Engineered Science
DOI: https://dx.doi.org/10.30919/es1805

The Feature Fusion Multi-Conv Networks using Best Feature Selection for Face Recognition

Thipwimon Chompookham,[1] Keerati Tongnate,[2] Siriwan Phiphiphatphaisit,[3] Siriwiwat Lata,[4] Niwat Angkawisittpan[5,*] and Sivarit Sultornsanee[6]

Abstract

This research paper presents a novel approach to face recognition through the implementation of Feature Fusion Multi-conv Networks, utilizing the best feature selection techniques. The experiments were conducted using the FPFV, the MIT-CBCL, and the UFI dataset, where they were employed for pre-trained models of state-of-the-art CNNs, specifically ResNet-50 and SE-ResNet-50. The feature fusion process was implemented on the TensorFlow platform, utilizing Google Colab's GPU support to enhance computational efficiency. The results indicate that the proposed method significantly improves face recognition accuracy compared to traditional techniques. Various recognizers, including KNN, MLP, LR, and SVM, were evaluated, showcasing high accuracy rates across different configurations. The results suggest that future research could focus on exploring additional feature extraction techniques and optimization methods to further enhance the robustness and performance of face recognition systems. This study contributes to the ongoing efforts in the field of computer vision and face recognition, aiming to develop faster, more accurate, and secure systems that can be practically implemented in real-world applications.

Keywords: Face recognition; Feature fusion; Multi-conv networks; Convolutional neural network.
Received: 05 April 2025; Revised: 27 September 2025; Accepted: 12 October 2025
Article type: Research article.

1. Introduction

The human face has a multidimensional structure and can convey much information to express an individual's facial features. Therefore, practical facial feature analysis is a challenging task that requires time and effort. Face recognition has become essential in smart cities as it is widely used in areas such as scanning identity verification,[1] face self-service cash machine verification system phone unlocking, finance, education, government data management, network information security,[2] early warning facial recognition is used in suspicious situations or tracking down suspects. These show the importance of technological progress and the need for accurate and fast facial recognition technology.[3]

However, face recognition in unconstrained environments, such as under varying lighting conditions, makeup, facial occlusion, and problems in comparing profiles with frontal images. These are commons in many applications and result in a significant decrease in face recognition accuracy. These problems are the weaknesses of face recognition systems. Many studies have used computer vision methods to generate hand-crafted features. The resulting features depend on the characteristics of the extraction and recognition methods, including:

1. Holistic methods include PCA, ICS, and DWT.
2. Local methods such as DGM and EBGM.
3. Local-Texture methods such as LBP, LPQ, and BSIF.
4. Deep learning-based methods.

It is important to consider combining and applying different methods to solve the face image recognition problem: faster, more accurate, secure, and miniaturized, and reducing the time to make it more practical.[4-6]

Recently, convolutional neural networks (CNNs) have demonstrated remarkable success in various computer vision tasks, particularly feature extraction and pattern recognition.[7] Feature extraction deep learning algorithms have been developed independently, with high feature extraction and learning capabilities. Some methods focus on extracting features from different facial parts using multi-CNNs or

[1]*Department of Computer Science, Faculty of Informatics, Mahasarakham University, Maha Sarakham, 44150, Thailand*
[2]*Faculty of Information Technology, Rajabhat Maha Sarakham University, Maha Sarakham, 44000, Thailand*
[3]*Faculty of Business Administration and Information Technology, Rajamangala University of Technology Khon Kaen Campus, Khon Kaen, 40000, Thailand*
[4]*Department of Communication Arts, Faculty of Informatics, Mahasarakham University, Mahasarakham, 44150, Thailand*

extracting features from profile facial images.[8] The proposed method outperforms traditional methods or hand-crafted features.[9-11] While Chen *et al*. presented the CS-NWALBP+HOG algorithm to Improve the performance of the Local Binary Pattern (LBP), Neighborhood Weighted Average LBP (NWALBP) takes into account the strong correlation between similar pixel pairs and introduces one novel operator called CS-NWALBP,[12] which not only smooths noise sensitivity but also reduces the original computational complexity. Qi *et al*. presented a face recognition method using a convolutional neural network, CNN, based on BiLSTM and attention,[9] focusing on the speed of traditional CNN and the problems of occlusion and transformation. The expression changes of facial recognition in practical applications show that the proposed method can improve accuracy and speed calculated significantly.

Moreover, Wu *et al*. considered the signal components with potential recognition in recognition based on 1D and 2D CNN.[5] Wen *et al*. presented a signal-to-image conversion method that converted 2-Dimensional (2D) images and eliminated the effect of handcrafted features[13] Experimental results showed that the proposed method performs better than other CNN methods. While additional research uses a 1-Dimensional (1D) model using the original signal as input data of a CNN.[14,15] In the research of Zan *et al*.,[16] the multi-dimension input convolutional neural network (MDI-CNN) is used to learn multi-dimensional joint features using a fusion model in CNN. The results show that the proposed model improves the recognition rate and has higher validity and robustness.

Recent developments in face recognition have witnessed significant advancements in the integration of Vision Transformers (ViTs) and attention mechanisms. Rodrigo *et al*. demonstrated that ViTs outperform CNNs in face recognition tasks, particularly in handling occlusions and distance variations. Similarly, Khalifa *et al*.[17] proposed attention-integrated multi-level CNNs that achieve superior performance while maintaining computational efficiency. The emergence of lightweight models for real-time applications has also gained momentum, with Lighter Face[18] showing remarkable improvements in detection speed while reducing computational demands by 85.4%. These recent advances highlight the evolution toward more robust, efficient, and practically applicable face recognition systems.

This paper presents a novel Feature Fusion Multi-Conv Networks (FFM-Nets) where we can extract smaller but still robust features with different CNN architectures. All features are divided into smaller sets. We then select all features using the best feature selection method, and they are sent for recognition using traditional machine learning techniques.

The proposed FFM-Nets framework consists of three distinct feature fusion components: Feature Fusion ResNet-50 Multi-conv Networks (FFRM-Nets), Feature Fusion SE-ResNet-50 Multi-conv Networks (FFSM-Nets), and Feature Fusion ResNet-50 and SE-ResNet-50 Multi-conv Networks (FFRSM-Nets). Experimental results demonstrate that FFRSM-Nets achieves significant performance improvements when combined with traditional machine learning recognizers, particularly Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Multi-Layer Perceptron (MLP). The highest accuracy achieved by FFRSM-Nets is 99.45% on the MIT-CBCL dataset when combined with the LR recognizer, while achieving an outstanding performance of 99.94% on the FPFV dataset with the LR recognizer. Furthermore, FFRSM-Nets demonstrates substantial improvements over existing methods, outperforming previous approaches by approximately 24% on the UFI (Cropped) dataset and 36% on the UFI (Large) dataset. The significant contributions of this research are summarized in the following:

1. We propose the deep learning architecture: ResNet and SENet with the addition of Conv1D and Conv2D Blocks, called Multi-conv Networks of CNN models. We include layers including ReLU, batch normalization, and dropout layer for extracting smaller features, which reduces the model size and training time.
2. In the experiment, we divide feature sets into sub-sets and use the best feature selection method to get the best features in each set fusion, called feature fusion Multi-conv Networks (FFM-Nets), for forwarding to recognition using four machine learning methods.
3. Comprehensive experiments using different datasets and comparing the proposed methods with state-of-the-art methods have been performed, indicating the efficacy of the proposed Feature Fusion Multi-Conv Networks (FFM-Nets) framework.
4. The proposed framework can be used to improve the performance of traditional face image recognition methods by integrating the robust feature fusion steps for face image recognition.

This work advances face recognition through several key innovations, including: 1) the integration of 1D and 2D convolutional blocks within CNN architectures creates multi-dimensional feature representations; 2) the fusion of ResNet-50 and SE-ResNet-50 features results in complementary learning; 3) a systematic best subset selection method that enhances recognition performance. These contributions address feature extraction and recognition accuracy challenges, demonstrated through comprehensive experiments across diverse datasets.

This paper has been organized as follows. Section 2 describes the proposed FFM-Nets framework for face image recognition. Three face image datasets and benchmarks are described in Section 3. Section 4 and 5 presents evaluation

[5]*Research Unit for Electrical and Computer Engineering Technology, Faculty of Engineering, Mahasarakham University, Maha Sarakham, 44150, Thailand*
[6]*College of Engineering, Northeastern University, Boston, Massachusetts, 02115, USA*
[*]*Email:* niwat.a@msu.ac.th (N. Angkawisittpan)

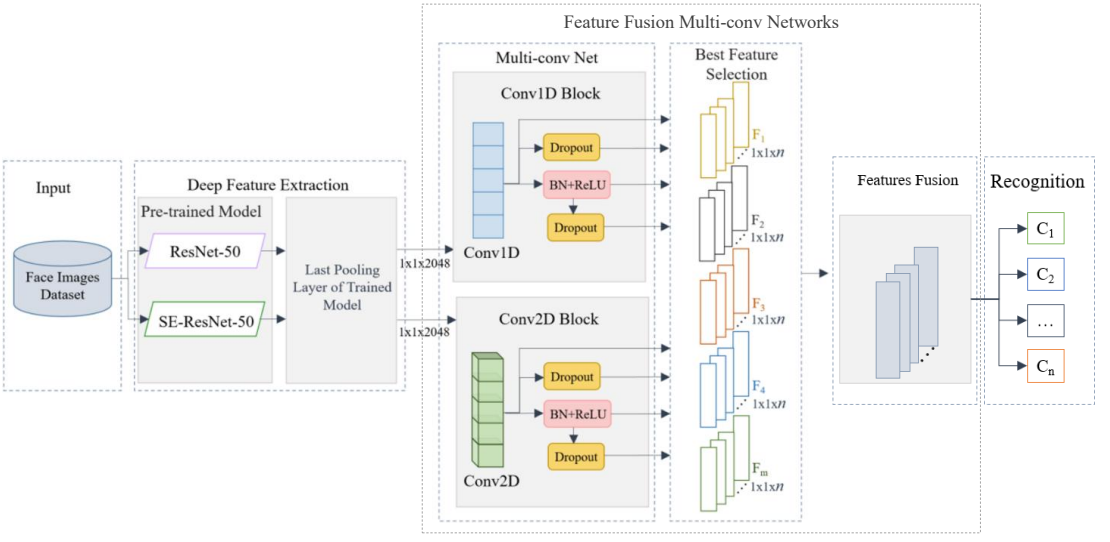


Fig. 1: Architecture of our proposed framework for face image recognition.

metrics and the experimental setup and results of face are recognition. The conclusions and directions for future work given in Section 6.

