FACE VERIFICATION ACROSS AGE PROGRESSION USING ENHANCED CONVOLUTION NEURAL NETWORK

Areeg Mohammed Osman[1] and Serestina Viriri[2]

[1]Faculty of Science and Technology, Sudan University of Science and Technology, Khartoum, Sudan
[2]School of Maths, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

ABSTRACT

This paper proposes a deep learning method for facial verification of aging subjects. Facial aging is a texture and shape variations that affect the human face as time progresses. Accordingly, there is a demand to develop robust methods to verify facial images when they age. In this paper, a deep learning method based on GoogLeNet pre-trained convolution network fused with Histogram Orientation Gradient (HOG) and Local Binary Pattern (LBP) feature descriptors have been applied for feature extraction and classification. The experiments are based on the facial images collected from MORPH and FG-Net benchmarked datasets. Euclidean distance has been used to measure the similarity between pairs of feature vectors with the age gap. Experiments results show an improvement in the validation accuracy conducted on the FG-NET database, which it reached 100%, while with MORPH database the validation accuracy is 99.8%. The proposed method has better performance and higher accuracy than current state-of-the-art methods.

KEYWORDS

Facial Aging, Face verification, GoogLeNet

1. INTRODUCTION

Facial aging defined from a computer vision perspective as a function of changing facial shape and texture over time [3]. Since childhood aging affects the human face in several aspects. However, creating robust face recognition systems is a challenge, especially when facial variations, such as different levels of illumination, poses, and facial expressions, are present in images.

Facial aging is a sophisticated process that affects the shape and texture of the human face [5], and such changes degrade the performance of automatic face verification systems. The challenge of these systems has based on the fact that facial aging involves both essential and unessential factors [30]. Facial aging also influences individual facial components (such as the mouth, eyes, and nose). Most problems in image recognition relate to identifying invariant facial features. This study examines the issue of designing a model and an appropriate schema capable of identifying an individual's features as they change due to aging and improving the performance by increasing the accuracy of face verification on the aging system. Besides, this study aims to determine whether the proposed method performs better in comparison to other methods and if it can identify individuals through images as they age Changes in facial appearance due to aging typically depend on several factors, including race; geographical location; eating habits; and

DOI: 10.5121/sipij.2020.11504

stress level, which makes the problem of recognizing faces across the aging process extremely difficult. In addition, no simple statistical model for analyzing appearance changes due to aging.

Faces affected by gradual variations due to aging may necessitate the periodical updating of facial image databases with more recent subject images for the success of facial verification systems. Since updating large databases periodically would be a difficult task, a better alternative would be to develop facial verification systems that can verify the identity of subjects from two face images of different ages. Understanding how age progression affects the similarity between two facial images of one subject is significant in such a task.

Facial recognition has categorized into two classes [5]: face identification and face verification. The former aims to recognize an individual from a gallery of facial images or videos and find the most similar one to the probe sample, while the latter identifies whether or not a given pair of facial images belong to the same subject.

This paper will consider how a subject could be recognized despite age changes over the years and other significant variations caused by lighting, expressions, poses, resolutions, and backgrounds. Face verification in aging subjects is a challenging process, as human aging is non-uniform. Besides, extracting textural and shape features from the images is another challenge.
Several methods have been used to extract facial features from images, hand-crafted descriptors have been used for a while to extract facial features [3], texture descriptor is suitable to extract general appearance facial changes, while shape descriptors are more suitable for face shape changes.

Some researchers study the effect of Local Binary Pattern (LBP) features [6,7] in face verification have achieved significant improvements. On the other hand, Histograms of Oriented Gradient (HOG) is a shape descriptor used to detect objects like cars and humans, was chosen for its advanced results in facial recognition [3].

Hand-crafted descriptors are not completely capable to represent the appearance of face [5], so we take deep learning methods into account. The primary advantage of deep neural network is to learn discriminative features through autonomous learning without supervision [4]. Thus, combine both hand crafted feature descriptors and convolutional neural network to take both sides advantages.

Convolutional Neural Network (CNN) is the method used in this paper that reveals significant results in age estimation, facial recognition, and object recognition in general. The architecture of CNNs consists of multiple layers, and each layer is responsible for performing a specific process based on the output of the previous layers. CNN's were considered deep networks if it has a large number of layers, a large database needed to optimize its parameters during the training process.

This paper consists of five sections: section 1 deals with the introduction; section 2 deals with previous studies on deep learning and the methods involved in the proposed methods; Section 3 describes the experiments carried out over the course of the study; section 4 summarizes the results; section 5 provides the conclusion.

