Comparative study of machine learning algorithms for face recognition

Harouna Naroua, Kadri Chaibou, Atsu Alagah Komlavi

Comparative study of machine learning algorithms for face recognition

Harouna NAROUA[1], Kadri CHAIBOU[2], Atsu ALAGAH KOMLAVI[*1,2]

[1]Department of Mathematics and Informatic, Abdou Moumouni University, Niamey, Niger
[2]Department of Mathematics and Informatic, Abdou Moumouni University, Niamey, Niger
[*1,2]Department of Mathematics and Informatic, Abdou Moumouni University, Niamey, Niger

*E-mail : alagahdejesus@yahoo.fr

Abstract

Background: The fundamental need for authentication and identification of humans through their physiological, behavioral or biological characteristics, continues to be used extensively to monitor and secure localities, property, financial transactions, etc. These biometric systems, especially those based on face characteristics, continue to attract the attention of researchers, major public and private services. In the literature, a multitude of methods have been deployed by different authors. Naturally an order of performance must be defined in order to be able to recommend the most effective method, because an error in an attempt to identify a person is of serious consequence. So, the main objective of the work carried out in 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, pre-processing, extraction of characteristics and finally classification. We will focus our comparison on the performance of artificial learning algorithms used in the last stage. These algorithms are: Support Vector Machines (SVM), Artificial Neural Networks (ANN), K Nearest Neighbors (KNN), Random Forests (RF), Logistic Regression (LR) and Naive Bayesian Classification (NB: Naive Bayes' Classifiers). The comparison criterion is the average performance, calculated using three performance measures: recognition rate, f-measure of the confusion matrix, and the area under the receiver operating characteristic (ROC) curve.

Results: Based on this criterion, the performance comparison of these machine learning algorithms, ranked SVM the best, with an average performance of 99.05%.

Discussion: Consideration of performance measure of all six classification techniques such as SVM, LR, ANN, RF, KNN, NB classifiers, we obtained respectively the following average performances: 99.05%, 98.91%, 98.89%, 97.14%, 96.82% and 95.89%. According to these results of which the worst is 95.89%, we can say that the machine learning algorithms for classification are one key step for face recognition. The comparison of these results with existing state-of-the-art methods is nearly the same. Authors achieved for the following classifiers: SVM, ANN, LR, RF, KNN and NB respectively the following accuracies: 98.80%, 97.25%, 96.44%, 95.56%, 95.14%, 94.82%. Finally, in depth discussion, we concluded that between all these approaches which are useful in face recognition, the SVM is the best classification algorithm.

Keywords

Face Recognition; Comparative Study; Machine Learning Algorithms.

I INTRODUCTION

We can all notice that our world is characterized by increasing technological development in all sectors. For example, the means of 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 (bank accounts, cash) is increasing day after day. Because of the failures of traditional solutions such as passwords, extra badges, ID cards, PINs and passwords, keys, which were once effective, currently there is increased interest in biometrics systems, with many applications for access control using these systems being developed and implemented. 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 enormous face databases. Finally, the evolution of facial recognition systems is mainly caused by scientific advances in machine learning with powerful data analysis algorithms. Consequently, facial recognition is becoming a reliable technology for identity verification. For this panoply of facial technologies presented and which is growing exponentially, it is more than necessary to perform a comparative study to assess the performances of these techniques. Moreover, it is necessary to choose one system that is the best. In this study, we are interested in knowing which classification system 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, such as facial recognition, has two phases: training and testing. The effectiveness of these systems depends on the quality of four additional stages within these two phases. These four additional stages are as follows: image acquisition, pre-processing, feature extraction and classification. In this article, we focus of the classification stage; there is much recent research that uses machine learning algorithms for the classification stage.

Three previous studies have achieved good results when using Support Vector Machine (SVMs) as a classifier to implement face recognition systems [13] [20] [27], as assessed by Principal Component Analysis (PCA), Local Binary Patterns, and Gabor filters. The recognition rate achieved by these studies was, 98.75% [13], 100.00% [20], and 97.65% [27] using the ORL database.

In 2018, Kak et al [26] applied discrete wavelet transform (DWT) and PCA techniques for pre-processing and features extraction, and a K Nearest Neighbors (KNN) classifier for classification. They obtained a 99.25% recognition rate on the ORL database. Other approaches using KNN have been tested on face features obtained with PCA [15] and using the ORL dataset. That previous researcher found a 92.47% recognition rate. In another study [19], Bala et al also studies facial recognition using KNN classifiers. Those authors developed the extraction stage with linear discriminant analysis (LDA), and this approach provided an interesting recognition rate of 93.70% on the ORL database of pictures.

