Performance Evaluation of Middle Block Convolutional Regularization Algorithm on CNN Architecture

Sanusi Abu and Fatma Susiluwati Mohd

June 22, 2022

Performance Evaluation of Middle block Convolutional Regularization Algorithm on CNN Architecture

Abu Sanusi Darma[1&2]
[1]University Sultan Zainal Abidin
Faculty of Informatics and Computing, Campus Besut,
22200, Terengganu, Malaysia.
[2]Al-Qalam University Katsina,
School of Natural & Applied Sciences, Department of Mathematical Sciences,
P.M.B. 2137
Katsina, Nigeria
darmasanusiabu@yahoo.com

Fatma Susilawati Binti Mohamad[1]
[1]University Sultan Zainal Abidin
Faculty of Informatics and Computing,
Fatma@unisza.edu.my

***Abstract*— The complexity of recognizing the human face by machines due to different variations in poses, illumination, age, facial expression, occlusion, personal appearance, and different cosmetic effects makes face recognition more challenging. However, this makes it difficult to implement a robust computational system. Our main goal is to enhance the current deep learning approaches for face recognition applications using an enhanced and efficient hybrid deep learning method that involves multi-layer CNN and SVM. The model is encompassed with a newly developed middle block convolutional regularization algorithm (MBCRA) for computational stability and convergence speed. The experimental results demonstrate that the multi-layer CNN+SVM has achieved 99.87% accuracy, and the comparative analysis shows that the proposed model is more resilient for face image classification under unconstrained settings.**

Keywords*— *Deep Learning, Face Recognition, Convolutional neural Network, and Support Vector Machine.

I. Introduction

The face recognition systems (FR) have progressed to become a frequently used form of biometric system, which is one of the most important and difficult areas of computer vision [1]. It is now the most active field of research demand, and it is an important technology in the fields of security, business applications, law enforcement, and many more. With more emphasis on its utilization for biometric analysis, human computing interaction (HCI), surveillance systems, and content-based coding of images [2]. Despite the availability of numerous presence biometric verification systems such as hand geometrics, iris scans, retinal scans, and fingerprints, human facial recognition applications will continue to be the most authenticating system due to the numerous features they possess. Features such as low cost, absence of physical contact between the user and the system, and user acceptance.

The system will continue to have huge significance in the area of security that provides intelligence services since the concern about proper security systems has reached its maximum point [3]. The face conveys information concerning the person's identity and state of emotion. Face recognition encompasses various biometric surveillance security systems such as (e.g., border control, suspect tracking, and terrorist identification) The technology is widely considered an important technology for digital verification and identification tasks [4].

A large amount of training and testing data is required for many CNN models used in face recognition systems. This increased network performance and led to a high recognition rate.

This paper introduces a new developed regularization method called MBCRA in a deep learning architecture for proper feature extraction and network stability. The deep learning architecture uses a hybrid model that involves a convolutional neural network and a support vector machine (CNN+SVM) to extract and classify facial features [5]. After the CNN has finished extracting features, the model employs the SVM as the final classifier to recognize the face because of its classification capability.

We used two publicly online available image databases: Label Faces in the Wild and the Caltech-101-ObjectCategories. We selected some of the face images from these two databases to reasonably build our solid dataset, which is suitable for the training and testing of the proposed model. The database was named AS_Darmaset, and the dataset has 5280 human face images of 135 individuals[6].

II. THE CONTRIBUTIONS OF THIS RESEARCH PAPER ARE AS FOLLOWS

A. This paper investigates the impacts of regularization techniques by adding Batch Normalization Layers, Dropout layers, and data augmentation techniques to improve CNN training, performance and to prevent network overfitting.

B. This is the first research work to incorporate MBCRA in the architecture of a multi-layer CNN+SVM to have the training and testing accuracies of 100% and 99.87%, respectively [7].

C. The study demonstrates that multi-layer CNN when combined with SVM on a large number of training and testing datasets is more robust than the basic standard CNN of a few layers of architectural models, which may effectively operate well with a small number of datasets and fewer network layers

III. REVIEW OF RELATED WORK

Due to numerous research studies on face recognition systems, this research paper tried to investigate various architectures for face recognition systems [12]. To develop and implement a robust face recognition system, many researchers in the computer vision community tend to think of new approaches based on deep learning techniques such as Convolutional Neural Network (CNN), Stock De-Noising Auto-Encoder (SDAE), Deep Belief Neural Network (DBN), etc. However, much of the current face recognition research has employed a different variant of the CNN architecture [13]. In this paper, we reviewed some of these exhibit choices to screen what is significant from relevant details, since CNN has been used as a powerful biometric system for solving numerous verification and identification problems. The deep learning model has been inspired by the biology of human vision [22].