2. Literature Review

Over the last few years, attention on the use of deep learning in computer vision and image processing has increased. The literature has shown several deep learning methods that have used for face verification system [9,11,19].

Generally, deep learning methods aim to learn hierarchical feature representations by building high-level features from low-level ones. There are three categories of deep learning methods: unsupervised, supervised, and semi-supervised. These methods have successfully applied in many visual analysis applications, such as object recognition [19] and human action recognition [19]. Recently, deep convolutional neural networks (DCNNs) have made significant improvements in object classification [9], facial analysis [19], and image super-resolution [23].
A study by Simone [1] on a large age-gap face verification task implemented a DCNN by including a feature injection layer to increase verification accuracy through learning a similarity measure of the external features. This method was evaluated according to the LAG (Large Age Gap) dataset and found to perform better than current state-of-the-art products.

El Khiyari and Wechsler [5] evaluated the use of CNN's in feature extraction for the automatic facial verification of subjects belonging to various age, ethnicity, and gender categories. For multiple demographic groups, biometric performance in facial verification was relatively lower in black female subjects 18-30 years old. Later, the VGG-Face convolutional neural network [4] was used to extract features by activation layers. The features distance between subjects and the similarity distances between their respective sets found to be the same. Its identification and verification performance was evaluated using both singleton and set similarity distances. On the other hand, Yanhai et al. [6] proposed an unsupervised learning model called PCN (PCA-Based Convolutional Network) consisting of two feature extraction steps and output steps. However, the feature extraction stage contains a convolutional layer that learns filters using Principal Component Analysis (PCA), and the output feature maps reduce images resolution. The nonlinear stage includes binary hashing and histogram statistics, and the output of all stages fed into a Support Vector Machine (SVM) classifier.

Digit recognition experiments were carried out based on the MNIST database, face recognition experiments on the Extended Yale Face database B, texture classification experiments on the CUReT database, and texture classification experiments on the Outex dataset. The results show that PCN performs competitively with and sometimes even better than current state-of-the-art deep learning models.

The proposed idea by Zhai et al. [25] was to combine both local binary pattern (LBP) histograms and 9-layer deep convolutional neural networks. This study confirmed that this fusion approach is performed better than current state-of-the-art methods. Besides, the approach provides hairstyle and facial expression features using models trained on the CACD and LFW datasets.

Hu et al. [7] proposed a discriminative deep metric learning (DDML) method for face verification in the wild by building a DDML neural network to perform nonlinear transformations of features such that the distance between two pairs of images belonging to the same person is less than a calculated small threshold value and vice versa. Their experiments on the LFW and YouTube Faces (YTF) datasets performed better than literature methods.

Finally, Moschoglou et al. [15] used the VGG-Face deep network and other state-of-the-art algorithms on a new manually compiled dataset called the AgeDB (Age-Database). This dataset is suitable for use with experiments on age-invariant recognition, face verification, age estimation, and facial age progression in the wild.

3. METHODOLOGY

This paper proposes a methodology for performing facial verification in the context of age progression using images of human faces. Figure 1 shows a flow diagram of the proposed methodology, which has divided into multiple steps. In the first step, facial images pre-processed through scaling, data augmentation, cropping, and normalization before training and testing. The second step feature extraction it's responsible for generating the CNN model based on the augmented dataset for training and testing to calculate the validation accuracy. The third step is to save the generated classifier to be used in the following step. Finally, in the prediction step, two pairs of images are used as test input images after being pre-processed, and the trained model and classifier then determine whether or not they belong to the same subject.