2. The proposed feature fusion multi-conv networks framework for face image recognition

Face recognition in an unconstrained environment: Due to environmental problems and limited data, it is difficult to improve the recognition performance using facial feature extraction with a single CNN.[8] Feature extraction of faces using multi-CNNs outperforms traditional methods or hand-crafted features. The 1D Convolutional (Conv1D) and 2D Convolutional (Conv2D) layers have emerged as an indispensable constituent, playing a pivotal role in extending the success of CNNs.[19,20]

In this section, we present a feature fusion multi-conv networks framework to solve the face recognition problem, as shown in (Fig. 1). The first part of our framework combines state-of-the-art CNN architectures and two pre-trained CNN models: ResNet-50 and SE-ResNet-50 for use in facial feature extraction. As a result, the last pooling layer of the CNN model is employed as the feature, called deep feature extraction. Details of CNNs are described in section 2.1. In the second part, the feature obtained from deep feature extraction undergoes multi-conv networks, which were Conv1D and Conv2D block, and then the best features obtained from the best selection are combined. This is described in section 2.2. Finally, the best features are sent for recognition using traditional machine learning techniques.

2.1 Deep feature extraction

In this section, we propose state-of-art CNN architectures: two pre-trained CNN models, ResNet-50 and SE-ResNet-50, for facial feature extraction. As a result, the last pooling layer of the CNN model is employed as the feature, called deep feature extraction. An overview of each CNN will now be described.

2.1.1 ResNet-50

ResNet-50, introduced by He *et al.* in 2016,[21] is an abbreviated "Residual Network with 50 layers," representing a convolutional neural network (CNN) architecture that has garnered considerable attention and prominence within computer vision. The ResNet-50 model is designed to tackle the intricate challenge of training neural networks with substantial depth. A seminal innovation integral to ResNet-50 incorporates residual learning, a paradigm marked by introducing shortcut or skip connections that circumvent one or more layers within the network. This architectural enhancement serves the crucial purpose of facilitating the effective training of exceptionally deep networks, mitigating the vanishing gradient predicament, and fostering the acquisition of more discriminative features.

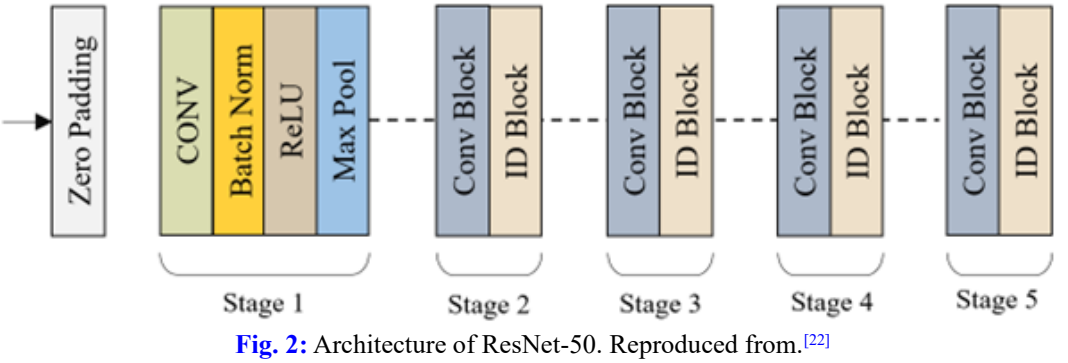


Fig. 2: Architecture of ResNet-50. Reproduced from.[22]

The empirical evidence attests to the superior performance of ResNet-50 across diverse image recognition tasks, encompassing applications such as image recognition and object detection. Consequently, ResNet-50 has emerged as a pivotal cornerstone in the evolution of deep learning models tailored for visual recognition tasks. It has an architectural structure, as shown in (Fig. 2).

The ResNet-50 framework has been embraced and modified by researchers and practitioners across diverse domains, extending its utility beyond image recognition to encompass applications such as medical image analysis, natural language processing, and audio signal processing. The pervasive adoption of ResNet-50 underscores its adaptability and the resilience inherent in the underlying residual learning paradigm. Consequently, ResNet-50 retains a pivotal role in the progression of deep learning models across a spectrum of domains, profoundly influencing the contemporary landscape of artificial intelligence research and applications.

2.1.2 SE-ResNet-50

SE-ResNet-50, as introduced by Hu *et al*. in 2018,[23] presents a novel attention mechanism aimed at augmenting the discriminative capacity of Convolutional Neural Networks (CNNs). The central innovation within SE-ResNet-50 resides in integrating a "squeeze-and-excitation" block (shown in (Fig. 3)) within each convolutional block, affording the network the capability to dynamically recalibrate the significance of distinct feature channels.

This attention mechanism facilitates SE-ResNet-50 in selectively prioritizing informative features while concurrently attenuating the impact of less pertinent ones, thereby resulting in an ameliorated feature representation and an overall enhancement in model performance. Empirical evidence attests to the noteworthy success of SE-ResNet-50 across diverse computer vision tasks, including but not limited to image recognition, object detection, and segmentation.The influence of SE-ResNet-50 surpasses conventional applications in image recognition, as evidenced by its adoption and adaptation by researchers to confront challenges across varied domains, including but not limited to medical imaging and autonomous driving. The attention mechanism introduced by SE-ResNet-50 has catalyzed subsequent endeavors in the construction of attention-based models, showcasing its efficacy in apprehending intricate dependencies inherent in input data. Consequently, SE-ResNet-50 has evolved into an indispensable reference in the progression of attention mechanisms within deep neural networks, thereby exerting a discernible impact on the formulation of subsequent architectures and contributing substantively to the ongoing advancements in the realm of deep learning.

2.1.3 Convolution1D

Convolutional Neural Networks (CNNs) have prominently featured at the forefront of deep learning applications, particularly within the domain of computer vision. The 1D Convolutional (Conv1D) layers have emerged as an indispensable constituent, playing a pivotal role in extending the success of CNNs to sequential data. Introduced as a specialized form of convolutional layer tailored for one-dimensional sequences, Conv1D has garnered widespread applicability in tasks spanning time-series analysis, speech recognition, and natural language processing. A seminal contribution to this field is exemplified by F. Chollet *et al*., in 2016,[19] which underscores the adaptability of Conv1D layers in capturing temporal dependencies inherent in sequential data, thereby showcasing their efficacy in diverse applications extending beyond the conventional realm of image-based CNNs.

The capacity of Conv1D layers to apprehend hierarchical dependencies within sequential data has resulted in substantive advancements, enhancing the comprehension of intricate patterns within temporal sequences. As Conv1D continues its trajectory of evolution, it maintains a pivotal status in the developmental trajectory of deep learning models, offering a potent instrument for extracting meaningful representations from one-dimensional data, as illustrated in (Fig. 4b).

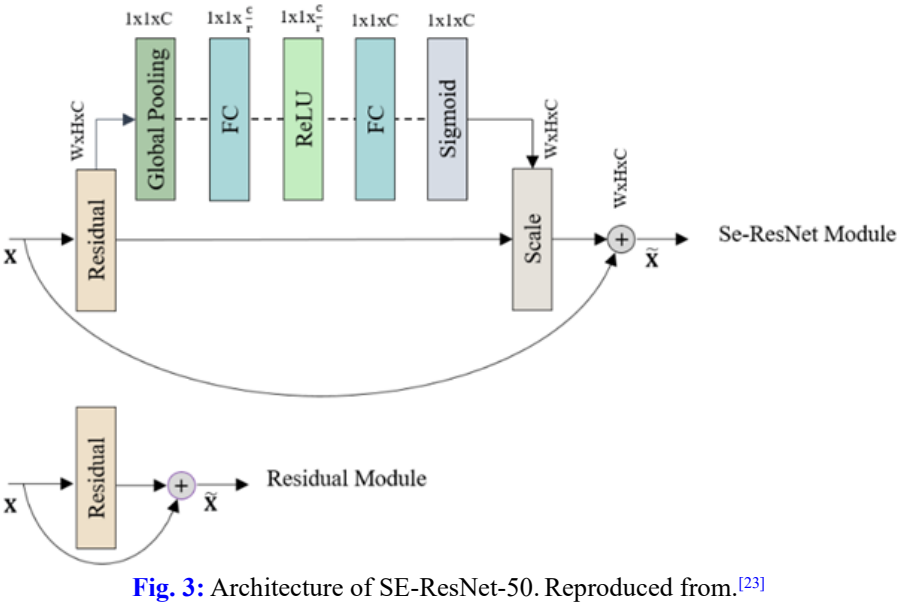


Fig. 3: Architecture of SE-ResNet-50. Reproduced from.[23]

