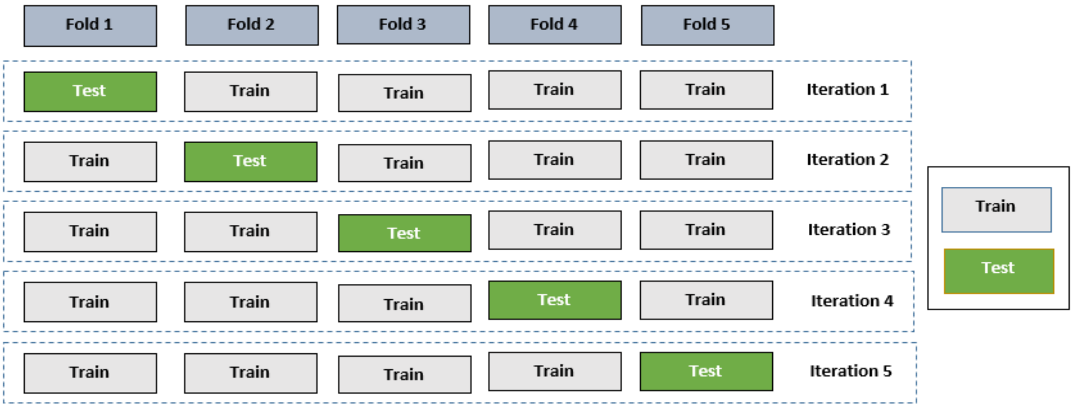
3. Training SNN wit Transfer Learning Models

A commonly used neural network for determining the similarity between two comparable inputs is called the SNN. This study explores the possibility of two individuals having similar facial features. To ensure the highest accuracy in facial recognition and identification, it is essential to incorporate the SNN with VGG-16 as the backbone. The SNN with VGG-16 as the backbone have been trained using the LFW dataset, which contains images of 1680 individuals with two or more distinct image available. The purpose is to identify the individuals. The LFW dataset is specifically created for studying the challenge of FR in real-life scenarios, where conditions are unpredictable and include various factors such as pose, lighting, and expression variations. These conditions are commonly encountered in everyday situations.

In order to use the LFW dataset as input for the SNN, it is necessary to preprocess it. The pre-processing steps include resizing the images. The images have been resized to dimensions of 62 pixels by 47 pixels (see equation 3.4). The images are resized to decrease the computational complexity and to ensure that the network focuses on the essential features of the faces. The specific size was chosen because the dimensions of the head are always wider than its width. Since this study using VGG-16 which is a color-based model, the step to convert images to grayscale has been omitted.

$$resizedimage = resize\big(image, (62,47)\big)(2.4)$$

Since the LFW dataset contains limited number of samples (13,000), the proposed study utilised the Stratified K-Fold cross-validator [26] through splitting the LFW dataset into 5 sets to increase the number of dataset images and to a void the overfitting, allowing us each fold has a chance to appear in the training set (k-1) times, making sure each observation in the dataset appears in the dataset and enabling the model to learn the fundamental information distribution better. The number of observations (n= 5). Figure 10 depicts the process of splitting the utilised dataset into 5 loops.



Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	
Test	Train	Train	Train	Train	Iteration 1
Train	Test	Train	Train	Train	Iteration 2
Train	Train	Test	Train	Train	Iteration 3
Train	Train	Train	Test	Train	Iteration 4
Train	Train	Train	Train	Test	Iteration 5

Figure 10. The Process of LFW Dataset Splitting

To evaluate the accuracy of the proposed FR approach, this study has utilised several metrics like accuracy, precision, f-measure, and recall as they are common metrics widely used [27]. False positive and false negative rates are another way to gauge the approximation error of the SNN model. These measures are well-known and commonly used in the existing research [28], [29]. moreover, computational complexity, and memory utilization were used to evaluate the efficiency of the proposed scheme based on computation theory as it is commonly practised by the related works [30].As seen in Equations 5 , 6 , 7 , 8 and 9 were used to calculate the detection accuracy, detection rate, precision, false positive rate, and the f measure, respectively.