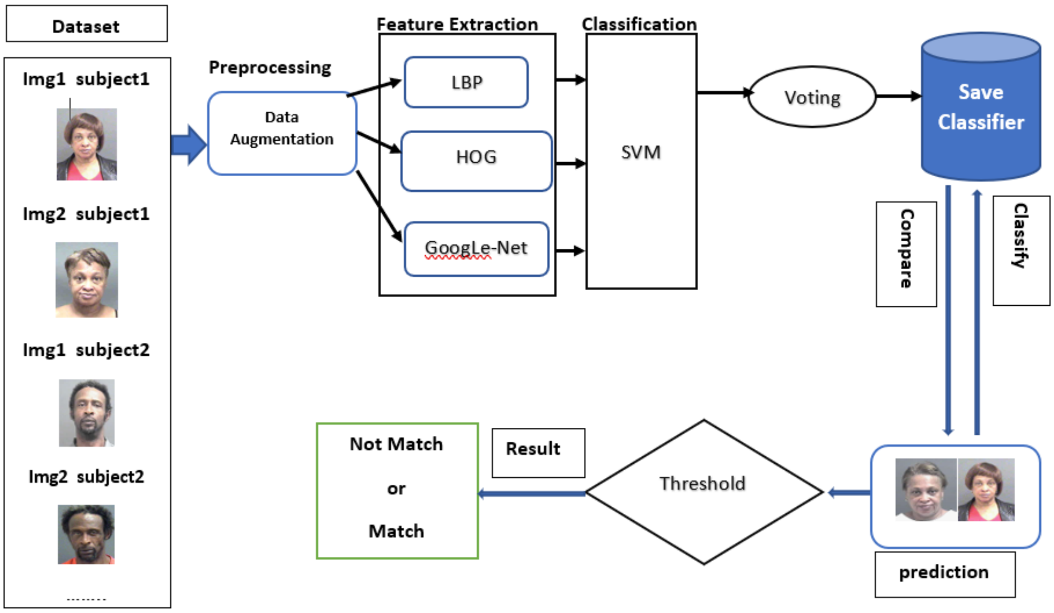


Figure 1. Proposed Methodology Framework

3.1. Facial Dataset

In this study, the FG-NET [17] and MORPH databases [20] were used to train and test the proposed model. FG-NET dataset is a standard benchmark, and a freely available dataset for facial recognition consists of 1002 images of 82 subjects ages between 0–6. MORPH dataset comprises 2798 images of 672 subjects that vary in age. The images were split into classes with a five years maximum of an age difference. Furthermore, the image datasets have divided into two groups: 80% of the set was randomly selected to train the CNN network, and the remaining 20% has used to test it.

3.2. Image Pre-processing

MORPH and FG-NET databases have pre-processed to improve the performance of the CNN model. Facial images were reflected and scaled to the standard input layer size of 224 ×224 and fed to the convolutional neural network using RGB color values to match the image input layer which requires input images 224× 224 ×3 in size, where 3 is the number of color channels.

3.3. Feature Extraction and Classification

Convolutional neural networks [14,12,18] are artificial neural networks that include both fully connected and connected layers known as convolutional layers. Other types of layers, such as pooling, activation, and normalization (rectified linear units) layers, are often observed in deep convolutional networks. CNN's have recently been more successful in both object classification [24] and automatic recognition than in handcrafted feature extraction.

3.3.1. Deep Convolutional Neural Networks and Transfer Learning

Training DCNNs from scratch is difficult, as it can require extensive computational resources and large amounts of training data. If these resources are not available, one can use a pre-trained network as a feature extractor and a classifier. The general architecture consists of a convolutional layer and a pooling layer followed by a fully connected layer [12].

In this work, the chosen architecture is the GoogLeNet model was proposed by [22], which is a convolutional neural network that was trained on images from ImageNet dataset [10]. The network is 22 layers deep and can classify images into 1000 object categories, including keyboard, mouse, pencil, and many animal classifications.

GoogLeNet model has almost 12 times fewer parameters (less than AlexNet model), as it reduces the number of parameters to 4 million [22] because it based on small convolutions. As a result, it has learned rich feature representations for a wide range of images, so it has become much more accurate. The network used a CNN inspired through the inception modules, the module range calculated then the fully connected layers has removed. Meanwhile, in the inception modules, there is a pooling layer to minimize the number of parameters involved. A shadow network and an auxiliary classifier added to facilitate better outputs. GoogLeNet has more layers (than AlexNet) due to its 9 inception modules which include the convolutional, pooling and softmax layers [22] in addition to concatenate processes

The idea behind the inception layer is to cover a larger area in the images while maintaining a precise resolution for small image information. The goal is to convolve different sizes in parallel from the most accurate detailing (1x1) to a larger one (5x5).

The most straight forward way to improve deep learning performance is to use more layers in the network and more data for training.

3.3.2. Histograms of Oriented Gradient (HOG)

HOG is a shape descriptor used to detect objects like cars and humans. It is firstly introduced by Dalal and Triggs to detect human [3]. The basic idea about HOG, the shape of objects, and appearance inside the image could be defined by the distribution of intensity gradients or edge directions. The image is dividing into cells, for each cell create a histogram to describe the distribution of the directions. Histograms are normalized and concatenating into a vector, which will be as large as the number of features and calculated as follows:

1. Gradient has to be computed by this equation [3].

$$g_x((X,Y) = I(X+1,Y) - I(X-1,Y) \qquad (1)$$

$$g_y((X,Y) = I(X,Y+1) - I(X,Y-1) \qquad (2)$$

2. Then Orientation θ and magnitude   are calculated as in the following formula.

$$m((X,Y) = \sqrt{\partial x(x,y)^2 + \partial y(x,y)^2} \qquad (3)$$

$$\theta(\mathrm{x},\mathrm{y}) = \mathrm{arctan}\frac{\partial y(x,y)}{\partial x(x,y)} \qquad (4)$$

3. Divide image orientation and magnitude into cells such that the number of cells in row and column is parameters to be chosen when implements HOG.

