Face Recognition using Convolutional Neural Network

Najwa Wanis, Abdelsalam Almarimi*
College of Electronic Technology Baniwalid, Libya
*Higher Institute of Engineering Technologies Baniwalid, Libya
Contact Information: E-mail: Belgasem_2000@yahoo.com

ABSTRACT

Biometric systems have emerged as a highly dynamic and critical domain within information security, garnering substantial attention. These systems leverage inherent human attributes, which are uniquely distinctive across individuals, such as the iris, voice patterns, facial characteristics, palm prints, retinal scans, gait, and fingerprints. Among these, the recognition of human faces represents a particularly significant biometric technique, possessing extensive applicability across numerous real-world scenarios. However, the complexity of face recognition is considerably amplified by variations in pose, illumination conditions, and the effects of aging. Consequently, these challenges have spurred an escalating demand for, and intensive research into, robust face recognition systems.

This study first undertakes a comprehensive investigation into the state-of-the-art in face feature extraction and classification methodologies. Subsequently, we propose a novel face recognition system predicated on deep learning principles. We carefully constructed a dataset and trained a GoogleNet-based model, employing advanced deep learning techniques. Our experimental results demonstrate a high degree of efficiency, achieving a 91.66% accuracy rate in facial recognition. The expansive potential of deep learning applications is evident, and by harnessing vast quantities of information through significant computational capacity, the precision of such systems can be further augmented. We argue that our contribution offers strong insights into the deep learning paradigm, encompassing the entire process from dataset construction to the careful preparation and deployment of models. This holistic approach aims to substantially reduce human effort.

KEYWORDS: Facial Recognition, Convolutional Neural Networks (CNN), Deep Learning, Feature Extraction.

1. INTRODUCTION

Facial recognition stands as a pivotal biometric methodology employed for the classification and identification of individual human faces. This biometric application facilitates the automated authentication or recognition of a person by comparing a digital facial image against a pre-existing database of learned images [1]. It is crucial to differentiate between facial recognition and facial detection; while facial detection systems are designed to ascertain the presence and spatial location of faces within an input image [2], the scope of this paper is exclusively focused on the recognition component, presuming that the detection phase has already been accomplished.

Typically, human facial recognition systems encompass two primary operational phases: feature extraction and face matching. Feature extraction is performed to derive salient information that enables the differentiation between faces of various individuals. Following this, in the face matching stage, the extracted features from an input face are rigorously compared against the pre-stored facial images within a dataset [3].

The techniques employed in facial recognition are broadly categorized into several distinct types: holistic methods, feature extraction methods, hybrid approaches, and deep learning techniques. Concurrently, a multitude of algorithms can be leveraged for the purpose of matching facial images, including but not limited to Euclidean Distance, Nearest Neighbor, and Structural Similarity [4].

One of the compelling advantages of facial recognition lies in its inherent ease of use, naturalness, and broad social acceptance. In stark contrast to other biometric modalities that may necessitate physical interaction with shared equipment, potentially exposing users to the transmission of pathogens or impurities from other individuals, human facial recognition is entirely non-intrusive and presents no associated health risks.

Originating from the recognition of faces in static images, facial recognition capabilities have progressively expanded to encompass both still images and dynamic video sequences. Approaches to human facial recognition are generally grouped into four overarching categories: holistic, feature-based, hybrid, and deep learning. Feature-based techniques involve the transformation of the input human face into a vector of geometric features. This is achieved by initially processing the input face to identify, extract, and characterize distinguishing facial attributes such as the mouth, nose, eyes, and other unique facial markers. Subsequently, these derived measurements are utilized by conventional numerical model recognition algorithms to facilitate face identification.

Despite significant advancements, facial recognition remains a challenging endeavor due to the inherent variability in illumination, pose, and the effects of aging. These complexities have consequently driven an increasing demand for, and extensive research into, robust facial recognition systems. Therefore, this paper addresses these critical issues by investigating the cutting-edge advancements in face feature extraction and classification techniques, culminating in the proposal of an enhanced human facial recognition system.

Persistent challenges within human facial recognition systems include fluctuating lighting conditions, diverse face orientations, variations in skin tones, and the physiological changes associated with human aging. Given these contributing factors, achieving a 100% precision rate in any recognition system remains an elusive goal. Hence, the automated recognition of human faces continues to represent a formidable problem with a wide array of practical applications. Hybrid approaches frequently integrate both local and global features to enhance the accuracy of facial image recognition.