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} (2.5)$$

$$DR = \frac{TP}{TP+FN} (2.6)$$

$$Precision = \frac{TP}{TP+FP} (2.7)$$

$$FPR = \frac{FP}{FP+TN} (2.8)$$

$$F1 = \frac{2\times Precision\times Recall}{Precision+Recall} (2.9)$$

where TP, TN, FP and FN denote the true positive, true negative, false positive and false negative respectively.

4. Experimental Results

The evaluation of the VGG16-based SNNs also yielded promising results, demonstrating its suitability for FR in the proposed model. As shown in the figure below, as the number of training epochs increased, the binary cross-entropy loss decreased, reaching lower values. This decrease in loss indicates that the network effectively learned the distinguishing features of each face, resulting in a minimal difference between the predicted and actual outputs. This effective learning makes the network as a robust tool for FR systems. Figure 11 depicts the Binary Cross Entropy Loss for the SNNs.

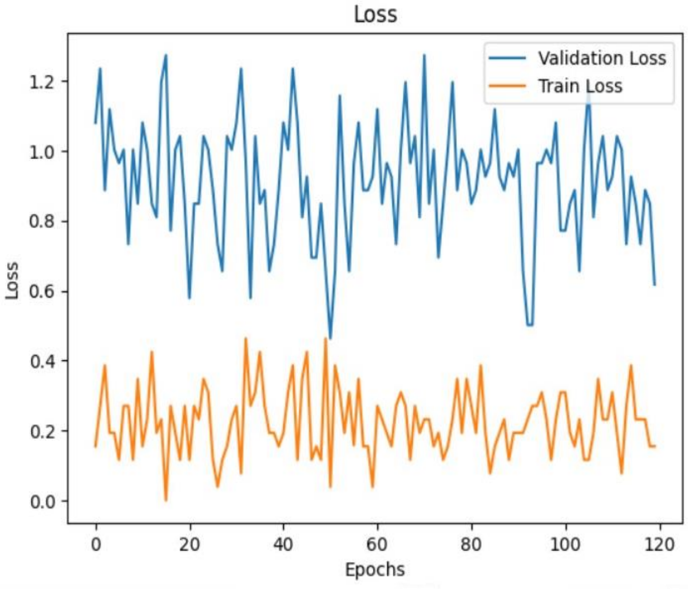


Figure 11. The Binary Cross Entropy Loss for Siamese Network

In addition, an accuracy curve was visualized showing a high accuracy score of 95% approximately on the validation set. This value signifies the ability of the model to correctly classify most of the input images. The high accuracy result renders the network suitable for the task of logging out the employees in the company. As seen in Figure 12 illustrates the SNN Accuracy with VGG16 as the backbone.

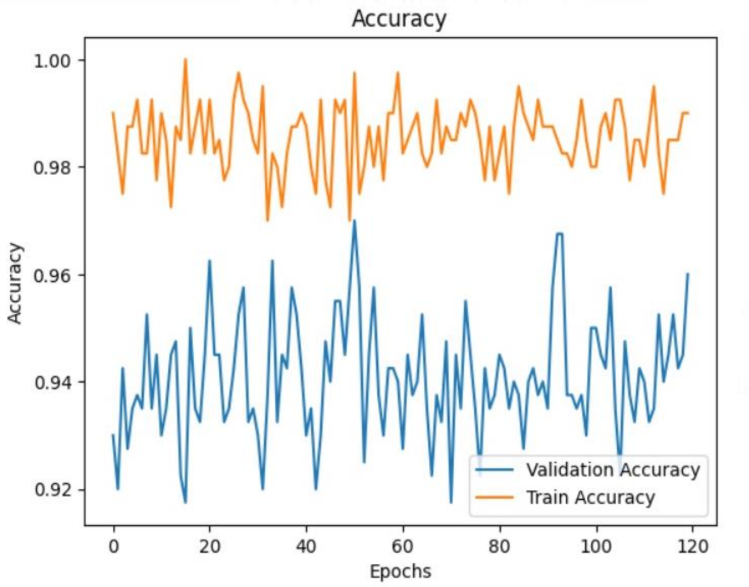


Figure 12. The Siamese Network Accuracy with VGG16 as the Backbone

Once the training of the SNN with VGG 16 was finished, it was tested on new test images to see how well the network performed. When presented with 3 images of the same individual but with different facial expressions, the network was still able to correctly identify the three images as the same person. This demonstrates the network's ability to accurately detect and group similar faces within the dataset. As seen in Figure 13 illustrates the test model on a Random Sample.