4. Orientations histogram computed for each block; then normalized by the formula below:

$$Hist_{norm}=\frac{Hist}{Hist+\in} \qquad (5)$$

5. Finally, Concatenated normalized histograms into a vector.

3.3.3. Local Binary Pattern (LBP)

Local Binary Pattern (LBP) is a texture descriptor [19], it's operated by dividing an image into multiple cells, any pixel in the center of the cell is compared to its eight neighbors, starting from the top-left direction. Starting clockwise manner if the pixel in the center is larger than its neighbors it is replaced by zero, otherwise, it replaces by one. After that, calculate the decimal value of all binary numbers, resulting in LBP code which replaced center pixel. To collect information over larger regions, select larger cell sizes. The LBP code for P neighbors situated on a circle of radius R is computed as follows [2]:

$$LBP_{P,R}(X,Y) = \sum_{p=0}^{p} S(g_p - g_c)^{2p} \qquad (6)$$

Where s (l)=1 if l ≥ 0 and 0 otherwise.

3.3.4. Support Vector Machine (SVM)

Instead of using a classification layer GoogLeNet as a classifier, we tried another classifier like Support Vector Machine (SVM) [31] to make a performance comparison. Precisely, the study included a linear multi-class SVM in order to constitute subjects/classes. The Multi-class SVM technique is to use a one-versus-all classification approach to represent the output of the k-th SVM as in (7).

$$a_k(\mathrm{x})=W^T\mathrm{x} \qquad (7)$$

The forecast class is:

$$arg_k(\mathrm{max})=a_k(\mathrm{x}) \qquad (8)$$

3.3.5. k-nearest neighbour (KNN)

The k-nearest neighbour (KNN) algorithm [32] is determine the k nearest neighbours to specific case by using Euclidean distance (or other measures). KNN returns the most common value among the k training examples nearest to the query.

Given a query instance x_q to be classified let x_1,.., x_k denote the k instances from training examples that are nearest to x_q return [32]:

$$\mathrm{f}(x_q) \rightarrow \mathrm{arg\ max} \sum_{i=1}^{k} \sigma\,(v, f(x_i)) \qquad (9)$$

Where δ(a ,b)=1 if a=b and δ(a ,b)=0

3.3.6. Majority Voting

Majority Voting [33] utilizes the standard class label values that have been retrieved from the predicted label array obtained through the classifier. It counts class with most than half occurrence from all feature extractors, finally, return class label as the final prediction as follows.

$$\mathrm{C(X)=mode}\{h_1(\mathrm{X}), h_2(\mathrm{X}), h_3(\mathrm{X})\} \qquad (10)$$

Where X is the class label, h(x) is a prediction array.

3.3.7. Euclidean distance and Threshold

The performance has evaluated using Euclidean distance [4], which measures the similarity between pairs of feature vectors. Given the two feature image vectors a and b, the similarity distance is the Euclidean distance calculated in the following way:

$$d(a,b) = ||a - b|| \qquad (11)$$

For two image feature sets, A $=\{a_1, a_2, \dots\dots, a_n\}$ and B $=\{b_1, b_2, \dots\dots, b_n\}$, we define the minimum similarity distances between the two sets as follows:

$$h_{\min}(A,B) = \min(d_{(a\in A \& b\in B)}(a,b) \qquad (12)$$

Euclidean distance takes the features vector returned by CNN network and calculate the distance between them and compared with threshold. If the result is less than threshold faces are considered for the same person otherwise it’s considered extra-personal.

3.4. Classifier Performance

All measures of performance based on four numbers obtained by applying the classifier to the test set. These metrics are false positives (FP), true positives (TP), true negatives (TN), and false negatives (FN). Thus, system validation accuracy has calculated as follows [13]:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (13)$$

The system validates the network in each iteration during the training process. The validation images were classified using the fine-tuned CNN network, and their classification accuracy has calculated.

4. EXPERIMENTS AND ANALYSIS

Images are collected from the MORPH dataset [20] and partitioned into multiple classes. Each class is composed entirely of subject’s images of various ages. Images pre-processing, feature extraction, and classification are performed as in previous sections. A common practice in transfer learning is to remove the top layers of a DNN and replace them with new different layers to handle the new dataset. In GoogLeNet, the top network layer is responsible for processing the output of many underlying convolutional layers, we add new layers to make predictions for the new task, then retrained the added layers from scratch and initialized with random weights.