The overarching objective of this research is to carefully investigate the state-of-the-art techniques in human facial recognition systems. This investigation aims to inform the design and implementation of a facial recognition system that will empower organizations and individuals to support rigorous security protocols, thereby deterring fraudulent activities. This is achieved by maintaining a comprehensive record of pre-stored facial images and granting authorized access upon successful recognition.

In this scholarly contribution, we first undertake a thorough examination of face feature extraction methods and matching algorithms through an exhaustive review of existing literature. Following this, we propose a novel model for facial recognition, which has been successfully implemented utilizing MATLAB. Section 2 provides an in-depth illustration of the literature review. Section 3 introduces the proposed model, specifically applied for the identification of facial images. Section 4 carefully presents the experimental work and the obtained results. Finally, we conclude our research endeavor and delineate several recommendations for future work in this promising field.

2. RELATED WORK

This section provides a comprehensive overview of prior research concerning face recognition methodologies. Numerous studies have explored diverse approaches to enhance the accuracy, robustness, and efficiency of facial recognition systems, addressing various challenges inherent in the field.

[5] introduced a hybrid methodology for face recognition, integrating Gabor wavelet transforms with Linear Discriminated Analysis (HGWLDA). Their technique involved convolving grayscale facial images with a bank of Gabor filters, which varied in orientation and scale. Classification and recognition were subsequently achieved through the application of the k-nearest neighbor (k-NN) algorithm, by comparing the features of the test face image with those in the training set.

Ojala et al. [6] developed a human face recognition system employing Discrete Cosine Transform (DCT) and Principal Component Analysis (PCA) techniques. The system utilized the minimum Euclidean distance as a metric for measurement and decision-making. The efficacy of this system was demonstrated across a variety of human face databases.

Barkan et al. [7] proposed a novel representation for human face images, leveraging an over-complete Local Binary Pattern (OCLBP). This representation constitutes a multi-scale, modified iteration of the traditional LBP technique. Concurrently, Sharma et al. [8] presented an efficient pose-invariant face recognition system founded on the PCA method and an ANFIS classifier. The PCA method was deployed for extracting image features, while the ANFIS classifier was developed to facilitate identification under diverse pose conditions. The performance of their proposed PCA-ANFIS system exhibited superiority over both ICA-ANFIS and LDA-ANFIS in the context of face recognition, with its assessment conducted using the ORL database.

Cho et al. [9] put forth an efficient hybrid face recognition system that judiciously combines both holistic and local features. The PCA method was employed to effectively minimize dimensionality. Following this, the Local Gabor Binary Pattern Histogram Sequence (LGBPHS) technique was utilized to execute the recognition stage, a strategy designed to mitigate the complexity typically associated with Gabor filters. The experimental findings unequivocally demonstrated an improved recognition rate when compared to conventional PCA and Gabor wavelet techniques, particularly under varying illumination conditions. The Extended Yale Face Database served as the benchmark for show outside the effectiveness of this system.

Kamencay et al. [10] devised a novel face recognition method incorporating SIFT features, alongside PCA and KNN techniques. The Hessian-Laplace detector, in conjunction with an SPCA descriptor, was utilized to extract local features. SPCA was specifically introduced for the identification of human faces, while the KNN classifier was employed to pinpoint the closest human faces from the trained features. The experimental outcomes indicated a recognition rate of 92% for the un-segmented ESSEX database and an impressive 96% for the segmented database.

Barnouti [11] presented a methodology that integrates PCA and a Back propagation Neural Network (BPNN) with DCT. The synergistic combination of BPNN with PCA facilitated more straightforward face recognition. DCT was additionally employed to reduce the dimensionality of face databases. The performance of this system was rigorously evaluated using the Face94 and Grimace databases.

Huang et al. [12] introduced a two-dimensional discrete wavelet transform (2D-DWT) method for face recognition, featuring an innovative patch strategy. A non-uniform patch approach for the top-level's low-frequency sub-band was proposed, utilizing an integral projection technique for the two top-level high-frequency sub-bands of 2D-DWT, based on the average image of all training samples. This particular patch approach is deemed more effective for preserving the integrity of local information and is more aptly suited to reflect the nuanced features of the human face image.

