ARIMA épisciences

Comparative study of machine learning algorithms for face recognition

Atsu ALAGAH KOMLAVI[*1], **Chaibou KADRI**[1], **Harouna NAROUA**[1]

[1]Department of Mathematics and Informatic, Abdou Moumouni University, Niamey, Niger

*E-mail : alagahdejesus@yahoo.fr

DOI : 10.46298/arima.9291
Submitted on 2 April 2022 - Published on 19 January 2024
Volume : **40** - Year : **2024**

Editors : *Mathieu Roche, Nabil Gmati, Clémentin Tayou Djamegni*

Abstract

Background: The fundamental need for authentication and identification of humans using their physiological, behavioral or biological characteristics, continues to be applied extensively to secure localities, property, financial transactions, etc. Biometric systems based on face characteristics, continue to attract the attention of researchers, major public and private services. In the literature, many methods have been deployed by different authors. The best performance must be found in order to be able to recommend the most effective method. So, the main objective of this article is to make a comparative study of different existing techniques.

Methods: A biometric system is generally composed of four stages: acquisition of facial images, preprocessing, extraction of characteristics and finally classification. In this work, the focus is on machine learning algorithms for classification. These algorithms are: Support Vector Machines (SVM), Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Random Forests (RF), Logistic Regression (LR), Naive Bayesian Classification (NB: Naive Bayes' Classifiers) and deep learning techniques such as Convolutional Neural Networks (CNN). The comparison criterion is the average performance, calculated using three performance measures: recognition rate, confusion matrix, and the Area Under Receiver Operating Characteristic (ROC) curve.

Results: Based on this criterion, the performance comparison of selected machine learning algorithms, shows that CNN is the best, with an average performance of 100.00% On ORL face database. However, on the YALE database, classical algorithms such as artificial neural networks have obtained the best performances, the highest being a rate of 100%.

Discussion: Deep learning techniques are very efficient in image classification as proven by the results on the ORL database. However, their inefficiency on YALE face database is due to the small size of this database which is inappropriate for some deep learning algorithms. But this weakness can be corrected by image augmentation techniques. The comparison of these results with existing state-of-the-art methods is nearly the same. Authors achieved performances of 94.82%, 95.79%, 96.15%, 96.44%, 97.27%, 98.52% and 98.95% for NB, KNN, RF, LR, ANN, SVM and CNN classifiers, respectively. Finally, in depth discussion, it is concluded that between all these approaches which are useful in face recognition, the CNN is the best classification algorithm.

Keywords
Face Recognition; Comparative Study; Machine Learning Algorithms.

I INTRODUCTION

The world is characterized by a significant technological development in all sectors. For example, transport and communication are very fast and very cheaper. That is why the world has become a global village. Some of the benefits of science and technology are unfortunately used to endanger people and their wealth. Due to this continuous development, the necessity to secure, to protect, and to control people and their wealth is increasing gradually. Because of the limitations of traditional solutions such as passwords, badges, ID cards, PINs, there is an increased interest in biometrics systems. Among the most used biometric modalities, the face has attracted many researchers for many reasons. First, a human face recognition system is non-intrusive. Second, the growth of technology, digital cameras and storage devices allow the management of many face databases. Finally, the evolution of facial recognition systems is mainly caused by advances in machine learning with powerful data analysis algorithms. Consequently, facial recognition is becoming a reliable technology for identity verification. For these many facial technologies, it is important to perform a comparative study to assess the performances of these techniques. Moreover, it is necessary to choose the best system. In this study, the question is to know which classification technique is the most efficient in face recognition, using a performance measure based on three specific indicators: accuracy rate, confusion matrix and area under the Receive Operating Characteristic (ROC) curve.

II BIOMETRIC SYSTEM TESTING

Any biometric system has two phases: training and testing. The effectiveness of these systems depends on the quality of four additional stages (to these two phases). These four stages are: image acquisition, preprocessing, feature extraction and classification. There are many recent researches that use machine learning algorithms. This section presents some of the most relevant recent works presented by researchers.

In 2003, Xiaoou and Xiaogang [1], proposed a face recognition approach combining a Bayesian probabilistic classifier and Gabor filter. To evaluate this system, two image databases were used: XM2VTS database on which authors found an accuracy rate of 97.10% and AR database which gave an accuracy rate of 93.30%. In 2006 and 2010, authors [2] [3] have performed artificial neural networks in face recognition on feature vectors obtained with Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Authors obtained satisfactory results on the ORL dataset. Alaa et al, got accuracy rates of 95% with the PCA and 97% with the LDA. Mayank et al, obtained an accuracy rate of 97.01% combining PCA and ANN. In 2012, some approaches were implemented using random forests and two facial feature extractors: Wavelet Gabor and Histogram of Oriented Gradients (HOG) [4]. The results obtained on the ORL database were: 95.10% and 95.70% for HOG and Gabor, respectively.

In 2015, work presented by Vaishali et al, [5] showed how to use random forests. They used a combination of PCA, DCT, and DWT for preprocessing and feature vector extraction on the ORL database. They obtained an accuracy of 96%. A study using Support Vector Machine (SVM) as a classifier and PCA as extractor has achieved good recognition rate of 98.75% [6].

Other works [7] produced good results by applying Naive Bayes' Classifiers. With face representations done using PCA, a recognition rate of 94.35% on the Yale B is obtained; against a recognition rate of 94.56% using LDA as features extractor.

In 2016, some approaches [8] have proposed to combine PCA and LDA for feature extraction. Before, they have done image preprocessing using histogram equalization, image size normalization, and conversion of rgb or coloured images into grey scale images. The evaluation of these systems based on Artificial Neural Network (ANN) gave an accuracy rate of 100.00% on the ORL database. In [9], authors have performed face recognition using PCA, Local Binary Pattern (LBP) and SVM to achieve better performance. For data acquisition they have chosen Yale and ORL databases. The face region from each image is extracted using Viola-Jones algorithm followed by image resizing and cropping into 70×70 pixel. Authors applied contrast to enhance images. The feature extraction is performed separately with PCA and LBP. For classification, SVM are used to create models and then evaluate them. However, PCA+SVM obtained a precision rate of 86.67% on Yale database and a precision rate of 100.00% on ORL database. And, LBP+SVM gives 100% accuracy for both; Yale and ORL database [9].

In 2017, Vanlalhruaia et al, [10] have proposed a Logistic regression algorithm for face recognition system. They converted color images to gray scale, and removed noises using Local Window Standard Deviation (LWSD) and then Adaptive Thresholding to get good quality binary image. Researchers detected from from binary image probable face region by vertical and horizontal profiling and extracted features from them to make the classification using Logistic Regression technique. Author achieved good recognition rate of 100% on FEI database which is freely distributed for research. Authors C. Zhou et al, [11] also exemplified the use of Logistic Regression classifier. They extracted the signatures using PCA, and obtained good recognition rates of 93.33% on Yale database [12] and 96% on ORL database.

In 2018, Lahaw et al, [13] have presented methods using PCA, LDA, Independent Component Analysis (ICA) and Discrete Wavelet Transform (DWT). The Two-Dimensional Principal Component Analysis (2D-DWT) is a multi level decomposition technique used to preprocess images. It decomposes original image into four sub-bands low-low (LL), low-high (LH), high-low (HL) and high-high (HH). The low-low (LL) sub-band is used as input image for feature extraction process based on ICA, PCA and LDA algorithms. These features were then classified by using the SVM. The models DWT+PCA+SVM, DWT+LDA+SVM, DWT+ICA+SVM obtained on the ORL database the following rates of recognition respectively: 96%, 96%, 94.50%.

Kak et al, [14] applied DWT and PCA techniques for preprocessing and features extraction, and K-Nearest Neighbors (KNN) classifier for classification. They obtained 99.25% recognition rate on ORL database. Other approaches using KNN have been tested on face features obtained with PCA [15] using ORL dataset. Researchers found 92.47% recognition rate. In [16], Bala et al, also studies facial recognition using KNN classifiers. They developed the extraction stage with LDA, and this approach provided an interesting recognition rate of 93.70% on ORL database.

Huda et al, [17] proposed a method using Random Forest classifier (RF). First of all, the system utilized Viola Jones Algorithm to detect face from image. LBP and HOG descriptors are applied simultaneously to extract feature vectors. Classification is done by Random Forest classifier (RF). This system obtained a precision of 97.82% on Mediu staff database (Mediu-S-DB).

In 2019, three groups of researchers implemented facial recognition systems using SVM methods. Laith et al, [18], started with image preprocessing using two techniques: Normalization and Contrast stretching. Then, they applied DWT and PCA to reduce image size to half, remove noise and extract features for classification. Pranati et al, [19], also used PCA technique to extract the features and reduce dimensionality of the images. Putta et al, [20], applied Gabor Wavelet to extract rotation and scale invariant features from normalized face image and used PCA for feature reduction. These three authors have done the classification with SVM. So, on three data sets: ORL, AR and Grimace, Putta et al, obtained the following accuracy rates respectively: 97.65%, 92.31% and 100.00%. The models of Laith et al, achieved an accuracy of 95.20% for Yale database and 96.25% for ORL database. Pranati et al, performed a good accuracy rate of 92% on ORL database.

Sri et al. [21] and Kadek et al. [22], applied 2D-PCA for face representation. In addition, conversion to grayscale, region of interest (ROI) and haar cascade segmentation were applied for image preprocessing by Ni Kadek et al. All classifications are done using KNN method. On ORL database, an accuracy rate of 96.88% is obtained with 2D-PCA+KNN method [21], while PCA+KNN method performed a recognition rate of 81% on a dataset containing 790 faces from 158 people taken from several angles [22].

In 2021, Muhammad et al, [23], performed a comparative study between four traditional machine learning algorithms using the ORL database : PCA, 1-Nearest Neighbor (1-NN), LDA, and SVM. Researchers extracted features from the datasets and created the models. Finally, they evaluated the performance of these models with 5-fold cross validation (n=5) technique. Systems based on LDA, 1-NN, PCA and SVM achieved 96%, 96.25%, 96.75% and 98% accuracy rates respectively [23].

In 2022, ZM Nabat et al, [24], work on two algorithms to implement a face recognition system : SVM method and Rain Optimization Algorithm (ROA) that is inspired by the raindrops. This method can find global extremum as well as local extrema if its parameters are correctly tuned. In their study, the goal of ROA is to optimize the regularization parameter C for the SVM and the spread σ of the Radial Basis Function (RBF). Thus, to evaluate the proposed system, they used the Yale face dataset and n-fold validation (n = 10). Authors obtained a recognition rate of 86% [24].

