Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

Feature Level Combination for Face Recognition Based on Convolutional Neural Networks

Jamal M. Alrikabi [a] , Kadhim H. Alibraheemi [b]

[a,b] Computer Science Department, College of Education for Pure Science, Thi-Qar University, Thi-Qar, Iraq.

[a] Email: jamal13_mah@utq.edu.iq, [b] Email: khalibraheemi@utq.edu.iq

ARTICLE INFO.

Article history:
Received: 25 /07/2021
Rrevised form: 06 /09/2021
Accepted : 20 /09/2021
Available online: 23 /09/2021

Keywords:

Deep learning,
Convolution neural network,
Feature combination,
Face detection,
Face recognition.

ABSTRACT

Face detection and recognition systems have recently achieved encouraging results using deep learning, especially Convolutional Neural Network (CNN). Face Recognition (FR) systems have many challenges in unconstrained environments that decrease the accuracy; for overcoming these challenges, a deep learning-based features combination has been proposed. The scheme performs feature-level combination by applying two pre-trained GoogLeNet and VggNet-16 models as deep feature extractors. First, faces are detected and aligned using the Multi-Task Convolutional Neural Networks (MTCNN) face detector. The deep features are extracted from a face image using each individually pre-trained CNN. Second, features obtained from GoogLeNet and VggNet-16 models are combined using the serial-feature combination method. Finally, a classification task is performed using a multiclass Support Vector Machine (SVM) classifier. Experiments on the following datasets: VggFace2, LFW, Essex, and ORL, indicate the efficacy of the proposed system as the combination of the two pre-trained CNN models improves performance. The combination strategy, in particular, yields an accuracy of 95.33% to 99.29% on all datasets**.** The proposed system was compared to existing models in terms of the LFW, and ORL datasets, the findings showed that the proposed system outperformed most current models in terms of accuracy.

MSC. 41A25; 41A35; 41A36

DOI : https://doi.org/10.29304/jqcm.2021.13.3.849

1. Introduction

In recent years, Personal identity become crucial. Biometrics is a preferred authentication method, as it is believed to be the most secure and complicated authentication method [1]. Biometric systems are evolving technologies that

∗Corresponding author: Jamal M. Alrikabi

Email addresses: jamal13_mah@utq.edu.iq.

Communicated by*: Dr. Rana Jumaa Surayh aljanabi.*

can be utilized in automated systems to uniquely and effectively identify persons, making them a viable alternative to more traditional approaches such as passwords [2]. According to a survey published in [3], customers prefer to utilize smartphone biometrics instead of passwords since they give an additional layer of security for emerging technologies such as Apple Pay [3]. Biometrics is a technique for automatically authenticating a person based on a physical or behavioral feature. Face, fingerprints, iris, and voice are all physical features. Several behavior features can be learned or gained, including keystroke dynamics, dynamic signature verification, and speaker verification [4].

Face recognition is a computer application capable of confirming or detecting an individual from a digital image or a video frame. One way to accomplish this is to compare the chosen facial feature from the image to face dataset [5]. Face recognition systems are used in various applications and situations, including personal identification, image film processing, psychology, computer interface, security systems, surveillance, law enforcement, smart cards, and entertainment systems. Face recognition systems, in general, consist of two phases: face detection and face recognition [6]. Generally, there are three methodologies for face recognition: holistic methods, Feature-based methods, and Hybrid methods. CNN falls under the feature-based approach [6].

Several face detection and recognition systems based on deep CNN have been suggested [7]. F. Tabassum et al. introduced a face recognition approach based on the combination of DWT with a Convolutional Neural Network. The results are combined using the entropy of detection probability and a fuzzy system. Experiments demonstrate that the combined method outperforms prior efforts using the unique algorithms [8]. S. Bajpai et al. the author presents a method for a face recognition system combining pre-trained Inception-Resnetv1 CNN architecture for extracting image features and sparse linear approximation for classification. The experiment shows that the approach performs better even in an unconstrained environment as compared to the existing methods [9]. Y. Yang et al. a new face matching approach called the SR-CNN model has been introduced a combination of the CNN, the rotation-invariant textures feature (RITF) vector, and the scale-invariant features transform (SIFT) vector. This combination approach achieves higher accuracy than the individual methods [10]. J. Li et al. the authors proposed a new approach for the FR system that combined CNN with principal component analysis. C2D-CNN integrates the features and then performs combination at the decision level, which results in a significant enhancement in the implementation of the system [11]. K. Patrik et al. The authors presented a seven-layer convolutional neural network that incorporates three well-known image identification techniques: LBPH, KNN, and PCA. These testing findings established the proposed facial recognition method's efficacy [12]. S. Guo et al. suggested a CNN-based SVM-based facial recognition system. CNN is utilized to extract features, and SVM is used to classify in this system. Experiments demonstrate the system's superiority. The testing demonstrated that the system acquired a high recognition rate after only a few hours of training [13].

This paper is organized in the following sequence: Section2 describes the Convolution Neural Networks. Section3, Introduces the proposed system. The Methodology and Experimental results are represented in Section4. Finally, some conclusions are made in Section5.

2. Convolution Neural Networks

CNN works on images by convolving them with a group of filters and passing the convolution result by the nonlinear dropping to get the classification of identity. The filter weights are learned to minimize classification loss. Each layer uses the previous layer's output to recognize top level features. The number of neurons in each layer of CNN is determined by its parameters [14]. The architecture of CNN is illustrated in figure (1). The CNN model is composed of multiple layers, with each layer receiving input in the form of a multi-dimensional array of numbers. The primary layers that comprise the CNN architecture [15] are the Input Layer, the Convolutional Layer, the Batch Normalization Layer, the RELU (Rectified Linear Unit) Layer, the Pooling Layer, and the Fully Connected Layer.