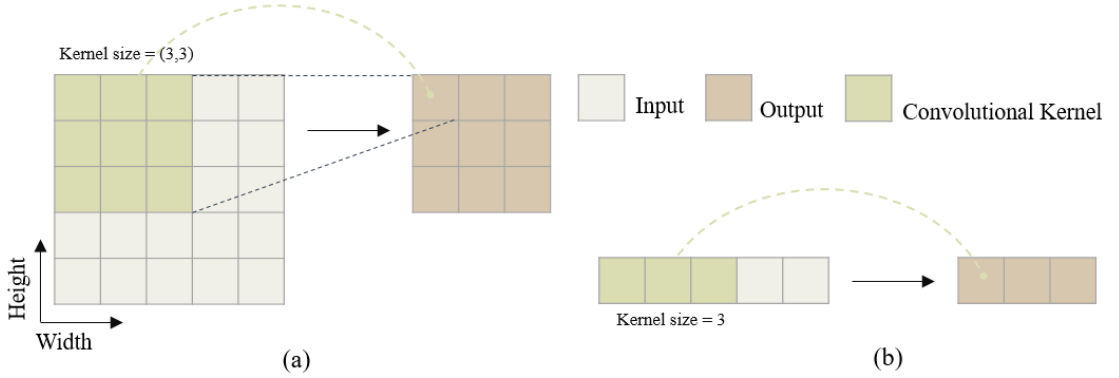


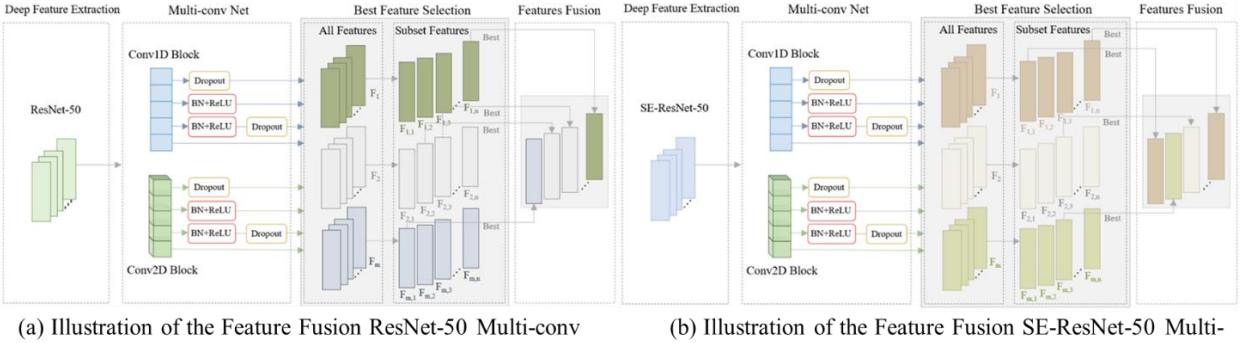
Fig. 4: The difference between the (a) 2D Convolution operation and (b) 1D Convolution operation. Reproduced from.[24]

2.1.4 Convolution2D

The 2D Convolutional (Conv2D) layers serve as foundational components within Convolutional Neural Networks (CNNs), assuming a pivotal role in the triumph of deep learning methodologies applied to computer vision tasks. As shown in (Fig. 4a), the incorporation of Conv2D layers facilitates hierarchical feature extraction, enabling neural networks to effectively capture local patterns and spatial relationships inherent in images. K. Simonyan and A. Zisserman introduced the VGGNet architecture,[20] which demonstrated the efficacy of employing multiple Conv2D layers with diminutive filter sizes to achieve exceptional performance in image recognition tasks. The adoption of Conv2D layers within VGGNet subsequently established a paradigm for subsequent deep learning models, solidifying Conv2D as a cornerstone in the realm of image feature extraction. The versatility of Conv2D in accommodating diverse scales and resolutions of input data renders it a versatile instrument for representing visual features. Researchers have delved into diverse variations and optimizations of Conv2D layers, thereby actively contributing to the continual advancement of CNN architectures. As Conv2D endures as an indispensable element in the toolkit of deep learning, its impact transcends the domain of image recognition, profoundly influencing the design of neural networks tailored for a diverse array of visual recognition tasks.

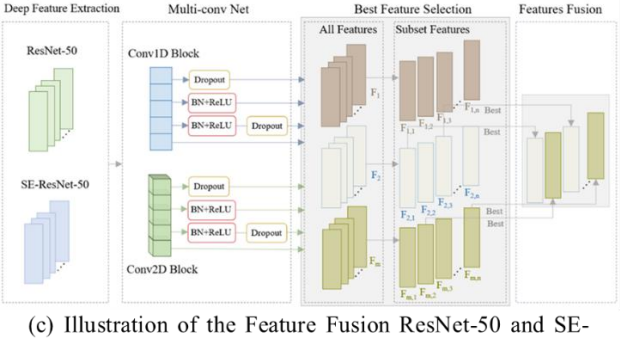
2.2 The feature fusion multi-conv networks

In this study, we propose the Feature Fusion Multi-Conv Networks (FFM-Nets), we divide the feature fusion framework into three sets: feature combined via Conv1D and Conv2D Blocks, where these blocks are via the Batch Normalization (BN) layer, ReLU activation function, dropout, and without dropout layer. The first framework is called Feature Fusion ResNet-50 multi-conv networks (FFRM-Nets), the second Feature Fusion SE-ResNet-50 multi-conv networks (FFSM-Nets), and the third Feature Fusion ResNet-50 and SE-ResNet-50 multi-conv networks (FFRSM-Nets) shown in (Fig. 5a-c), respectively.



(a) Illustration of the Feature Fusion ResNet-50 Multi-conv networks (FFRM-Nets).

(b) Illustration of the Feature Fusion SE-ResNet-50 Multi-conv networks (FFSM-Nets).



(c) Illustration of the Feature Fusion ResNet-50 and SE-ResNet-50 Multi-conv networks (FFRSM-Nets).

Fig. 5: Illustration of FFM-Nets. (a) The FFRM-Nets, (b) the FFSM-Nets, and (c) FFRSM-Nets.

As shown in (Fig. 5), in the best feature selection section, we divide all found features from the Multi-conv Net block extraction into subsets and then select the features using the best subset selection method. That aims to discover the subset of independent variables that results in the best prediction (Y), which is done by considering possible combinations of independent variables. It is widely used in all research areas, including computer science and medicine.[25] If we have an independent variable to consider to find the best subset selection: Step 1 Start by considering all possible feature datasets with variable 1, variable 2, variable 3 ... m variables. Step 2, then choose the best feature of size m. Step 3, finally, the feature set with the best overall results, with the best subset section, selects 1 model from 2n possible models and determines the best dataset in each set. The best data set might be the data set that predicts the outcome with the least error or determines the best model of the data set by combining multiple data sets.

The best subset selection method was chosen for several reasons: 1) it systematically evaluates feature combinations to find optimal subsets, 2) it reduces overfitting by eliminating redundant features, and 3) it improves computational efficiency during inference. To manage computational cost, features from Conv1D and Conv2D blocks are divided into smaller subsets. The selection process uses cross-validation accuracy as the criterion. It justifies the computational overhead through the performance improvements demonstrated in our experimental results, including reduced model complexity and faster training times while maintaining accuracy.

3. The face image datasets

In this study, we use the data set from three benchmark face image datasets, including the Frontal to Profile Face Verification (FPFV) in the Wild, the MIT-CBCL face, and the Unconstrained Facial Images (UFI). The details of each face image dataset were as follows.

3.1 The frontal to profile face verification (FPFV) dataset

The FPFV dataset is the "Celebrities in Frontal-Profile data set",[10] comprising 10 frontal and 4 profile images of 500 people. In this experiment, we selected only the frontal images to create a frontal-to-frontal face verification dataset. The dataset contains celebrity face images that present challenges for face recognition systems under varying conditions.

3.2 The MIT-CBCL face dataset

The MIT-CBCL face contains images of 10 people.[26] The 3D models are rendered under varying pose and illumination conditions to build a large set of synthetic images. The conditions for creating synthetic training data include face rotation from 0 to 34 degrees and adjusting ambient lighting and directed light source. The test set consisted of 200 images of each person under various poses and illumination conditions.

3.3 The unconstrained facial images (UFI)

The Unconstrained Facial Images (UFI) is a collection of real photographs selected from a large group of pictures provided by the Czech News Agency (CTK).[27] This data is divided into two datasets: 1) the UFI (Cropped) is a picture of a person that has been cropped to a size of 128x128 pixels, This set contains images of 605 people with an average of 7.1 images per person in the training set and one in the test set, and 2) the UFI (Large) have the total number of the subjects in this partition is 530 and an average number/person of training images is 8.2. The size of images in this partition is 384 x 384 pixels.

3.4 Dataset ethics and licensing

All datasets were used in compliance with their respective licensing agreements for research purposes only. The FPFV dataset follows the protocol by Sengupta *et al.*,[10] the MIT-CBCL dataset is used under an MIT academic license with proper attribution,[26] and the UFI dataset is freely available for research under a University of West Bohemia license.[27] No commercial use or redistribution was performed.

4. Performance evaluation

The degree of fitting and model performance evaluation are crucial aspects in assessing the effectiveness of machine learning algorithms.[28] The evaluation metrics used for face image recognition to measure the performance of the proposed method,[29,30] as follows: Eq. (1-5).

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$\text{Precision}=\frac{TP}{TP+FP} \tag{2}$$

$$\text{Sensitivity}=\frac{TP}{TP+FN} \tag{3}$$

$$\text{Specificity}=\frac{TN}{TN+FP} \tag{4}$$

$$\text{F1-Score}=\frac{2\ X\ TP}{2\ X\ TP+FP+FN} \tag{5}$$

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

We present the graph of the receiver operating characteristic (ROC) curve that is used to comprehensively measure each performance index.[31,32] In the ROC, the vertical axis is the true positive rate (TPR), which is equivalent to sensitivity, and the horizontal axis is the false positive rate (FPR), defined in the following: Eq. (6).

$$\text{FPR} = 1\text{-Specificity} \tag{6}$$

The ROC curves help in comparing performance by comparing the area under the curve of the proposed models, which can be defined between different classes where a greater area under the curve indicates higher efficiency.

Table 1: The hyper-parameter and the best values of each recognition models.