The first ten layers are frozen while the training process to slow down the learning process, but learning layers that have not been frozen are slowed down during the learning process by setting the initial learning rate to a small value of 10. The result of these processes is rapid learning in the new layers, slowing of learning in the middle layers, and not learning in the previously frozen layers.

The convolutional layers of the network extract image features using the network's last learnable layer, and the final classification layer classifies the input images. Thus, these two layers in GoogLeNet contain information on how to combine the features that the network extracts into class probabilities, a loss value, and predicted labels.

To retrain a pre-trained network to classify new images, we replace these two layers with new layers adapted to the new dataset. The last learnable layer in the network was replaced by a new layer with an output size of 672, the weight of the learning rate factor was equal to 10, and the classification layer was replaced by a new layer with an input and output size equal to 672, which is the number of classes (subjects) in the dataset. To verify the facial images, we calculate the minimum distance between each pair of images using equation No. 2, where each image in the test dataset was compared with an image in the gallery set to see whether they belong to the same person or not.

The images pairs with the lowest value are identical (belong to the same person), and the other pairs are not identical. At this point, the network must be enhancing to improve its performance and achieve high accuracy. Accordingly, we changed the training parameters and solved the overfitting problem as described in subsequent sections.

4.1. Training Parameters

Training parameters are kept consistent unless otherwise specified in transfer learning techniques. No need to train the model for many epochs when using transfer learning, the number of epochs have set to 40, and the mini-batch size has set to 100.The ReLU activation function was used in all weight layers, and the initial learning rate was set to 0.001. A fully connected layer was added with the number of outputs equal to 672. The learning rate factor for the connected layer has been increased to 20 so that the network can learn faster, the verification frequency has been set to 3, and the learning rate drop factor has been set to 0.3.

4.2. Number of Epochs

An epoch is one pass through all the data in the training set [1], it's one of the training parameters to be considered during training the network. The number of epochs might be high or low. Knowing the optimal value depends on the database used, organization techniques, and network depth. If the number of epochs is low, the network will be under-learned, but if it is high, the model becomes overfitted.

Figure 2 shows the validation accuracy of the model over increasing numbers of epochs. In this experiment, the optimal number of epochs is 30, where the achieved validation accuracy was 99.8%.

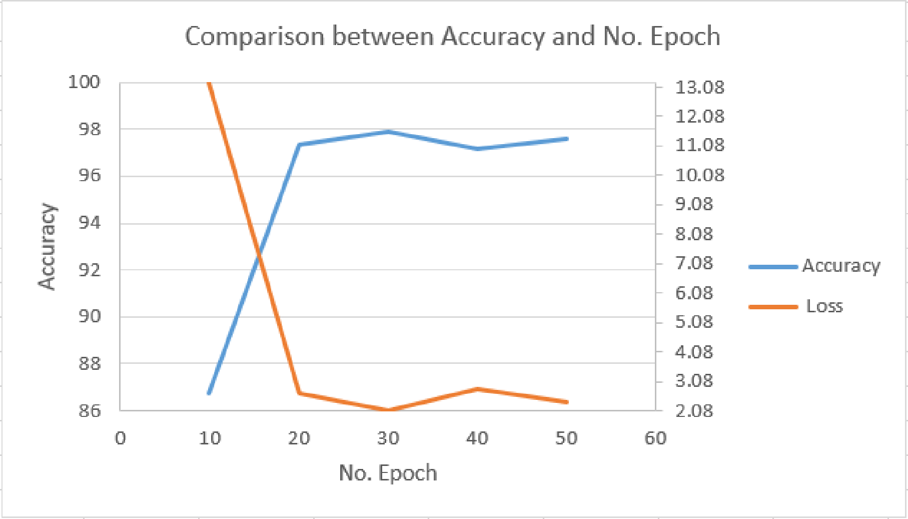


Figure 2. Suitable Number of Epochs to Maximize Accuracy and Minimize Validation Loss

One of the most challenging problems in machine learning is overfitting, which occurs when a model learns details and noise in the training data, also, when the validation accuracy is lower than the training accuracy, which affects model performance. To overcome the overfitting problem, we use a dropout and data augmentation [16].

4.3. Data Augmentation

Data augmentation helps prevent the network from overfitting by memorizing the exact details of the training images [24]. Beginning, we reflected each image horizontally, then horizontal and vertical translations were applied to input images in the [-30, 30] range. Finally, images have scaled in horizontal and vertical directions, thus allowing the classifier to trained on additional views of objects.