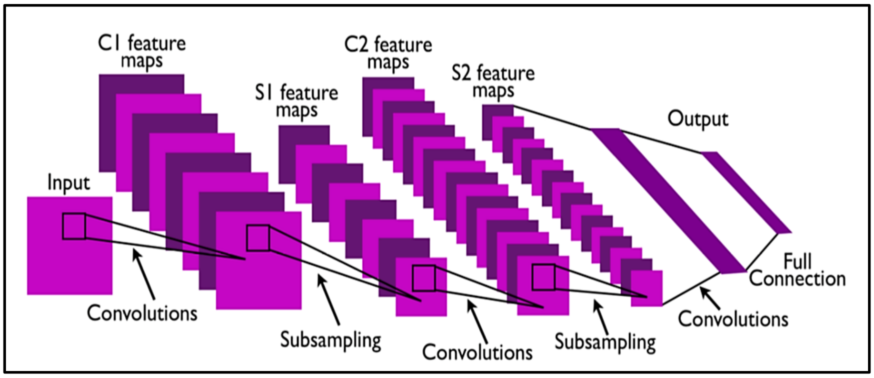

Figure (1): The architecture of typical convolutional neural networks.

2.1 Pre-trained CNN models

Many CNN models have been trained on large datasets such as the ImageNet dataset for image recognition purposes. These models can then be employed to recognize a different task without training from scratch [15]. Face representation heavily influences the performance of convolution neural networks in face recognition [16] and is currently a hot topic in face recognition research. In this paper, we used two pre-trained CNN as deep feature extractors. These CNN were GoogLeNet and VggNet-16. The pretrained CNN networks were utilized to extract for appropriate features from face image to use it for the classification step.

2.1.1 Pre-trained VggNet-16 Model

VggNet, presented by Simonyan et al. [17], as the second convolution neural networks that won the Image-Net competition in 2014, with a top-5 error of 7.3 percent as illustrated in Figure (2). This network is characterized by its simplicity, using only 3×3 convolutional layers stacked on top of each other in increasing depth. In these convolution and max-pooling layers, the filters we use are of the size 3*3 instead of 11*11 in AlexNet and 7*7 in ZF-Net. Reducing volume size is handled by max pooling. Two fully connected layers, each with 4,096 nodes are then followed by a SoftMax classifier. This architecture takes as input a 224-by-224-by-3 image.

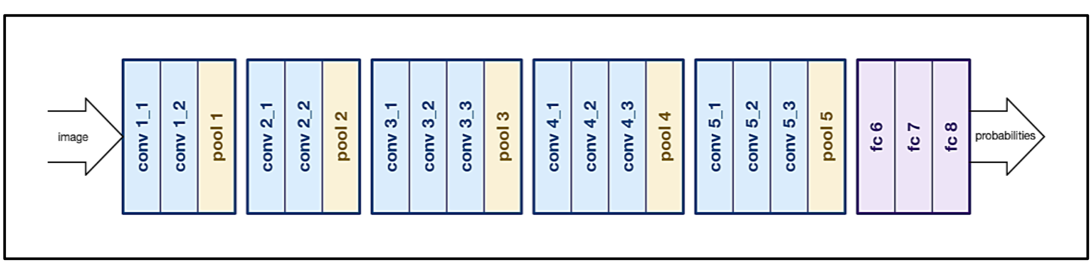

Figure (2): An illustration of the architecture of the VggNet-16 model [15].

2.1.2 Pre-trained GoogLeNet Model

GoogLeNet [18] is a CNN proposed by google that won the ImageNet Challenge in 2014 (ILSVRC) for classification and detection tracks. The salient property of GoogLeNet is an inception module, as shown in Figure (3), that introduces sparsity and multiscale information in one block. Functionally it is equivalent to a small network inside a large network. GoogLeNet architecture contained 22 layers and 40 million parameters. This network was able to achieve the least top-5 error rate of 6.67%. Because GoogLeNet is not a sequential CNN, it can increase its width and depth without putting the system under computational strain. This architecture takes as input a 224-by-224-by-3 image.

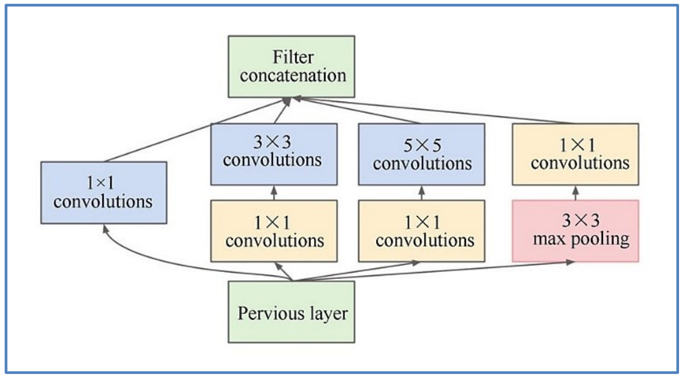

Figure (3): An illustration of the architecture of the inception module [15].

3. The Proposed System

In this study, we suggest a deep learning-based features extraction and combination for face detection and recognition. The scheme performs feature-level combinations by applying two pre-trained convolutional neural networks, GoogLeNet and VggNet-16. First, the features are extracted from the face image using each individually pre-trained convolutional neural network. Second, features obtained from GoogLeNet and VggNet-16 are combined using the feature-level combination method. Finally, for the classification process, the classifier of the support vector machine is utilized to perform the task of classification.

