تحليل مقارنه لأداء معماريات الشبكات العصبية التلافيفية في التعرف على الوجه (رؤى من قواعد بيانات LFW وORL)

[1]محمد مصدق [*]، [2]كمال بشير، [3]أمين مبارك الأمين، [4]محمد أحمد محمد
[4،2،1] جامعة كرري (السودان)، [3] جامعة النيلين (السودان)

Comparative Analysis for the Performance of Convolutional Neural Network Architectures in Facial Recognition (Insights from LFW and ORL Databases)

[1] Mohamed Mosadag *, [2] Kamal Bashir, [3] Amin Mubark Alamin, [4] Mohammed Ahmed Mohammed
[1] https://orcid.org/0009-0003-6020-2624, [2] https://orcid.org/0000-0002-1820-6010, [4] https://orcid.org/0009-0006-6036-8765
[1, 2, 4] Karary University (Sudan), m.mosadag@gmail.com, kamalbashir1@yahoo.com, seedaanet@gmail.com
[3] Alneelain University (Sudan), Aminibrahim2006@gmail.com

Received: 28 /01 /2025 Accepted: 09/04/2025 Published: .01/06 /2025

الملخص:

يُستخدم تقنية التعرف على الوجه القائمة على تقنيات الذكاء الاصطناعي والتعلم الآلي والتعلم العميق بشكل متكرر في أنظمة الأمان من خلال المصادقة البيومترية للوجه، وإنفاذ القانون، واعتقال المشتبه بهم والمجرمين باستخدام المراقبة وكشف الوجه، وأنظمة مراقبة الحضور والغياب في الوقت الفعلي، وحالات استخدام مختلفة أخرى. لقد قدمت تقنيات التعلم العميق (DL) مؤخرًا مساهمات مهمة في تقنية التعرف على الوجه، وخاصة الشبكة العصبية التلافيفية (CNN) ؛ والتي تُستخدم على نطاق واسع لمعالجة الصور نظرًا لدقتها العالية في نمذجة الظواهر المعقدة. هناك شبكات عصبية ملتوية متنوعة تتميز بهندستها المعمارية. في هذه المقالة، يتم تحليل أداء أربع هياكل معمارية: LeNet و: AlexNet و ResNet و DenseNet لتحديد أي هيكل معماري هو الأفضل، يتم استخدام مقياسين للأداء للمقارنة؛ الدقة والوقت. تم تدريب هذه الهياكل المعمارية واختبارها على قواعد بيانات LFW وORL. كانت أفضل الهياكل المعمارية أداءً هيDenseNet ، بدقة 100٪ على قاعدة بيانات وجه ORL واستغرق تنفيذها أقل وقت. يمكن أن يكون لهذا إمكانات جيدة لتطوير أنظمة وتطبيقات تستخدم التعرف على الوجه في عملها.

كلمات مفتاحية: التعرف على الوجه.، التعلم العميق.، الشبكات العصبية التلافيفية.، تحليل مقارنه.

Abstract:

Facial recognition technology based on artificial intelligence, machine learning techniques and deep learning techniques is frequently used in security systems through facial biometric authentication, law enforcement, arresting suspects and criminals using surveillance and face detection, attendance monitoring systems on real-time, and other various use cases. Deep learning (DL) techniques have lately made important contributions to face recognition technology, particularly the convolutional neural network (CNN); which is

---

Mohamed Mosadag et al.

most widely used for image processing due to its high accuracy for modeling complex phenomena. There are diverse CNNs that are characterized by their architecture. In this article, four architectures are performance analyzed: LeNet, AlexNet, ResNet, and DenseNet. To determine which architecture is the best, two performance measures are used for comparison; accuracy and time. These architectures were trained and tested on the LFW and ORL databases. The best performing architecture was DenseNet, with an accuracy of 100% On ORL face database and it took the least time to implement. This can have good potential to be developed Systems and applications that use facial recognition in their work.

Keywords: Face Recognition; deep learning; Convolutional Neural Networks; Comparative analysis.