4.4. Dropout

One way to solve overfitting is to add dropout to weight layers. At each iteration, neurons randomly selected for removal from the network. The number of neurons omitted from a layer is called the dropout rate [21], which is set manually. When the dropout rate value is high, we get a better regularization to prevent overfitting but slow down the learning process. Therefore, the dropout rate value must be balanced so that it is suitable for both overfitting and the learning process.

In table 1, we tested the model using different dropout values to decide which value would best to prevent overfitting.

Table 1. The Optimal Dropout Rate Value.

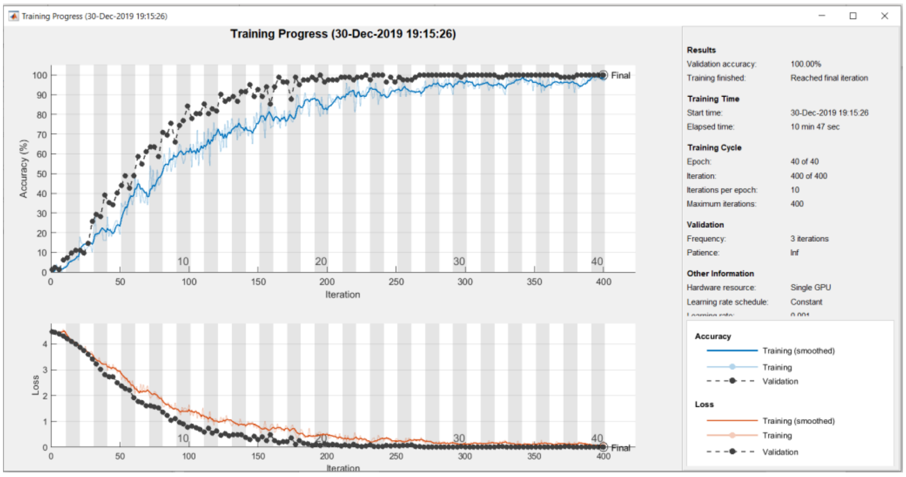
Dropout	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
Loss	15.48	2.08	13.84	10.86	13.1

The lowest validation loss value was the optimal value obtained with a dropout rate of 0.3. We evaluated the performance of our method against previous works by comparing verification accuracy with several other results. As observed in table 2, our method outperforms state-of-the-art methods.

Table 2. Comparison of Proposed Model and Current State-of-the-Art Methods.

Approach	Dataset	Method	Accuracy
B.Simone [1]	LAG Datset	Siamense DCNN Injection	85.75%
El Khiyari, H., et al. [4]	FG-NET Dataset	VGG-Face	0.16 (EER)
Moschoglou, S., et al. [15]	AgeDB Dataset	VGG-Face	93.4%
Zhai et al. [25]	LFW Dataset	DCNN + LBPH	91.40%
Proposed Method	**MORPH** **FG-NET**	**GoogLeNet,** **LBP,HOG**	**99.8%** **100%**

The performance of the model is evaluated using the receiver operating characteristic (ROC) curves. False accept errors were reported using the false accept rate (FAR), which is the percentage of negative pairs labeled as positive. False reject errors were calculated using the false reject rate (FRR), which is the percentage of positive pairs classified as negative. The ROC curves represent the tradeoffs between the FARs and FRRs of different values [7].



Results

Validation accuracy:	100.00%
Training finished:	Reached final iteration

Training Time

Start time:	30-Dec-2019 19:15:26
Elapsed time:	10 min 47 sec

Training Cycle

Epoch:	40 of 40
Iteration:	400 of 400
Iterations per epoch:	10
Maximum iterations:	400

Validation

Frequency:	3 iterations
Patience:	Inf

Other Information

Hardware resource:	Single GPU
Learning rate schedule:	Constant
Learning rate:	0.001

Figure 3. Training Progress for the Model

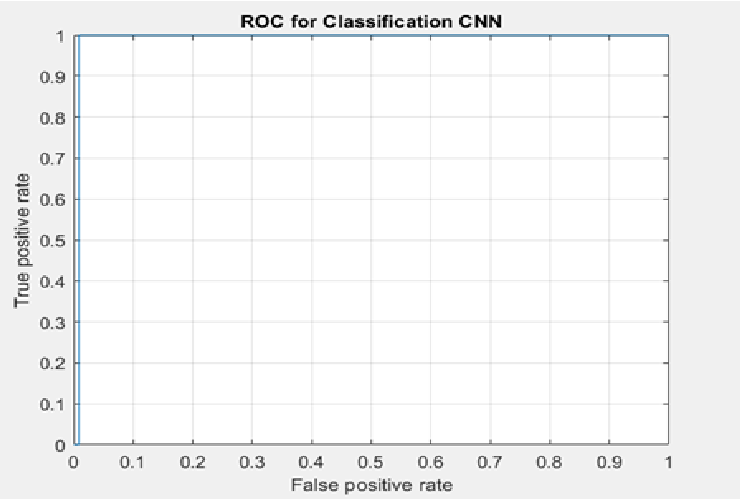


Figure 4. ROC Curve for Classification

An example of a successfully classified subject's images labelled as a match with a 5-year age difference is shown in figure 5, and an example of a misclassified subject is shown in figure 6.