The proposed scheme for solving the problem of face detection and recognition based on deep learning illustrated in Figure (4) is divided into the following steps:

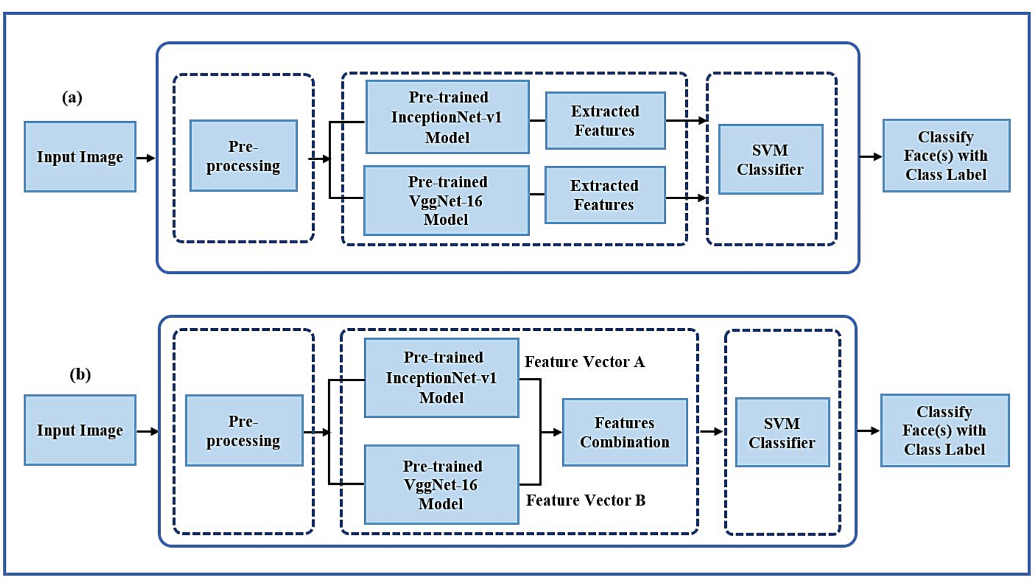

Figure (4): The central diagram for the proposed system approaches: (a) pre-trained CNN model as feature extractor and SVM classifier (b) feature combination from two pre-trained CNN models and SVM classifier.

3.1 Pre-processing Step:

The pre-processing stage is critical before entering the face images into the deep CNN network. In this study, the pre-processing stage, in general, contains the following steps: First, Image enhancement was performed using the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm that restricts so that the contrast value can be restricted in order to produce an image that has better contrast and lighting. Second, the primary step in any face recognition system is Face detection. The MTCNN Detector is used to detect the location of a face from the background in an input image.

In addition, the MTCNN detector identifies five facial landmark positions, namely, the right and left corners for the mouth, the eyes, and the tip of the nose, that are used for face alignment. Face alignment is used to increase recognition accuracy. Second, following the detection and alignment step, the face area is cropped from the input image. Third, The VggNet-16 and GoogLeNet networks receive RBG images at a size of 224-by-224, so the size of the images must be resized to 224-by-224 for all training and test images. Additionally, if the dataset contains gray images, we transform all gray images into RGB images.

3.2 Deep Feature Extraction Stage

After some pre-processing, the image of the face is delivered to the deep feature extractor model. The extracted features must be robust against the variations in face pose, illumination, expression, aging, and other challenging. In this study, this aim was achieved by using two pre-trained convolutional neural networks GoogLeNet and VggNet-16. The VggNet-16 model consists of three fully connected layers (fc8, fc7, and fc6). the "fc7" and "fc6" layers output is a vector with 4096-dimensional. The last two layers in the GoogLeNet model are the global average pooling layer "pool5" and the fully connected layer "fc1000". the "Pool5" layer output is a vector with 1024-dimensional.

The different layers in GoogLeNet and VggNet-16 networks extract different levels of features. Most of the extracted features from the initial layers are composed of colors and edges. In the next layers, different filters are used this enables the network to build more complicated features in the following layers. Additionally, the final layers such as "fc7" in the VggNet-16 network and "Pool5" in the GoogLeNet network learn a high-level set of the features extracted by the prior layers Which are located at the beginning of the network.

In this study, the features were taken from the VggNet-16 network's fully connected layer "fc7" and the "pool5" layer in the GoogLeNet network, because the best recognition accuracy is achieved with features from the fully connected layer "fc7" and the "pool5" layer, where High level layer features are an abstract of low level layer features and are more discriminatory for classification tasks.

3.3 Classification stage

In the classification stage, the extracted features are used to train a Classifier. In this work, an SVM classifier has been used to classify the image features because it is very efficient in image classification. Even though SVM was initially developed for binary classification, it can be successfully extended to be applied to multiclass classification problems. The Error-Correcting Output Code (ECOC) is a commonly used framework to model multiclass classification problems. Fundamentally, there are two methods for multiclass classification, "one-against-one" and "one-against-all". In this study, image classification is implemented by combining the SVM classifier and ECOC framework.

The "fitcecoc" function has been used for training a multiclass model for SVM on the extracted features. The function returns a trained ECOC model. The present work uses K binary SVM Algorithm with a "one-verses-all" coding approach, K refers to a distinct class. A fast Stochastic Gradient Descent solver is used for training by setting the fitcecoc function's "Learners" parameter to "Linear" since we use linear kernel function. This helps speed up the training when working with high-dimensional CNN feature vectors. To predict label "predict" function is used which returns a predicted class label for the predictor data in the table or matrix based on the ECOC model.