Logistic regression algorithm is used a lot in face recognition systems; the literature provides evidence for this [11] [23]. In 2014, C. Zhou et al, exemplified its use. With signatures extracted using PCA, the authors obtained good recognition rates such as 93.33% with the Yale database

[2] and 96% by using ORL database. In 2017 Vanlalhruaia et al, made some improvements by binarizing the images before extracting the features. They achieved a 100% recognition rate.

In 2012, some approaches were implemented using random forests and two facial feature extractors: wavelet Gabor and histogram of oriented gradients (HOG) [9]. 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, [16] showed another way to use random forests. On the front end, they used a combination of PCA, DCT, and DWT for preprocessing and feature vector extraction on the ORL database. They obtained an acceptable accuracy of 96%.

In Tang (2003) [3], a combination of the Gabor filter and the Bayesian naive classifier is presented. To evaluate this system, two image databases were used: the XM2VTS database on which the authors found an accuracy rate of 97.10% and the AR database which gave an accuracy rate of 93.30%. Other works [14] have clearly reported good results by applying the Naive Bayes' Classifiers. With a face representation that is 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.

Many authors [4] [7] have put into practice the use of artificial neural networks in face recognition on feature vectors obtained with the PCA and then with the LDA. The authors obtained satisfactory results on the ORL dataset. Alaa Eleyan et al, got accuracies rate of 95.00% with the PCA and 97.00% with the LDA. Mayank Agarwal et al, for their part, obtained an accuracy rate of 97.01%. Certain approaches [18] have effectively proposed a combination of PCA and LDA as a features extractor. The evaluation of these systems based on ANNs gave an accuracy rate of 100.00% on the ORL dataset.

A summary of the results produced by these authors in the recent literature gives the following histogram which contains the average recognition rates per machine learning algorithm. This analysis ranks SVM classifiers at the top, with a satisfactory average recognition rate of 98.80%. They are followed, in order, by artificial neural networks, logistic regression, random forests, k-nearest neighbors, and Naive Baye's Classifiers which provided 97.25%, 96.44%, 95.56%, 95.14%, 94.82% as average recognition rate respectively. This ranking will be compared to the one that will be established during the experiments.

III MACHINE LEARNING ALGORITHMS

Machine Learning is defined as the science (or art) of programming computers to learn from data [22]. It can be subdivided into two main kinds of learning: supervised and unsupervised. And when learning is supervised, it can be for regression or classification. In this article, we will focus on classification algorithms which we will study, because they are the fundamental elements of our comparative analysis. We can divide machine learning algorithms into two main groups based on their purpose: supervised learning, unsupervised learning. There are also two main groups of supervised machine learning problems, called classification and regression. In classification, the goal is to predict a class label, which is a choice from a large dataset. So, classification stage plays an outstanding role in identifying human faces. There are many different types of classification methods as such: Support vector machines (SVM), K Nearest Neighbours (KNN), Logistic regression (LR), Random forest (RF), Naive Baye's Classifiers (NB), Artificial neural networks (ANN).

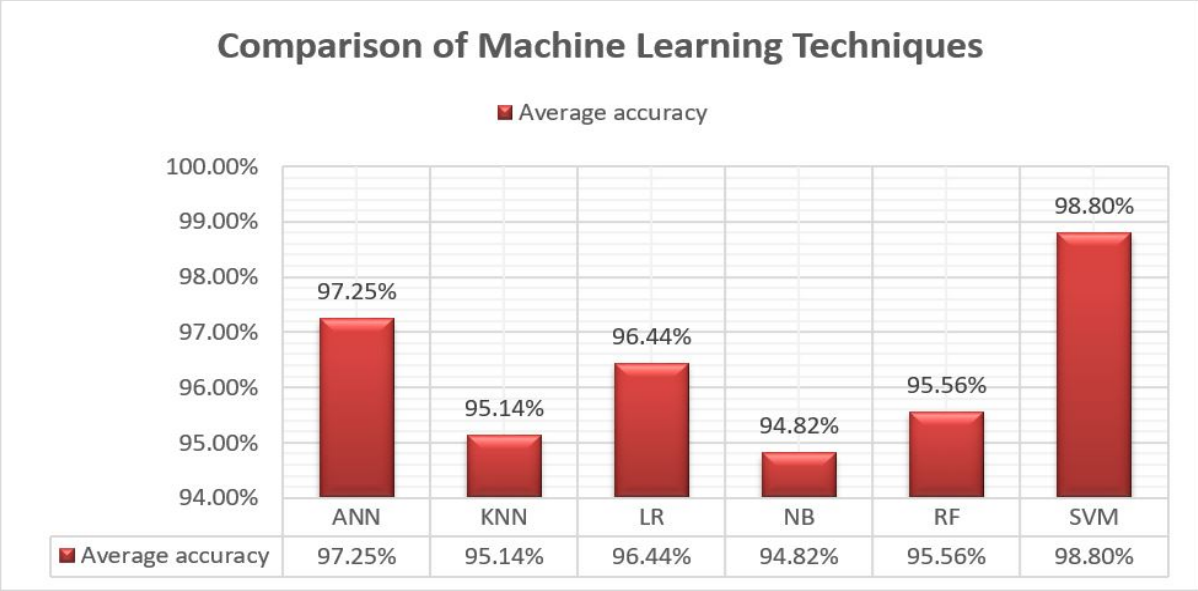


Figure 1: Performance of some classifiers in recent literature.