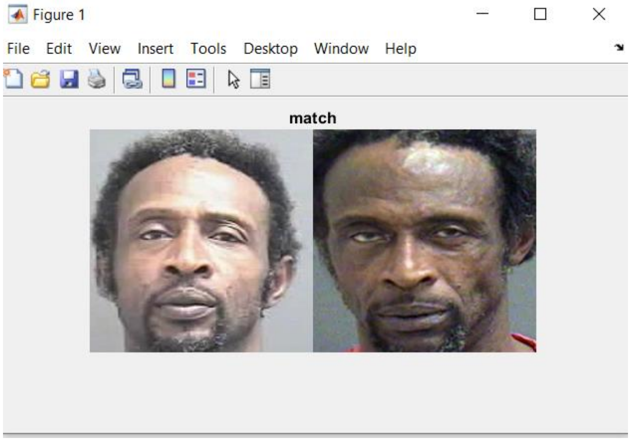


Figure 5. Example of a Successfully Classified Subject's Images

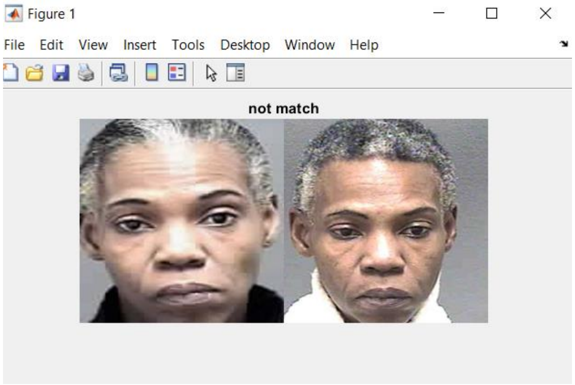


Figure 6: Example of a Misclassified Subject's Images

As it is shown in table 3 using GoogLeNet, HOG, and LBP for feature extraction and SVM for classification by a majority voting is the best result than GoogleNet, in FG-NET the best results is 100% which is more best than MORPH. When comparing KNN with SVM as classifier, we find that SVM has better performance than KNN as it produces high accuracy.

Table 3. Results using Morph Dataset

Method	**GoogLeNet for Extraction and Classification**	**GoogLeNet for Extraction and SVM for Classification**	**GoogLeNet and LBP**	**Majority Voting GoogLeNet, SVM, LBP**	**Majority Voting GoogLeNet, KNN, LBP**	**Majority Voting GoogLeNet, HOG, LBP**
Training Accuracy	100%	100%	100%	100%	100%	100%
Testing Accuracy	98.96%	93.30%	95.56%	96.42%	90.28%	**99.8%**

Table 4. Results using FG-Net Dataset

Method	**GoogLeNet For Extraction and Classification**	**GoogLeNet for Extraction and SVM for Classification**	**GoogLeNet and LBP**	**Majority Voting GoogLeNet, SVM, LBP**	**Majority Voting GoogLeNet, KNN, LBP**	**Majority Voting GoogLeNet ,HOG,LBP**
Training Accuracy	100%	100%	100%	100%	100%	100%
Testing Accuracy	94%	95.4%	51%	99.2%	94.67%	**100%%**

5. CONCLUSIONS

This paper addresses the challenge of facial verification on aging subjects using a transfer learning method based on deep learning. A CNN model (GoogLeNet) and a face dataset. (MORPH) was used to pre-train a convolutional neural network to extract features from facial images. Better results were observed when optimal dropout rate and numbers of epochs were used. Euclidian distance was used to determine whether or not pairs of images belonged to the same person. The model's performance was evaluated using a training and validation set, and we were able to show that GoogLeNet yields a better performance than other state-of-the-art methods.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Bengio, Y. (2009). Learning deep architectures for AI. Now Publishers Inc.

[2] Bouadjenek, N., Nemmour, H., & Chibani, Y. (2016, November). Writer's gender classification using HOG and LBP features. In International Conference on Electrical Engineering and Control Applications (pp. 317-325). Springer, Cham.

[3] Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886-893). IEEE.