Gonzalez et al. [23] presented the deep CNN in 2012 as a method to analyses the ImageNet database and get better results. Several CNN designs have been presented in the literature [24] and have been used to achieve better results than the current state of the art. In their research, Ahmad et al. [2] created a MATLAB-based CNN for face recognition. To minimize network training time, the proposed Convolutional Neural Network Method is capable of receiving new inputs by training the last two layers of the model. The experiment used face images of 40 people from the AT & T database and 10 people from the JAFTE database, and the result was 100 percent accuracy in less than a minute of training time.

Y. Li et al. [15] proposed a CNN-SVM hybrid technique to recognize human faces. The CNN is used for feature extraction, while the Support Vector Machine (SVM) will detect face images more effectively using the input of facial features extracted by the Convolutional Neural Network (CNN). The experiment results show that the model is more efficient, with a greater recognition rate and a shorter training time. Parkhi et al. [25] developed a facial recognition system employing a very deep CNN and corresponding training system to obtain higher face recognition accuracy when compared to the current trend on the public benchmark. This study evaluates how a large-scale database (of 2.6 million images) spanning over 2.6 thousand individuals could be developed by semi-automatic annotation with a human in the loop. Mohamed [26] aimed to increase the accuracy of a 2D face recognition system by learning discriminative features using a CNN of 15 layers. The network is trained using the stochastic gradient descent technique. Their research suggests that the Face96 database has a 99.6% accuracy rate.

Benkaddour et al. [24] proposed a deep learning-based hybrid algorithm for face identification and recognition. A convolutional neural network (CNN), a support vector machine (SVM), and principal component analysis are used in the technique. The CNN was employed for feature extraction, while the SVM is being used as a classifier. To minimize the dimensionality of facial traits, they applied the Principal Component Analysis (PCA) technique.

The findings of their experiment reveal that their proposed technique has resulted in a significant improvement in model recognition accuracy. Najm et al. [27] proposed integrating and exploiting CNN for human face recognition and categorization in a variety of applications. In their paper, they present a deep learning-based CNN approach with fuzzy logic to get a higher degree of exactness in the facial grin. The predictive feature of the human face can be exploited for a criminal investigation of the social analytics-based application using this method [16].

IV. MATERIALS AND METHODS

A. Methods

This research aims to achieve better network performances and human face recognition with a higher recognition rate, without system overfitting and with less computational load [14]. Image registration, powerful feature extraction, and classification models are the key points to use in achieving the above aims. To achieve the aims, we employed the use of a hybrid deep learning method that involved multilayer CNN+SVM [17]. In addition, a new middle block convolution regularization algorithm (MBCRA) was developed to enhance stabilized network training and processing speed [28]. The multilayer CNN is for feature extraction and classification. The SVM in this model is responsible for further feature extraction and final classification, utilizing the face characteristics extracted by the CNN. This hybrid system is capable of extracting more features than the standard CNN [20].

The SVM has several advantages in addressing high-dimensional pattern recognition and nonlinear classification applications. On the other hand, the CNN model can receive images as direct input, manage them for rotation, scaling, image distortion, and translation, and can automatically extract useful face characteristics [29]. The research also adopted the mini-batch stochastic gradient descent training technique to train the proposed models. A batch size of 400 was used for the training of the models. This means that 400 samples from the training dataset will be used to estimate the error gradient before the network weights are updated[18].

The strength of our model is based on adding the new MBCRA to the model architecture and a step of data augmentation technique after the normal pre-processing step, to generate and build a suitable dataset for the training and testing of the proposed model [19]. The database is derived from the two public benchmark databases, namely the Caltech 101_ObjectCategory database and the Label Faces in the Wild database.

B. System Model Design

Most of the very simple CNN structures developed by numerous researchers originate from LeNet-5, which contains a very simple topology of layers [2]. LeNet-5 encompasses seven (7) layers, all of which contain trainable parameters (weights), with an input size of 32x32 pixel images [30]. This research proposes to study and develop an enhanced deep learning approach for face recognition [21]. The research proposes to study the architectural design of a simple LeNet-5 convolutional neural network and to modify the hybrid CNN+SVM developed by [2] by adding more layers to their original 7-layer architecture. The structural topology of our multi-layer CNN+SVM will have multiple layers of architecture [32]. The improvement of the model structure is achieved by adding more layers and network features of MBCRA that ensure a high recognition rate while maintaining network training stability, performance, and processing speed[33].