Benradi et al, [25], performed a comparative study of machine learning algorithms in the field of face recognition. For image acquisition, they worked with two databases: ORL database and Sheffield face database which contains 564 images of 20 individuals where each image has an identical size of 220×220 pixels and a 256-bit grayscale. They applied feature extraction using Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Features from Accelerated Segment Test (FAST) and LBP. Then, classification is done using SVM, KNN, PCA, and 2D-PCA methods to create the prediction models. Results showed that the proposed techniques named SIFT+SVM, LDA+KNN, PCA, 2D-PCA performed, respectively, on ORL Dataset the following accuracy rates: 99.16%, 96%, 92.50% and 96.25% [25]. Also, On Shiefilled Dataset, the accuracy rates of these methods (SIFT+SVM, LDA+KNN, PCA, 2D-PCA) are, respectively 99.44%, 96%, 27.11%, 43.10%.

Many researchers on facial recognition use new techniques of Deep Learning : Convolutional Neural Networks (CNN) [26] [27]. The biggest advantages of CNN techniques are that they are typically composed of two basic parts: feature extraction and classification. LeNet-5 was one among the earliest CNN which promoted the event of deep learning [28].

Some works [29] on face recognition are implemented by simplifying a CNN structure to get a more efficient one. The proposed CNN has fused together the convolution and sub sampling to reduce the number of layers required. This architecture has four layers C1, C2, C3 and F4 (Output Layer). C1 layer has 5 feature maps, 14 feature maps in C2 layer and 60 feature maps in C3 layer. F4 layer has 40 feature maps. The design has a reduced size of feature maps, so as not to require padding in the convolution process. After some preprocessing steps, authors created the model and evaluated it using AT&T (ORL) database and 10 subjects from JAFFE database. On test dataset the classification accuracy reached 100%.

Meenakshi et al, [30] proposed a new CNN architecture because of the effect of pose and illumination changes, any occlusions, facial expressions, etc. The developed method is built by varying feature maps in convolutional layers C1, C2 and C3, with the aim of finding the most efficient architecture. This architecture has an input layer of 32×32 size, subsampling layer and fully connected layer. To evaluate this model, several tests are conducted on ORL database after image resizing to 32×32 pixels. Among different architectures, a best accuracy of 98.75% is obtained with 15-90-150 architecture [30].

Zhiming et al, proposed a model for face recognition [26]. Figure 1 shows the architecture of this model based on two main points: the number of hidden layer neurons and the number of convolutional layer feature maps. This CNN architecture includes, in order, input layer, convolution layer, pooling layer, convolution layer, pooling layer, fully connected layer and Softmax regression classification layer. So, they defined the structure C1-C2-H, where C1 is the number of feature maps in the first convolutional layer, C2 is the number of feature maps in the second convolutional layer, and H represents the number of hidden layer neurons. After multiple experimental tests, Zhiming et al, found the optimal model 36-76-1024; and they obtained a recognition rate reaches 100% [26] on ORL database.

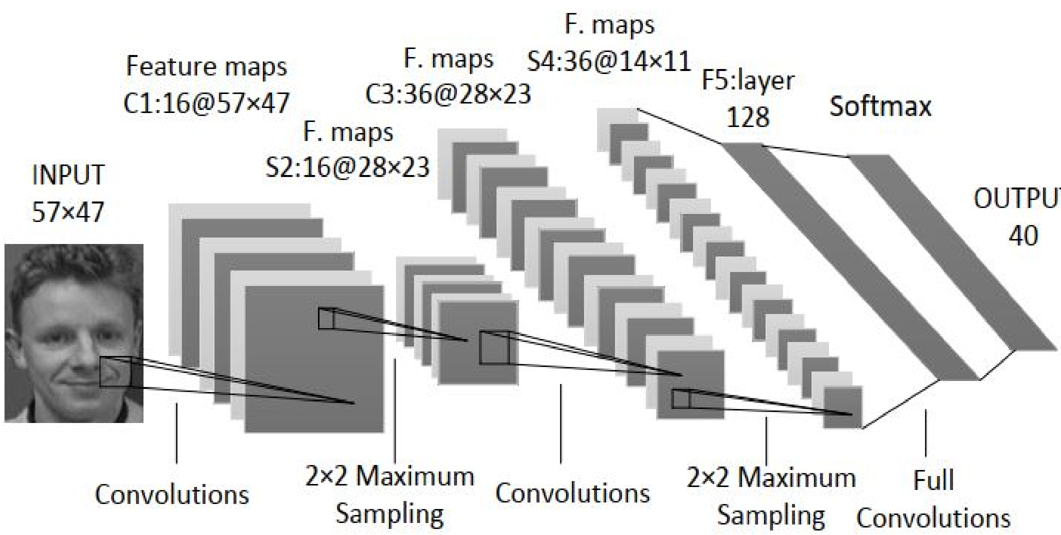


Figure 1: CNN infrastructure model designed for experiment. [26].

Yohanssen et al, [27], used Residual Networks-50 to perform a face recognition system. Residual Networks is a CNN that achieves 3.57% error on the ImageNet test set and won the first place on the ILSVRC 2015 classification task [31]. The contribution of this research paper is to determine effectiveness ResNet architecture using different configurations of hyperparameters such as the number of hidden layers, the number of units in the hidden layer, batch size, and learning rate.

To evaluate the model, they used a dataset of 1050 images divided into a training and testing datasets, giving 80% of images for training data and 20% for testing data. Yohanssen et al, have found that learning rate of 0.0001, epoch of 100 and step per epoch of 150 give a model with accuracy rate of 99%.

In 2022, a facial recognition approach based on Resnet-152v2 has been proposed by Arshi and Virendra in several steps [32]. Firstly, authors performed the grayscale images of dataset and converted images from pgm format into jpg format. Secondly, the original dataset of 400 images is splitted using a ratio of 70:30, into two sub datasets known as training dataset and testing dataset. The next step is to create the model using the three parameters that are Optimizer, Loss function and metrics. The model employs Adam as an optimizer and an epoch number sets to 25. The proposed approach produced 97% of face recognition accuracy on AT&T dataset.

Tables 1, 2, 3, 4, 5 and 6 show averages of some good performances of some machine learning techniques such as : SVM, ANN, KNN, LR, RF and CNN.

Authors	**Methods**	**Database**	**Accuracy (%)**
Sumita et al, 2016	PCA+SVM	ORL	100
Sumita et al, 2016	LBP+SVM	ORL	100
Putta et al, 2019	GABOR+PCA+SVM	ORL	100
Benradi et al, 2022	SIFT+SVM	ORL	99.16
Hadi et al, 2015	PCA+SVM	ORL	98.75
Muhammad et al, 2021	SVM	ORL	98
Putta et al, 2019	GABOR+PCA+SVM	ORL	97.65
Laith et al, 2019	DWT+PCA+SVM	ORL	96.25
Lahaw et al, 2018	DWT+PCA+SVM	ORL	96
Average Accuracy	—-	—-	98.42

Table 1: Average accuracy of some works using SVM in the recent literature.

Authors	**Methods**	**Database**	**Accuracy (%)**
Alaa et al, 2006	PCA+ANN	ORL	95
Alaa et al, 2006	LDA+ANN	ORL	97
Mayank et al, 2010	PCA+ANN	ORL	97.1
Gurleen et al, 2016	PCA+LDA+ANN	ORL	100
Average Accuracy	—-	—-	97.27

Table 2: Average accuracy of some works using ANN in the recent literature.

Authors	**Methods**	**Database**	**Accuracy (%)**
Kak et al, 2018	DWT+PCA+KNN	ORL	99.25
K.S.Maheswari et al, 2015	PCA+KNN	ORL	92.47
Bala et al, 2016	LDA+KNN	ORL	93.7
Muhammad et al, 2021	KNN	ORL	96.25
Benradi et al, 2022	LDA+KNN	ORL	96
Average Accuracy	—-	—-	95.53

Table 3: Average accuracy of some works using KNN in the recent literature.

Authors	Methods	Database	Accuracy (%)
C.Zhou et al, 2015	PCA+Logistic Regression	ORL	96
C.Zhou et al, 2015	PCA+Logistic Regression	ORL	93.33
Average Accuracy	—-	—-	94.66

Table 4: Average accuracy of some works using Logistic Regression in the recent literature.

Authors	Methods	Database	Accuracy (%)
Abdel Ilah et al, 2012	HOG+RF	ORL	95.1
Abdel Ilah et al, 2012	Gabor+RF	ORL	95.70
Vaishali et al, 2015	PCA+DCT+DWT+RF	ORL	96
Average Accuracy	—-	—-	95.60

Table 5: Average accuracy of some works using RF in the recent literature.

Authors	Methods	Database	Accuracy (%)
Meenakshi et al, 2019	DL	ORL	98.75
Zhiming et al, 2019	DL	ORL	100
Arshi and Virendra, 2022	DL-ResNet152V2	ORL	97
Average Accuracy	—-	—-	98.53

Table 6: Average accuracy of some works using Deep Learning (DL) in the recent literature.

Results of the state of the art algorithms are summarized in the Figure 2 (that contains average recognition rates per algorithm). This analysis placed the Convolutional Neural Networks (CNN) at the top of the ranking, with a recognition rate of 98.53%. They are followed, in order, by Support vector machines (SVM), Artificial Neural Networks (ANN), Random Forests (RF), K-Nearest Neighbors (KNN) and Logistic Regression (LR) which provided 98.42%, 97.27%, 95.60%, 95.53%, 94.66% as average recognition rate respectively. These results will be compared with those of experimental research.

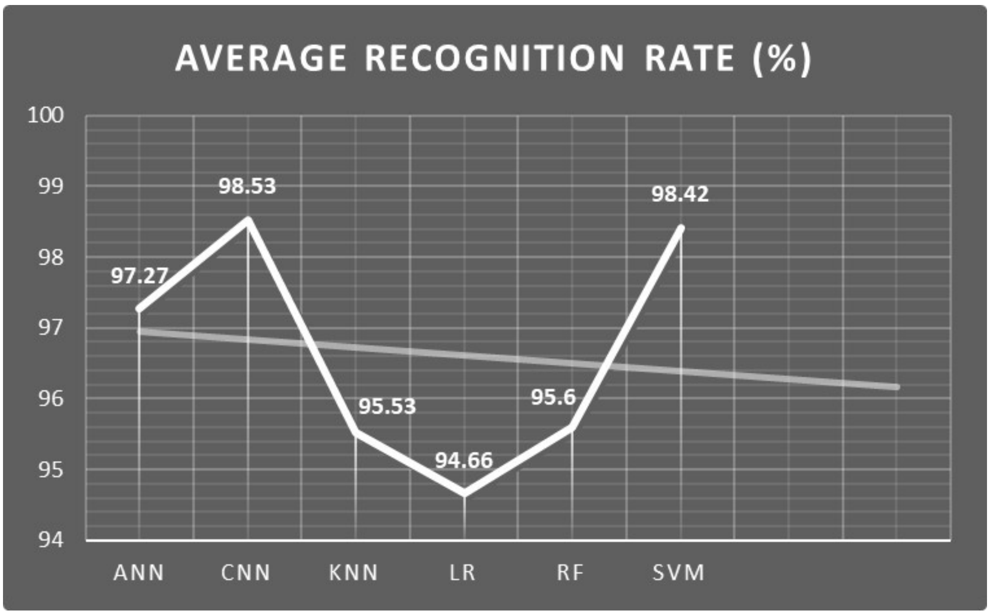


Figure 2: Performance of machine learning algorithms in recent literature.