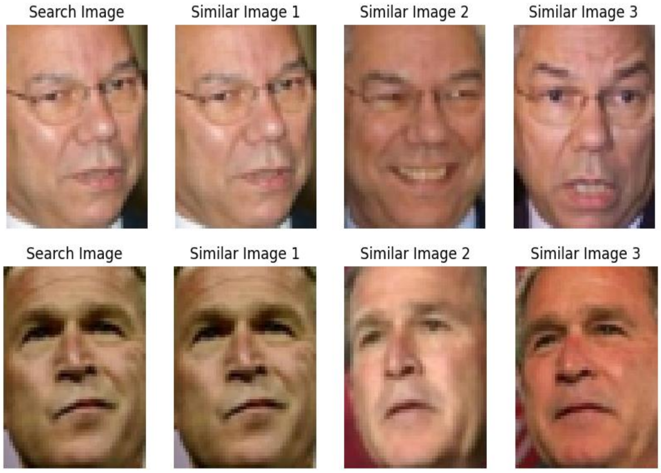


Figure 13. The test Model on Random Sample

As seen in Figure 14 below illustrates the results which had been achieved during training the proposed system using the suggested SNN with Efficient New algorithm. The achieved results explain that the proposed system has gain 78% as a validation accuracy. While as seen in Figure 15 depicts the results gained from training the proposed SNN with ConvNext, it had ganied 93% asa validation accuracy and As seen in Figure 16 show SNN with ResNet-50 has 95 %.