4. The Methodology and Experimental Results

The proposed face detection and recognition system in this study use several approaches that can be grouped into two strategies:

A. **Strategy-1:** Two pre-trained CNN were used separately as a deep features extractor and an SVM classifier.

- **Approach-1:** Pre-trained GoogLeNet as a deep features extractor and an SVM classifier.

- **Approach-2:** Pre-trained VggNet-16 as a deep features extractor and an SVM classifier.

B. **Strategy-2:** Feature-level combination from the pre-trained GoogLeNet and VggNet-16 models followed by an SVM classifier.

4.1 The Experimental Setups

4.1.1 The Hardware Tools

The experiments were executed using a laptop with a processor Intel (R) Core (TM) i7- 10750 QH CPU @ 2.60 GHz, and RAM of 16 GB and GPU of Nvidia GeForce RTX 2060.

4.1.2 The Software Tools

Matlab 2020b install on Windows 10 Pro 64-bit operating system platform to assess the proposed approach and implement feature extraction and classification task. Matlab toolbox and support packages that will be used include the following:

- Deep Learning Toolbox™
- Machine Learning Toolbox™
- Computer Vision Toolbox™
- Neural Network Toolbox™
- Image Processing Toolbox™
- Deep Learning Toolbox™ for VggNet-16 Model
- Deep Learning Toolbox™ for GoogLeNet Model

4.2 Description of Datasets

Detailed descriptions of all image-based datasets that were utilized in all experiments can be found in this section. Four datasets were used for this purpose, namely, VggFace2 [19], Essex [21], LFW [20], and ORL [22]. Table (1) summarizes the information contained in each dataset that was used in this study. Figure (5) illustrates some of the datasets face images.

Table (1): Details about the datasets used in the experiments.

Dataset	Dataset Description				
	Total Images	Subjects	Images/ person	Size of Images	Type of Images
VggFace2	280	20	14	137x180	JPEG
LFW	280	20	14	250x250	JPEG
ORL	200	20	10	92x112	JPEG
Essex	280	20	14	196x196	JPEG

4.2.1 VggFace2 Dataset

The VggFace2 dataset contains approximately 3.31 million images that have been divided into 9131 classes, each of which represents a different person's identity. With large variations in the profession, ethnicity, illumination, age, and pose, the VggFace2 database is an excellent choice for training deep learning models on face-related tasks. The average resolutions are (137x180) pixels for each image.

4.2.2 LFW dataset

The LFW was created to investigate face recognition problems in an unconstrained environment, including variations in focus, hairstyles, camera quality, clothing, age, gender, lighting, ethnicity, background, race, facial expression, posture, color saturation, and others. It consists of 13,233 face images gathered from the internet. Each image has a resolution of 250 by 250 pixels.

4.2.3 Essex dataset

This dataset consists of 384 people (male and female), each with 20 samples. Most of the data are for students aged (18-20 years) and others for older individuals, and therefore some individuals with beards and wore glasses. Each image with resolution (196x196) pixels. Grimace and Faces 96 are more complex, increasing the complexity of the data (variance, appearance, taps, and background).

4.2.4 ORL Dataset

This database contains 400 images of faces from 40 distinct individuals, each with ten different images. Images were captured at various times and with variable lighting, various facial expressions (open eyes, closed eyes, without smiling, smiling), facial detail (with glasses, without glasses), and head-posing (rotating up to a maximum of 20°, tilting). Each image is (92-by-112) pixels in size.

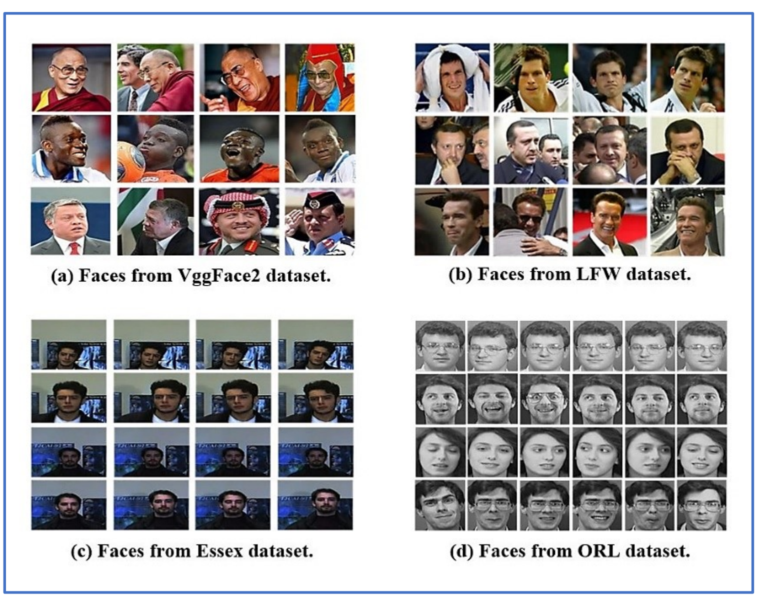

Figure (5): A sample of face images contained in used datasets.

4.3 Performance Evaluation

The performance of recognition is measured in terms of recognition accuracy. Accuracy is expressed as a percentage of correct labels divided by the total number of testing images. The accuracy is determined using the confusion matrix. Accuracy is calculated by adding the true positive and negative values and dividing them by the total number of samples. Its percentages are calculated by multiplying them by 100%.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{1}$$

In this equation, FP stands for false positives, FN stands for false negatives, TN stands for true negatives, and TP stands for true positives.

4.5 The Experimental Results

This part offers all the experiments and findings that evaluated the effectiveness of image-based face detection and recognition system using two strategies based on different standard databases.