C. The Topology of the Proposed Multilayers CNN

The Architectural Designed of the proposed multilayers Convolutional neural network consists of a single input layer (1) labeled as INPT-L, six (6) convolutional layers labeled as C1, C2, C3, C4, C5 and C6 respectively. The structure has four (4) Rectifiers Linear Unit (ReLu) layer label as R1 to R4, four (5) Batch normalization layers' label as BN1 to BN5, four (4) Max pooling layers MP1 to MP4, one (3) Dropout layer label as DO-L, two (2) fully connected layers, and SVM classification layer [34]. Each layer in this topology is a connected layer of linear mapping of a different number of face image data [31].

V. DATASETS

The data acquisition process is concerned with the methods used to obtain a large number of training and testing datasets for the proposed model. In this paper, we proposed collecting a sufficient number of datasets from two publicly available databases, namely the Caltech-101-Object-Category database and the Label Face in the Wild (LFW) Database (shown in figure 4). Caltech's-101-Object-Category database has 450 facial photos of 27 different persons. The photographs are 325 × 495 pixels in Jpeg format, with varying expressions, backgrounds, and lighting, but this study advocated cropping and reducing each face image to 64x64 pixels.

Fig 1. Shows the sample face images from the Caltech 101_ObjectCatogary database.

The LFW database, on the other hand, comprises 13,233 target face images of 5749 different people. There are 1680 people in the database who have two or more photos. The remaining 4069 people have only one image in the database [38]. Figure 5 illustrates the image samples of LFW.

The images are available in JPEG format with a resolution of 250 x 250 pixels. The majority of the images are in color, with only a few in grayscale. In this study, we recommended using 4,200 selected face images from this database. For the training and testing of the proposed models, all of these images are in the color scheme, resized to 64x64 pixels, and in their original jpg format.

Fig. 2. Shows the sample face images from Label Faces in the Wild in resize scale of (64x64) pixel.

In this study, we proposed using 5280 human face images of 132 different people. Each person has 40 different face images. In doing this, we used dynamic data augmentation and preprocessing techniques to generate many synthetic face images from the face images of each individual [35].

We created 1,080 face images of 27 people using the Caltech-101-objects-categories database and 4,200 images of 105 people using the LFW. This gives us a total of 5,280 face images. The 5280 images were divided into 5280/100 x 70 = 3,696, which is 70%, and 5280/100 x 30 = 1,584, which is 30%. for both training and testing purposes. The first 70% is for training, while the remaining 30% is for testing [36].

TABLE I. TABLE 1. DATABASE COMPARISON

DATASETS	PEOPLE	IMAGES
Caltech 101_objects_Categories	27	450
Label Faces in the Wild	5749	13233
AS_Darmaset for Proposed model	132	5280
Face96 Dataset for 9 Layers CNN & SVM by [15]	200	2200
FERET Dataset for 15 Layers CNN by [26]	152	3040

From the table I, the proposed AS_Darmaset database has the largest collection of 5280 face images outside the LFW.

VI. EXPERIMENTS

This section covers the experimental analysis done on the proposed deep learning model of CNN+SVM for face recognition systems. The experiment was carried out with the proposed AS Darma_set of 5280 face images of 135 people. This database was created by combining some of the face images selected from the two publicly available datasets. Caltech 101_ObjectCategory and Label Faces Databases in the Wild[37].

The experiment was carried out to determine the advantages of using a sophisticated regularization technique for deep learning algorithm to improve the performance of the face recognition system. We compared the accuracy and loss errors caused by the proposed model with that of three other deep learning approaches for face recognition system. To aid in the selection of the best and most powerful model in terms of performance and processing speeds. Furthermore, the study investigated the ability of the new developed regularization approach of MBCRA used in the proposed model architecture to increase network training stability, speed and performance.

VII. EXPERIMENT WITH HYBRID MODEL OF MULTI-LAYERS CNN+SVM COUPLE WITH MBCRA

We first used the proposed model for segmentation, feature extraction, and classification, all on a hybrid deep learning model [39]. MATLAB R2018b was used in developing and implementing the four deep learning models. It was also employed to carry out the experiments. Because it is the appropriate programming software for engineering and artificial intelligence systems.

The architecture of the proposed models is designed to run on the HP Elite Book 854w Mobile Workstation. The system has an Intel Core i7 M620 @ 2.67GHz processor with 8.00GB of inbuilt physical memory. Four experiments were carried out. The first experiment was for the proposed multilayer CNN+SVM, the second was for 15-layer CNN, the third was for 9-layer CNN with SVM, and the fourth was for 7-layer CNN. The experiments were conducted on the AS_Darmaset, which has six different classes of face image variations. Cosmetic effects, facial expression, occlusion, faces of older age, pose variation, and faces of younger age are six classes of face image variations. Each of these classes contains 880 face images. Each class comprises face images of 22 people, and each person has 40 face images of size 64x64 pixel. For this experiment, the proposed model's architecture was outfitted with a newly constructed MBCRA, a regularization method involving two convolutional layers, single batch normalization layer, and a P = 45% dropout. For proper network training, computational stability and convergence speed, the method was integrated with the pre-activation batch normalization method.