	Hyper-Parameter Values		Best Value
SVM	C	[0.1,1,10,100]	10
	Gamma	[0.0001,0.01,0.1,1]	0.0001
	Kernel	[rbf, poly, linear, precomputed, sigmoid]	rbf
KNN	N neighbors	[1, 2, 3,...,30]	4
	Weights	[uniform, distance]	distance
MLP	Hidden layer sizes	[10, 20, 30, 50, 100, 200]	200
	Activation	[identity, logistic, tanh, relu]	tanh
	Solver	[sgd, adam]	sgd
	Alpha	[0.0001, 0.001, 0.01, 0.1]	0.1
	Learning rate	[constant, adaptive]	constant
LR	Penalty	[l1, l2]	l2
	C	np.logspace(-3,3,7)	100
	Solver	[lbfgs, newton-cg, liblinear]	lbfgs

5. Experimental Setup and Results

5.1 Experimental Setup

In this section, we implemented the feature fusion of multi-conv networks with the TensorFlow platform running on Google Colab GPU support for all the experiments. As explained in section 2. We used a pre-train model of two CNN architectures; ResNet-50 and SE-ResNet-50, to train and extract features from face images. Second, the features sent to the multi-conv net are Conv1D and Conv2D blocks. Those features have output shapes equal to 1,024, 512, and 256, and kernel sizes are 3, 5, and 7, respectively. The batch normalization layer (BN) and rectified linear unit (ReLU) activation function and dropout layer were combined into the last layer of the multi-conv blocks. Third, we divide the feature into subsets to select the best features, called the best feature selection. Finally, the robust features introduced are recognized using four machine learning methods: Support Vector Machine (SVM), K-Nearest Neighbors Algorithm (KNN), Multi-Layer Perceptron (MLP), and Logistic Regression (LR). In the experiment, the best parameters using the Grid search method as shown in Table 1.

5.2 Recognition performance on the FPFV dataset

5.2.1 Performances evaluation of deep feature extraction and the multi-conv networks

In the experiment, we trained the FPFV dataset using a pre-trained model of the two state-of-the-art CNNs: ResNet-50 and SE-ResNet-50. We proposed the deep feature extraction method to extract the feature from the last pooling layer of the trained model. Second, in the experiment, we prepared the FPFV dataset using a pre-trained model of the two state-of-the-art CNNs: ResNet-50 and SE-ResNet-50. The features sent to the Multi-conv net have multiple output shapes and kernel sizes. The deep feature extraction method extracted a dimension of the features. The number of parameters and size of features were reported in Table 2. It shows that ResNet-50 has an original output shape of 2,048 with 23,561,152 parameters and a feature size of 78.1 MB. When we add the multi-conv net, it has approximately 6-62% more parameters when using conv1D blocks and 20-436% when using conv2D blocks, respectively, while the feature size is reduced by about 49-87%. When looking at the SE-ResNet-50 model, the original output shape 2,048 has 26,092,144 parameters, and the feature size is 78.1 MB. When we add the multi-conv net, the parameters increase by approximately 6-56% with conv1D blocks and 18-393% with conv2D blocks. Feature size has been reduced by approximately 49-87%.

Table 3 presents the accuracy results on 10-fold cross-validation to evaluate the result of the FPFV dataset. We used the recognition rate (accuracy) and standard deviation to measure the performance of each feature extraction method. The experimental results found that the recognition accuracy with SE-ResNet-50 + SVM increased by approximately 1.49-1.62% when adding Conv1D block, and the accuracy increased by approximately 1.42-1.64% when adding Conv2D block, while ResNet-50 + SVM showed smaller improvements of 0.09-0.10% with Conv1D blocks. The experimental results found that recognition The LR method has the highest accuracy compared to other methods, as shown in (Fig. 6).

Moreover, when considering other recognizers, we found that although the accuracy of using Conv1D and Conv2D blocks did not increase, the accuracy was as high as the original. Meanwhile, the model size decreased by approximately 50-87%, and the training time of the model decreased by about 16-75%, as shown in (Fig. 7). From the results of the experiments, we decided to use Conv1D and Conv2D block size 512 together with the BN layer and ReLU

Table 2: Comparison of parameters and sizes of features.

Model	Output Shape	Multi-conv net	Kernel Size	No. of Parameters	Size of Features
ResNet-50	2,048	-	-	23,561,152	78.1MB
	1,024	Conv1D	3	29,853,632	39.1MB
			5	34,047,936	
			7	38,242,240	
		Conv2D	3	42,436,544	
			5	75,990,976	
			7	126,322,624	
	512	Conv1D	3	26,707,392	19.5MB
			5	28,804,544	
			7	30,901,696	
		Conv2D	3	32,998,848	
			5	49,776,064	
			7	74,941,888	
	256	Conv1D	3	25,134,272	9.8MB
			5	26,182,848	
			7	27,231,424	
		Conv2D	3	28,280,000	
			5	36,668,608	
			7	49,251,520	
SE-ResNet-50	2,048	-	-	26,092,144	78.1MB
	1,024	Conv1D	3	32,384,624	39.1MB
			5	36,578,928	
			7	40,773,232	
		Conv2D	3	44,967,536	
			5	78,521,968	
			7	128,853,616	
	512	Conv1D	3	29,238,384	19.5MB
			5	31,335,536	
			7	33,432,688	
		Conv2D	3	35,529,840	
			5	52,307,056	
			7	77,472,880	
	256	Conv1D	3	27,665,264	9.8MB
			5	28,713,840	
			7	29,762,416	
		Conv2D	3	30,810,992	
			5	39,199,600	
			7	51,782,512	

Table 3: Comparison of the accuracy of the four recognizers with each feature extraction method.

		Accuracy (%)						
		Original		Conv1D Blocks		Conv2D Blocks		
		Output Shape						
	Model	2048	1024	512	256	1024	512	256
ResNet-50	KNN	99.32±0.001	99.21±0.002	99.21±0.002	98.97±0.002	99.15±0.002	99.15±0.002	98.90±0.002
	MLP	98.53±0.004	98.45±0.003	98.22±0.004	97.67±0.004	98.75±0.003	98.48±0.003	97.96±0.004
	LR	99.64±0.002	99.61±0.002	99.58±0.002	99.38±0.003	99.58±0.002	99.56±0.002	99.48±0.002
	SVM	99.42±0.002	99.52±0.002	99.51±0.002	99.28±0.002	99.49±0.002	99.46±0.002	98.88±0.002
SE-ResNet-50	KNN	99.46±0.002	99.29±0.002	99.31±0.003	99.20±0.003	99.39±0.003	99.33±0.003	99.13±0.003
	MLP	98.63±0.006	98.51±0.004	98.35±0.004	97.99±0.005	98.86±0.003	98.68±0.003	98.44±0.005
	LR	99.72±0.002	99.71±0.002	99.64±0.002	99.60±0.003	99.71±0.002	99.64±0.002	99.57±0.003
	SVM	98.04±0.004	99.66±0.001	99.58±0.002	99.53±0.002	99.68±0.002	99.61±0.001	99.46±0.002

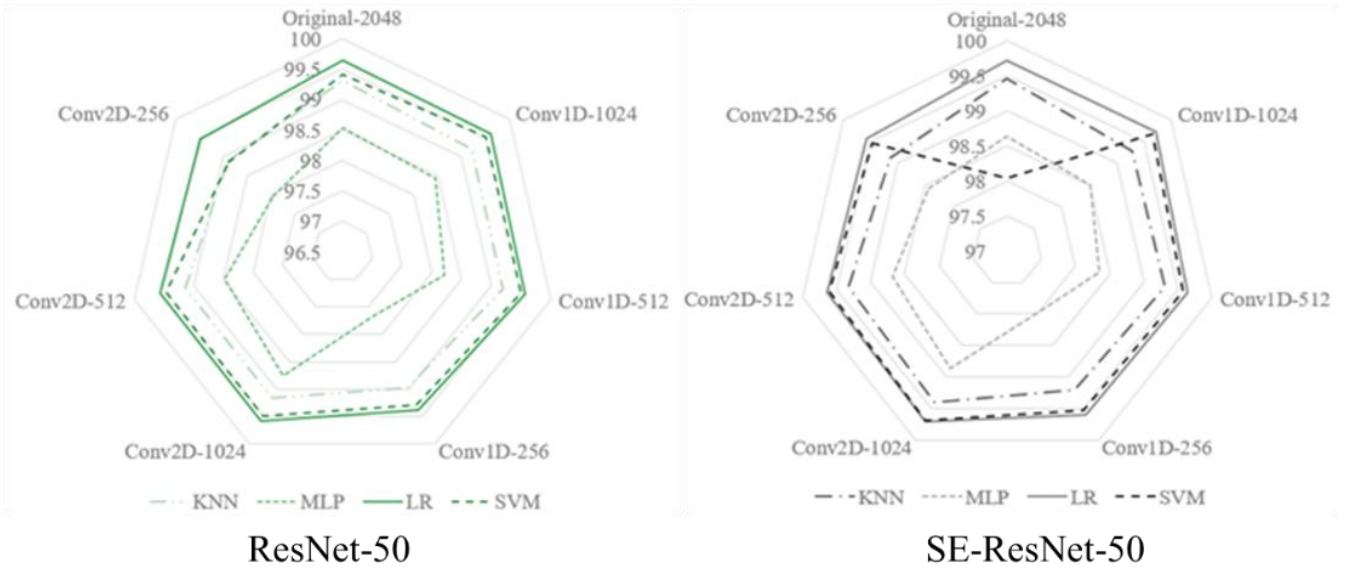


Fig. 6: Recognition results of four recognizers when adding conv1D and conv2D blocks and original features.