Collectively, these studies underscore the diverse array of methodologies and the persistent efforts aimed at advancing face recognition systems, effectively addressing various inherent challenges and consistently improving performance across a spectrum of datasets and operational conditions.

3. METHODOLOGY

3.1 The System Architecture

The fundamental premise of this research methodology centers on the development of a robust model for human face recognition, primarily leveraging the capabilities of a Convolutional Neural Network (CNN). The proposed system is carefully structured into six distinct operational stages.

The overarching inspiration for this research methodology is the design of a model rooted in convolutional neural networks specifically tailored for the accurate recognition of human facial images. As previously stated, the proposed system is delineated into six sequential stages, as depicted in Figure 1. The initial phase involves the resizing and subsequent division of human images into designated sets for training, validation, and testing. Subsequently, the prepared image dataset is fed through the model, ultimately leading to the classification of the images.

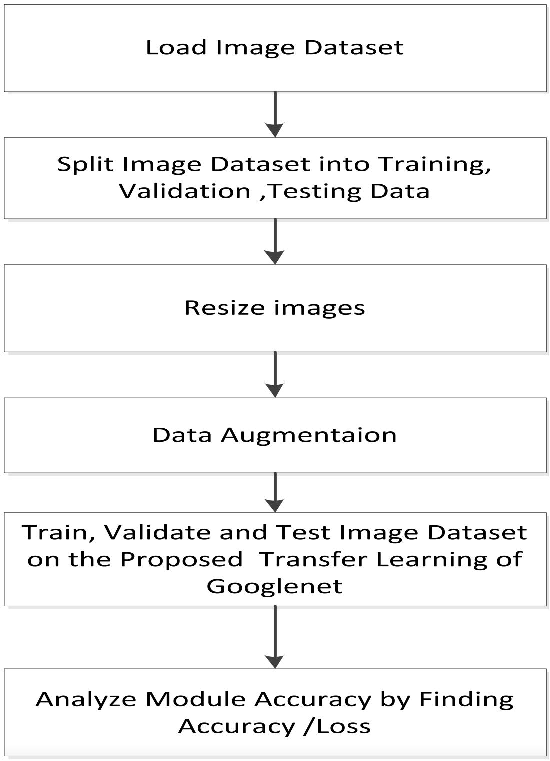

Figure 1: Block diagram of the proposed method.

3.2 Dataset Creation

The creation of a comprehensive and representative dataset constitutes a paramount task within the realm of deep learning, as it directly influences the efficacy and outcome of the trained model. For the purpose of this study, a unique and individualized dataset was carefully compiled by acquiring facial images of various celebrities. This undertaking specifically involved the collection of faces from four distinct individuals. Upon the completion of the data collection, all photographic assets were carefully collated, with stringent measures taken to ensure that no two facial images of the same person were identical, thereby enhancing the dataset's integrity.

3.3 Image Pre-Processing

All images within the compiled dataset undergo a standardized pre-processing procedure. They are uniformly resized to a resolution of 224x224 pixels with three color channels (224x224x3), a configuration specifically chosen to align with the input layer requirements of the neural network architecture.

3.4 Dataset Splitting

For the purpose of classification by the convolutional neural network, the validated images within the dataset are systematically bifurcated into two primary categories. A substantial 70% of the images are allocated for the training phase, while the remaining 30% are reserved for validation. In practical, this translates to 28 images designated for training and 12 images for validation as shown in Figure 2.

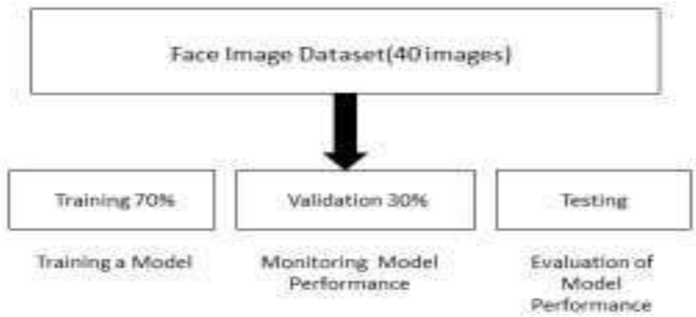

Figure 2: Dataset Splitting.