TABLE II. RESULT OF THE LABEL COUNT

Label Name	Label Count
Facial Expressions	880
Makeup	880
Occlusion	880
Old Age Faces	880
Pose	880
Younger Age Faces	880

The proposed model yielded the label count result shown in Table II. The model can figure out how many images are in each class. The label count is a table that contains the labels for each class as well as the number of images [8]. The model is capable of detecting, categorizing, and calculating the number of images in each of the six classes.

TABLE III. TRAINING PLOT DETAILS FOR THE PROPOSED CNN TRAINING FROM SCRATCH.

Parameters	Values
Trained Acc.	100.00%
Testing Accuracy	99.87%
Status	Completed
Processing Time	28 min 08 sec
Epoch	80
Iteration	880
Frequency	90

Table III. displays the data received from the experiment's training plot. It displays the training accuracy, validation accuracy, training loss, and validation loss in the figure 7 below.

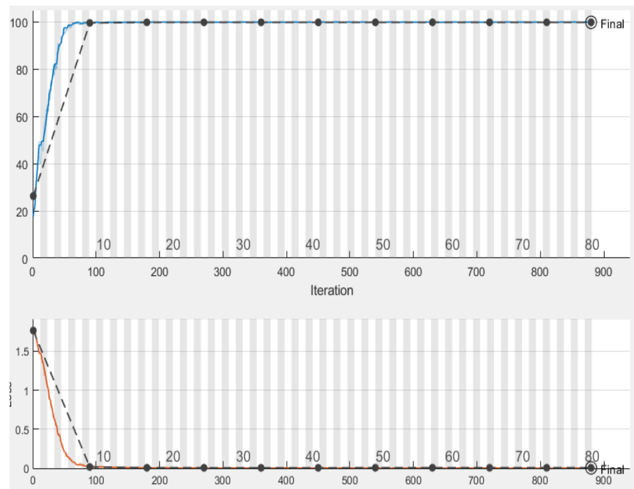

Fig. 3. Training progress plot graph of the proposed 22 layers CNN.

Fig. 3 is the training progress graph. The first graph at the top represents training accuracy (classification accuracy). The x-axis represents a scale from 0 to 80 epochs and iterations from 0 to 880, respectively, while the y-axis represents the accuracy values scale from 0% to 100%. The loss function (cross-entropy loss errors) is represented in the second graph at the bottom. The gradient is estimated using the iteration indicated in this graph. A light blue line and a dark blue line represent classification accuracy. This accuracy was obtained by implementing a smooth approach for training accuracy. The classification accuracy of the entire validation dataset is defined by the interrupted black dotted line. The number of iterations of each epoch is represented by the dotted black markings (points) on the black dashed line. We have set the model to 90 iterations per epoch on the graph.

It is clear that the model began to converge around epoch 9, iteration 90, with training accuracy nearing 100%. and validation accuracy reaching 99.62%. The network continued to converge successfully until epoch 28, iteration 300, where the training accuracy dropped to 99.75% and the validation accuracy increased to 99.87%. With the power of the batch normalization and dropout in the MBCRA and the dropout in the fifth convolutional unit, the training accuracy regained its convergence rate at epoch 32 and iteration 350 with 100.00% accuracy and the validation accuracy rate 99.87% was maintained. In the last epoch, 80 and iteration 880, the training and validation accuracy of 100.00% and 99.87%

remained unchanged. All this was achieved using the regularization operation performed by MBCRA in 28 minutes and 8 seconds of processing time. The loss function is presented at the bottom of the second graph. The light orange line indicates the training loss, whereas the dark dotted line represents the validation loss.

The broken line represents the loss on each mini-batch as well as the loss on the validation dataset [9]. The number of images used for the training and validation is 70% for the training that was selected randomly, while the remaining 30% is used for the validation (testing). The training and validation loss functions (the light orange and dark dotted lines) have converged to the minimum error at the last epoch of 80 and iteration 880 with a training loss value of 0.0020 and a validation loss value of 0.0075, respectively [10]. This shows that both the middle blocks convolutional regularization algorithm (MBCRA) and the pre-activation batch normalization algorithm (PABNA) used in the model architecture work well, since the proposed batch normalization, dropout, and ReLU activation function hybrid model consistently gains very low training and validation loss error rates and achieves immediate convergence speed.