III MACHINE LEARNING ALGORITHMS

Machine Learning is defined as the science of programming computers to learn from data [33]. It can be subdivided into supervised and unsupervised learning. Supervised learning is defined as the use of labeled data and the creation of prediction models. It can be divided into two types of problems: classification and regression. In this paper, machine learning techniques studied are Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Logistic Regression (RF), Random forest (RF), Naive Baye's Classifiers (NB), Artificial Neural Network (ANN) and Convolutional Neural Networks (CNN).

3.1 Support vector machines (SVM)

Support vector machines are powerful supervised machine learning algorithms, capable of performing binary classifications for linear or non-linear problems. They were introduced by Vladimir Vapnik in 1965 [9]. In addition to performing linear classification, the extension to non-linear problems using kernel trick was discovered in 1995 by Vapnik and Corinna [6] [34]. Given training samples (x_i, y_i) where x_i are observations and $y_i \in \{-1, +1\}$ are labels; this dataset is linearly separable when there exists a hyperplane whose equation is a linear function in x as:

$$f(x) = w^*.x + b^* \tag{1}$$

where w is a weight vector, b is the bias and x is the problem variable. To decide to which class a data item x_i belongs, we just take the sign of the decision function [35], thus:

- if $w^*.x_i + b^* > 0$ then x_i belongs to the positive label class +1
- if $w^*.x_i + b^* < 0$ then x_i belongs to the negative label class -1.

There have been many various hyperplanes that are able to separate the data points. The basic idea of SVM methods is to obtain the best function f(x) that maximizes the margins of the two classes. The margin is the distance between the hyperplane and the observations closest to the hyperplane (support vectors: the points that are closest to the hyperplane). The optimal hyperplane is selected so as to maximize the margin. And, to obtain the largest possible margin between the hyperplane and the support vectors, the minimization of the following function is needed :

$$L = \frac{1}{2}\|w\|^2 \tag{2}$$

under the constraints $y_i f(x_i) \geq 0$ which verify that the data (x_i, y_i) are well classified.

3.2 K-Nearest Neighbours (KNN)

The K-Nearest Neighbors algorithm (KNN) is one of the simplest classifiers in supervised machine learning [36] [15]. This lazy learner technique doesn't technically train a model [37] [38]. So, the training stage doesn't require any computation. It consists of storing the training dataset with little or no processing. However, the test stage requires an intensive distance calculation between a test data point and all training data points, in order to find the k nearest neighbors of this test data. The distance is so important for the KNN technique. The most often used distance, is the Minkowski distance of order q where q is an integer. Between two data points $U = (U_1, \ldots, U_n)$ and $V = (V_1, \ldots, V_n)$, this distance is defined by the following formula :

$$d_q(U, V) = (\sum_{i=1}^{n} |U_i - V_i|^q)^{\frac{1}{q}} \tag{3}$$

It is a generalization of both the Manhattan distance (order $q = 1$) and the Euclidean distance (order $q = 2$). The formulas of these two particular distances are defined by:

- Manhattan distance is the distance a car would drive in a city (Manhattan) :

$$d_1(U, V) = \sum_{i=1}^{n} |U_i - V_i| \tag{4}$$

- Euclidean distance is the length of line segment between the two points :

$$d_1(U, V) = \sum_{i=1}^{n} (U_i - V_i)^2 \tag{5}$$

The KNN algorithm is widely used to determine to which class a new data belongs to. As shown in Figure 3, the predicted class is typically the class that is the most voted in the k nearest neighbors of this data (majority vote of its neighbors) [34].

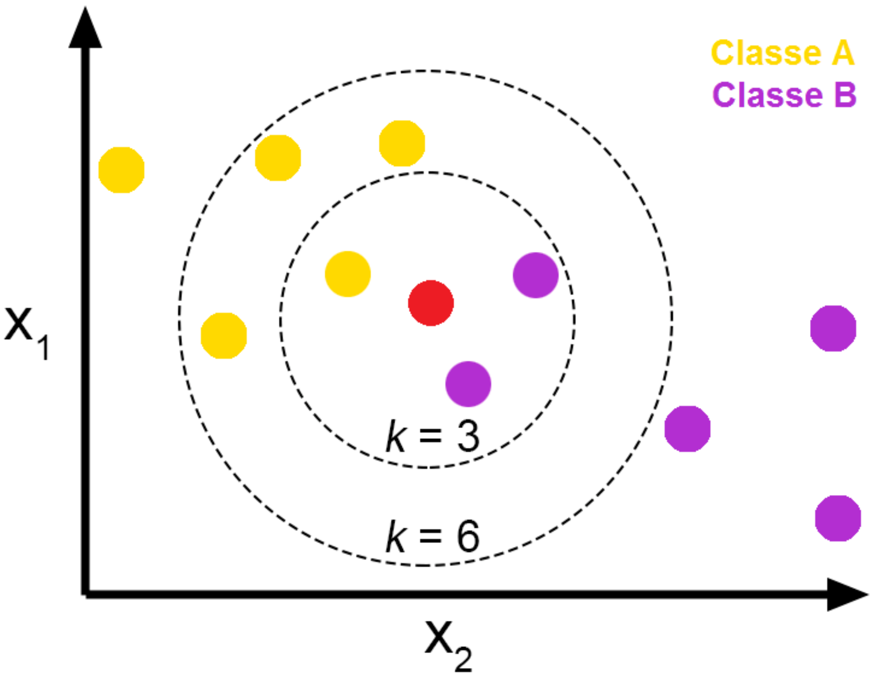


Figure 3: K-Nearest Neighbours Algorithm.

3.3 Logistic regression (LR)

The logistic regression is actually a widely used supervised classification technique [37]. This originally binary, and later multiclass classification algorithm began its emergence in 1989 with Hosmer and Lemeshow [11]. Logistic regression uses a function that takes as argument a linear combination of the input variables and models the probability of belonging to a class. This logistic regression model computes a weighted sum of the input features, and a linear model like u(x) is included in a logistic function $\rho(u)$, sigmoid function that returns values between 0 and 1 [34] [38]. These functions are defined by these following formulas:

$$U(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p \tag{6}$$

$$\rho(u) = \frac{1}{1 + e^-u} \tag{7}$$

$$\rho(u) = \frac{1}{1 + e^-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p)} \tag{8}$$

Instead of providing the result directly, the logistic function $\rho(u)$, provides the logistics of the linear model $u(x)$.

3.4 Random Forest (RF)

As shown in Figure 4, Random Forests is a collection of trees-based models trained on random subsets of the training data. This algorithm is an ensemble of decision trees, usually trained with the bagging method (or sometimes pasting). A method is called bagging when sampling is performed with replacement. The first version of this technique was created in 1995 by Tin Kam Ho. Then, in 2001 [34], an extension of the decision tree forests was developed by Leo Breiman and Adele Cutler who registered the Random Forests as a trademark in 2006. The Random forests can be used for both classification and regression tasks. And, for both supervised learning settings, a collection of labeled observations known as the training set like $D = (x_i, y_i), \dots, (x_n, y_n)$ was provided, each x_i is a vector and y_i its true target variable.

- Classification:

For classification tasks, the prediction of the random forest is the most dominant class among the predictions made by the individual trees. Consider T trees in the forest, and consider that a class called m has received a number of votes called V_m defined by the following formula:

$$V_m = \sum_{t=1}^{T} I(\hat{y}_t == m) \tag{9}$$

where, $\hat{y}_t$ is the prediction of the t-th tree on a particular random subsets of the training data. The function $I(\hat{y}_t == m)$ takes on the value 1 if the condition is met, else it is zero. The final class predicted by the algorithm is the class with the most votes as shown in the following formula where $m \in \{1, \dots, M\}$:

$$\hat{y} = \underset{m \in \{1,...,M\}}{\arg\max} \; V_m \tag{10}$$

- Regression:

For regression tasks, the prediction of the random forest is the mean or average of the predictions made by the individual trees. Consider T trees in the forest, and each predicts $\hat{y}_t$, then the final prediction $\hat{y}$ is:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t \tag{11}$$

Algorithm 1 Random Forest for Classification (RFC)

for $b \leftarrow 1, B$ **do**
  (a) Draw a bootstrap sample Z^* of size N from the training data.
  (b) Grow a random forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
  (I) Select m variables at random from the p variables.
  (II) Pick the best variable/split-point among the m.
  (III) Split the node into two daughter nodes.
end for
Output the ensemble of trees $\{T_b\}_1^B$
Make prediction at new point x:
Let $\hat{C}_b(x)$ be the class prediction be the class prediction of the bth random forest tree. Then $\hat{C}_r f^B(x) =$ majority vote $\{\hat{C}_b(x)_1^B\}$.

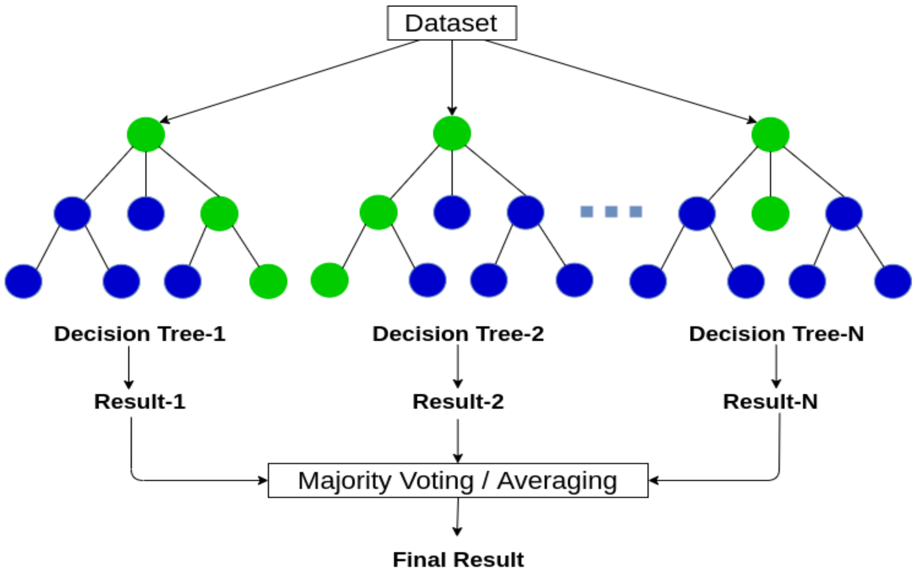


Figure 4: Random Forest Algorithm.