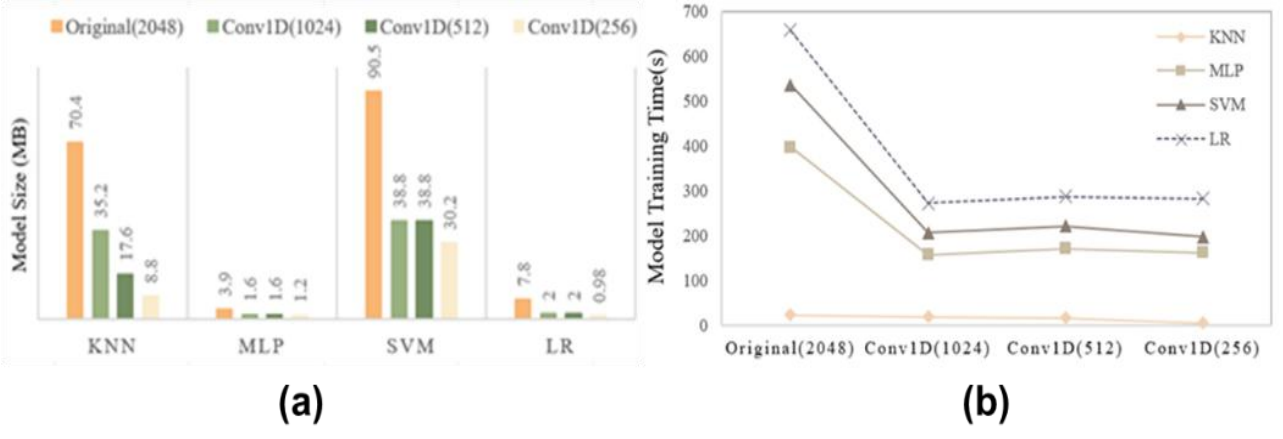


Fig. 7: (a) the size of the model (b) the training time.

activation function and dropout layer to experiment with dropout rates of 0.1-0.5. The results showed that a dropout of 0.5 had the highest accuracy.

5.2.2 Performance of feature fusion from the Best Feature Selection

To study Feature Fusion multi-conv Networks (FFM-Nets), we divide the feature fusion framework into three sets: feature combined via Conv1D and Conv2D Blocks, where these blocks are via the BN layer, ReLU activation function, dropout, and without dropout layer. The first framework is called Feature Fusion ResNet-50 multi-conv networks (FFRM-Nets), the second Feature Fusion SE-ResNet-50 multi-conv networks (FFSM-Nets), and the third Feature Fusion ResNet-50 and SE-ResNet-50 multi-conv networks (FFRSM-Nets). From the experimental results, we get different features from deep features, namely ResNet-50 and SE-ResNet-50, through the feature fusion framework, then select the best feature subset from all subsets for recognition.

In Table 4, the experimental results show that the FFRSM-Nets combined with LR algorithm significantly outperforms the other techniques and provides a high accuracy of 99.94%. Furthermore, the ROC curve is shown in (Fig. 9). Additionally, accuracy increases when combined with SVM and MLP, with an average accuracy increase of 0.27-0.43%compared to the original ResNet-50, while accuracy increases by 0.22-1.65% compared to the original SE-ResNet-50, as shown in (Fig. 8).

Table 4: The experimental results use the features obtained from the feature fusion by the best subset selection on the FPFV dataset.

Model	Original		Fusion Feature		
	ResNet-50	SE-ResNet-50	FFRM-Nets	FFSM-Nets	FFRSM-Nets
KNN	99.32±0.001	99.46±0.002	99.28±0.002	99.41±0.002	99.41±0.002
MLP	98.53±0.004	98.63±0.006	98.69±0.003	98.74±0.004	98.96±0.004
LR	99.64±0.002	99.72±0.002	99.63±0.002	99.69±0.002	99.94±0.001
SVM	99.42±0.002	98.04±0.004	99.28±0.003	99.64±0.002	99.69±0.002

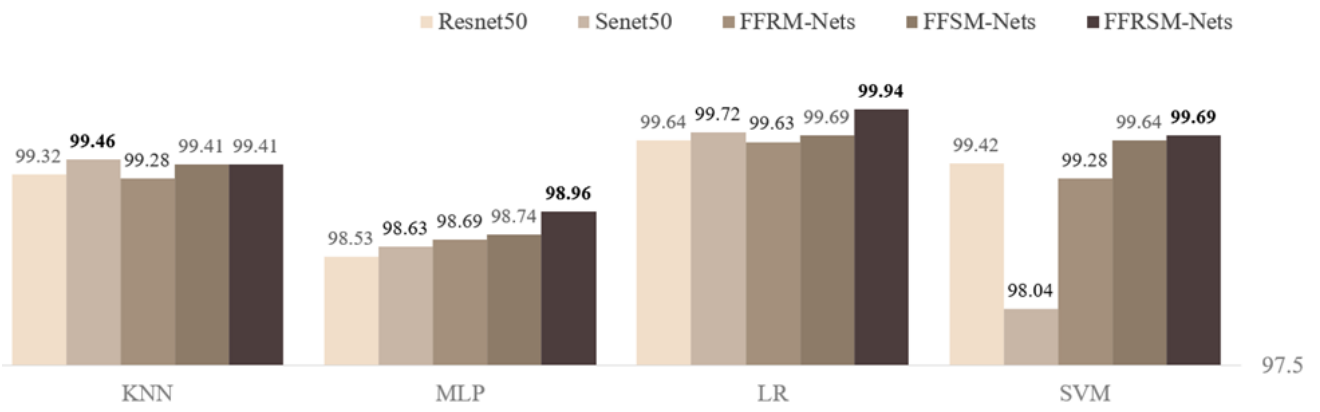


Fig. 8: Performance evaluation of four recognizer models consisted KNN, MLP, LR, and SVM that extract features and feature fusion five different nets.

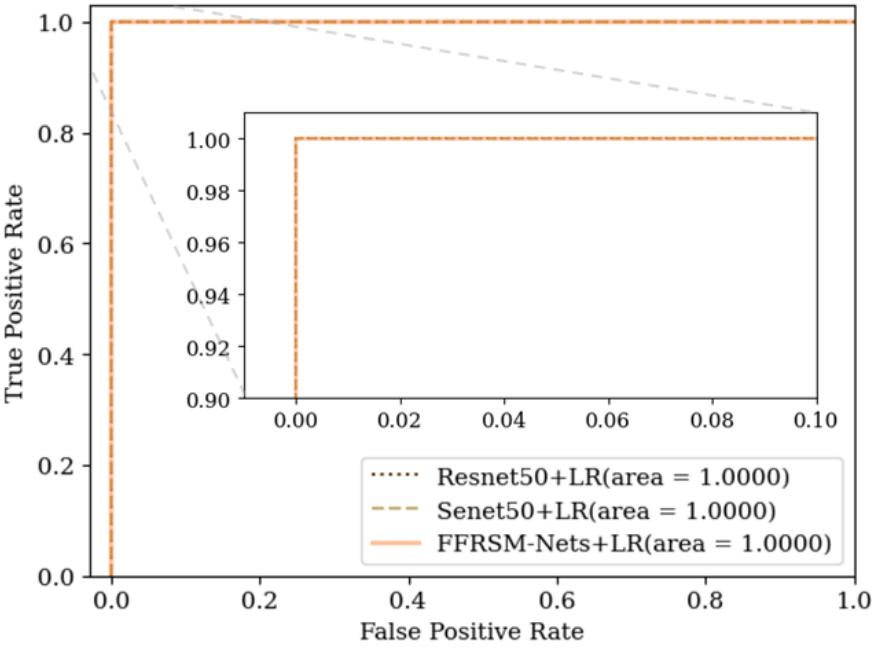


Fig. 9: ROC curve of the FFRSM-Nets fusion feature, ResNet-50, and SE-ResNet-50 combine the LR recognizer on the FPFV dataset.

We presented the confusion matrix of the presented methods, as shown in (Fig. 10).

5.2.3 Comparison of the proposed method on the FPFV dataset with the other existing methods.

The experimental results shown in Table 5. Show that the Fusion Feature, in which the FFRSM-Nets combined the SVM and LR recognizers, yielded better performance than other techniques on the FPFV dataset with an accuracy of 99.69% and 99.94%, respectively. The comparative results are shown in Table 5. The results show that the proposed method outperforms the CCFF-GAN combined ResNet-50 methods by approximately 0.33% on The FPFV dataset.

5.3 Recognition performance on the face recognition benchmark datasets

The previous section evaluated the Feature Fusion multi-conv Networks (FFM-Nets) on the FPFV dataset and found that the proposed method assigned the best features from subset feature selection for use in the recognizer methods. In this section, we experiment with the other dataset to ensure that the proposed method will give us the best features and increase recognition efficiency. The performance of the proposed method is shown as follows.

5.3.1 Recognition performance on the UFI (Cropped)

This section used pre-trained CNN models (ResNet-50 and SE-ResNet-50) trained on the UFI datasets. We extract the feature from the last pooling layer of the trained model. This dataset is divided into two sets: 1) the UFI (Cropped), we divided 4,316 images as a training and 605 images as a test set, and 2) the UFI (Large), we divided 4,350 images as a training and 530 images as a test set. Second, we put our feature into the proposed fusion framework, which includes FFRM-Nets, FFSM-Nets, and FFRSM-Nets. Finally, we trained the best features from the subset feature using the machine learning

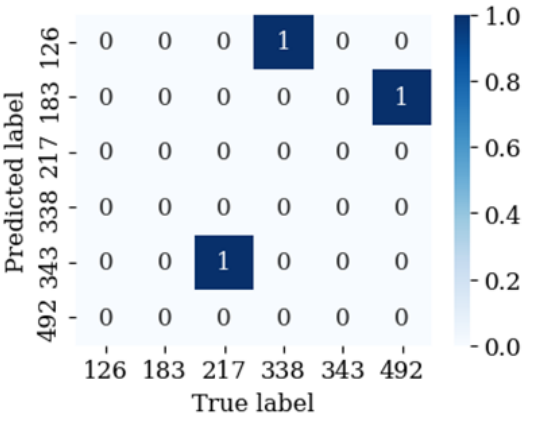


Fig. 10: The confusion matrix of the FFRSM-Nets features fusion combines the LR recognizer on the FPFV dataset.