1. Introduction

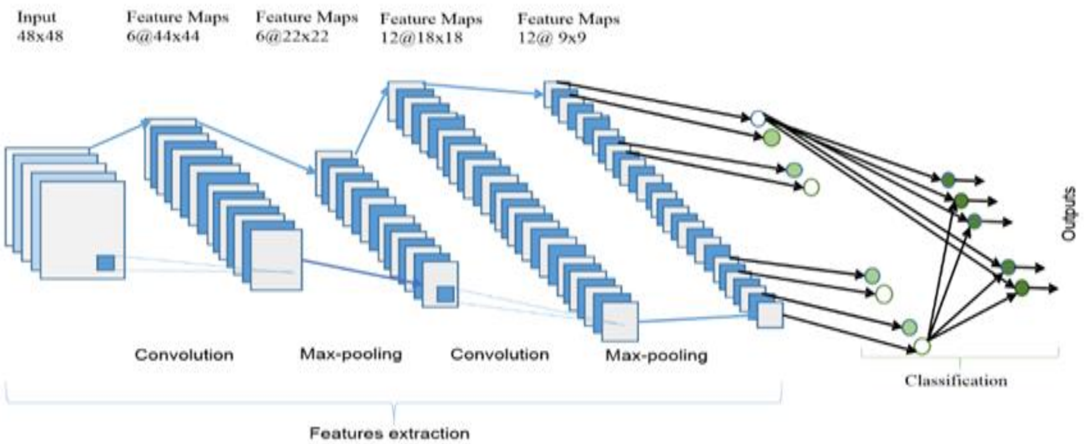
Facial recognition technology is the practice of using systems that combine hardware and software components to recognize people by looking at their faces in a simulation of the capabilities of the human mind. These systems mostly came in two levels: a method for picking up or extracting the characteristics of face (Features) is applied followed by the process of pattern categorization (Classification). In recent years, after research reached its peak in the use of traditional methods for facial recognition, attention was directed to modern technologies, namely Deep Learning techniques (DL). It is a specialized subfield of machine learning which in turn a subfield of artificial intelligence (Bashir & Mosadag, 2025; Bashir, Pirasteh, Abdelrhman, Mosadag, & Mohammed, 2024; Khalil, Bashir, & Mosadag, 2025). It relies on multi-layer neural networks to deep analyse large data sets (hence the name "Deep") (Sarker, 2021). Each layer of the network extracts different features from the data, enabling the system to learn and make decisions in a way that simulates the same way the human brain works. DL is especially viable for assignments such as picture and speech recognition. It uses multiple layers of processing to accomplish sorting data into different ranks of feature extraction. This developing technology has altered the study landscape of facial recognition and dramatically improved cutting-edge performance and aided in the development of effective real-world applications (Mosadag, Bashir, & Fattouh, 2024; Mosadag, Mohammed, Bashir, & Mubark, 2025).

The importance of CNN is referenced in a variety of settings regarding image recognition, object detection, image segmentation etc. There are a lot of Deep CNNs accessible, but how they are implemented will determine how many are used. In case of face recognition systems, numerous articles based on neural networks have been written. But most works exclude comparative studies on the effectiveness of different CNN architectures and often do superficial work on implementation in face recognition (Cuevas-Rodriguez et al., 2023).

2. Preliminaries

A convolutional neural network is one of the various types of artificial neural networks (ANN) which are specifically used for recognition tasks. When looking at the comparison aspects, feature extraction is done manually when using machine learning, while deep learning (DL) does not require manual feature extraction (Data (images) were fed automatically into CNN algorithms for analyzing, instead of feeding manually extracted features) (Cuevas-Rodriguez et al., 2023). If we ask about the general shape of CNN networks, it is built from three main neural layers, which are convolutional layers, pooling layers, and fully connected layers (Krizhevsky, Sutskever, & Hinton, 2017) (see Figure 1). These various types of layers play distinct functions. Although the general architecture is very similar, a number of CNN versions with design differences have been used in the recent literature.

Figure 1: General CNN architecture



In this paper effective data analysis and comparative on different architectures is carried out in order to gain the understanding of performance parameter of (accuracy) among the CNN architectures: LeNet, AlexNet, ResNet, and DenseNet; based on two databases of face Image (LFW and ORL).

2.1 LeNet

Is the first CNN a pioneering 7-level proposed by LeCun et al (LeCun, Kavukcuoglu, & Farabet, 2010). That classifies digits, was applied by several banks to recognize hand-written numbers on checks (cheques) digitized in 32x32 pixel refer to greyscale image. Its general structure consists of three (3) convolutional layers, two (2) pooling layers, and two (2) fully connected layers (total 7 layers). Processing high-resolution images like ORL database images requires larger, more convolutional layers, and availability of advanced computing resources.