The first strategy analyzed the results when image features were extracted using pre-trained GoogLeNet and pre-trained VggNet-16 separately, followed by the SVM classifier to classify the image features. The second strategy analyze the effectiveness of using a deep feature level combination between two features extracted from the CNN GoogLeNet and the CNN VggNet-16 and using the SVM classifier.

4.5.1 The Analysis Results for pre-trained GoogLeNet with SVM

This method combines the GoogLeNet Convolutional Neural Network model with SVM to achieve its results. The pre-processing stage included image enhancement using the CLAHE algorithm, detecting the face or faces in the image and aligning them using the MTCNN algorithm. Then a rescaling operation is performed to change each face image size that was detected and aligned, to 224-by-224 which is the input size for GoogLeNet.

We used 80% of the data for training and 20% for testing and using random sampling to prevent the results from biasing. We extracted features in this experiment from the global average pooling layer "Pool5". The Support Vector Machines SVM is used as a classifier to implement the task of classification following that. The outcomes of this approach are estimated using all of the databases mentioned earlier in Section (4-2). Figure (6) shows the recognition accuracy for all datasets using GoogLeNet with an SVM classifier.

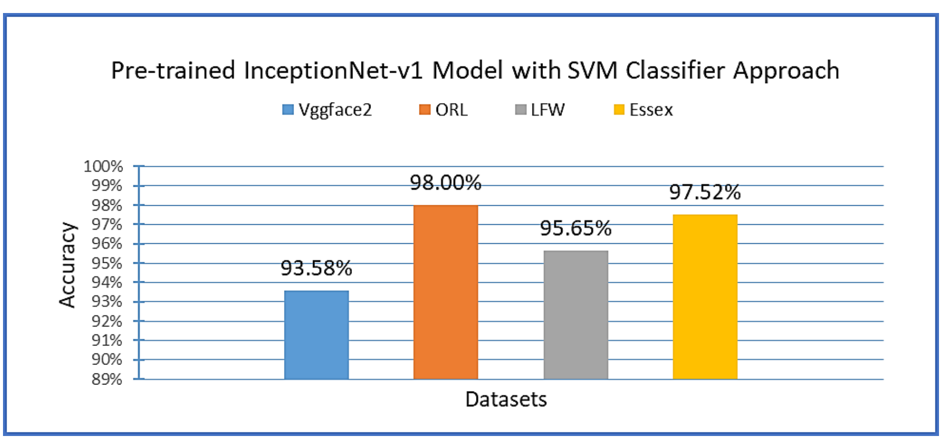

Figure (6): The recognition accuracy for approach-1.

According to the results of the ORL dataset, the GoogLeNet model achieved greater accuracy of recognition relative to the rest of the datasets of 98 percent. In addition, the GoogLeNet network obtained recognition accuracy of 97.52 percent on the Essex dataset, 95.65 percent on the LFW dataset, and 93.58 percent on the Vggface2 dataset.

4.5.2 The analysis of Results for pre-trained VggNet-16 with SVM

This method combines the VggNet-16 Convolutional Neural Network model with SVM to achieve its results. The pre-processing stage included image enhancement using the CLAHE algorithm, detecting the face or faces in the image and aligning them using the MTCNN algorithm. Then a rescaling operation is performed to change each face image size that was detected and aligned, to 224-by-224 which is the input size for VggNet-16.

We divided the data into train and test by 80% and 20% using random sampling to avoid biasing the results. We extracted features in this experiment from the fully connected layer "fc7". The SVM is used as a classifier to implement the task of classification following that. The outcomes of this approach are estimated using all of the databases that were mentioned earlier in section (4-2). Figure (7) shows the recognition accuracy for all datasets using VggNet-16 with an SVM classifier.

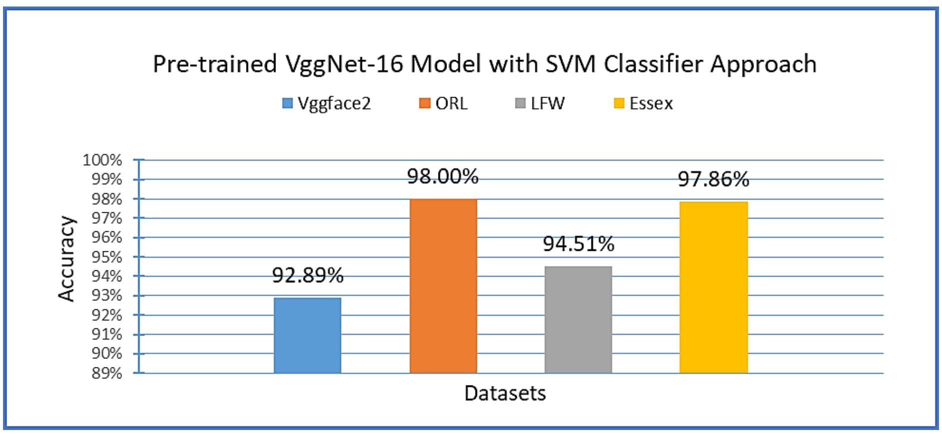

Figure (7): The recognition accuracy for approach-2.

The various experiments were performed in this approach and the obtained results are shown in Table (4-2). It's clear from the experiments carried out on all datasets that the VggNet-16 network obtained a greater recognition accuracy relative to the rest of datasets of 98 percent on the ORL dataset. In addition, the VggNet-16 network obtained recognition accuracy of 97.86 percent on the Essex dataset, 94.51 percent on the LFW dataset, and 92.89 percent on the Vggface2 dataset according to the results.

4.5.3 The Experimental results for Features combination with SVM

The experiments in this strategy examined the performance of feature-level combination between two vectors of deep features extracted from the GoogLeNet CNN model and the VggNet-16 CNN model to make comparisons of results with an individual approach when using both GoogLeNet and VggNet-16 each separately, as well as the performance of SVM for the task of classifications.