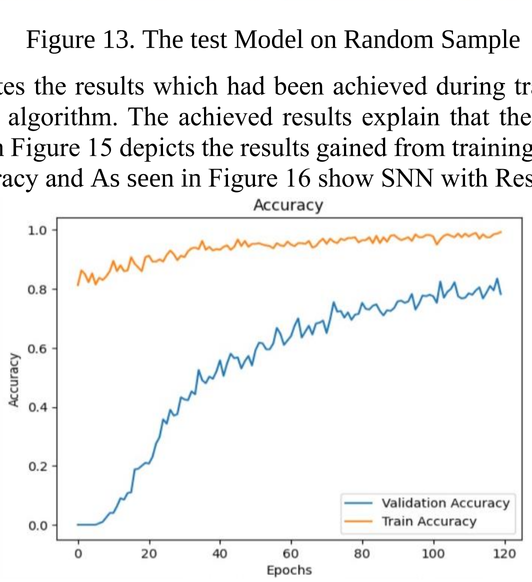


Figure 14. Siamese Network with Efficient

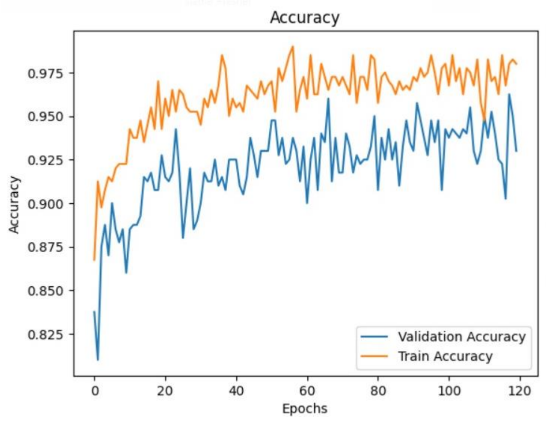


Figure 15. SNN with ConvNextTiny Algorithm

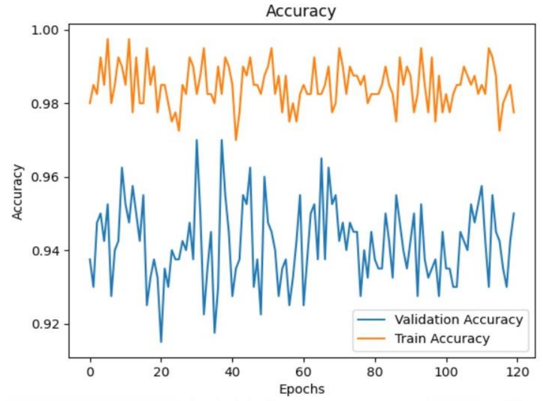


Figure 16. SNN with ResNet50 Algorithm

In this research, a SNN with VGG 16 as a backbone has been used for the FRmodel. These hybrid techniques have achieved an accuracy of 96% on the test set, which in comparison to other studies is a better level of accuracy, or at least equally as good to them. For instance, As seen in Table 1, the accuracy score of the proposed SNN with VGG 16 as a backbone is (96.00) which is much higher accuracy than some methods such as DLB (88.50), CFN+APEM (87.50), L-CSSE-KSRC (92.02), and SiameseFace1 (94.80). Other methods such as weighted PCA-EFMNet (95.00) and Siamese-VGG (95.62) show good results, that are still lower in accuracy compared to the proposed SNN with VGG 16 as a backbone. The only method that surpasses the proposed hybrid techniques in accuracy is CosFace, achieving 99.73 accuracy. However, despite achieving a lower accuracy compared to CosFace, the proposed SNN is faster, more lightweight, and suitable for running on simple hardware.

Table 4.1. SNN with VGG 16 as a Backbone Performance in Comparison to Other Methods

Reference	Method	Face Recognition ACC (%)
Chong et al [31]	DLB	88.50
Xiong et al [32]	CFN+APEM	87.50
A. Majumdar, R. Singh [33]	L-CSSE+KSRC	92.02
J. Zhang, X. Jin, [34]	SiameseFace1	94.80
B. Ameur, M. Belahcene [35]	Weighted PCA-EFMNet	95.00
M. Heidari and K.Fouladi-Ghaleh [16]	Siamese-VGG	95.62
H. Wang *et al.* [36]	CosFace	99.73
SNN with VGG 16 as a backbone		**96.00**
SNN with EfficientNetB0 as a backbone		**78.00**
SNN with ConvNextTiny as a backbone		**93.00**
SNN with ResNet50 as a backbone		**95.00**

To keep this study more robust and Efficient, two more comparative study have been utilized with VGG 16, these are Efficient loss and ConvNext networks. When evaluating the SNN's performance, it's essential to consider both accuracy and loss simultaneously. In this research, VGG16 serves as the backbone architecture. To keep the proposed method more robust, VGG16 have been benchmarked with ConvNext and EfficientNet. The performance of these networks are evaluated based on metrics such as accuracy and loss, providing insights into their effectiveness for FR tasks. By comparing the performance of different backbone architectures, this research aims to identify the most suitable model for enhancing the efficiency of FR approach. As seen in Figure 17 and 18 illustrate comparative studies between VGG16 , ResNet50 and Efficient loss and ConvNext networks. Based on the graph in Figure 17, it is not possible to definitively say which model is the best. However, the train VGG16 model appears to have the lowest loss. Based on the accuracy graph in the image, the Validation VGG16 Accuracy appears to be the highest throughout the epochs, which suggests it might be the best model for this task.

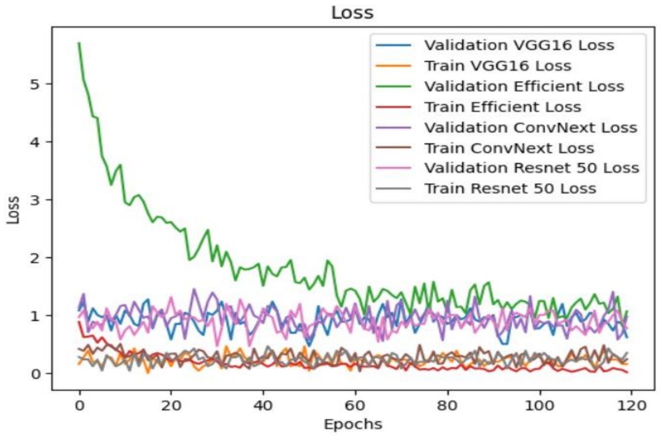


Figure 17. Comparative Study of validation Process Between VGG16 Lass, Efficient Loss and ConvNext loss and ResNet50