2.2 AlexNet

In 2017, Designed by Krizhevsky et al. Its structure is close to the structural structure of LeNet networks, but it is characterized by greater depth, with an increase in the number of filters for each layer, in addition to direct stacking of a number of layers convolved on top of each other (Krizhevsky et al., 2017). AlexNet essentially outflanked all the earlier competitors and won the challenge by diminishing the top-5 mistake from 26% to 15.3%. The moment put top-5 blunder rate, which was not a CNN variety, was around 26.2%. The architecture of it, consists of one input layer, five convolutional layers, seven nonlinear activation (function ReLU activation layers), three max-pooling layers, two normalization layers, two fully connected layers, one softmax layer and one output layer. The repeated of layers in network (back propagated to earlier layers), this repeated may make gradient infinitely small. This problem is called vanishing gradient, it makes Deep networks hard to train.

2.3 ResNet

At the ILSVRC 2015, the so-called Residual Neural Network (ResNet) introduced a novel architecture with "skip connections" (Named for a solution devised to solve the vanishing gradient problem. This treatment is done by adding the original inputs to the outputs of the convolution block) and features strong normalization by batch (He, Zhang, Ren, & Sun, 2016). These skip connections, which are often referred to as closed recurrent units or closed units, strongly resemble recently implemented successful RNN elements; it outperforms human-level execution with a top-5 mistake rate of 3.57% on this particular dataset.

2.4 DenseNet

In 2017, to solve the chronic problem of vanishing gradients, Huang et al. came up with the idea of two main axes of a densely connected convolutional network: The first axis involves linking each layer to each subsequent layer, and the second axis involves linking feature maps through aggregation rather than through summation (G. Huang, Liu, Van Der Maaten, & Weinberger, 2017). DenseNet is c consist of a series of densely connected blocks and intermediary transition layers. A dense block could be a bunch of layers associated to everyone past. All of the previous layers' highlight maps are used as inputs for each layer, and each subsequent layer's own highlight maps are used as inputs. The gradients of the initial input signal and the misfortune function are directly accessible to each layer. Batch Normalization, ReLU activation, and 3 × 3 Convolution (BN–ReLU–Conv) make up a single layer. Batch normalization, 1×1 convolution, and average pooling make up a transition layer.

3. Related Works

Numerous facial recognition researchers employ novel Deep Learning methods - Convolutional Neural Networks (CNN) (LeCun et al., 2010). The principal benefits of Convolutional Neural Network (CNN) methodologies reside in their fundamental architecture, which generally consists of two primary components: the extraction of features and the subsequent classification process. LeNet constituted one of the first CNNs to support deep learning (Syazana-Itqan, Syafeeza, & Saad, 2016).

Meenakshi et al, in 2019 suggested a new CNN architecture due to the impact of occlusions, facial expressions, changes in lighting and posture, etc. (Meenakshi, Jothi, & Murugan, 2019). The four main components of the developed method's architecture are C1, C2, C3, and F4 (Output Layer). The C1 layer comprises five distinct feature plots, while the C2 layer encompasses fourteen feature plots, and the C3 layer contains sixty feature plots, F4 layer has 40 feature maps. In order to avoid the need for padding during the convolution process, the feature plots in the design are smaller. This architecture consists of a 32 × 32 input layer, a subsampling layer, and a fully linked layer. Following a series of pre-processing procedures, the researchers developed the model and subsequently assessed its performance utilizing the AT&T (ORL) database alongside ten participants from the JAFFE database.

Zhiming et al. put forward a theoretical framework for the facial recognition that same year (Xie, Li, & Shi, 2019). This framework is predicated upon two fundamental aspects: the quantity of neurons within the hidden layers and the count of feature maps present in the convolutional layers. Sequentially: an input layer, a convolutional layer, a pooling layer, another convolutional layer, a subsequent pooling layer, a fully connected layer, and a Softmax regression classification layer are the layers that make up this CNN design. Thus, they established the structure C1-C2-H, in which H stands for the whole quantity of hidden layer neuronal cells, C1 for the amount of feature plots in the first convolutional layer, and C2 for the amount of plots maps in the second convolutional layer. A good recognition rate obtained after multiple experimental tests on ORL database.