In these experiments, the preprocessing stage included image enhancement using the CLAHE method then detecting the face or faces in the image and then aligning them using the MTCNN algorithm then changing the size of the face images that were detected and aligned, to the appropriate size for input to both GoogLeNet network and VggNet-16 network, where the face images were resized to 224 by 224. Then the data were separated into two groups: training data (80%) and test data (20%) using random sampling .

The outcomes of this strategy when two deep models, GoogLeNet and VggNet-16, are combined at the feature level are shown in figure (8). This strategy achieved a maximum performance with a recognition accuracy of 99.29 % on the ORL and 99.18 % on the Essex Face datasets, respectively. Also, the feature combination approach obtained high recognition accuracy of 97.42 % on the LFW dataset and achieved recognition accuracy of 95.33 % on the VggFace2 dataset.

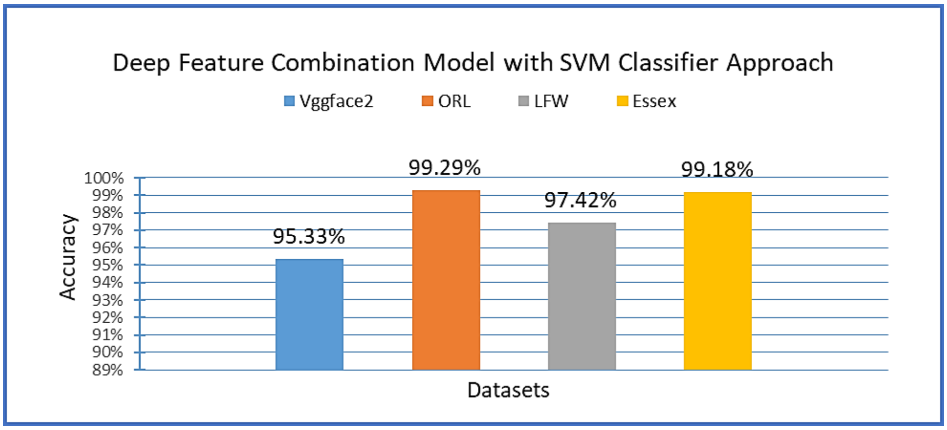

Figure (8): The recognition accuracy for strategy-2.

4.6 Comparison of Strategy-1 with Strategy-2

This section compares the results obtained when a single approach or a combination approach is used. The outcomes obtained when using a single approach or a combination of two models, GoogLeNet and VggNet-16, with all datasets are provided in Table (2). The results of all approaches GoogLeNet-SVM, VggNet-SVM, and combination-SVM with all datasets are displayed in Figure (10) according to each approach and Figure (9) according to each dataset.

Table (2): Comparing the single approach and the deep combination approach between GoogLeNet and VggNet-16.

CNN Model	Experiment performed on			
	Vggface2	ORL	LFW	Essex
GoogLeNet +SVM	93.58%	98%	95.65%	97.52%
VggNet-16 +SVM	92. 89%	98%	94.51%	97.86%
Feature Combination +SVM	95.33%	99.29%	97.42%	99.18%

We can observe that the highest recognition accuracy for all approaches with all datasets among all results is 99.29% and 99.18 achieved when using the deep feature combination approach on the ORL dataset and the Essex dataset, respectively. Also, we can note that the highest accuracy achieved by the individual GoogLeNet model is 98% on the ORL dataset. Also, the highest accuracy obtained for the individual VggNet-16 model was also 98% on the ORL dataset.

The deep feature combination approach achieved higher results in challenging datasets that contain face images taken under general conditions, such as the Vggface2 and LFW datasets, as compared to the individual VggNet-16 model and individual GoogLeNet model because the feature combination approach produces better results when dealing with large and complex datasets that contain challenges. Moreover, the feature combination approach outperforms GoogLeNet and VggNet-16 individual methods in all datasets.

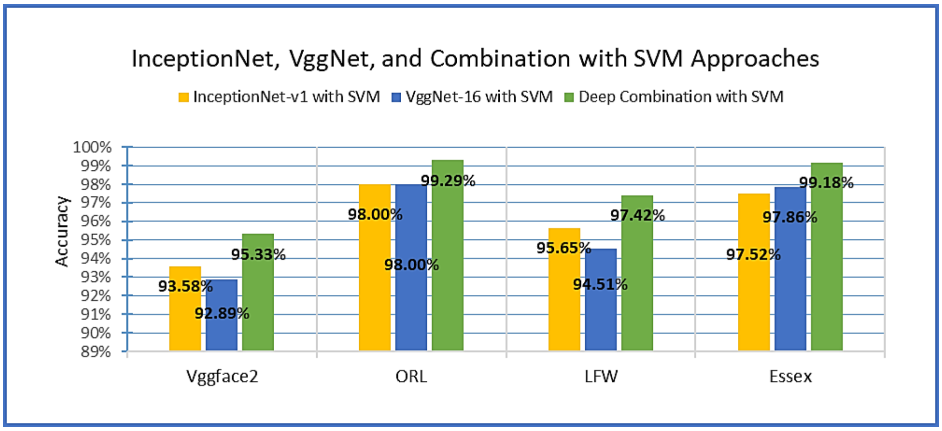

Figure (9): The mean recognition accuracy of all approaches according to each dataset.