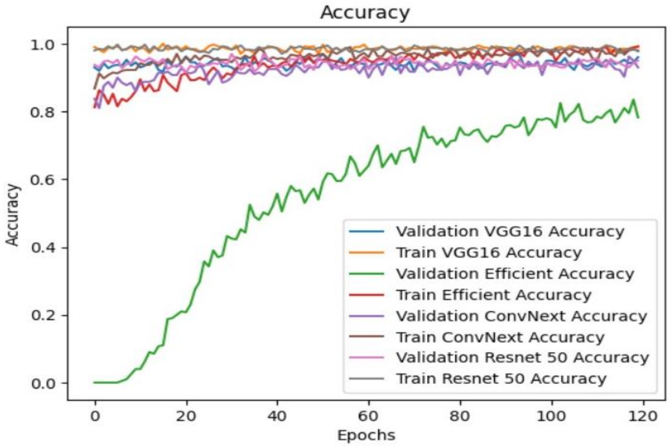


Figure 18. Comparative Study of Validation Process Between VGG16 Accuracy, Efficient Accuracy and ConvNext Accuracy

5. Conclusion

Many studies have been reported on FRin complex scenarios remains unresolved. A face detection and recognition-based system was implemented for recognition SNN with transfer learning models as the backbone. Various transfer learning architectures such as VGG-16, EfficientNet, RestNet50, and ConvNext have used for the experiments. 5-fold cross validation has been used to evaluate the performance of the architectures on LFW dataset. Comparative results show that EfficientNet, RestNet50 and ConvNext backbones achieved 78% accuracy, 95% and 93 % accuracy respectively. VGG-16 produced the best accuracy in FRwith 96%. This illustrates its capability to identify and distinguish individuals in different poses, lighting conditions, and facial expressions. For the future studies, better models can be developed to improve the accuracy of the system. Also the real-time improvements can be made to make system with fast recognition response.

References

[1] S. G. Bhandari, S. Rodrigues, P. C. Thejas, and B. S. Nausheeda, “ANALYSIS OF FRUSING LBPH ALGORITHM: A REVIEW,” *Redshine Arch.*, vol. 2, 2023.

[2] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, “Past, present, and future of face recognition: A review,” *Electronics*, vol. 9, no. 8, p. 1188, 2020.

[3] T. Gerig *et al.*, “Morphable face models-an open framework,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 75–82.

[4] Y.-S. Lim, S.-H. Lee, S.-J. Cheong, and Y.-H. Park, “A long-distance 3D FRarchitecture utilizing MEMS-based region-scanning LiDAR,” in *MOEMS and Miniaturized Systems XXII*, 2023, vol. 12434, pp. 87–91.

[5] S. Koley, H. Roy, S. Dhar, and D. Bhattacharjee, “Illumination invariant FRusing fused cross lattice pattern of phase congruency (FCLPPC),” *Inf. Sci. (Ny).*, vol. 584, pp. 633–648, 2022.

[6] aAmal A. Moustafa, A. Elnakib, and N. F. F. Areed, “Age-invariant FRbased on deep features analysis,” *Signal, Image Video Process.*, vol. 14, pp. 1027–1034, 2020.

[7] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*, 2016, pp. 499–515.

[8] Z. Chen, X. Feng, and S. Zhang, “Emotion detection and FRof drivers in autonomous vehicles in IoT platform,” *Image Vis. Comput.*, vol. 128, p. 104569, 2022.

[9] T. Sabharwal and R. Gupta, “Deep facial recognition after medical alterations,” *Multimed. Tools Appl.*, vol. 81, no. 18, pp. 25675–25706, 2022.

[10] Torrey, L., & Shavlik, J. (2010). Transfer learning. In Handbook of research on machine learning applications and trends: algorithms, methods, and techniques (pp. 242-264). IGI global.

[11] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[12] D. Chicco, “Siamese neural networks: An overview,” *Artif. neural networks*, pp. 73–94, 2021.