Table 5: Recognition performance on The FPFV dataset with the existing methods.

References	Methods	Accuracy (%)
[10]	Deep Features	96.40±0.69
[33]	PIM	99.44±0.36
[34]	PF-cpGAN	98.88±1.56
[35]	QMagFace	98.74± −−
[36]	CCFF-GAN+ResNet-50	99.61±0.23
The proposed technique	FFRSM-Nets +SVM	99.69±0.002
	FFRSM-Nets +LR	99.94±0.001

methods of four recognizers, including SVM, KNN, MLP, and LR.

As shown in Table 6, the experimental results indicated that training LR recognizer combined feature fusion the FFSM-Nets on CNN model consistently obtained better performance than without using the convolution layer are conv1D and conv2D blocks. The result showed that the FFRSM-Nets combined with the LR recognizer achieved an accuracy of 98.99% on the UFI (Cropped). Furthermore, the ROC curve is shown in (Fig. 11). From the experimental results, we can see that the FFSM-Nets combined LR recognizer has an accuracy increase of approximately 0.02-0.2% compared to the original feature and an accuracy increase of 0.04% and 0.09% compared to FFRM-Nets and FFSM-Nets respectively.

Table 6: The experimental results use the features obtained from the feature fusion by the best subset selection on the UFI (Cropped).

Model	Original		Feature Fusion		
	ResNet-50	SE-ResNet-50	FFRM-Nets	FFSM-Nets	FFRSM-Nets
KNN	97.93	97.65	97.81	97.78	97.94
MLP	96.99	97.14	94.49	95.65	95.65
LR	98.97	98.79	98.95	98.90	98.99
SVM	98.65	97.80	95.73	98.67	98.67

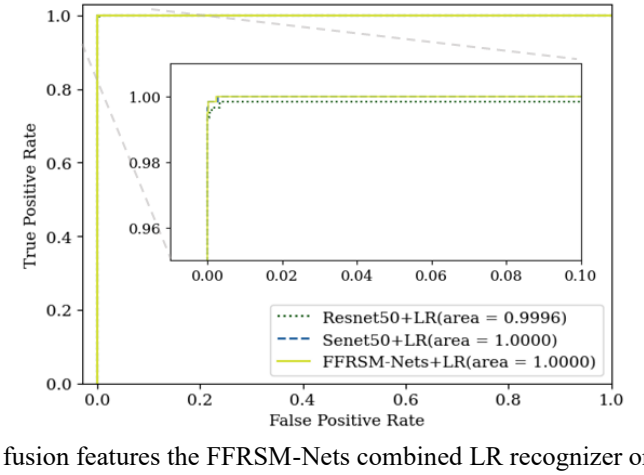


Fig. 11: ROC curve of fusion features the FFRSM-Nets combined LR recognizer on the UFI (Cropped) dataset.

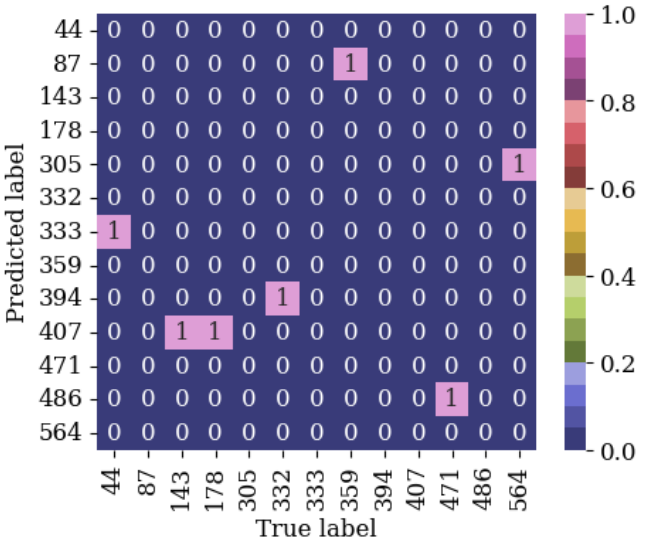


Fig. 12: The confusion matrix of fusion features the FFRSM-Nets combined LR recognizer on the UFI (Cropped) dataset.

We presented the confusion matrix of the fusion features of the FFRSM-Nets combined LR recognizer, as shown in (Fig. 12).

5.3.2 Recognition performance on the UFI (Large)

As shown in Table 7, the experimental results indicated that training the KNN recognizer combined feature fusion FFRSM-Nets on the CNN model using conv1D and conv2D blocks consistently obtained better performance. The result showed that the FFRSM-Nets combined KNN recognizer achieved an accuracy of 79.62% on the UFI (Large). Furthermore, the ROC curve is shown in (Fig. 13).

Table 7: The experimental results use the features obtained from the feature fusion by the best subset selection on the UFI (Large) dataset.

Model	Original		Fusion Feature		
	ResNet-50	SE-ResNet-50	FFRM-Nets	FFSM-Nets	FFRSM-Nets
KNN	78.96	78.94	78.84	78.77	79.62
MLP	73.05	72.01	67.49	70.07	70.07
LR	72.43	51.05	65.24	52.15	65.24
SVM	78.02	72.83	69.28	78.24	78.24

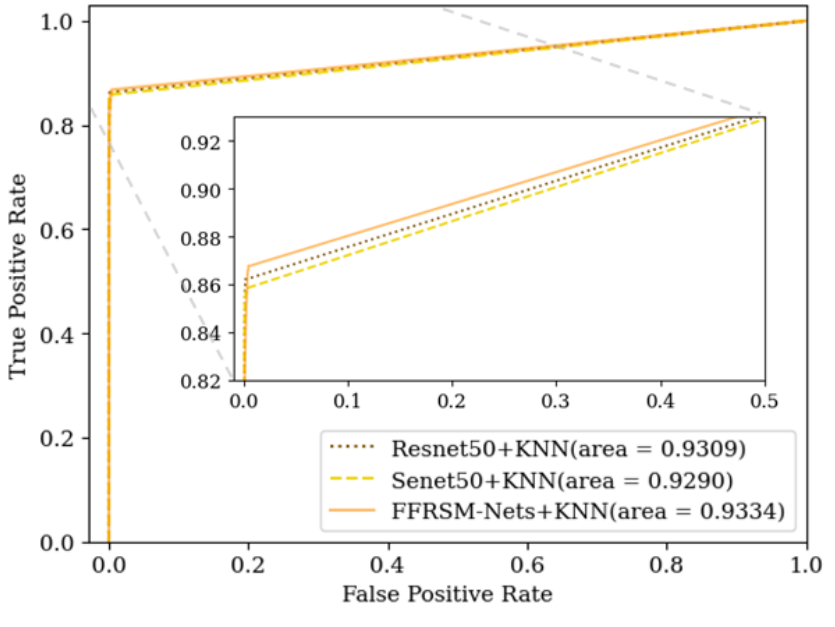


Fig. 13: ROC curve of feature fusion FFRSM-Nets combines KNN recognizer on the UFI (Large) dataset.

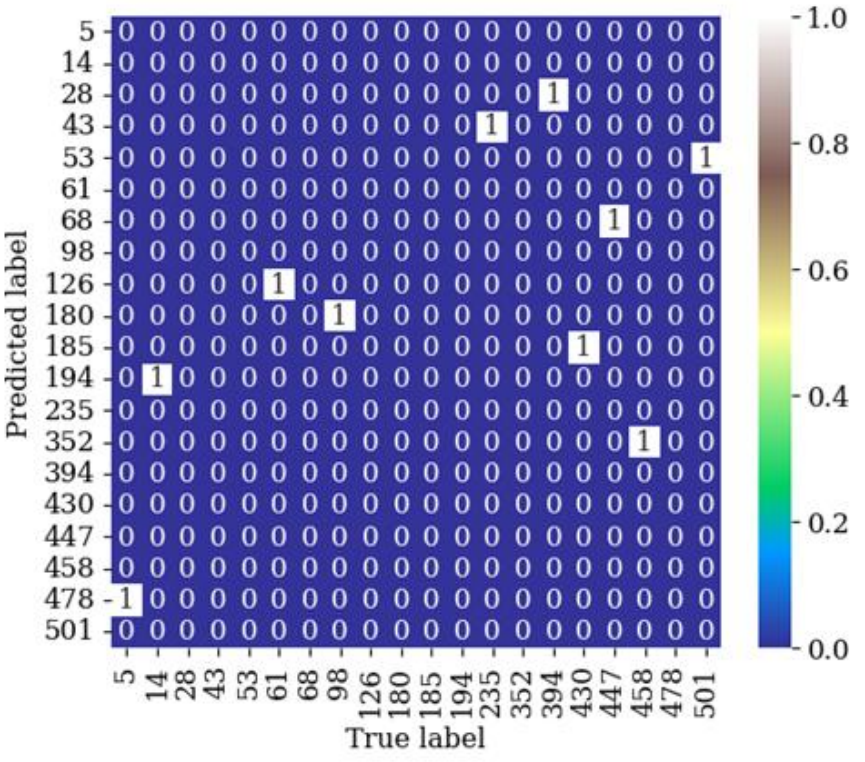


Fig. 14: The confusion matrix of the example misrecognition feature fusion FFRSM-Nets combines the KNN recognizer on the UFI (Large) dataset.

We presented the confusion matrix of the example misrecognition fusion features of the FFRSM-Nets combined KNN recognizer, as shown in (Fig. 14).