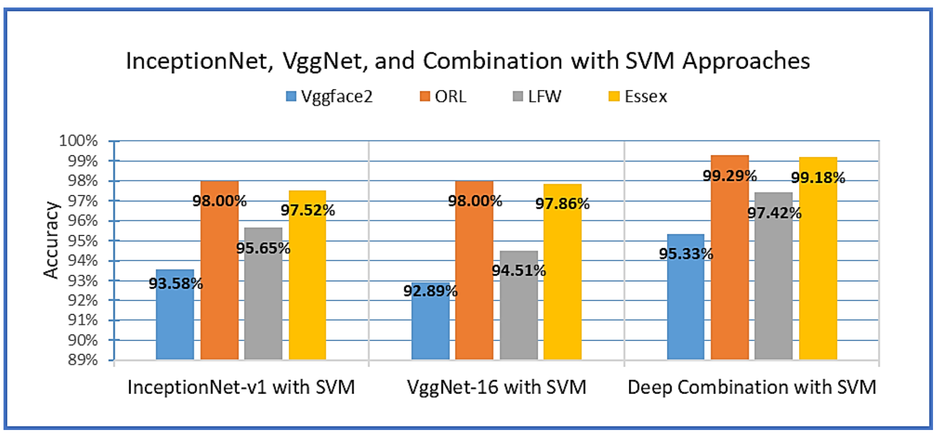

Figure (10): The mean recognition accuracy of all approaches according to each approach.

For the testing time, figure (4.17) and table (4.5) show the individual GoogLeNet model takes less time than the individual VggNet-16 model with all datasets. The feature-level combination approach between pretrained GoogLeNet and VggNet-16 CNN models, on the other hand, takes longer than the individual GoogLeNet and VggNet-16 models separately with all datasets.

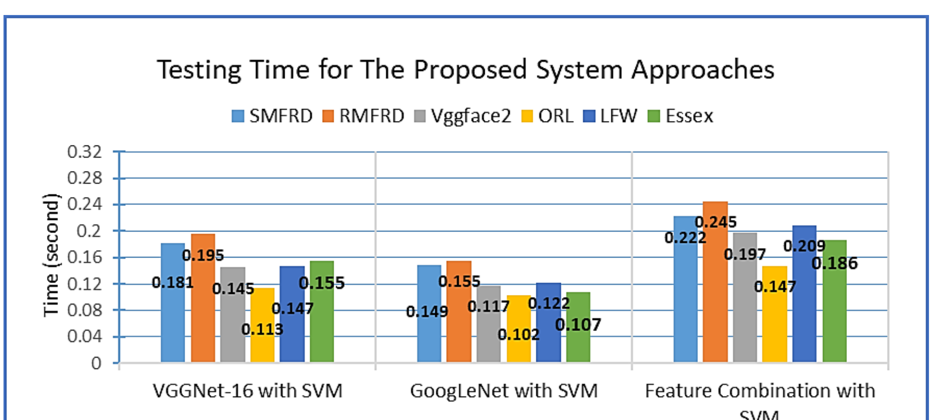

Figure (4.17): The testing time for the single approach and the combination approach.

Table (4.5): Comparison of the testing time for the single approach with the combination Approach.

CNN Model	Experiment performed on (Testing Time in second)					
	SMFRD	Vggface2	LFW	ORL	RMFRD	Face96
VGGNet-16 + SVM	0.141	0.175	0.161	0.155	0.166	0.153
GoogLeNet + SVM	0.112	0.107	0.121	0.113	0.137	0.125
Combination +SVM	0.223	0.212	0.199	0.201	0.241	0.192

4.7 Comparison with the other Models

This section compares the performance with different face recognition techniques, where this comparison is based on the LFW dataset and ORL datasets. This study aims to evaluate the face detection and recognition system using pre-trained convolutional neural networks with different datasets. Some of them are simple datasets, and others are complex datasets that contain challenges in unconstrained environments.

The recognition accuracy for the feature combination approach in this research is higher than others by using LFW and ORL datasets. Table (3) illustrates the proposed approach achievement with the other models.

Table (3): Comparing the proposed system and other models.

References	Model	Dataset	Recognition Accuracy
F. Tabassum et al. (2020) [8]	DWT+CNN	LFW	93.34 %
S. Bajpai et al. (2020) [9]	Inception-ResNet-v1 + LSA	LFW	95.41 %
Y. Yang et al. (2018) [10]	SR-CNN + The Random Forest	LFW	96.07 %
J. Li et al. (2018) [11]	Deep C2D-CNN	LFW	95.15 %
P. Kamencay et al. (2017) [12]	CNN- PCA- LBPH- KNN	ORL	98.30 %
S. Chen et al. (2016) [13]	CNN + SVM	ORL	97.50 %
The Proposed System (2021)	**Feature Combination + SVM**	**LFW**	**97.42 %**
		ORL	**99.29 %**

5. Conclusion

In this work, a face detection and recognition system based on deep learning is presented. Three different deep learning approaches were applied to investigate the pre-trained CNN architectures for face recognition. The hybrid approach to face detection and recognition is based on a combination of two deep feature extractor schemes, this method leads to a robust FR algorithm under challenges where this method achieves excellent recognition accuracy reaches 99.29%. Preprocessing of the dataset before feature extraction, such as face alignment and image enhancement, is very important. The accuracy of the proposed deep feature combination approach is better than the individual GoogLeNet approach and the individual VggNet-16 approach since the combined features contain richer and discriminant information than the input raw features. Features extraction using the pre-trained GoogLeNet model gives better performance than the pre-trained VggNet-16 model and this performance depends on the network architecture. Using the GoogLeNet model as a feature extractor gives a feature vector with 1024-dimensional, which reduces computational time complexity compared with the VggNet-16 model, which gives a feature vector with 4096-dimensional. The findings displayed that the proposed system outperformed most current models in terms of accuracy. We plan to increase recognition and classification accuracy in the future. To accomplish this, several data combinations will be performing such as score level and decision level fusion. Test other CNN models for better performance such as InceptionNet-v4, ResNet-101, ResNeXt, and DenseNet.