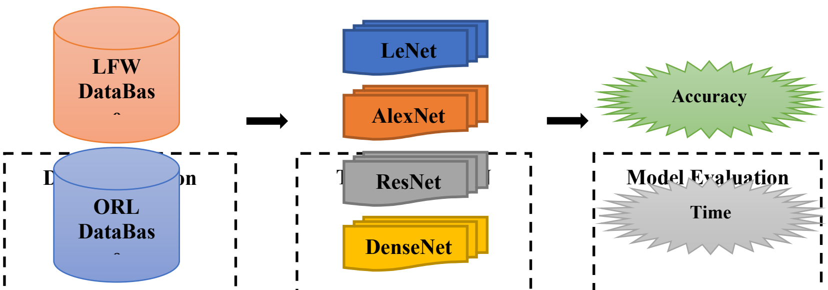
In 2024, employed Residual Networks to implement a facial recognition system (Komlavi, Chaibou, & Naroua, 2022). This research study serves by measuring the viability of the ResNet architecture utilizing multiple parameter setups, like batch size, learning rate, number of hidden layers, and the total amount of units in the hidden layer. They used a database of 1050 photos, split into training and testing database, with 80% of the images used for training and 20% for testing, in order to assess the framework.

Arshi and Virendra suggested a multi-step facial recognition method in 2022 based on Resnet (Husain & Vishvakarma, 2022). The researchers Firstly turned the pgm files into jpg files and processed the databases grayscale photos. Furthermore, the 400-image original dataset is divided into two smaller sets of data, called to as both the examination and training datasets, utilizing a 70:30 ratios, utilizing the optimizer, loss function, and metrics as the three parameters. Adam is used as the optimizer in this model, and the epoch number is set at 25. The proposed approach produced good accuracy on AT&T dataset.

4. Proposed Methodology

CNN is a deep learning method which has the ability to perform both feature extraction and classification stages. Figure 2 illustrates the methodology proposed to implement the four CNN architectures model for the facial recognition system, encompassing: data acquisition (databases), Training of CNN, and model evaluation & Validation.

Figure 2: Diagram of the proposed methodologies



4.1 Data Acquisition

The input of a face recognition system is always an image or video stream. An identification or confirmation of the subject or subjects that are shown in the picture or video is the result. For recognizing things like faces in photos, routine methods like Viola and Jones (HAAR Cascade) and HOG are employed (Saxen et al., 2019). However, these approaches necessitate a significant amount of feature engineering work. With the help of deep learning techniques such overload can be reduced and we do not require lots of feature engineering. To train and test the system to recognize a specific face, database of facial images that is used. The database is created by using images captured from cameras, such as CCTV cameras, or submitted by users. For implementation and evaluation of facial recognition models in this study, two databases of facial images are used:

4.1.1 Labeled Faces in the Wild (LFW)

It is an image database containing pictures of faces designed to evaluate face recognition algorithms in more realistic and uncontrolled environments, Collected from the web specifically to study the problem of unrestricted facial recognition (G. B. Huang, Mattar, Berg, & Learned-Miller, 2008). It contains a large number of images (13,000 images of 5,749 subjects collected from the web, with variations in lighting, exposure, and background). The LFW database was developed and maintained by researchers at the University of Massachusetts, Amherst. It was released for research purposes to advance facial verification. The original database contained four different sets of LFW images as well as three different types of "aligned" images suitable for testing the robustness and performance of facial recognition systems in real-world scenarios.

4.1.2 Olivetti Research Laboratory (ORL)

Contains 400 images of 40 separate subjects (10 images per person) with different lighting, facial expressions, and facial details, Samaria and Harter's presentation of parameterizing a stochastic model for the identification of human faces, Collected from 1992 to 1994 in the laboratory (Samaria & Harter, 1994). It is mainly used for controlled experiments in face recognition. Ideal for testing algorithms under controlled conditions with changes in pose and expression.

4.2 Training of CNN

To execute the experiments in this study, a good computer and updated programming software's was used. The overall specification was as follows: Laptop Specifications, HP notebook, Window 10, processor (CPU) Intel Core i5 and 5200 U 5th Generation, System Ram 2 GB, Storage 1 Tetra HDD, Camera Front. As a software language, Python 3.9.0 was used to apply our algorithms and conduct the whole experiment.