A. *Model Performance Evolution Matrix*

Sometimes the deep learning models give out an accuracy of 90%, 91%, up to 99%, and so on, but this is not what we need to depend on from the models' given accuracy. Because sometimes that doesn't reflect the actual truth of the result. To appraise the true performance of the proposed hybrid multilayer CNN+SVM architecture on each of the six classes of face variation, this study adopted an evaluation metric evaluation measures of accuracy, precision, recall, and f1-score to evaluate and prove the good performance of the proposed model on the multi-class classification of six human face image versions. The model accuracy is achieved by dividing the correctly classified samples by all the samples.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

$$F_1\text{-Score} = \frac{2(\text{Precision})*(\text{Recall})}{\text{Precision}+\text{Recall}} \tag{4}$$

Where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively [11].

TABLE IV. PERFORMANCE EVALUATION RESULT FOR THE PROPOSED MULTILAYERS CNN+SVM BASED ON 6 CLASSES OF VARIATIONS

Deep Learning Models		**Performance Evaluation**			
	Classes of Variations	**Accuracy**	**Precision**	**Sensitivity (Recall)**	**F_1-Score**
Multylayer CNN+ SVM	Facial Expression	1.00	1.00	1.00	1.00
	Makeup	1.00	1.00	1.00	1.00
	Occolusion	1.00	1.00	1.00	1.00
	oldAge	0.9988	1.00	0.9923	0.9961
	Pose	1.00	1.00	1.00	1.00
	YongerAge	1.00	1.00	1.00	1.00
	Average	0.9998	1.00	0.9987	0.9994

Table IV shows the performance evaluating results of the proposed hybrid multilayer CNN+SVM based on each class of face image variations (uncontrolled conditions) that involve the classes of facial expiration, facial makeup, occlusion, old age, pose, and young age variations. The results were obtained using four measurement standards, including accuracy, precision, recall, and f1-score, respectively. From the table, it can be seen that the best value among all the measurements is precision. So the result shows higher precision and a higher recall. This implies that the proposed model avoids making excessive mistakes during classification, indicating that it makes reliable predictions and is robust. The evaluation table clearly shows that the suggested multilayer CNN+SVM has a robust feature extraction and classification capability, with an average accuracy of 0.9998, precision of 1.00, recall of 0.9987, and f1-score of 0.9994.

The f1-score shows that both precision and recall are in balance, as the f1-score has a value of 0.9994, which is close to 1. Generally, the proposed model has better performance in terms of all the measurement standards. Above all, the proposed model has good performance for face recognition applications under different uncontrolled conditions. This mean the proposed model shows resilience to all the six classes of face image variations that involves facial expressions, facial makeup, occlusion, age-related variation, and pose.

B. *Comparative Analysis on the Four Models Performance Evaluation Results*

In this section, the study explicates the details of the performance evaluation results obtained from the four deep learning architectures. To test and compare the performance of the proposed Multilayer CNN+SVM architecture with that of the other three deep learning models. The study still used the four performance evaluation metrics of accuracy, precision, recall, and F1-score. The loss error performance graph curve in figure 14 was also used to demonstrate the good performance of the proposed hybrid model.

$$\text{Loss} = \frac{1}{m} * \sum_{i=1}^{m} yi \; log f(xi) \qquad (5)$$

Where m is the value of training images xi and yi are the input and expected output respectively, $f(xi)$ denoted the real output.

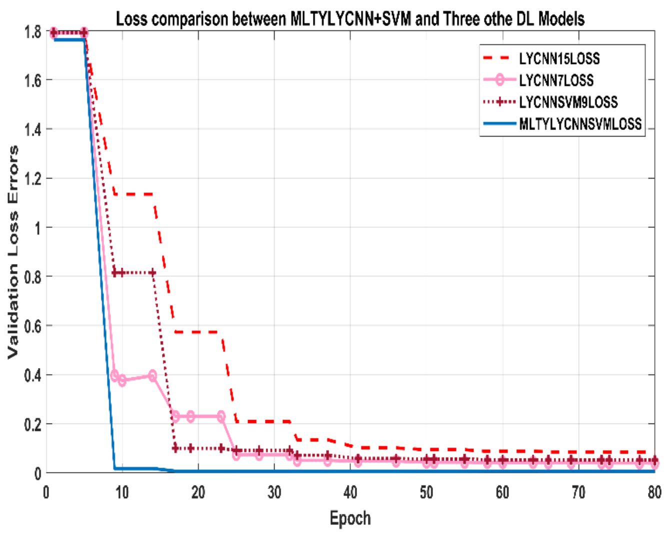

Fig. 4. Validation Loss Comparison between the proposed MLTYLY CNN and Three Deep Learning Model Architectures.