3.5 Naive Baye's Classifiers (NB)

Naive Baye's Classifier is a supervised classification algorithm based on two fundamental ideas: the first is the naive assumption that all features are conditionally independent of each other, and the second is the use of Bayes' theorem which is discovered by Thomas Bayes in the 18th century and defined by [34] [38]:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{12}$$

A and B are events; $P(A)$ is the probability of observing event A, and $P(B)$ is the probability of observing event B. $P(A|B)$ is the conditional probability of observing A given that B was observed [38].

In machine learning, the Naive Baye's Classifier is based on Bayes' theorem that is reworded and given the following more natural formula for a classification task [37] :

$$P(Y|X_1, X_2, \ldots, X_j) = \frac{P(X_1, X_2, \ldots, X_j|Y)P(Y)}{P(X_1, X_2, \ldots, X_j)} \tag{13}$$

Where, $P(Y|X_1, X_2, \ldots, X_j)$ is the posterior, the probability that an observation is class y given the observation's values for the j features, $X_1, X_2, \ldots, X_j$. $P(X_1, X_2, \ldots, X_j|Y)$ is the likelihood of an observation's values for features, $X_1, X_2, \ldots, X_j$, given their class, y. P(y) is the prior, the probability of class y before looking at the data. $P(X_1, X_2, \ldots, X_j)$ is the marginal probability and is the probability of observing a particular feature in the training set; it is constant for all dataset. That is why the comparison of an observation's posterior values for each possible class, is focused on the numerators of the posterior for each class. For each observation, the class with the greatest posterior numerator becomes the predicted class, ŷ.

3.6 Artificial neural networks (ANN)

Artificial Neural Networks (ANN) are a family of machine learning techniques which are inspired by the biological neural networks that constitute the human brain [8]. By using a simple node called formal neuron carrying a value between 0 and 1, called activation, it is possible to create a network composed of neurons organized in three different layers: the input layer, the hidden layers and the output layer shown in Figure 5. The neurons of a layer are connected to the neurons of the adjacent layers by links carrying weights which play an important role in the training of the network. There is a diversity of architectures and training algorithms for neural networks. The Multi-Layer Perceptron (MLP) invented by Frank Rosenblatt at the Cornell Aeronautical Laboratory in the late 1950s [38], is considered as the first simple and complete neural network [34]. This network, based on the concept of gradient backpropagation as a training algorithm, becomes one of the most used and productive model in various domains, such as image recognition, data classification.

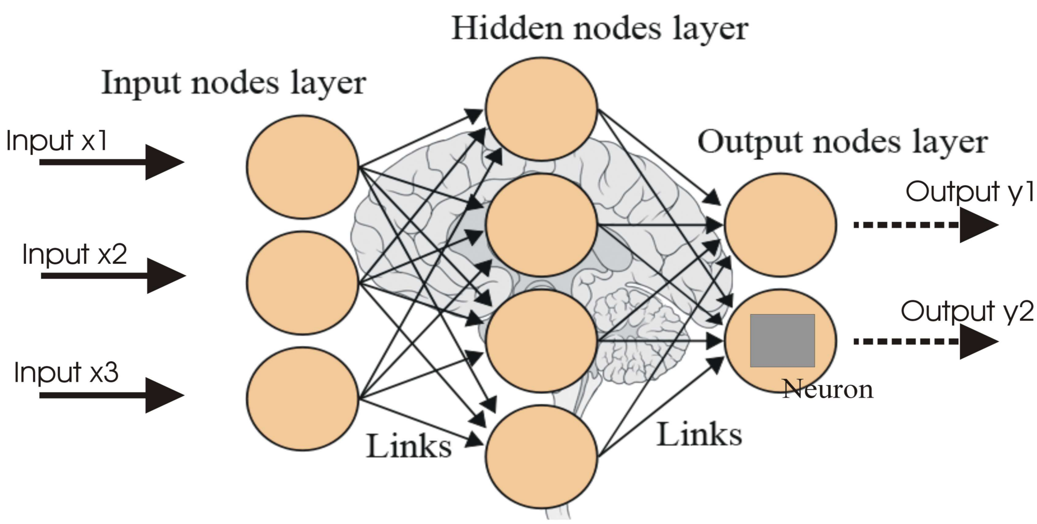


Figure 5: Artificial Neural Network Structure.

3.7 Convolutional Neural Networks (CNN)

A convolutional neural network (CNN or convnet) is one of the various types of artificial neural networks (ANN) which are specifically used for image classification tasks. Compared to machine learning based on manual feature extraction, deep learning (DL) is free of feature extraction. And, instead of feeding manually extracted features, images were fed automatically into DL algorithms such as Convolutional Neural Networks (CNN) that are powerful techniques for analysing visual imagery. [39]. Generally, a CNN consists of three main neural layers, which are convolutional layers, pooling layers, and fully connected layers [40]. These different kinds of layers play different roles. There are Several CNN architectures that were used in the recent literature.

LeNet is the first Convolutional Neural Network (CNN) proposed by Yann et al, in 1998 [41]. It was used mainly to recognize digits and handwritten numbers on bank checks. It is composed of seven (7) layers including three (3) convolution layers, two (2) pooling layers, two (2) fully connected layers [41]. Input size what LeNet accepted is $32 \times 32 \times 1$ and this one refer to greyscale image. The ability to process higher resolution images like ORL dataset requires larger and more convolutional layers, and the availability of computing resources.

Proposed by Krizhevsky et al, AlexNet was close to LeNet, but was deeper, with more filters per layer, and with stacked convolutional layers directly on top of each other [42]. It won the ImageNet visual recognition competition (ILSVRC-2012) in 2012 by achieving a top-5 error of 15.3%. AlexNet architecture consists of 1 input layer, 5 convolutional layers, 7 nonlinear activation function ReLU activation layers, 3 max-pooling layers, 2 normalization layers, 2 fully connected layers, 1 softmax layer and 1 output layer.

When there are more layers in network, as the gradient is backpropagated to earlier layers, repeated multiplication may make this gradient infinitely small. It is called vanishing gradient problem and it makes Deep networks hard to train. That is why the ResNet introduced the concept of skip connection by adding the original input to the output of the convolution block, to solve this vanishing gradient problem. ResNet, short for Residual Network was introduced in 2015 by He et al,[43] . It won the first place in the ILSVRC 2015 classification competition with a top-5 error rate of 3.57%. This block is a stack of layers so that an input x of the block is directly added to the output of the block in the network : Y = F(X)+X. There are many variants of ResNet architecture like ResNet-18, ResNet-34, ResNet-50, ResNet-152.

To solve the vanishing gradients problem, in 2017, Huang et al, introduced Densely connected convolutional network (DenseNet) with two main ideas: connect each layer to every subsequent layer and connect feature maps through concatenation not through summation [44]. DenseNet consist of a stack of dense blocks and transition layers. A dense block is a group of layers connected to all their previous layers. For each layer, the feature maps of all the preceding layers are used as inputs, and its own feature maps are used as input for each subsequent layer. Each layer has direct access to the gradients of the loss function and the original input signal. A single layer consists of a Batch Normalization, a ReLU activation and 3×3 Convolution (BN–ReLU–Conv). A transition layer is made of Batch Normalization, 1×1 Convolution layer and Average pooling layer.

IV METHODS

4.1 Face recognition databases

For implementation and evaluation of facial recognition systems, face image databases are needed. The Olivetti Research Ltd (ORL) database of faces contains 400 images from 40 distinct subjects was introduced by Samaria and Harter in Parameterization of a stochastic model for human face identification [45]. It has been gathered from 1992 to 1994 in AT & T laboratory of Cambridge University in England [6] [20]. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The size of each image is 92×112 pixels, with 256 grey levels per pixel.

The Yale Face Database is designed at the CVC (Center for Computational Vision and Control) of Yale University in a fully controlled environment [9]. It contains 165 grayscale images in GIF format of 15 individuals [46]. There are 11 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.

Labeled Faces in the Wild (LFW) is an image database containing face photographs, collected from the web especially for studying the problem of unrestricted face recognition [47]. LFW dataset was developed and maintained by researchers at the University of Massachusetts, Amherst. It was released for research purposes to make advancements in face verification. So, 13233 images of 5749 people were detected and centered by the Viola Jones face detector. And, 1680 of the people pictured distinctly appear in two or more photos in the data set. LFW includes different sets of images, including the original and three types of aligned images.

To facilitate future face detection research, some authors [48] introduced the WIDER FACE dataset, which is 10 times larger than existing datasets. The WIDER (Web Image Dataset for Event Recognition) FACE dataset is a face detection benchmark dataset. It contains 61 event categories and around 50574 images annotated with event class labels. Images have a high degree of variability in scale, pose and occlusion as depicted in the sample images.

4.2 Image processing and feature extraction techniques

Image pre-processing and feature extraction techniques are mandatory for any image-based applications. They influence the performance of systems. In this work, several efficient techniques are applied. Discrete Wavelet Transform (DWT) is widely used in image compression. Principal Component Analysis (PCA) is generally considered as the main approach to reduce the dimension of dataset. It was introduced by M. A. Turk and M. P. Pentlanden in 1991 [49] [50]. Linear Discriminant Analysis (LDA) appeared in the work of Belhumeur in 1997 at Yale University in the USA [50]. It is a good dimension reduction technique always used before a classification algorithm. GABOR filter was introduced by Dennis Gabor in 1946, then improved in 1980 by John G. Daugman [51]. This filter is used to extract useful local facial features from an image. It is efficient and robust to illumination variations and geometric transformations such as facial expression.