Data are taken from ORL & LFW image databases (G. B. Huang et al., 2008; Samaria & Harter, 1994). The most popular split ratio, 80:20, is used to separate these databases into two separate categories. In other words, 80% of the database is used for training, while 20% is used for testing.

4.3 Model Evaluation and Validation

Two performance measures are simultaneously used to evaluations which are accuracy rate and the time taken for implementation. We will use the accuracy measure as a primary measure to evaluate the model or method used, and the time measure as a secondary measure that can be used if the accuracy measure is same. The matrix of confusion for the binary category problem is presented in Table 1 to provide more clarity regarding accuracy. The true class label is shown in the first column, while the predicted class label is shown in the second and third columns. True Positive (TP) and True Negative (TN) indicate the number of correctly classified positive and negative samples, while False Negative (FN) and False Positive (FP) indicate the number of incorrectly classified positive and negative samples, respectively. As performance indicators, overall classification accuracy has traditionally been extensively evaluated. Verifying the performance and validity of the results will be done by testing on the part of the image database 20% that was not used in training.

Volume 6, Issue 22 (2025) p 474 - 485

Table 1: Confusion matrix for a two-class problem

	Predicted as positive	**Predicted as negative**
Actual positive class	True Positive (TP)	False Negative (FN)
Actual negative class	False Positive (FP)	True Negative (TN)

The metric of binary accuracy can be delineated in the following manner:

$$(TP+TN) / (TP+TN+FP+FN)$$

5. Results and Discussion

In this section, the results and a performance comparison of CNN architectures have been presented based on two databases of face Image (LFW and ORL).

According to the CNN architectures applied LeNet, AlexNet, ResNet and DenseNet. The batch size has been configured to 32, while parameters such as periods have been configured to 50 or just 25. Table 2 presents the accuracy results and the average of it, for the different CNN architectures in generally for the two databases.

Table 2: The Average performance results of different CNN architectures

CNN Architectures	**Accuracy on LFW**	**Accuracy on ORL**	**Average (%)**
LeNet	89.97	100	**94.98**
AlexNet	76.67	100	88.33
ResNet	80	99.05	89.52
DenseNet	67	100	83.50

As shown in Table 2 and Figure 3 we find that the LetNet architecture achieved the highest average performance with 94.98%, whereas the DenseNet architecture came in last place, scoring the lowest average performance with 83.50%. While the other two architectures achieved results in between the previous results.

Figure 3: Comparative based on Average performances of CNN architectures

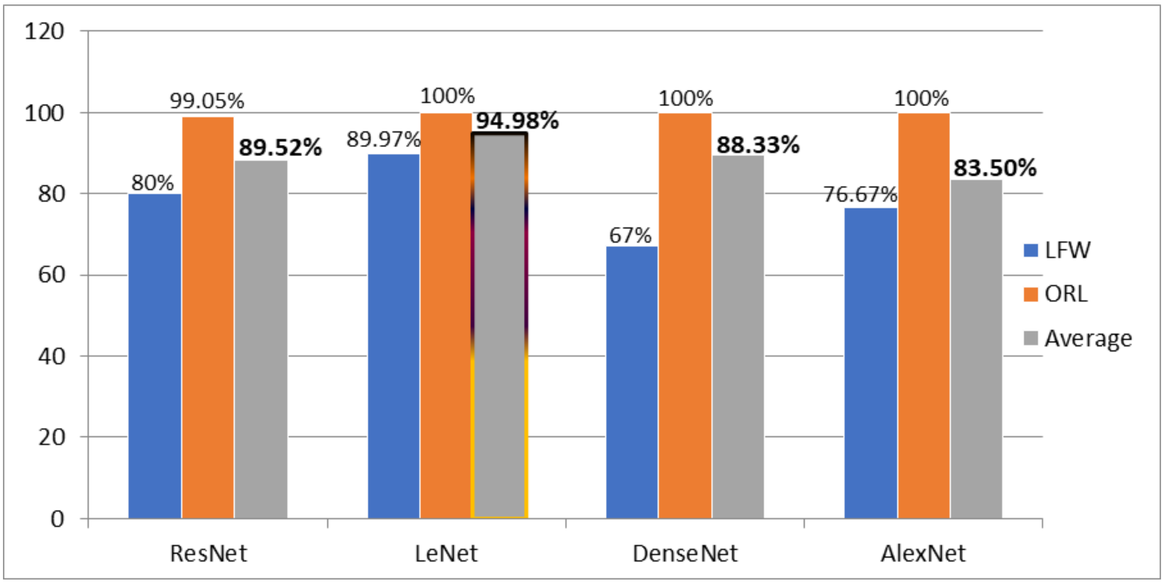


Table 3 presents the accuracy results of different CNN architecture and the time taken for execution each of them on the LFW and ORL databases.