5.3.3 Comparison of the Proposed Method on the UFI Dataset with the Other Existing Methods

From the experimental results shown in Table 6 and Table 7, it can be seen that the Fusion Feature, in which FFRSM-Nets combined SVM and LR recognizers, yielded better performance than other techniques on the UFI (Cropped) dataset with an accuracy of 98.67% and 98.99%. The FFRSM-Nets combined SVM and KNN recognizers yielded better performance than other techniques on the UFI (Large) dataset, with an accuracy of 78.24% and 79.62% respectively. The comparative results are shown in Table 8. The results show that the proposed method outperforms previous methods by approximately 24% on the UFI (Cropped) and 36% on the UFI (Large).

Table 8: Accuracy of the proposed method on UFI dataset in comparison with other approaches.

Dataset	References	Methods	Accuracy (%)
The UFI (Cropped)	[27]	POEMHS	67.11
	[37]	M-BNCC	74.55
	[38]	GL-POEM	74.20
	The proposed technique	FFRSM-Nets+SVM	98.67
		FFRSM-Nets+LR	98.99
UFI (Large)	[27]	FS-LBP	43.2
	The proposed technique	FFRSM-Nets+SVM	78.24
		FFRSM-Nets+KNN	79.62

Table 9: The experimental results use the features obtained from the feature fusion by the best subset selection on the MIT-CBCL.

Model	Original		Feature Fusion		
	ResNet-50	SE-ResNet-50	FFRM-Nets	FFSM-Nets	FFRSM-Nets
KNN	91.75	98.84	91.91	98.48	98.48
MLP	95.44	92.15	97.28	97.42	98.04
LR	91.72	98.68	94.69	.9945	.9945
SVM	89.57	75.88	97.81	98.43	98.73

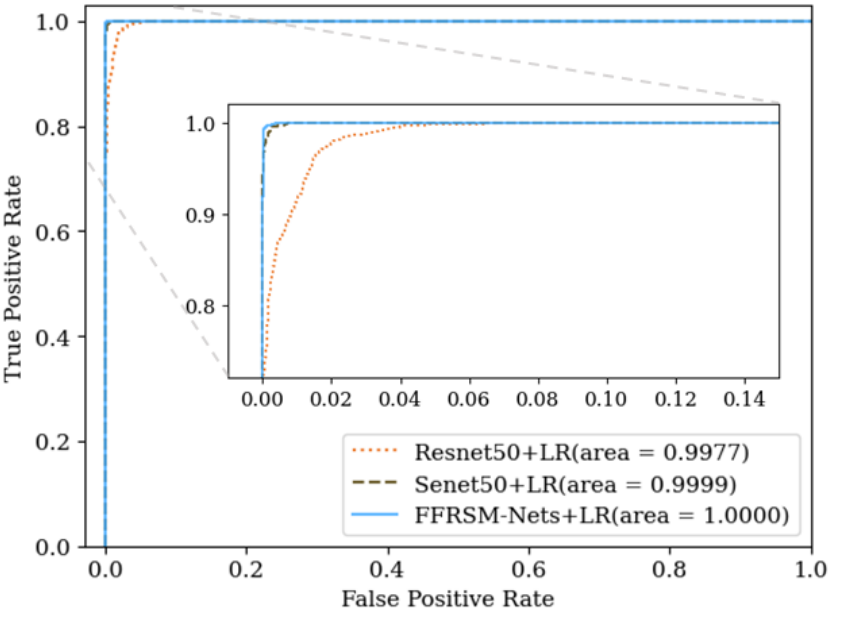


Fig. 15: ROC curve of feature fusion FFRSM-Nets combines LR recognizer on the MIT-CBCL dataset.

5.3.4 Recognition Performance on the MIT-CBCL. In this experiment, we tested the proposed fusion framework on the MIT-CBCL dataset, which has 3,240 training-synthetic images and 2,000 test images. The proposed feature fusion results are shown in Table 9.

From Table 9, the experimental results indicated that training the LR recognizer combined feature fusion FFRSM-Nets consistently obtained better performance. The result showed that the FFRSM-Nets combined LR recognizer achieved an accuracy of 99.45% on the MIT-CBCL. Furthermore, the ROC curve is shown in (Fig. 15). Considering the results, it indicates that the LR recognizer combined feature fusion FFRSM-Nets the accuracy better than by approximately 7% when using ResNet-50, 0.7% SE-ResNet-50 features, and around 4.7% when using FFRM-Nets.

We presented the confusion matrix of the fusion features of the FFRSM-Nets combined with the LR recognizer on the MIT-CBCL dataset, as shown in (Fig. 16).

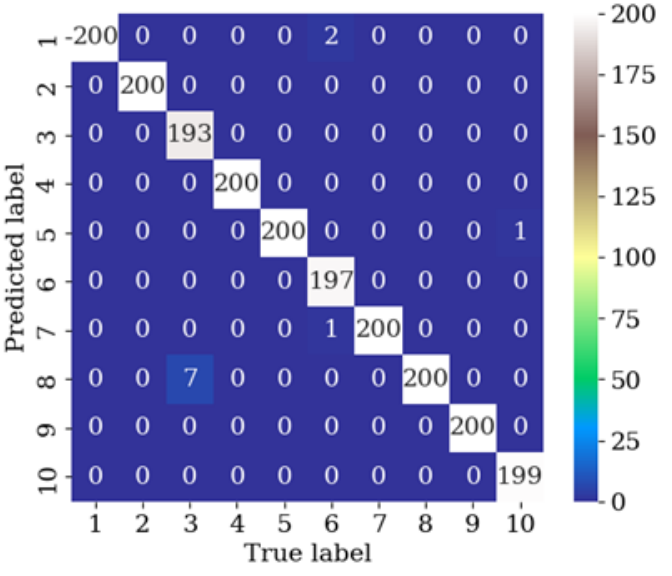


Fig. 16: The confusion matrix of feature fusion FFRSM-Nets combines the LR method on the MIT-CBCL dataset.

Table 10: Accuracy of the proposed method on the MIT-CBCL dataset in comparison with other approaches.

References	Methods	Accuracy (%)
[26]	Component-based face recognition with 3D morphable models + SVM	88
The proposed technique	FFRSM-Nets +MLP	98.04
	FFRSM-Nets +SVM	98.73
	FFRSM-Nets +LR	99.45

5.3.5 Performance on the MIT-CBCL.

From the experimental results shown in Table 10, it can be seen that the Fusion Feature, in which FFRSM-Nets combined MLP, SVM, and LR recognizers, yielded better performance than other techniques on the MIT-CBCL dataset with an accuracy of 98.04%, 98.73%, and 99.45% respectively. The results show that the proposed method outperforms previous methods by approximately 11.45% on the MIT-CBCL dataset.

6. Conclusion

This study presents feature fusion Multi-conv Networks (FFM-Nets) in face image recognition. Combining the best features from ResNet-50 and SE-ResNet-50 CNN models, we experimentally evaluated the FFM-Nets for extracting smaller features, resulting in a smaller model size and reduced model training time. We divide the feature fusion framework into three sets: feature combined via Conv1D and Conv2D Blocks, where these blocks are via the BN layer, ReLU activation function, dropout, and without dropout layer. The first framework is called Feature Fusion ResNet-50 Multi-conv Networks (FFRM-Nets), the second framework is called Feature Fusion SE-ResNet-50 Multi-conv Networks (FFSM-Nets), and the third framework is called Feature Fusion ResNet-50 and SE-ResNet-50 Multi-conv Networks (FFRSM-Nets) respectively. From the experimental results, we get different features from deep features, namely ResNet-50 and SE-ResNet-50. All features are then divided into subsets, and robust features are selected using the best feature selection method. Finally, the robust features introduced are recognized using four machine learning methods: SVM, KNN, MLP, and LR.

The results show that the fusion feature, in which the FFRSM-Nets combined the LR recognizers, has the best performance of 99.94% (Table 4) and yields better performance than other techniques. The results show that the proposed method outperforms the previous work on the FPFV dataset by approximately 0.33% (Table 5). While the fusion feature, in which FFRSM-Nets combined LR recognizers, yielded better performance than other techniques on the UFI (Cropped) dataset with an accuracy of 98.99% (Table 6), The FFRSM-Nets combined KNN recognizers yielded better performance than other techniques on the UFI (Large) dataset with an accuracy of 79.62% (Table 7), The results show that the proposed method outperforms previous methods by approximately 24% on the UFI (Cropped) and 36% on the UFI (Large) (Table 8), and the MIT-CBCL dataset it was found that FFRSM-Nets combined LR recognizers yielded better performance with an accuracy of 99.45% (Table 9). The results show that the proposed method outperforms previous methods by approximately 11.45% (Table 10). Experiments using different datasets and comparing the proposed method with state-of-the-art methods have been performed, indicating the efficacy of the proposed FFM-Nets framework. It can be used to improve the performance of traditional methods by integrating the robust features of face image recognition. The proposed approach shows potential for specific real-world applications mentioned in the introduction. The feature size reduction of up to 87% achieved through Multi-conv networks enable deployment on resource-constrained devices while maintaining accuracy. The improved performance demonstrated across diverse datasets supports reliability requirements for identity verification and financial authentication scenarios. The consistent improvements across different challenging conditions suggest practical viability for security applications. However, implementation challenges include the computational overhead during the feature selection phase, which may require offline optimization for deployment scenarios. The modest accuracy improvements necessitate careful cost-benefit analysis for specific applications where marginal gains justify the additional complexity.

In future work, several potential improvements can be explored to optimize this research approach. Future research should investigate advanced CNN architectures such as Vision Transformers and Efficient Net to improve feature extraction capabilities further and explore novel attention mechanisms and adaptive feature fusion strategies to enhance framework robustness. The integration of evolutionary algorithms and swarm intelligence for feature selection optimization presents promising research directions, while developing ensemble and parallel network architectures could significantly boost recognition performance while maintaining computational efficiency. Additionally, expanding the evaluation to larger-scale real-world datasets and optimizing the framework for real-time applications would enhance its practical utility, particularly improving performance under challenging conditions such as extreme lighting variations and occlusions. These improvements would contribute to developing more robust, efficient, and practically applicable face recognition systems for diverse real-world scenarios.