3.4 Data Augmentation

Stands as an exceptionally efficient technique employed to mitigate the phenomenon of "over fitting" in deep convolutional neural networks, a common issue arising from the constraints of limited training samples. This process effectively approximates the data possibility gap by intelligently manipulating existing input samples. Such manipulations encompass a variety of techniques, including random cropping, scale transformations, horizontal flipping, and the introduction of controlled noise disturbances. Fundamentally, as long as the quality, quantity, and diversity of the data within the dataset are judiciously augmented, the overall efficiency and performance of the model can be significantly enhanced and improved.

3.5 Feature Extraction and Classification using CNN

Convolution Neural Networks (CNNs), a distinguished class of artificial neural networks, have ascended to a central position in numerous computer vision tasks. Their inherent capabilities have attracted widespread interest and adoption across a diverse spectrum of domains. CNNs are inherently designed to automatically and adaptively acquire spatial hierarchies of features. This learning process is facilitated through back propagation, utilizing an array of fundamental building blocks, such as pooling layers, convolution layers, and fully connected layers.

3.6 CNN Pre-Trained Models

GoogleNet represents a prominent example of pre-trained CNN models, characterized by its 22 layers and a groundbreaking structural innovation termed the "Inception Module." This module is composed of a sophisticated arrangement of convolutions and max-pooling operations, with numerous such modules integrated throughout the GoogLeNet architecture. A key architectural distinction is the replacement of traditional fully-connected layers with parallel convolutions that operate concurrently on the same input layer. The 1x1 convolutions positioned at the base of the module serve to reduce the number of inputs, thereby significantly decreasing computational costs. Furthermore, this design effectively captures the associated features of an input image within the same section. Concurrently, 3x3 and 5x5 convolutions are employed to respond to image patterns at larger scales. The feature maps generated by all these convolutions are then concatenated to form the final output. This innovative architecture notably secured the winning position at the ILSVRC 2014 image classification challenge [13].

4. CNN ARCHITECTURE

The architectural framework of a Convolutional Neural Network (CNN) is fundamentally comprised of several integral layers, including but not limited to convolution layers, pooling layers, and fully connected layers. A conventional CNN configuration typically involves the iterative repetition of a stack consisting of multiple convolution layers and a subsequent pooling layer, which are then succeeded by one or more fully connected layers. The sequential process through which input data is transformed into output across these various layers is formally referred to as forward propagation.

4.1. Convolution Layer

The convolution layer serves as the primary and most fundamental module within the CNN architecture, specifically tasked with the crucial function of feature extraction. Convolution itself is a specialized form of linear operation utilized for this very purpose. It involves the application of a small array of numerical values, commonly known as a kernel, across the input, which is itself an array of numbers referred to as a tensor. At each specific location within the tensor, an element-wise product is computed between each element of the kernel and the corresponding element of the input tensor. These products are then summed to derive the output value for the corresponding position in the output tensor, which is termed a feature map. This intricate procedure is applied numerous times with multiple kernels to delineate an arbitrary number of distinct feature maps. Furthermore, each individual feature map can be carefully formulated with respect to a multitude of input maps [14].

4.2. Pooling Layer

A pooling layer is designed to execute a classic down-sampling operation, which effectively diminishes the in-plane dimensionality of the feature maps. This reduction serves multiple critical purposes: it helps to establish translation invariance to minor shifts and distortions within the input, and it significantly reduces the number of subsequent trainable parameters required by the network. Consequently, pooling layers do not possess any learnable parameters themselves; however, hyper parameters such as filter size, and padding are integral to pooling operations, bearing a conceptual resemblance to their roles in convolution operations.

4.3 Max Pooling

Max pooling represents the most widely adopted and popular form of pooling operation. In this process, patches are extracted from the input feature maps, and the maximum numerical value within each respective patch is outputted, while all other values within that patch are discarded. In practical applications, a max pooling operation typically employs a filter of size 2x2 with a stride of 2.