Fig. 4 shows the comparison line curve for validation loss errors between the proposed hybrid multilayer CNN (solid blue line) and three other deep learning architectures that involve 15-layer CNN (dash red line), 9-layer CNN+SVM (dotted marron line), and 7-layer CNN (solid bobble pink line). The proposed hybrid model architecture has the lowest and most stable loss error of the four model architectures. More specifically, the proposed model starts to be more stable at epoch 17 and lasts up to epoch 80. This means the model has an earlier convergence at epoch 17 with an error value of 0.0075, while it continues to be stable to the last epoch with an error value of 0.0075. These results show that the proposed model has the representational power to overcome the problem of gradient disappearance since the validation loss values have continued to have the same loss values around zero values right from the beginning.

This indicates that the regularization effect of both the MBCRA and PABNM is strong enough to prevent the proposed model from overfitting. While the other three models still exhibit dramatic fluctuations up to the 80th epoch, more specifically, the 15-layer CNN has a higher error value of 0.0872. This model has an overlapping loss error; it conjectures that the model has nowhere near the representational power to classify any of the face images in the six classes of the dataset correctly compared to the other three models due to its overlapping error. It indicated that the model is not stable and has a higher overfitting problem. Generally, it can be stated that the proposed hybrid model has good performance in classifying all the six classes of face variations.

TABLE V. COMPARATIVE ANALYSIS OF FACE RECOGNITION ABILITY OF DIFFERENT DEEP LEARNING MODELS FOR OVERALL PERFORMANCES

	Performance Evaluation				
Deep Learning Models	**Average Accuracy**	**Average Precision**	**Average Recall**	**Average F_1-Score**	**Processing Time**
Proposed MLYCNN+ SVM	0.9998	1.00	0.9987	0.9994	28: 08
15-LYCNN [17].	0.9964	0.9868	0.9870	0.9874	28:55 sec
9-LYCNN (Guo et al, 2017).	0.9957	0.9872	0.9873	0.8218	30:49 sec
7-LYCNN [19].	0.9867	0.9641	0.9809	0.9717	21:29 sec

The performance evaluation results for each of the four deep learning models are shown in Table V. From the table, it can be seen that the best value among all the measurements is the average precision of the proposed hybrid multilayer CNN+SVM, with a value of 1.00. This signified that the model is avoided making a lot of mistakes during classification, so it indicates that the proposed model is making correct predictions and is robust. In this section,

the evaluation table gives a clear indication that the proposed multilayer CNN+SVM has a robust feature extraction and classification capability when compared with the other three models. It can be seen that the proposed model has a high average accuracy of 0.9998, average precision of 1.00, average recall of

0.9987, and an average f1-score value of 0.9994 The f1-score indicates that precision and recall are balanced since it has a value of 0.9994, which is close to 1.

The lowest performance result, on the other hand, goes to the 7-layer CNN, which has an average accuracy of 0.9867, an average precision of 0.9641, a recall of 0.9809, and an f1-score of 0.9717. Overall, the proposed model performs better across all measurement standards. Furthermore, the proposed model performs well in face recognition applications under a variety of uncontrolled conditions, most notably in the categories of facial expressions, cosmetics, occlusion, old age, posture, and younger age. From the comparative analysis table above, it can be noticed that the proposed MLYCNN+SVM has a training completion time of 28 min 08 seconds, which is 0.47 sends faster than the 15-layer CNN and 2 min 41 seconds faster than the 9-layer CNN. Though the proposed model is 6 min 76 seconds slower than the 7-layer CNN, which has a training completion time of 21 min 29 seconds. When considering the number of their internal connected layers, the completion time of the proposed model is almost the same as that of the 7 layers of CNN.

VIII. CONCLUSION

This research presents an improved and efficient framework for human face recognition applications based on a hybrid deep learning technique involving multilayer convolutional neural networks (CNN) and support vector machines (SVM). The research analyzes the effects of regularization strategies by incorporating a new MBCRA into the proposed model's design.

The study shows that multi-layer CNN+SVM is more effective when using a batch normalization layer in between a convolutional layer and a dropout, with a 45% probability of dropout in the middle convolutional unit. This is important when using large training and testing datasets.

The performance evaluation result obtained from each of the four models shows that the proposed model is more robust for face image classification under all the six unconstrained conditions when compared with the 15-layer CNN, which has a problem in classifying face images under occlusion with an accuracy of 0.9950 and young age variation with an accuracy of 0.9940. When compared with 7-layer CNN and 9-layer CNN, these two deep learning models have the problem of classification under facial expressions, older age, and young age variations. Deep learning methods with multiple convolutional connections can extract more complicated facial characteristics.

Acknowledgement

We would like to appreciate and applaud the efforts of the University Sultan Zainal Abidin, particularly the Faculty of Informatics and Computing, as well as the UNISA Project Management Centre (CRIEM), for academic and financial support of this research. We also thank the management of Al-Qalam University Katsina, Katsina State, Nigeria. For providing the fellowship support to the Corresponding Author.