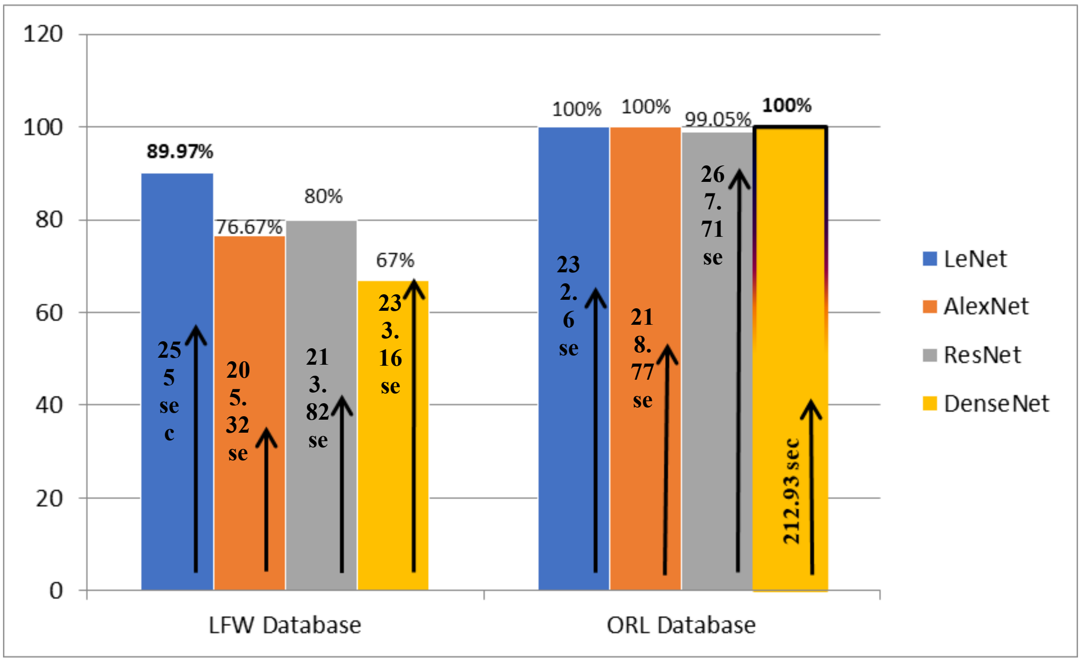
Table 3: The performance results of different CNN architectures based on LFW&ORL database

CNN Architectures	LFW database		ORL database	
	Accuracy (%)	Time (seconds)	Accuracy (%)	Time (seconds)
LeNet	89.97	225	**100**	232.6
AlexNet	76.67	205.32	**100**	218.77
ResNet	80	213.82	99.05	267.71
DenseNet	67	233.16	**100**	**212.93**

The performance accuracy based on LFW database reached good values: 67% for DenseNet, 76.67% for AlexNet, 80% for ResNet and 89.97%. The highest performance was for the LeNet architecture compared to others, which scored (89.97%) in 225 seconds, it emerges as the most successful technique, whereas DenseNet is the least successful.

On the other hand, based on ORL database we got more satisfactorily performed by gaining excellent accuracy reached (100%) with three architectures as shown in Table (3): 100% for DenseNet, 100% for AlexNet and 100% for LeNet. Except ResNet is the least effective with 99.05%.

Figure 4: Comparative performances of CNN architectures based on LFW&ORL databases



With regard to the findings compiled Table (3) and Figure (4), we clearly show all the methods of facial recognition score high accuracy results in the case of ORL database than LFW database. It is important to take into account that the DenseNet architecture came in first place among all, achieving a score of 100% in the shortest implementation time, which was 212.93 seconds.