4.4 Global Average Pooling

Global average pooling is another noteworthy pooling operation that performs an extreme variant of down-sampling. In this method, a feature map characterized by a given height and width is down-sampled into a 1x1 array. This is achieved by simply computing the average of all elements within every feature map, while crucially maintaining the original depth of the feature maps. This operation is generally executed only once, typically just prior to the fully connected layers [15]. The principal advantages conferred by the application of global average pooling include:

- A substantial reduction in the total number of trainable parameters.
- The inherent capability to enable the CNN to process inputs of arbitrary dimensions.

4.5 Fully Connected Layer

The output feature maps generated by the final convolution or pooling layer are customarily transformed into a one-dimensional (1D) array. This 1D array is then connected to one or more fully connected layers, also commonly referred to as dense layers. Within these layers, each input node is interconnected with every output node through a learnable weight. Once the features have been carefully extracted by the convolution layers and subsequently down-sampled by the pooling layers, they are then mapped by a subset of fully connected layers to produce the ultimate outputs of the network. These outputs typically represent probabilities for each class in classification tasks. The final fully connected layer generally possesses a number of output nodes equivalent to the total number of classes. Furthermore, each fully connected layer is invariably succeeded by a nonlinear activation function.

5. EXPERIMENTAL RESULTS

This section delineates the practical implementation and subsequent analysis of the proposed system, particularly focusing on the Graphical User Interface (GUI) model developed within this research. The entire face recognition system, including the design of its intuitive GUI, was carefully implemented utilizing the MATLAB toolbox. This GUI was specifically engineered to facilitate seamless user interaction, enabling the application and visualization of the system’s intricate processes. Its primary advantage lies in providing an accessible and user-friendly tool for operators, devoid of any inherent design complexities. The efficacy of the proposed system was rigorously evaluated using a dataset comprising 52 images. A substantial portion, specifically 40 images, was allocated for the training phase, while the remaining 12 images were reserved for the testing phase. The objective was to determine the system’s proficiency in accurately identifying previously undefined facial images. Out of the 12 images subjected to testing, 11 were correctly classified, yielding a commendable accuracy rate of 91.66%. Conversely, 8.33% of the 12 images could not be successfully detected by the system.

$$false\ detected\ percentage = \frac{Total\ false\ detected}{Total\ images} * 100$$

$$= \frac{1}{12} * 100 = 8.33\%\ ,$$

$$True\ detected\ percentage = \frac{Total\ true\ detected}{Total\ images} * 100$$

$$= \frac{11}{12} * 100 = 91.66\%$$

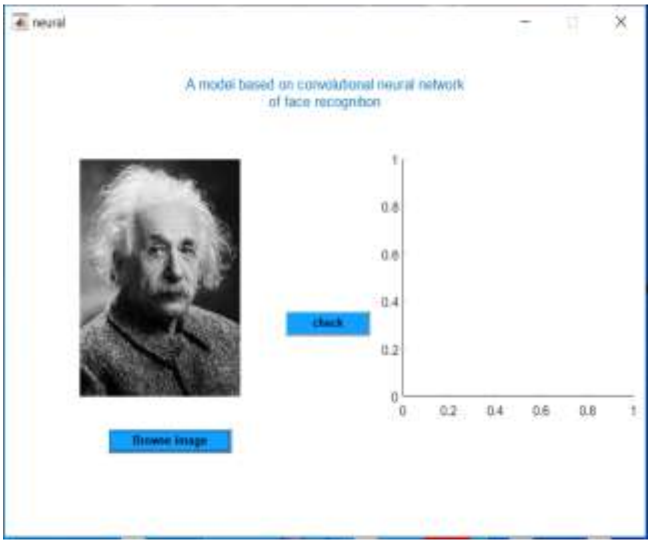

Figure 3: Selecting image.

5.1 The Training Stage

Subsequent to the precise definition of the network structure, the training parameters were carefully specified. The maximum number of epochs was set to 6, with each epoch signifying a complete training cycle across the entirety of the training dataset. The training regimen commenced with an initial learning rate of 3e-4. To ensure robust monitoring of network accuracy throughout the training process, validation data and a validation frequency were designated. The software systematically trained the network on the designated training data, concurrently calculating and reporting the accuracy on the validation data at regular intervals. Figure 3 graphically depicts the training progression of the model, while Figures 4-6 show case of the results obtained from recognizing three distinct facial images.