V METHODOLOGY

For face recognition systems, an experimental methodology in different steps is used. The first step consists of preparing data using ORL and YALE image databases. These databases are divided into two distinct groups using the most common split ratio 80:20. That is 80% of the dataset goes into the training set and 20% of the dataset goes into the testing set. In the second step, algorithms such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Gabor filter associated with PCA, Discrete Wavelet Transform (DWT) associated with PCA, are used to preprocess and extract the face feature vectors. In the third step, a classification prediction model was formed from the Train part by applying one of the classification methods used (SVM, KNN, RF, LR, NB, and ANN). After this step, models are trained, and saved for evaluation using the testing dataset and some performance measures. The diagram of the Figure 6 explains the experimental methodology. For Convolutional Neural Networks (CNN), after preparing data, all images are converted into jpg format and resized following 100×100 resolution. Then, data were normalized by subtracting the mean of each feature and a division by the standard deviation.

According to the CNN architecture applied such as AlexNet, DenseNet, LeNet and ResNet, models have employed accuracy as a metric to view accuracy score, Adam as an optimizer and categorical cross entropy as a loss function. Parameters like epochs is set to 25 or 50 and the bacthsize is set to 32.

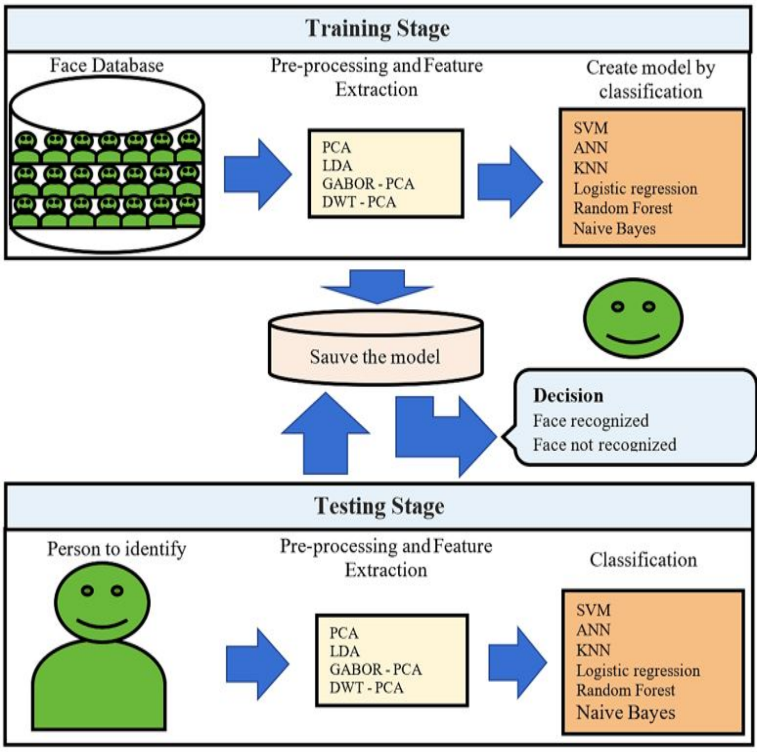


Figure 6: Diagram of the experimental methodology.

VI PERFORMANCE MEASURES

Three performance measures are simultaneously used to evaluate these machine learning models. First, the recognition rate which is the quotient of the number of correct classifications by the total number of face images tested [15].

Second, the confusion matrix that counts how many times an observation from a class A was classified into a class B [34]. Several measures are defined based on the confusion matrix. One is the recall, also called sensitivity, which is the True Positive rate. Another measure often associated with this matrix, the precision, which is the rate of correct predictions among the positive predictions. The simultaneous use of these two indicators gives a good measure called F-measure: harmonic mean of precision and recall. It is defined as:

$$F - measure = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{14}$$

The ROC curve plots graphically the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings [34]. In machine learning for model comparison, researchers focus, most often, on the use of the Area Under the Curve (AUC). In this paper, the average value of these three performance measures is used.

VII RESULTS AND DISCUSSION

The performance evaluation of the different methods was carried out on two ORL [45] and YALE [46] face databases. To better present the classification results, metrics used are F-measure, area under the ROC curve (AUC) and accuracy rate. The final performance of each algorithm is the average value of the accuracy rate, the f-measure and the area under the ROC curve (AUC), defined by the following formula:

$$Performance = \frac{Accuracy + F\_measure + AUC}{3} \tag{15}$$

7.1 Classification methods based on ORL Database

Figures 7, 8, 9, 10, 11 and Table 8, shows the different classification results of six machine learning traditional classifiers (SVM, KNN, RF, LR, ANN and NB algorithms) and four deep learning networks (AlexNet, LeNet, DenseNet and ResNet).

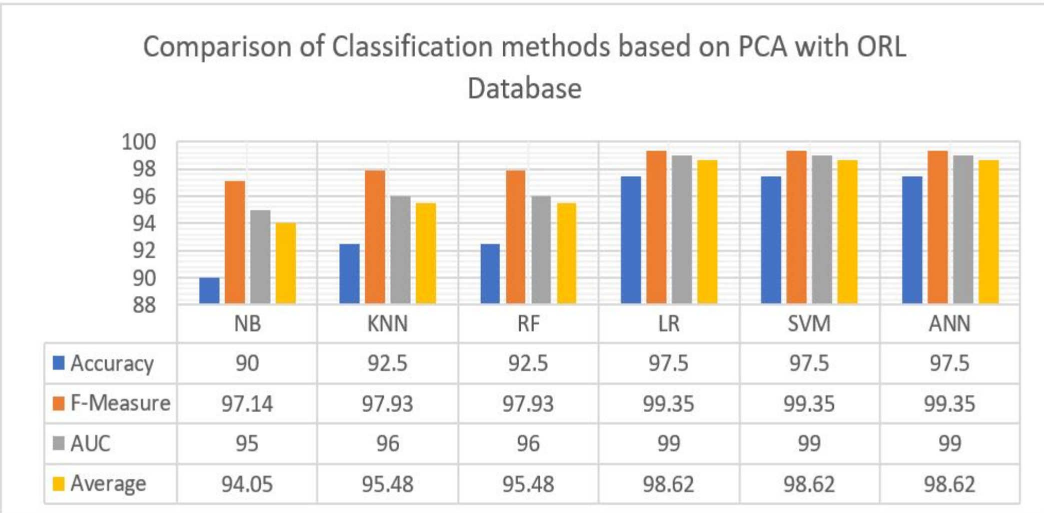


Figure 7: Performance of some classifiers on ORL database using PCA.

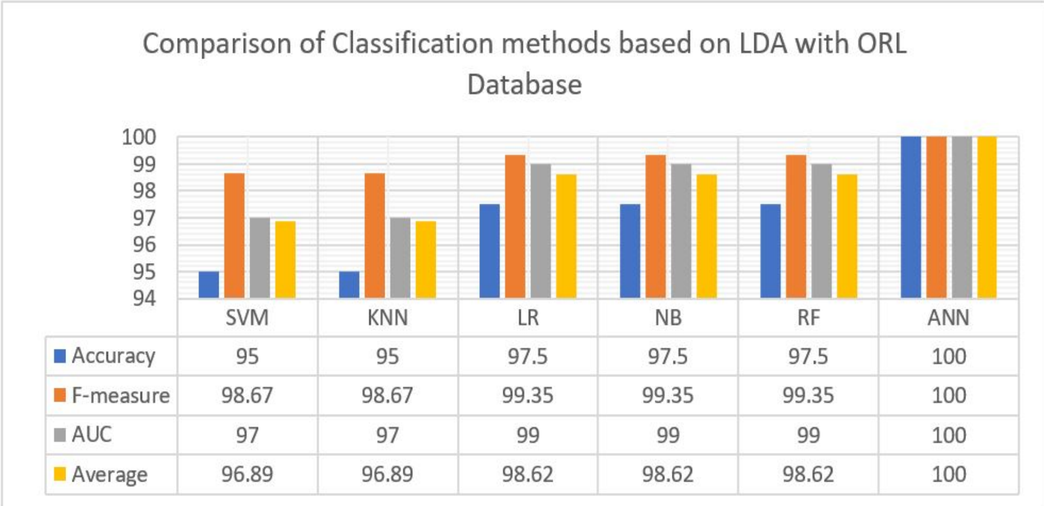


Figure 8: Performance of some classifiers on ORL database using LDA.

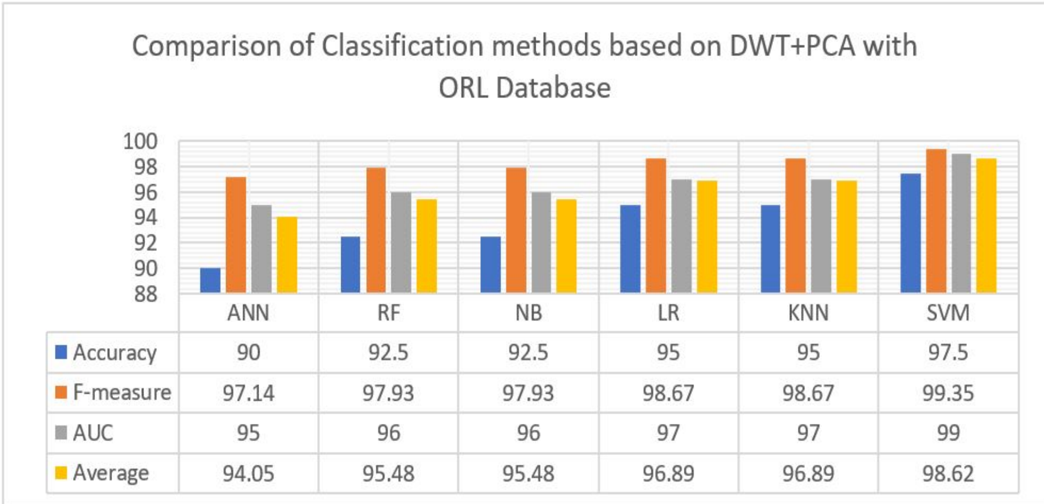


Figure 9: Performance of some classifiers on ORL database using DWT and PCA.

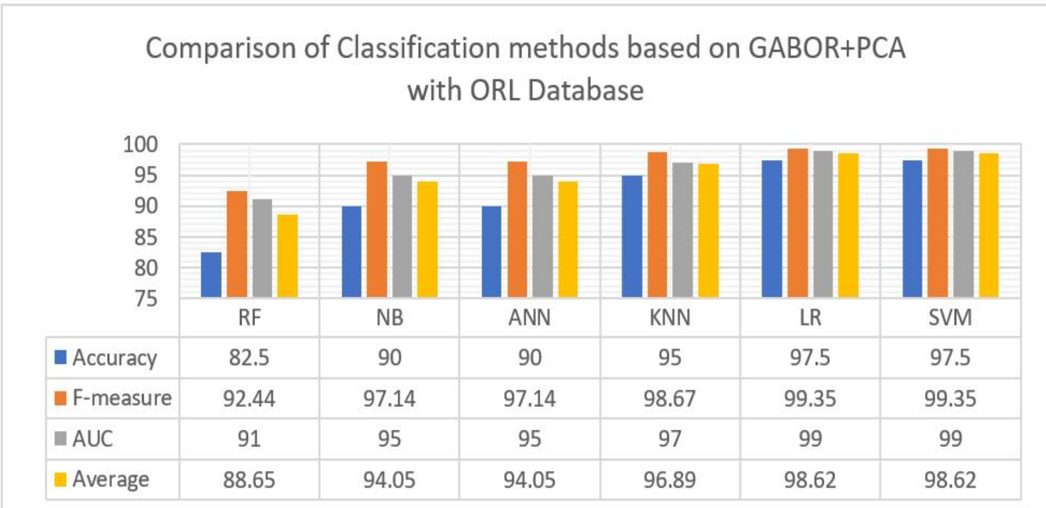


Figure 10: Performance of some classifiers on ORL database using GABOR and PCA.