References

[1] A. Vinay, V. S. Shekhar, K. N. B. Murthy, and S. Natarajan, "Face Recognition Using Gabor Wavelet Features with PCA and KPCA - A Comparative Study," *Procedia Comput. Sci.*, vol. 57, pp. 650–659, 2015, DOI: 10.1016/j.procs.2015.07.434.

[2] S. Ahmad Radzi, "A MATLAB-based Convolutional Neural Network Approach for Face Recognition System," *J. Bioinforma. Proteomics Rev.*, vol. 2, no. 1, pp. 1–5, 2016, DOI: 10.15436/2381-0793.16.009.

[3] M. Chihaoui, W. Bellil, A. Elkefi, and C. Ben Amar, "Face Recognition Using HMM-LBP Face Recognition Using HMM-LBP," no. May 2018, 2016, DOI: 10.1007/978-3-319-27221-4.

[4] M. Bennamoun, Y. Guo, and F. Sohel, "Feature Selection for 2D and 3D Face Recognition," *Wiley Encycl. Electr. Electron. Eng.*, no. January, pp. 1–28, 2015, DOI: 10.1002/047134608x.w8257.

[5] L. Chen, Y. H. Wang, Y. D. Wang, and D. Huang, "Face recognition with statistical local binary patterns," *Proc. 2009 Int. Conf. Mach. Learn. Cybern.*, vol. 4, no. February, pp. 2433–2439, 2009, DOI: 10.1109/ICMLC.2009.5212189.

[6] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," no. Section 3, pp. 41.1-41.12, 2015, DOI: 10.5244/c.29.41.

[7] and C. L. Z. Tsung-Yi Lin1, Michael Maire2, Serge Belongie1, James Hays3, Pietro Perona2, Deva Ramanan4, Piotr Doll´ar5, "Microsoft COCO," *Eccv*, no. June, pp. 740–755, 2014.

[8] A. J. Champandard, "FaceNet: A Unified Embedding for Face Recognition and Clustering http://arxiv.org/abs/1503.03832 (Google Research) dlearn pic.twitter.com/fFotqHa1HC," *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, pp. 815–823, 2015, 2.

[9] S. D. Abu and F. S. Mohamad, "Approaches Of Deep Learning In Persuading The Contemporary Society For The Adoption Of New Trend Of AI Systems: A Review,"

Researchgate.Net, vol. 9, no. 12, pp. 163–177, 2020

[10] Z. Sufyanu, U. Sultan, Z. Abidin, F. S. Mohamad, and A. A. Yusuf, "Feature extraction methods for face recognition," no. February 2017, 2016.

[11] T. Satepuri and P. C. Reddy, "A Survey on Facial Expression Recognition Techniques," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 5, pp. 980–984, 2019, doi: 10.26438/ijcse/v7i5.980984.

[12] T. Zia *et al.*, "Face recognition with Bayesian convolutional networks for robust surveillance systems," *EURASIP J. Image Video Process.*, 2019, doi: 10.1186/s13640-019-0406-y.

[13] A. Muniasamy, A. Alasiry, and S. Arabia, "Deep Learning : The Impact on Future eLearning," pp. 188–199.

[14] A. M. TURING, "MIND," vol. VOL. LIX., pp. 433–460, 1950.

[15] S. Guo, S. Chen, and Y. Li, "Face Recognition Based on Convolutional Neural Network and Support Vector Machine *," no. August 2016, DOI: 10.1109/ICInfA.2016.7832107.

[16] F. Mahmud, S. Afroge, M. Al Mamun, and A. Matin, "PCA and back-propagation neural network-based face recognition system," *2015 18th Int. Conf. Comput. Inf. Technol. ICCIT 2015*, no. July 2017, pp. 582–587, 2016, DOI: 10.1109/ICCITechn.2015.7488138.

[17] M. Abuzneid, "IMPROVING HUMAN FACE RECOGNITION USING DEEP LEARNING BASED IMAGE REGISTRATION AND MULTI-CLASSIFIER APPROACHES," 2018.

[18] P. Duan, G. Mao, C. Zhang, and B. Zhang, "Applying DCOP to user association problem in heterogeneous networks with Markov chain based algorithm," *arXiv*, no. 6, 2017.

[19] M. A. Abuzneid, A. Mahmood, and S. Member, "Enhanced Human Face Recognition Using LBPH Descriptor, Multi-KNN, and Back-Propagation Neural Network," *IEEE Access*, vol. 6, pp. 20641–20651, 2018, DOI: 10.1109/ACCESS.2018.2825310.