With each successive iteration, the model's comprehension of the dataset progressively deepens, a trend clearly discernible from the training graph. The loss values exhibit a drastic reduction after an initial series of iterations, following which the slope of the loss curve gradually diminishes. Notably, the losses converge to minimal values after approximately one hundred iterations. Concurrently, the Similarly, accuracy is low during the nascent iterations, demonstrably peaks after around 100 training rounds, signifying optimal model performance.

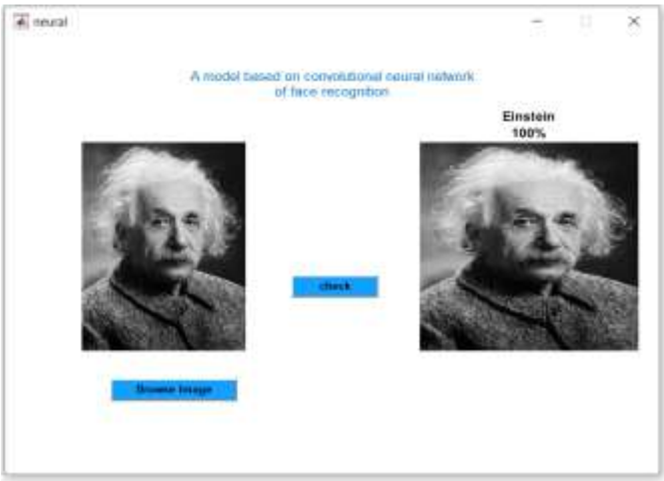

Figure 4: Result of Face Recognition.

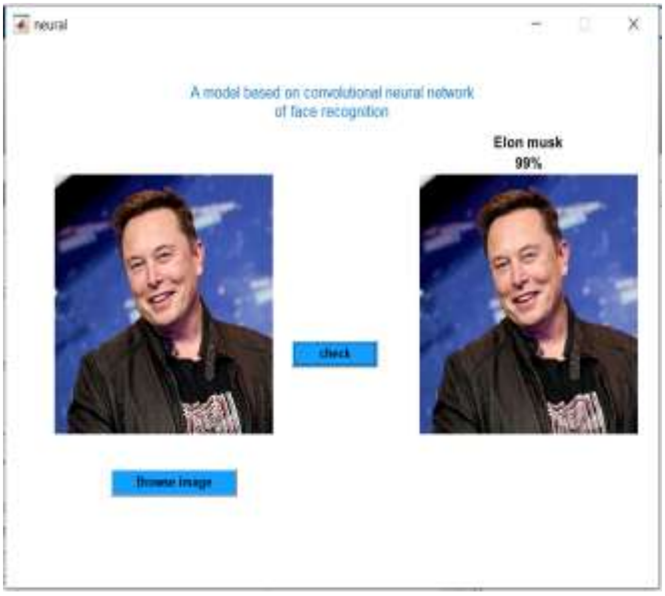

Figure 5: Result of face recognition.

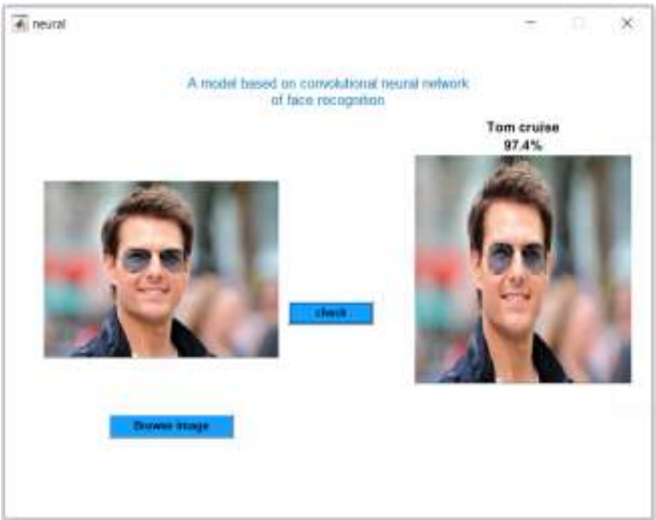

Figure 6: Result of face recognition.

6. CONCLUSION AND FUTURE WORK

6.1. Conclusions

Human facial identification systems constitute a fundamental and increasingly vital component of image processing applications, with their significance as a research domain continually expanding. The proposed system, developed within this research, offers a diverse array of practical applications, encompassing person verification, crime prevention, video observation, and various other security-centric operations.