The Table 7 summarizes the average of the three performance measures for different classifiers associated with the different preprocessing and extraction techniques. Overall, the performances of machine learning models are satisfactory, with ORL database. Therefore, SVM has a superior performance, globally. It resulted into an average score of 98.19%. It is immediately followed by logistic regression (LR) with an average of 97.76%, then artificial neural networks (ANN) with an average of 96.68%. In this work Random Forest algorithm achieved the lowest performance, 94.56%. From the discussion it is clear that, the SVM classifier has comparatively higher performance than the other traditional machine learning classifiers (RF, NB, KNN, ANN, LR). Thus, using the same criteria, a comparison of the performances of SVM and CNN is interesting.

ALGORITHM	**PCA**	**LDA**	**DWT+PCA**	**GABOR+PCA**	**Average (%)**
RF	95.48	98.62	95.48	88.65	94.56
NB	94.05	96.89	95.48	94.05	95.12
KNN	95.48	96.89	96.89	96.89	96.54
ANN	98.62	100	94.05	94.05	96.68
LR	98.62	96.89	96.89	98.62	97.76
SVM	98.62	96.89	98.62	98.62	98.19

Table 7: Comparative Analysis of Face Recognition Approaches on ORL Database using different classifiers with average values of F-measure, accuracy and AUC.

As shown in Figure 11, based on the model performances, LeNet (100%), AlexNet (100%), DenseNet (100%) and ResNet (98.04%) performed more satisfactorily compared to the value of the tested metric of SVM (98.19%), on ORL database. It is therefore recommended to apply LeNet, AlexNet, DenseNet and ResNet for face recognition.

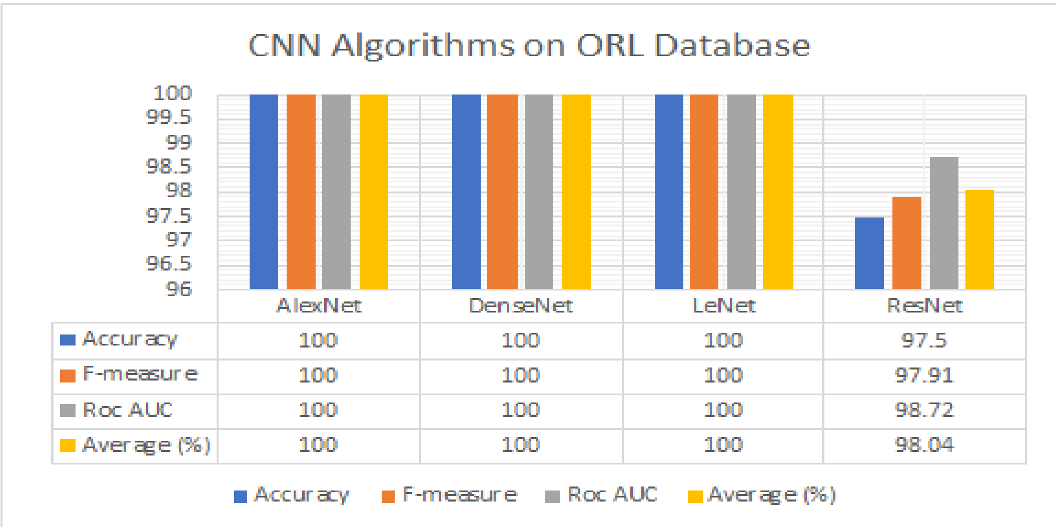
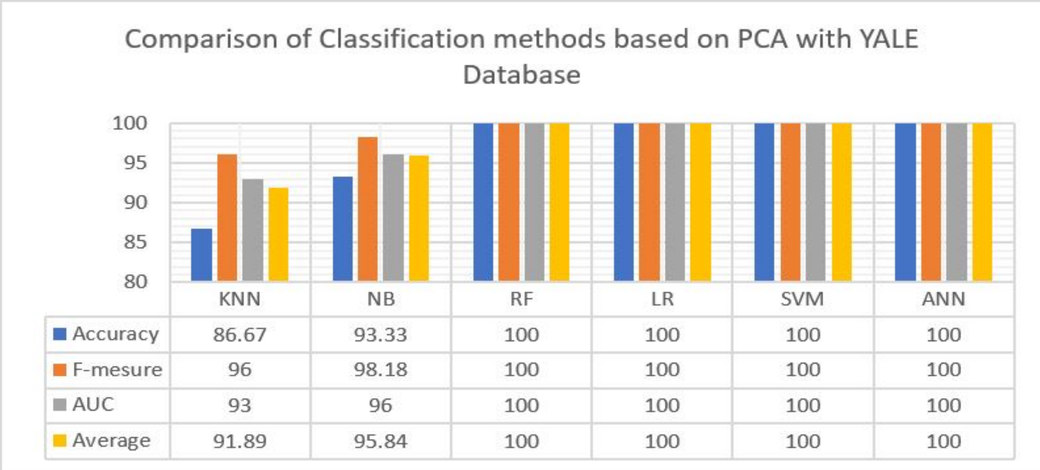


Figure 11: Performance of some CNN algorithms on ORL database.

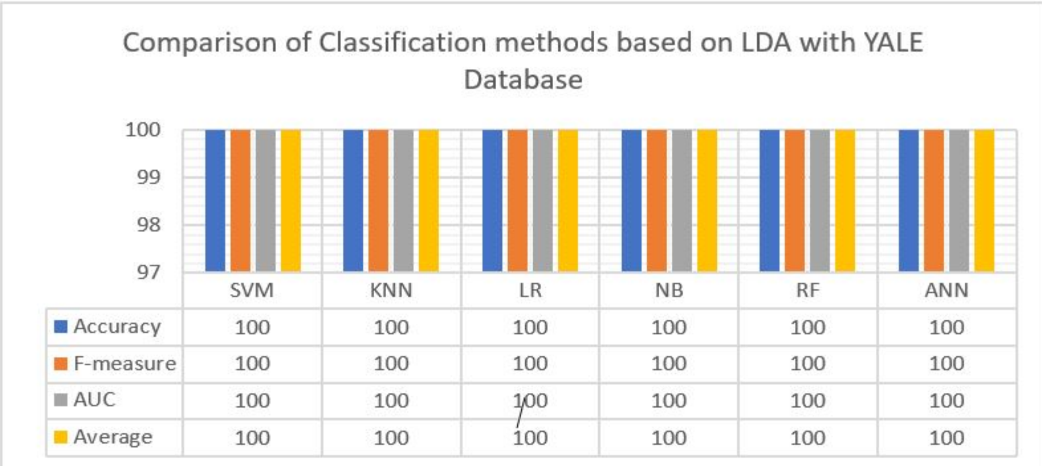
7.2 Classification methods based on YALE Database

Results of all techniques are represented in Figures 12, 13, 14, 15, 16 and Table 8.



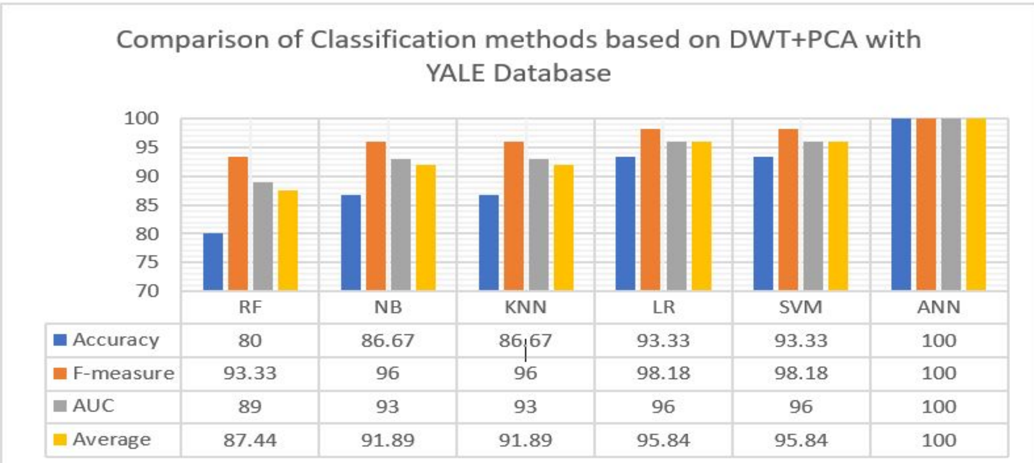
	KNN	NB	RF	LR	SVM	ANN
Accuracy	86.67	93.33	100	100	100	100
F-mesure	96	98.18	100	100	100	100
AUC	93	96	100	100	100	100
Average	91.89	95.84	100	100	100	100

Figure 12: Performance of some classifiers on YALE database using PCA.



	SVM	KNN	LR	NB	RF	ANN
Accuracy	100	100	100	100	100	100
F-measure	100	100	100	100	100	100
AUC	100	100	100	100	100	100
Average	100	100	100	100	100	100

Figure 13: Performance of some classifiers on YALE database using LDA.



	RF	NB	KNN	LR	SVM	ANN
Accuracy	80	86.67	86.67	93.33	93.33	100
F-measure	93.33	96	96	98.18	98.18	100
AUC	89	93	93	96	96	100
Average	87.44	91.89	91.89	95.84	95.84	100

Figure 14: Performance of some classifiers on YALE database using DWT and PCA.