[20] L. Guo, Y. Liao, D. Luo, and H. Liao, "Face image classification using appearance and texture features," *ICCASM 2010 - 2010 Int. Conf. Comput. Appl. Syst. Model. Proc.*, vol. 3, no. Occam, pp. 476–480, 2010, doi: 10.1109/ICCASM.2010.5620850.

[21] M. C. Fysh and M. Bindemann, "Human-Computer Interaction in Face Matching," *A Multidiscip. Jounal*, no. 1551–6709, pp. 1–19, 2018, doi: 10.1111/cogs.12633.

[22] P. H. LeCun Yann, Le'on Bottou, Yoshua Bengio, "A

B7CEDGF HIB7PRQTSUDGQICWVYX HIB edCdSISIXvg5r ` CdQTw XvefCdS."
[23] T. F. Gonzalez, "Handbook of approximation algorithms and metaheuristics," *Handb. Approx. Algorithms Metaheuristics*, pp. 1–1432, 2017, doi: 10.1201/9781420010749.
[24] M. K. Benkaddour and A. Bounoua, "Feature extraction and classification using deep convolutional neural networks, PCA and SVC for face recognition," *Trait. du Signal*, vol. 34, no. 1–2, pp. 77–91, 2017, doi: 10.3166/TS.34.77-91.
[25] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition - Abstract only," *Procedings Br. Mach. Vis. Conf. 2015*, no. Section 3, pp. 41.1-41.12, 2015, [Online]. Available: http://www.bmva.org/bmvc/2015/papers/paper041/index.html.
[26] S. S. Mohamed, W. A. Mohamed, A. T. Khalil, and A. S. Mohra, "Deep Learning Face Detection and Recognition," *Intl J. Electron. Telecommun.*, no. June, pp. 1–7, 2019.
[27] O. A. H. Hayder Najm, Hayder Ansaf, "A N E FFECTIVE I MPLEMENTATION OF F ACE R ECOGNITION U SING," 2019.
[28] U. Bakshi and R. Singhal, "A SURVEY ON FACE DETECTION METHODS AND FEATURE EXTRACTION TECHNIQUES OF FACE RECOGNITION," vol. 3, no. 3, pp. 233–237, 2014.
[29] M. A. Otair and W. A. Salameh, "Efficient training of backpropagation neural networks," *Neural Netw. World*, vol. 16, no. 4, pp. 291–311, 2015.
[30] H. M. Eraqi, M. N. Moustafa, and J. Honer, "End-to-End Deep Learning for Steering Autonomous Vehicles Considering Temporal Dependencies," no. Nips, pp. 1–8, 2017.
[31] A. R. Syafeeza, M. Khalil-Hani, S. S. Liew, and R. Bakhteri, "Convolutional neural network for face recognition with pose and illumination variation," *Int. J. Eng. Technol.*, vol. 6, no. 1, pp. 44–57, 2014.
[32] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, "Taigman_DeepFace_Closing_the_2014_CVPR_paper," doi: 10.1109/CVPR.2014.220.
[33] V. Sze, Y. H. Chen, T. J. Yang, and J. S. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," *Proc. IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017, doi: 10.1109/JPROC.2017.2761740.
[34] W. Wang and Y. Yang, "Development of convolutional neural network and its application in image classification: a survey," *Opt. Eng.*, vol. 58, no. 04, p. 1, 2019, doi: 10.1117/1.oe.58.4.040901.
[35] M. A. Ponti, L. S. F. Ribeiro, T. S. Nazare, T. Bui, and J. Collomosse, "Everything You Wanted to Know about Deep Learning for Computer Vision but Were Afraid to Ask," *Proc. - 2017 30th SIBGRAPI Conf. Graph. Patterns Images Tutorials SIBGRAPI-T 2017*, vol. 2018-Janua, no. October, pp. 17–41, 2017, doi: 10.1109/SIBGRAPI-T.2017.12.
[36] J. Wolfe, X. Jin, T. Bahr, and N. Holzer, "Application of softmax regression and its validation for spectral-based land cover mapping," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 42, no. 1W1, pp. 455–459, 2017, doi: 10.5194/isprs-archives-XLII-1-W1-455-2017.
[37] S. H. S. Basha, S. R. Dubey, V. Pulabaigari, and S. Mukherjee, "Impact of fully connected layers on performance of convolutional neural networks for image classification," *Neurocomputing*, vol. 378, no. October, pp. 112–119, 2020, doi: 10.1016/j.neucom.2019.10.008.
[38] 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," pp. 1–11.
[39] R. D. Safiyah, Z. A. Rahim, S. Syafiq, Z. Ibrahim, and N. Sabri, "Performance Evaluation for Vision-Based Vehicle Classification Using Convolutional Neural Network," vol. 7, pp. 86–90, 2018.