[4] El Khiyari, H., & Wechsler, H. (2017). Age invariant face recognition using convolutional neural networks and set distances. Journal of Information Security, 8(03), 174.

[5] El Khiyari, H., & Wechsler, H. (2016). Face verification subject to varying (age, ethnicity, and gender) demographics using deep learning. Journal of Biometrics and Biostatistics, 7(323), 11.

[6] Gan, Y., Liu, J., Dong, J., & Zhong, G. (2015). A PCA-based convolutional network. arXiv preprint arXiv:1505.03703.

[7] Hu, J., Lu, J., & Tan, Y. P. (2014). Discriminative deep metric learning for face verification in the wild. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1875-1882).

[8] Huang, D., Shan, C., Ardabilian, M., Wang, Y., & Chen, L. (2011). Local binary patterns and its application to facial image analysis: a survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 41(6), 765-781.

[9] Huang, G. B., Lee, H., & Learned-Miller, E. (2012, June). Learning hierarchical representations for face verification with convolutional deep belief networks. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2518-2525). IEEE.

[10] ImageNet.http://http://www.image-net.org. Accessed: 29 March 2019.

[11] Ji, S., Xu, W., Yang, M., & Yu, K. (2012). 3D convolutional neural networks for human action recognition. IEEE transactions on pattern analysis and machine intelligence, 35(1), 221-231.

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[13] Le, Q. V., Zou, W. Y., Yeung, S. Y., & Ng, A. Y. (2011, June). Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In CVPR 2011 (pp. 3361-3368). IEEE.

[14] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[15] Moschoglou, S., Papaioannou, A., Sagonas, C., Deng, J., Kotsia, I., & Zafeiriou, S. (2017). Agedb: the first manually collected, in-the-wild age database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 51-59).

[16] Nowlan, S. J., & Hinton, G. E. (1992). Simplifying neural networks by soft weight-sharing. Neural computation, 4(4), 473-493.

[17] Panis, G., Lanitis, A., Tsapatsoulis, N., & Cootes, T. F. (2016). Overview of research on facial ageing using the FG-NET ageing database. Iet Biometrics, 5(2), 37-46.

[18] Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition.

[19] Ranzato, M. A., Huang, F. J., Boureau, Y. L., & LeCun, Y. (2007, June). Unsupervised learning of invariant feature hierarchies with applications to object recognition. In 2007 IEEE conference on computer vision and pattern recognition (pp. 1-8). IEEE.

[20] Ricanek, K., & Tesafaye, T. (2006, April). Morph: A longitudinal image database of normal adult age-progression. In 7th International Conference on Automatic Face and Gesture Recognition (FGR06) (pp. 341-345). IEEE.

[21] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

[22] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9).

[23] Taylor, G. W., Fergus, R., LeCun, Y., & Bregler, C. (2010, September). Convolutional learning of spatio-temporal features. In European conference on computer vision (pp. 140-153). Springer, Berlin, Heidelberg.

[24] Van Dyk, D. A., & Meng, X. L. (2001). The art of data augmentation. Journal of Computational and Graphical Statistics, 10(1), 1-50.

[25] Zhai, H., Liu, C., Dong, H., Ji, Y., Guo, Y., & Gong, S. (2015, June). Face verification across aging based on deep convolutional networks and local binary patterns. In International conference on intelligent science and big data engineering (pp. 341-350). Springer,Cham.

[26] Nimbarte, M., & Bhoyar, K. K. (2020). Biased face patching approach for age invariant face recognition using convolutional neural network. International Journal of Intelligent Systems Technologies and Applications, 19(2), 103-124.

[27] Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. J Big Data 6(60):1–48.

[28] Kamarajugadda KK, Polipalli TR (2019) Age-invariant face recognition using multiple descriptors along with modified dimensionality reduction approach. Multimed Tools Appl.

[29] Rafique, I., Hamid, A., Naseer, S., Asad, M., Awais, M., & Yasir, T. (2019, November). Age and Gender Prediction using Deep Convolutional Neural Networks. In 2019 International Conference on Innovative Computing (ICIC) (pp. 1-6). IEEE.

[30] Kasim, N. A. B. M., Rahman, N. H. B. A., Ibrahim, Z., & Mangshor, N. N. A. (2018). Celebrity Face Recognition using Deep Learning. Indonesian Journal of Electrical Engineering and Computer Science, 12(2), 476-481.

[31] Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.

[32] Elkan, C..: Evaluating classifiers. University of San Diego, California, retrieved B,250. (2012).

[33] A.A. Ross, K. Nandakumar and A. Jain, “Handbook of multibiometrics (Vol. 6)”, Springer Science & Business Media, 2006, pp.73-82.