In this research endeavor, we successfully conceived, designed, and implemented a robust human face recognition system, fundamentally rooted in the principles of deep learning. The main operational process of such a system has been carefully described and elucidated. The ultimate phase of recognition involves a comparative analysis between the detected face, extracted from its background during the initial face detection step, and the repository of recognized faces carefully maintained within a specialized database. To arrive at a conclusive decision, several sophisticated comparison techniques were employed, primarily utilizing common measures of similarity or distance.

Our empirical findings clearly verified that training the classifier with a sufficient volume of facial images yields a remarkably high rate of accuracy in image recognition. Specifically, the proposed system achieved an impressive accuracy rate of 91.66% on average, with 11 out of 12 tested images being correctly identified. The expansive potential inherent in deep learning applications, coupled with the capacity to process vast amounts of data through significant computational power, promises further enhancements in accuracy. This, in turn, will substantially reduce human effort required for maintaining tough security protocols and effectively prohibiting fake activities.

6.2 Future Work

For subsequent and related research endeavors, we propose and recommend the exploration of the following relevant subjects:

- **Enhanced Detection and Expression Recognition:** The current system possesses considerable potential for enhancement by integrating advanced capabilities for both facial detection and the recognition of nuanced facial expressions.
- **Improved System Efficiency via Larger Databases:** The overall efficiency and realism of the system’s results can be significantly augmented through the judicious utilization of a more extensive and diverse database for training and validation.
- **Transition to Open-Source Platforms:** We strongly advise researchers with an interest in similar investigative pursuits to consider adopting Python as their primary development environment. Python’s open-source nature offers considerable advantages, particularly when contrasted with commercial software like MATLAB, which may necessitate the acquisition of additional, often costly, toolboxes that are neither freely available nor easily obtainable.

REFERENCES

1. Hegade, P. P., Nishanth, R., Manikantan, K., & Ramachandran, S. (2013). DWT-based Face Recognition using Morphological Edge Detection as a pre-processing technique. In *Nirma University International Conference IEEE*, pp. 1-6.

2. Li, S. Z., & Jain, A. K. (Eds.). (2011). *Handbook of Face Recognition* (2nd ed.). Springer.

3. Barbu, T. (Ed.). (2011). *Face Analysis, Modeling and Recognition Systems*. InTech. doi: 10.5772/1837.

4. Parmar, D. N., & Mehta, B. B. (2013). Face Recognition Methods & Applications. *International Journal of Computer Technology & Applications*, 4(1), 84-86.

5. Fathima, A. A., Ajitha, S., Vaidehi, V., Hemalatha, M., Karthigaiveni, R., & Kumar, R. (2015). Hybrid approach for face recognition combining Gabor Wavelet and Linear Discriminant Analysis. In *Proceedings of the 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)*.

6. Moussa, M., Hmila, M., & Douik, A. (2018). A Novel Face Recognition Approach Based on Genetic Algorithm Optimization. *Stud. Inform. Control*.

7. Barkan, O., Weill, J., Wolf, L., & Aronowitz, H. (2013). Fast high dimensional vector multiplication face recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1960-1967.

8. Sharma, R., & Patterh, M. S. (2015). *A new pose invariant face recognition system using PCA and ANFIS*.

9. Cho, H., Roberts, R., Jung, B., Choi, O., & Moon, S. (2014). An efficient hybrid face recognition algorithm using PCA and GABOR wavelets. *Int. J. Adv. Robot. Syst*.

10. Kamencay, P., Zachariasova, M., Hudec, R., Jarina, R., Benco, M., & Hlubik, J. (2013). A novel approach to face recognition using image segmentation based on spca-knn method. *Radioengineering*.

11. Barnouti, N. H. (2016). Face Recognition using PCA-BPNN with DCT on Face94 and Grimace. *International Journal of Computer Application*.

12. Li, W., & Huang, Z. H. (2015). Non-uniform face recognition via 2D-DWT. *Computer Science*.

13. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*.

14. Oloko-Oba, M., & Viriri, S. (2019). *Diagnosing Tuberculosis Using Deep Convolutional Neural Network*.

15. Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.