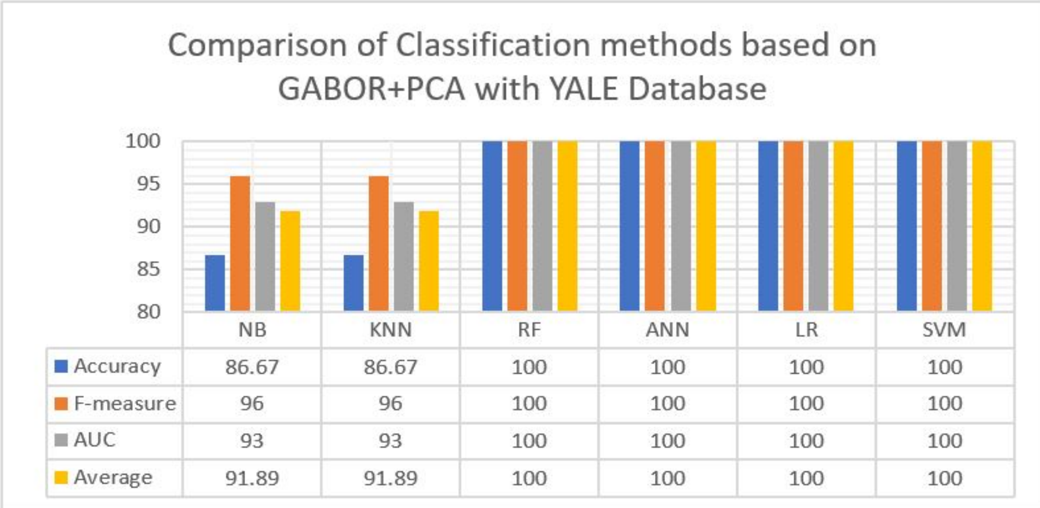


Figure 15: Performance of some classifiers on YALE database using GABOR and PCA.

According to the experiments summarized in Table 8 and Figure 16, results showed that classical ML techniques performed well on YALE database. The overall performances are above 93.90% for six traditional machine learning algorithms. However, ANN model (100%) is securing higher performance rate compared to all techniques such as SVM (98.96%), LR (98.96%), RF (96.86%), NB (94.05%), KNN (93.91%). It is observed that deep learning techniques are weak compared to traditional machine learning techniques based on manual feature extraction, in identification in the case of YALE database. The performance rates reached values of 69.97% for DenseNet, 82.28% for AlexNet, 84.33% for ResNet and 97.37% for LeNet which becomes the most effective method while DenseNet is the most ineffective as shown in Figure 16.

ALGORITHM	PCA	LDA	DWT+PCA	GABOR+PCA	Average (%)
RF	100	100	87.44	100	96.86
NB	95.84	100	91.89	91.89	94.05
KNN	91.89	100	91.89	91.89	93.91
ANN	100	100	100	100	100
LR	100	100	95.84	100	98.96
SVM	100	100	95.84	100	98.96

Table 8: Comparative Analysis of Face Recognition Approaches on YALE Database using different classifiers with average values of F-measure, accuracy and AUC.

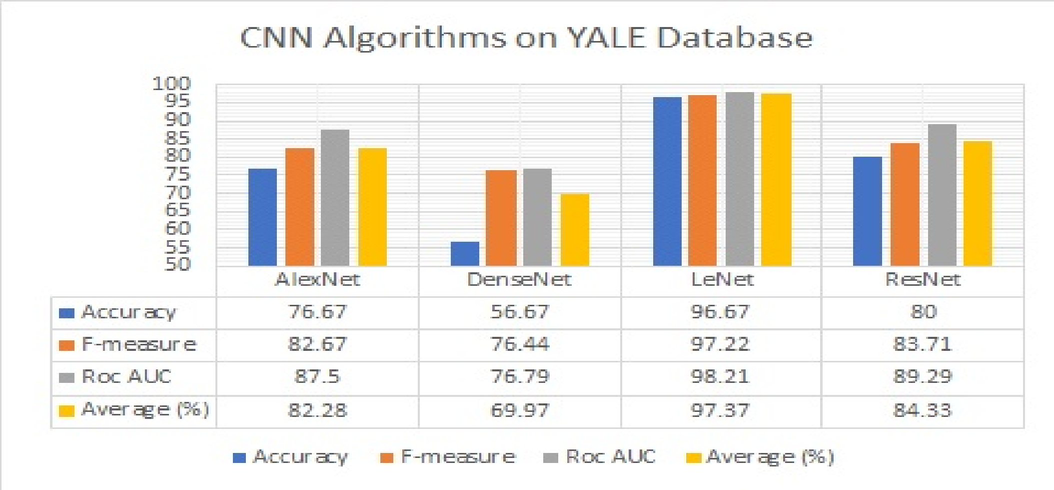


Figure 16: Performance of some CNN algorithms on YALE database.

It is noticed that on the YALE database, the results are good for the classical machine learning algorithms; while they are very bad especially for the deep learning techniques. This is due to the size of this database. Large number of samples make a dataset suitable especially for applying deep learning algorithms [52]. DL is the way to follow for image classification problems with relatively large datasets. For DL model, a small dataset like YALE face database, could be expanded with more diverse images, in order to allow it to generalize better. This technique is called image augmentation [39]. Thus, as an example, the original YALE database contains 165 images. After applying the image enhancement techniques with parameters such as rotation_range (40), shear_range (0.2), zoom_range (0.2), horizontal_flip (True) and brightness_range (0.5, 1.5), this database reached to 1155 images. Then, this preprocessed dataset was split into training and testing subsets with proportions of 80% (945 images) and 20% (210 images). On these datasets, the order of the classification results for the tested deep learning algorithms, from high to low, was DenseNet (96.83%) , ResNet (94.43%) and finally AlexNet (92.16%), as shown in Figure 17. These results prove that the deep learning techniques are also efficient on YALE database. It could be concluded that deep learning techniques almost always outperform classical algorithms when given reasonably large dataset. But with small size datasets traditional machine learning algorithms are preferable.

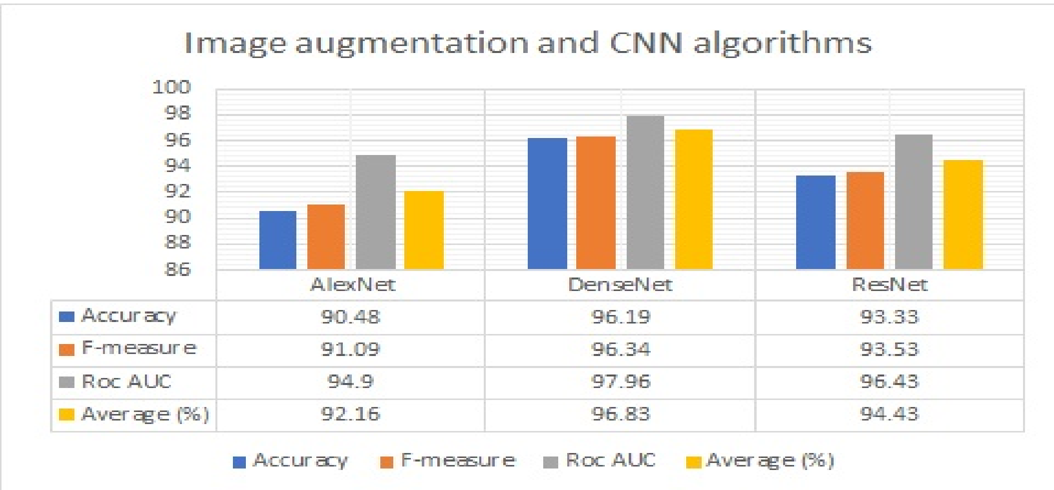


Figure 17: Performance of some CNN algorithms on YALE database with image augmentation.

Also, hybrid methods exist. It is possible to use a hybrid approach to perform face recognition task. Some examples of experimental works use convolutional neural network and support vector machine [53] or a convolutional neural network combined with feature extraction techniques [54]. CNN can be used as a feature extractor and then algorithms as support vector machine applied for classification task. Soad and al, obtained accuracies between 94% and 100% for all models using this hybrid approach on various datasets as Georgia Tech face dataset, FEI faces, GTAV face, YouTube face, LFW, F_LFW, ORL, and DB_Collection [53]. Another hybrid model based on the combination of feature extractor plus CNN can provide excellent performance. Hicham and al, performed this technique with two databases : ORL and Sheffield faces. Their results show that the hybrid model named SIFT(scale invariant feature transform)+CNN achieves an overall accuracy of 100% [54]. So, these experimental results show that the hybrid approaches yielded an excellent performance. In the future, these techniques will be considered.

VIII CONCLUSION

The context is favorable for the development of facial recognition systems. Technological evolution, powerful machines of this twenty-first century, the advent of artificial intelligence are the important factors which allow researchers to focus their work in this flied. There were many approaches using machine learning algorithms in the recent literature with different results [15] [11] [4] [1] [35]. That is why, this paper made a comparative study of these approaches in order to determine the more efficient technique for face recognition. Using a specific methodology, many systems are implemented with different supervised machine learning techniques such as: SVM, LR, ANN, RF, KNN, NB and CNN. For evaluation of each technique on the ORL and Yale databases, three performance measures are used simultaneously by calculating their average value: recognition rate, f-measure of the confusion matrix, and area under the ROC curve. The results obtained are good for both classical machine learning models and deep learning models. It is concluded that deep learning outperforms other techniques if the data size is large. So, on ORL database RF, NB, KNN, ANN, LR, SVM, AlexNet, DenseNet, LeNet and ResNet produced, respectively, the following average performances: 94.56%, 95.12%, 96.54%, 96.68%, 97.76%, 98.19%, 100%, 100%, 100% and 98.04%.

A ACKNOWLEDGEMENTS

Scholarship funders - International Development Research Centre (IDRC) and Swedish International Development Cooperation Agency (SIDA) Scholarship Programme - Artificial Intelligence for Development (AI4D) Africa Scholarship Fund Manager - Africa Center for Technology Studies (ACTS)

B BIOGRAPHY

REFERENCES

[1] Xiaogang Wang and Xiaoou Tang. "Bayesian Face Recognition Using Gabor Features". In: *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications* (2003), pages 70–73.

[2] Alaa Eleyan and Hasan Demirel. "PCA and LDA Based Face Recognition Using Feedforward Neural Network Classifier". In: *Springer-Verlag* 4105 (2006), pages 199–206.

[3] Mayank Agarwal, Nikunj Jain, Manish Kumar and Himanshu Agrawal. "Face Recognition Using Eigen Faces and Artificial Neural Network". In: *International Journal of Computer Theory and Engineering* 2.4 (2010), pages 624–629.

[4] Abdel Ilah Salhi, Mustapha Kardouchi and Nabil Belacel. “Fast and efficient face recognition system using random forest and histograms of oriented gradients”. In: 2012, pages 1–11.

[5] Vaishali W. Parate and Pramod Patel. “PCA, DCT and DWT based Face Recognition System using Random Forest Classifier”. In: *International Journal of Digital Application and Contemporary research* 3.6 (2015), pages 1–7.