6. References

[1] N. Ortiz, R.D. Hernández, R. Jimenez, **"Survey of biometric pattern recognition via machine learning techniques,"** Contemp. Eng. Sci. 11(34), 1677–1694, 2018.

[2] K. Sundararajan, D.L.Woodard, **"Deep learning for biometrics : a survey,"** . ACMComput. Surv. 51(3), 2018.

[3] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, **"CloudID: Trustworthy cloud-based and cross enterprise biometric identification,"** Expert Systems with Applications. 42(21): p. 7905-7916, 2015.

[4] M.O. Oloyede, S. Member, G.P. Hancke**, "Unimodal and multimodal biometric sensing systems: a review, "** IEEE Access 4, 7532–7555,2016.

[5] M. Taskiran, N. Kahraman, and C. E. Erdem, **"Face recognition: Past, present and future (a review), "**Digit. Signal Process. A Rev. J., vol. 106, p. 102809, 2020.

[6] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, **"Face recognition systems: A survey,"** Sensors (Switzerland), vol. 20, no. 2, 2020.

[7] M. Chihaoui, A. Elkefi, W. Bellil, and C. Ben Amar, **"A survey of 2D face recognition techniques, "**Computers, vol. 5, no. 4, pp. 1–28, 2016.

[8] F. Tabassum, M. Imdadul Islam, R. Tasin Khan, and M. R. Amin, **"Human face recognition with combination of DWT and machine learning,"** J. King Saud Univ. - Comput. Inf. Sci., no. xxxx, 2020.

[9] S. Bajpai and G. Mishra, **"Real Time Face Recognition with limited training data: Feature Transfer Learning integrating CNN and Sparse Approximation,"** 2021.

[10] Y. X. Yang, C. Wen, K. Xie, F. Q. Wen, G. Q. Sheng, and X. G. Tang, **"Face recognition using the SR-CNN model,"** Sensors (Switzerland), vol. 18, no. 12, 2018.

[11] J. Li, T. Qiu, C. Wen, K. Xie, and F. Q. Wen, **"Robust face recognition using the deep C2D-CNN model based on decision-level fusion,"** Sensors (Switzerland), vol. 18, no. 7, pp. 1–27, 2018.

[12] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, **"A new method for face recognition using convolutional neural network,"** Adv. Electr. Electron. Eng., vol. 15, no. 4 Special Issue, pp. 663–672, 2017.

[13] S. Guo, S. Chen, and Y. Li, **"Face recognition based on convolutional neural network & support vector machine,"** 2016 IEEE Int. Conf. Inf. Autom. IEEE ICIA 2016, no. August, pp. 1787–1792, 2017.

[14] L. Alzubaidi et al., **"Review of deep learning: concepts, CNN architectures, challenges, applications, "**future directions, vol. 8, no. 1. Springer International Publishing, 2021.

[15] L. Deng and D. Yu, **"Deep Learning: Methods and Applications,"** Found. Trends® Signal Process., vol. 7, no. 3–4, pp. 197–387, 2014.

[16] P. Kamencay, M. Benco, T. Mizdos, and R. Radil, **"A new method for face recognition using convolutional neural network,"** Adv. Electr. Electron. Eng., vol. 15, no. 4 Special Issue, pp. 663–672, 2017.

[17] A. Simonyan, Karen and Zisserman, **"Very deep convolutional networks for large-scale image recognition,"** arXiv Prepr. arXiv1409.1556, 2014.

[18] C. Szegedy, W. Liu, Y. Jia et al., **"Going deeper with convolutions,"** In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[19] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, **"Vggface2: A dataset for recognizing faces across pose and age,"** in 2018 13th IEEE, pp. 67-74, 2018, [Online]. Available: https://www.robots.ox.ac.uk/~vgg/data /vgg_face2/. [Accessed: 23-Sept-2020].

[20] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, **"Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments,"** 2007, [Online]. Available: https://hal.inria.fr/inria-00321923. [Accessed: 18-Sept-2020].

[21] Libor Spacek's Facial Images Databases **"Face 96 Image Database,"** [Online]. Available: http://www.cmp.felk.cvut.cz/spacelib/faces/faces96.html. [Accessed: 22-Apr-2020].

[22] **"ORL face database**,**"** [Online]. Available: http://www.uk.research.att.com /facedatabase.html. [Accessed: 06-Apr-2020].

[23] S. Turky, A. Ahmed AL-Jumaili, and R. Hasoun, "Deep Learning Based On Different Methods For Text Summary: A Survey", *JQCM*, vol. 13, no. 1, pp. Comp Page 26-35, Mar. 2021.

[24] F. Tahir Al-azawi and A. Abdulrahman, "Face Detection By some Methods based on MATLAB", *JQCM*, vol. 12, no. 4, pp. Comp Page 12 -17, Nov. 2020.

[25] T. Hameed Obaida, "Face Recognition By Using Nearest Feature Midpoint Algorithm", *JQCM*, vol. 9, no. 1, pp. 144-152, Aug. 2017.

[26] Y. Mohammed and E. Saleh, "Investigating the Applicability of Logistic Regression and Artificial Neural Networks in Predicting Breast Cancer", *JQCM*, vol. 12, no. 2, pp. Math Page 63-73, Jul. 2020.

[27] N. Jarah, "Deep Learning In Wireless Sensor Network", *JQCM*, vol. 13, no. 1, pp. Comp Page 11 -17, Feb. 2021.