6. Conclusion and Future Works

From the discussion and results of comparison it is clear and proves that all architectures of CNN work very well with ORL database than LFW. This is due to the LFW database; although it is big and there are a lot of instances, but the distribution of samples between the subjects is irregular. Sometimes you may find a subject that contains one image. In addition, the existing images are of different dimensions. But in the case of ORL database, the distribution of samples between the subjects is regular and the images are same in dimensions and size. It could be concluded that CNN always works efficiently when given reasonably large dataset with regular distribution of samples between the subjects.

Finally, among all of these architectures that are helpful for facial recognition, it is determined that, the DenseNet is the best architecture and recommended to apply it to build facial recognition systems, focusing on the necessity of training on the largest number of samples to provide the desired recognition.

In future works we should: explore the performance of these architectures on more diverse and larger datasets, Exploring hybrid models that combine the strengths of different CNN architectures, and continually exploring and developing new CNN architectures.

References

Bashir, K., & Mosadag, M. (2025). A Novel Resampling Technique for Imbalanced Classification in Software Defect Prediction by a re-sampling method with filtering. *Conclusions in Engineering, 1*(1), 8-17.

Bashir, K., Pirasteh, S., Abdelrhman, H., Mosadag, M., & Mohammed, A. (2024). An Enhanced Feature Selection Approach for Breast Cancer Prediction Using a Hybrid Framework. *Journal of Karary University for Engineering and Science*.

Cuevas-Rodriguez, E. O., Galvan-Tejada, C. E., Maeda-Gutiérrez, V., Moreno-Chávez, G., Galván-Tejada, J. I., Gamboa-Rosales, H., . . . Celaya-Padilla, J. M. (2023). Comparative study of convolutional neural network architectures for gastrointestinal lesions classification. *PeerJ, 11*, e14806.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). *Labeled faces in the wild: A database forstudying face recognition in unconstrained environments.* Paper presented at the Workshop on faces in'Real-Life'Images: detection, alignment, and recognition.

Husain, A., & Vishvakarma, V. P. (2022). *Face recognition method based on residual convolution neural network.* Paper presented at the IOP Conference Series: Materials Science and Engineering.

Khalil, W., Bashir, K., & Mosadag, M. (2025). Early Detection of Chronic Kidney Disease (CKD) Using Machine Learning Algorithms. *East Journal of Computer Science, 1*(2), 1-9.

Komlavi, A. A., Chaibou, K., & Naroua, H. (2022). Comparative study of machine learning algorithms for face recognition. *Revue Africaine de Recherche En Informatique et Mathématiques Appliquées, 40*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

LeCun, Y., Kavukcuoglu, K., & Farabet, C. (2010). *Convolutional networks and applications in vision.* Paper presented at the Proceedings of 2010 IEEE international symposium on circuits and systems.

Meenakshi, S., Jothi, M. S., & Murugan, D. (2019). Face recognition using deep neural network across variationsin pose and illumination. *International Journal of Recent Technology and Engineering (IJRTE), 8*(1S4), 289-292.

Mosadag, M., Bashir, K., & Fattouh, S. A.-D. (2024). Classifying of Isolated handwritten Arabic Characters (on-line) Using Machine Learning Methods. *Journal of Karary University for Engineering and Science (JKUES), 3*(1), 1-9. doi:https://doi.org/10.54388/jkues.v3i1.251

Mosadag, M., Mohammed, M. A., Bashir, K., & Mubark, A. (2025). A Comparative Analysis for the Performance of LFW and ORL Databases in Facial Recognition. *Conclusions in Engineering, 1*(1), 58-63.

Samaria, F. S., & Harter, A. C. (1994). *Parameterisation of a stochastic model for human face identification.* Paper presented at the Proceedings of 1994 IEEE workshop on applications of computer vision.

Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN computer science, 2*(6), 1-20.

Saxen, F., Werner, P., Handrich, S., Othman, E., Dinges, L., & Al-Hamadi, A. (2019). *Face attribute detection with mobilenetv2 and nasnet-mobile.* Paper presented at the 2019 11th international symposium on image and signal processing and analysis (ISPA).

Syazana-Itqan, K., Syafeeza, A., & Saad, N. (2016). A MATLAB-based convolutional neural network approach for face recognition system. *J. Bioinf. Proteomics Rev, 2*(1), 1-5.

Xie, Z., Li, J., & Shi, H. (2019). *A face recognition method based on CNN.* Paper presented at the Journal of Physics: Conference Series.