[6] Parvaneh Shayghan Gharamaleki and Hadi Seyedarabi. “Face recognition using Eigen faces, PCA and support vector machines”. In: *European Journal of Applied Engineering and Scientific Research* 4.3 (2015), pages 24–30.

[7] Telgaonkar Archana H. and Deshmukh Sachin. “Dimensionality Reduction and Classification through PCA and LDA”. In: *International Journal of Computer Applications (0975 – 8887)* 122.17 (2015), pages 100–104.

[8] Gurleen Kaur and Harpreet Kaur. “Efficient Facial Recognition Using PCA-LDA Combination Feature Extraction with ANN Classification”. In: *International Journal of Advanced Research in Computer Science and Software Engineering* 6.7 (2016), pages 258–263.

[9] Bhumika Pathya and Sumita Nainan. “Performance Evaluation of Face Recognition using LBP, PCA and SVM”. In: *SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE)* 3.4 (2016), pages 85–88.

[10] Vanlalhruaia, Yumnam Kirani Singh and Neetu Singh. “Binary Face Image Recognition using Logistic Regression and Neural Network”. In: *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. Chennai, India: IEEE, 2017, pages 3883–3888.

[11] Q. Z. Changjun Zhou Lan Wang and X. Wei. “Face recognition based on PCA and logistic regression analysis”. In: *Optik* 125.20 (2014), pages 5916–5919. ISSN: 0030-4026.

[12] Peter N. Belhumeur, Joao P. Hespanha and David J. Kriegman. “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.7 (1997), pages 711–720.

[13] Lahaw Zied Bannour, Essaidani Dhekra and Seddik Hassene. “Robust Face Recognition Approaches Using PCA, ICA, LDA Based on DWT, and SVM Algorithms”. In: *2018 41st International Conference on Telecommunications and Signal Processing (TSP)*. 2018, pages 1–5.

[14] Shakir Fattah Kak, Firas Mahmood Mustafa and Pedro R. Valente. “Discrete Wavelet Transform with Eigenface to Enhance Face Recognition Rate”. In: *Academic Journal of Nawroz University* 7 (Dec. 2018), pages 9–17.

[15] K.S.Maheswari and C.Harinath Babu. “A Color Face Recognition Using PCA and KNN Classifier”. In: *International Journal of Advanced Technology and Innovative Research* 9.13 (2015), pages 2408–2412.

[16] Manju Bala, Priti Singh and Mahendra Singh Meena. “Face Recognition Using Linear Discriminant Analysis”. In: *International Journal of Electrical and Electronics Research* 4.2 (2016), pages 96–103.

[17] Mady Huda and Shadi M. S. Hilles. “Face recognition and detection using Random forest and combination of LBP and HOG features”. In: *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*. 2018, pages 1–7.

[18] Al-Aubi Shaimaa A. and Fleah Laith R. “A Face Recognition System Based on Principal Component Analysis-Wavelet and Support Vector Machines”. In: *Cihan University-Erbil Scientific Journal* 3.2 (2019), pages 14–20.

[19] Pranati Rakshit, Rajit Basu, S. Paul, Sonali Bhattacharyya, Jhumpa Mistri and Ira Nath. "Face Detection using Support Vector Mechine with PCA". In: *2nd International Conference on Non-Conventional Energy: Nanotechnology & Nanomaterials for Energy & Environment (ICNNEE) 2019 (Archive)* (2020).

[20] Putta Sujitha, Venkatramaphanikumar S., Krishna Kishore K. V. "Scale Invariant Face Recognition with Gabor Wavelets and SVM". In: volume 7. 5S4. 2019.

[21] Sutarti Sri, Putra Anggyi and Sugiharti Endang. "Comparison of PCA and 2DPCA Accuracy with K-Nearest Neighbor Classification in Face Image Recognition". In: *Scientific Journal of Informatics* 6 (May 2019), pages 64–72.

[22] Ni Kadek Ayu Wirdiani, Praba Hridayami, Ni Putu Ayu Widiari, Komang Diva Rismawan, Putu Bagus Candradinata and I Putu Deva Jayantha. "Face identification based on K-nearest neighbor". In: *Scientific Journal of Informatics* 6.2 (2019), pages 150–159.

[23] Muhammad Shakeel Faridi, Muhammad Azam Zia, Zahid Javed, Imran Mumtaz and Saqib Ali. "A Comparative Analysis Using Different Machine Learning: An Efficient Approach for Measuring Accuracy of Face Recognition". In: *International Journal of Machine Learning and Computing* 11.2 (2021), pages 115–120.

[24] Zahraa Modher Nabat, Mushtaq Talib Mahdi and Shaymaa Abdul Hussein Shnain. "Face Recognition Method based on Support Vector Machine and Rain Optimization Algorithm (ROA)". In: *Webology* 19.1 (2022), pages 2170–2181.

[25] Benradi hicham, Chater Ahmed and Lasfar Abdelali. "Face recognition method combining SVM machine learning and scale invariant feature transform". In: *10th International Conference on Innovation, Modern Applied Science and Environmental Studies (ICIES 2022)*. Volume 351. 2022, pages 1–5.

[26] Zhiming Xie, Junjie Li and Hui Shi. "A Face Recognition Method Based on CNN". In: *Journal of Physics: Conference Series* 1395.1 (Nov. 2019), page 012006.

[27] Yohanssen Pratama, Lit Malem Ginting, Emma Hannisa Laurencia Nainggolan and Ade Erispra Rismanda. "Face recognition for presence system by using residual networks-50 architecture". In: *International Journal of Electrical and Computer Engineering (IJECE)* 11.6 (2021), pages 5488–5496.

[28] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pages 2278–2324.

[29] Syazana-Itqan K., Syafeeza A. R., and Saad N. M. "A MATLAB-based convolutional neural network approach for face recognition system". In: *J. Bioinf. Proteomics Rev* 2.1 (2016), pages 1–5.

[30] Meenakshi S., Siva Jothi M. and Murugan D. "Face Recognition using Deep Neural Network Across Variationsin Pose and Illumination". In: *International Journal of Recent Technology and Engineering (IJRTE)* 8.1S4 (2019), pages 289–292.

[31] He Kaiming, Zhang Xiangyu, Ren Shaoqing and Sun Jian. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pages 770–778.

[32] Arshi Husain and Virendra P. Vishvakarma. "Face Recognition Method Based on Residual Convolution Neural Network". In: *IOP Conference Series: Materials Science and Engineering* 1228.1 (Mar. 2022), page 012005.

[33] Aurélien Géron. *Machine Learning avec Scikit-Learn*. Paris: Dunod, 2017.

[34] Chloé-Agathe Azencott. *Introduction au Machine Learning*. Malakoff (France): Dunod, 2018.

[35] Ab Waheed Lone, Syed Sarver Hussain and Mohd Abdul Ahad. "Face Recognition by SVM Using Local Binary Patterns". In: *International Journal on Computer Science and Engineering* 9.5 (2017), pages 22–26.

[36] Eko Setiawan and Adharul Muttaqin. "Implementation of K-Nearest Neighbors Face Recognition on Low-power Processor". In: *TELKOMNIKA* 13.3 (2015), pages 949–954.

[37] Chris Albon. *Machine Learning with Python Cookbook*. Gravenstein Highway North, Sebastopol: O'Reilly Media, Inc., 2018.

[38] Gavin Hackeling. *Mastering Machine Learning with scikit-learn*. Birmingham - Mumbai (UK): Packt Publishing Ltd., 2014.

[39] Nusrat Jahan, Paulo Flores, Zhaohui Liu, Andrew Friskop, Jithin Mathew, Zhao Zhang. "Detecting and Distinguishing Wheat Diseases using Image Processing and Machine Learning Algorithms". In: *An ASABE Meeting Presentation* 2000372 (2020), pages 2–10.

[40] Yanming Guo and Yu Liu and Ard Oerlemans and Songyang Lao and Song Wu and Michael S. Lew. "Deep learning for visual understanding: A review". In: *Neurocomputing* 187 (2016). Recent Developments on Deep Big Vision, pages 27–48. ISSN: 0925-2312.

[41] Lecun, Y. and Bottou, L. and Bengio, Y. and Haffner, P. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pages 2278–2324.

[42] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems*. Edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger. Volume 25. Curran Associates, Inc., 2012.

[43] Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun. "Deep Residual Learning for Image Recognition". In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385.

[44] Huang, Gao and Liu, Zhuang and Van Der Maaten, Laurens and Weinberger, Kilian Q. "Densely Connected Convolutional Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017, pages 2261–2269.

[45] Samaria, F.S. and Harter, A.C. "Parameterisation of a stochastic model for human face identification". In: *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*. 1994, pages 138–142.

[46] Belhumeur, P.N. and Hespanha, J.P. and Kriegman, D.J. "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19.7 (1997), pages 711–720.

[47] Gary B. Huang and Manu Ramesh and Tamara Berg and Erik Learned-Miller. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*. Technical report 07-49. University of Massachusetts, Amherst, Oct. 2007.

[48] Shuo Yang, Ping Luo, Chen Change Loy and Xiaoou Tang. "WIDER FACE: A Face Detection Benchmark". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016.

[49] Matthew A. Turk and Alex P. Pentland. "Face recognition using eigenfaces". In: *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1991, pages 586–591.

[50] Nicolas Morizet, EA Thomas, Florence Rossant, Frédéric Amiel and Amara Amara. "Revue des algorithmes PCA, LDA et EBGM utilises en reconnaissance 2D du visage pour la biometrie". In: *Institut Superieur d'Electronique de Paris (ISEP), Departement d'Electronique* (2006), pages 1–11.

[51] Faten Bellakhdhar, Kais Loukil and Mohamed ABID. "Face recognition approach using Gabor Wavelets, PCA and SVM". In: *International Journal of Computer Science Issues (IJCSI)* 10.3 (2013), pages 201–207.

[52] D. Radovanović and S. Đukanovic. "Image-Based Plant Disease Detection: A Comparison of Deep Learning and Classical Machine Learning Algorithms". In: *2020 24th International Conference on Information Technology (IT)*. 2020, pages 1–4.

[53] Soad Almabdy and Lamiaa Elrefaei. "Deep Convolutional Neural Network-Based Approaches for Face Recognition". In: *Applied Sciences (MDPI)* 9 (2019), pages 1–21.

[54] Hicham Benradi, Ahmed Chater and Abdelali Lasfar. "A hybrid approach for face recognition using a convolutional neural network combined with feature extraction techniques". In: *IAES International Journal of Artificial Intelligence (IJ-AI)* 12.2 (2023), pages 627–640.