[13] N. Serrano and A. Bellogín, “Siamese neural networks in recommendation,” Neural Comput. Appl., pp. 1–13, 2023.

[14] Z. S. Naser, H. N. Khalid, A. S. Ahmed, M. S. Taha, and M. M. Hashim, “Artificial Neural Network-Based Fingerprint Classification and Recognition.,” *Rev. d’Intelligence Artif.*, vol. 37, no. 1, 2023.

[15] U. Ruby and V. Yendapalli, “Binary cross entropy with deep learning technique for image classification,” *Int. J. Adv. Trends Comput. Sci. Eng*, vol. 9, no. 10, 2020.

[16] M. Heidari and K. Fouladi-Ghaleh, “Using Siamese networks with transfer learning for FRon small-samples datasets,” in *2020 International Conference on Machine Vision and Image Processing (MVIP)*, 2020, pp. 1–4.

[17] S. Tammina, “Transfer learning using vgg-16 with deep convolutional neural network for classifying images,” *Int. J. Sci. Res. Publ.*, vol. 9, no. 10, pp. 143–150, 2019.

[18] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv Prepr. arXiv1409.1556*, 2014.

[19] McDermott, J. (2021). Hands-On Transfer Learning with Keras and the vgg16 Model.

[20] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR.

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[22] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11976-11986).

[23] A. Jalal and U. Tariq, “The LFW-gender dataset,” in *Computer Vision–ACCV 2016 Workshops: ACCV 2016 International Workshops, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13*, 2017, pp. 531–540.

[24] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database forstudying FRin unconstrained environments,” 2008.

[25] K. Sohn, H. Lee, and X. Yan, “Learning structured output representation using deep conditional generative models,” *Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.

[26] Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Journal of the royal statistical society: Series B (Methodological), 36(2), 111-133.

[27] X. Li, Y. Xiang, and S. Li, “Combining convolutional and vision transformer structures for sheep face recognition,” *Comput. Electron. Agric.*, vol. 205, p. 107651, 2023.

[28] P. Grother, M. Ngan, and K. Hanaoka, *FRvendor test (fvrt): Part 3, demographic effects*. National Institute of Standards and Technology Gaithersburg, MD, 2019.

[29] J. J. Howard, E. J. Laird, R. E. Rubin, Y. B. Sirotin, J. L. Tipton, and A. R. Vemury, “Evaluating proposed fairness models for FRalgorithms,” in *International Conference on Pattern Recognition*, 2022, pp. 431–447.

[30] M. Zulfiqar, F. Syed, M. J. Khan, and K. Khurshid, “Deep FRfor biometric authentication,” in *2019 international conference on electrical, communication, and computer engineering (ICECCE)*, 2019, pp. 1–6.

[31] S.-C. Chong, A. B. J. Teoh, and T.-S. Ong, “Unconstrained face verification with a dual-layer block-based metric learning,” *Multimed. Tools Appl.*, vol. 76, pp. 1703–1719, 2017.

[32] C. Xiong, L. Liu, X. Zhao, S. Yan, and T.-K. Kim, “Convolutional fusion network for face verification in the wild,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 3, pp. 517–528, 2015.

[33] A. Majumdar, R. Singh, and M. Vatsa, “Face verification via class sparsity based supervised encoding,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1273–1280, 2016.

[34] J. Zhang, X. Jin, Y. Liu, A. K. Sangaiah, and J. Wang, “Small Sample FRAlgorithm Based on Novel Siamese

Network.," *J. Inf. Process. Syst.*, vol. 14, no. 6, 2018.

[35] B. Ameur, M. Belahcene, S. Masmoudi, and A. Ben Hamida, "Weighted PCA-EFMNet: A deep learning network for Face Verification in the Wild," in *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2018, pp. 1–6.

[36] H. Wang *et al.*, "Cosface: Large margin cosine loss for deep face recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5265–5274.

Conflict of Interest Notice

Authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval

It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefited from are stated in the bibliography.

Availability of Data and Material

Not applicable.

Plagiarism Statement

This article has been scanned by iThenticate ™.