Acknowledgments

This research project was financially supported by Mahasarakham University, Thailand.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supporting Information

Not applicable.

CRediT Statement

Thipwimon Chompookham: Conceptualization, Writing - Original draft, Writing - Review and editing, Resources. **Keerati Tongnate:** Writing - Review and editing. **Siriwan Phiphiphatphaisit:** Writing - Review and editing. **Siriwiwat Lata:** Writing - Review and editing. **Niwat Angkawisittpan:**

Conceptualization, Supervision, Resources, Project administration, Writing - Review and editing, Funding acquisition. **Sivarit Sultornsanee:** Writing - Review and editing.

References

[1] Y. Kortli, M. Jridi, A. Al Falou, M. Atri, Face recognition systems: a survey, *Sensors*, 2020, **20**, 342, doi: 10.3390/s20020342.

[2] Y. Hu, H. An, Y. Guo, C. Zhang, T. Zhang, L. Ye, The development status and prospects on the face recognition, *4th International Conference on Bioinformatics and Biomedical Engineering*, Chengdu, China, IEEE, June 18-20, 2010, 1-4, doi: 10.1109/ICBBE.2010.5517197.

[3] L. Li, X. Mu, S. Li, H. Peng, A review of face recognition technology, *IEEE Access*, 2020, **8**, 139110-139120.

[4] Y. Duan, J. Lu, J. Zhou, UniformFace: learning deep equidistributed representation for face recognition, *CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, IEEE, June 15-20, 2019, 3410-3419, doi: 10.1109/CVPR.2019.00353.

[5] Y. Li, X. Wu, J. Kittler, L1-(2D)2PCANet: a deep learning network for face recognition, *Computer Vision and Pattern Recognition*, 2018, doi: 10.48550/arXiv.1805.10476.

[6] X. Yin, X. Yu, K. Sohn, X. Liu, M. Chandraker, Feature transfer learning for deep face recognition with under-represented data, *Computer Vision and Pattern Recognition*, 2018, doi: 10.48550/arXiv.1803.09014.

[7] S. P. Patil, R. S. Apare, R. H. Borhade, P. N. Mahalle, Automated dyslexia screening using children's handwriting in English language using convolutional neural network and bidirectional long short-term memory model, *Engineered Science*, 2024, **32**, 1345, doi: 10.30919/es1345.

[8] I. Adjabi, A. Ouahabi, A. Benzaoui, A. Taleb-Ahmed, Past, present, and future of face recognition: a review, *Electronics*, 2020, **9**, 1188, doi: 10.3390/electronics9081188.

[9] X. Qi, C. Wu, Y. Shi, H. Qi, K. Duan, X. Wang, A convolutional neural network face recognition method based on BiLSTM and attention mechanism, *Computational Intelligence and Neuroscience*, 2023, **2023**, 2501022, doi: 10.1155/2023/2501022.

[10] S. Sengupta, J.-C. Chen, C. Castillo, V. M. Patel, R. Chellappa, D. W. Jacobs, Frontal to profile face verification in the wild, *Winter Conference on Applications of Computer Vision*, Lake Placid, NY, USA, IEEE, March 7-10, 2016, 1-9, doi: 10.1109/WACV.2016.7477558.

[11] E. Zhou, Z. Cao, Q. Yin, Naive-deep face recognition: touching the limit of LFW benchmark or not, *Computer Vision and Pattern Recognition*, 2015, doi: 10.48550/arXiv.1501.04690.

[12] T. Chen, T. Gao, S. Li, X. Zhang, J. Cao, D. Yao, Y. Li, A novel face recognition method based on fusion of LBP and HOG, *IET Image Processing*, 2021, **15**, 3559-3572, doi: 10.1049/ipr2.12192.

[13] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, *IEEE Transactions on Industrial Electronics*, 2018, **65**, 5990-5998, doi: 10.1109/TIE.2017.2774777.

[14] D. Peng, Z. Liu, H. Wang, Y. Qin, L. Jia, A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains, *IEEE Access*, 2018, **7**, 10278-10293.

[15] T. Zan, Z. Liu, H. Wang, M. Wang, X. Gao, Control chart pattern recognition using the convolutional neural network, *Journal of Intelligent Manufacturing*, 2020, **31**, 703-716, doi: 10.1007/s10845-019-01473-0.

[16] T. Zan, H. Wang, M. Wang, Z. Liu, X. Gao, Application of multi-dimension input convolutional neural network in fault diagnosis of rolling bearings, *Applied Sciences*, 2019, **9**, 2690, doi: 10.3390/app9132690.

[17] A. Khalifa, A. A. Abdelrahman, T. Hempel, A. Al-Hamadi, Towards efficient and robust face recognition through attention-integrated multi-level CNN, *Multimedia Tools and Applications*, 2025, **84**, 12715-12737, doi: 10.1007/s11042-024-19521-0.

[18] Y. Shi, H. Zhang, W. Guo, M. Zhou, S. Li, J. Li, Y. Ding, LighterFace model for community face detection and recognition, *Information*, 2024, **15**, 215, doi: 10.3390/info15040215.

[19] F. Chollet, Xception: deep learning with depthwise separable convolutions, *Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, IEEE, July 21-26, 2017, 1800-1807, doi: 10.1109/CVPR.2017.195.

[20] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, *Computer Vision and Pattern Recognition*, 2015, doi: 10.48550/arXiv.1409.1556.

[21] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, *Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, IEEE, June 27-30, 2016, 770-778, doi: 10.1109/CVPR.2016.90.

[22] S. Mukherjee, The Annotated ResNet-50, *Blog post*, 2022, https://medium.com/data-science/the-annotated-resnet-50-a6c536034758 (accessed August 18, 2024).

[23] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, *Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, IEEE, June 18-23, 2018, 7132-7141, doi: 10.1109/CVPR.2018.00745.

[24] F. Wang, J. Feng, Y. Zhao, X. Zhang, S. Zhang, J. Han, Joint activity recognition and indoor localization with WiFi fingerprints, *IEEE Access*, 2019, **7**, 80058-80068, doi: 10.1109/ACCESS.2019.2923743.

[25] J. Zhu, C. Wen, J. Zhu, H. Zhang, X. Wang, A polynomial algorithm for best-subset selection problem, *Proceedings of the National Academy of Sciences of the United States of America*, 2020, **117**, 33117-33123, doi: 10.1073/pnas.2014241117.

[26] B. Weyrauch, B. Heisele, J. Huang, V. Blanz, Component-based face recognition with 3D morphable models, *Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA, IEEE, June 27 - July 2, 2004, 85, doi: 10.1109/CVPR.2004.315.

[27] L. Lenc, P. Král, Unconstrained facial images: database for face recognition under real-world conditions, *Advances in Artificial Intelligence and Its Applications. Cham: Springer*, 2015,

349-361, doi: 10.1007/978-3-319-27101-9_26.
[28] S. Suherman, A. H. Rambe, N. Panjaitan, A. S. Alfakeeh, Time-shifted gramian angular field and recursive plot convolutional neural network to adapt solar cell dataset overfitting in hybrid power generation, *Engineered Science*, 2024, **31**, 1222, doi: 10.30919/es1222.
[29] S. A. Hicks, I. Strümke, V. Thambawita, M. Hammou, M. A. Riegler, P. Halvorsen, S. Parasa, On evaluation metrics for medical applications of artificial intelligence, *Scientific Reports*, 2022, **12**, 5979, doi: 10.1038/s41598-022-09954-8.
[30] M. Hossin, S. M. N, A review on evaluation metrics for data classification evaluations, *International Journal of Data Mining and Knowledge Management Process*, 2015, **5**, 1-11, doi: 10.5121/ijdkp.2015.5201.
[31] K. Hajian-Tilaki, Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation, *Caspian Journal of Internal Medicine*, 2013, **4**, 627-635.
[32] R. Kumar, A. Indrayan, Receiver operating characteristic (ROC) curve for medical researchers, *Indian Pediatrics*, 2011, **48**, 277-287, doi: 10.1007/s13312-011-0055-4.
[33] J. Zhao, Y. Cheng, Y. Xu, L. Xiong, J. Li, F. Zhao, K. Jayashree, S. Pranata, S. Shen, J. Xing, S. Yan, J. Feng, towards pose invariant face recognition in the wild, *Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, IEEE, June 18-23, 2018, 2207-2216, doi: 10.1109/CVPR.2018.00235.
[34] F. Taherkhani, V. Talreja, J. Dawson, M. C. Valenti, N. M. Nasrabadi, PF-cpGAN: profile to frontal coupled GAN for face recognition in the wild, *International Joint Conference on Biometrics*, Houston, TX, USA. IEEE, September 28 - October 1, 2020, 1-10, doi: 10.1109/IJCB48548.2020.9304872.
[35] P. Terhörst, M. Ihlefeld, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, A. Kuijper, QMagFace: simple and accurate quality-aware face recognition, *Winter Conference on Applications of Computer Vision*, Waikoloa, HI, USA, IEEE, January 2-7, 2023, 3473-3483, doi: 10.1109/WACV56688.2023.00348.
[36] Z. Zhang, R. Liang, X. Chen, X. Xu, G. Hu, W. Zuo, E. R. Hancock, Semi-supervised face frontalization in the wild, *IEEE Transactions on Information Forensics and Security*, 2020, **16**, 909-922, doi: 10.1109/TIFS.2020.3025412.
[37] J. Gaston, J. Ming, D. Crookes, Unconstrained face identification with multi-scale block-based correlation, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2017, 1477-1481.
[38] L. Lenc, P. Král, improving face recognition methods based on POEM features, *Proceedings of the 12th International Conference on Agents and Artificial Intelligence*, Valletta, Malta. SCITEPRESS - Science and Technology Publications, February 22-24, 2020, 538-545, doi: 10.5220/0008950305380545.