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. These very effective methods for general purpose pattern recognition [20], were introduced by Vladimir Vapnik in 1965. The extension of these algorithms to non-linear problems, thanks to the kernel trick, was discovered in 1995, still by Vapnik in association with Corinna Cortes [13] [25]. Another extension that has made SVMs very popular is their ability to solve multiclass problems such as face recognition [21]. 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 [21], thus:

- if $w^*.x_i + b^* > 0$ then x_i belongs to the +1 positive label class,
- 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. The main purpose of SVMs is to obtain the best function f(x) that maximizes the margins of the two classes. The margin of a separating hyperplane is the distance of this hyperplane to the nearest data. To obtain the largest possible margin between the hyperplane and the support vectors, we need to minimize the following function :

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

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 yet most commonly used classifiers in supervised machine learning [17] [15]. The KNN is a lazy learner, because it doesn't technically train a model to make predictions [24] [12]. So, the training stage doesn't require any computation. It only and simply consists of storing the training dataset with little or no processing. However, the test stage of this classifier requires an intensive distance calculation between a test data point and all training data points, in order to know 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 :

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

Named after the German mathematician Hermann Minkowski , 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 :

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

In pattern recognition, the KNN algorithm is widely used to determine to which class a new data belongs. 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) [25].

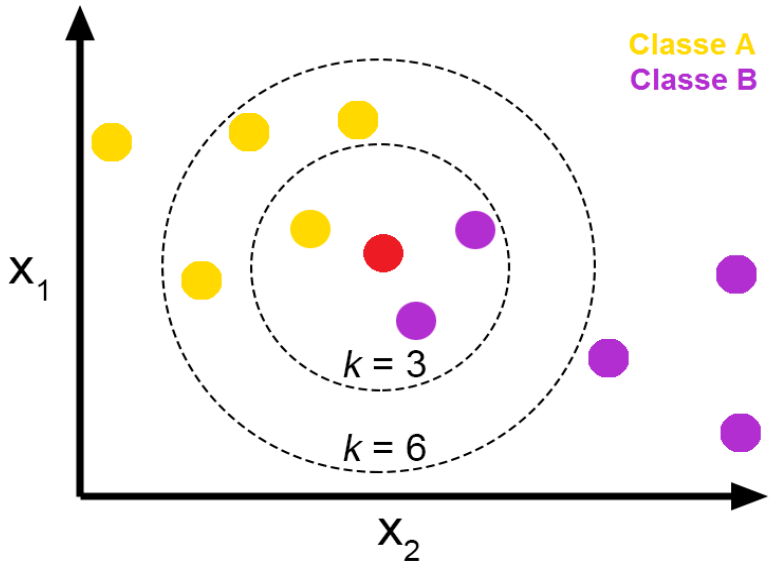


Figure 2: K Nearest Neighbours Algorithm.

3.3 Logistic regression (LR)

The logistic regression is actually a widely used supervised classification technique [24]. This originally binary, and later multi-class, 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 [25] [12]. 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)

Random forests or random decision forests, as the name implies is a collection of trees-based models trained on random subsets of the training data. This supervised learning algorithm is an ensemble of decision trees, usually trained with the bagging method (or sometimes pasting), that a combination of learning models increases the overall result. A method is called bagging (short for bootstrap aggregating) is when sampling is performed with replacement. The first version of this technique was created in 1995 by Tin Kam Ho. Then in 2001 [[25], 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 $D = (x_i, y_i), \ldots, (x_n, y_n)$ 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 $m \in \{1, \ldots, M\}$:

$$V_m = \sum_{t=1}^{T} I(\hat{y}_t == m) \tag{10}$$

- 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 predict $\hat{y}_t$, 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)

1: **for** $b \leftarrow 1, B$ **do**
2: (a) Draw a bootstrap sample Z^* of size N from the training data.
3: (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.
4: (I) Select m variables at random from the p variables.
5: (II) Pick the best variable/split-point among the m.
6: (III) Split the node into two daughter nodes.
7: **end for**
8: Output the ensemble of trees $\{T_b\}_1^B$
9: Make prediction at new point x:
10: 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 $\{\hat{C}_b(x)_1^B\}$.

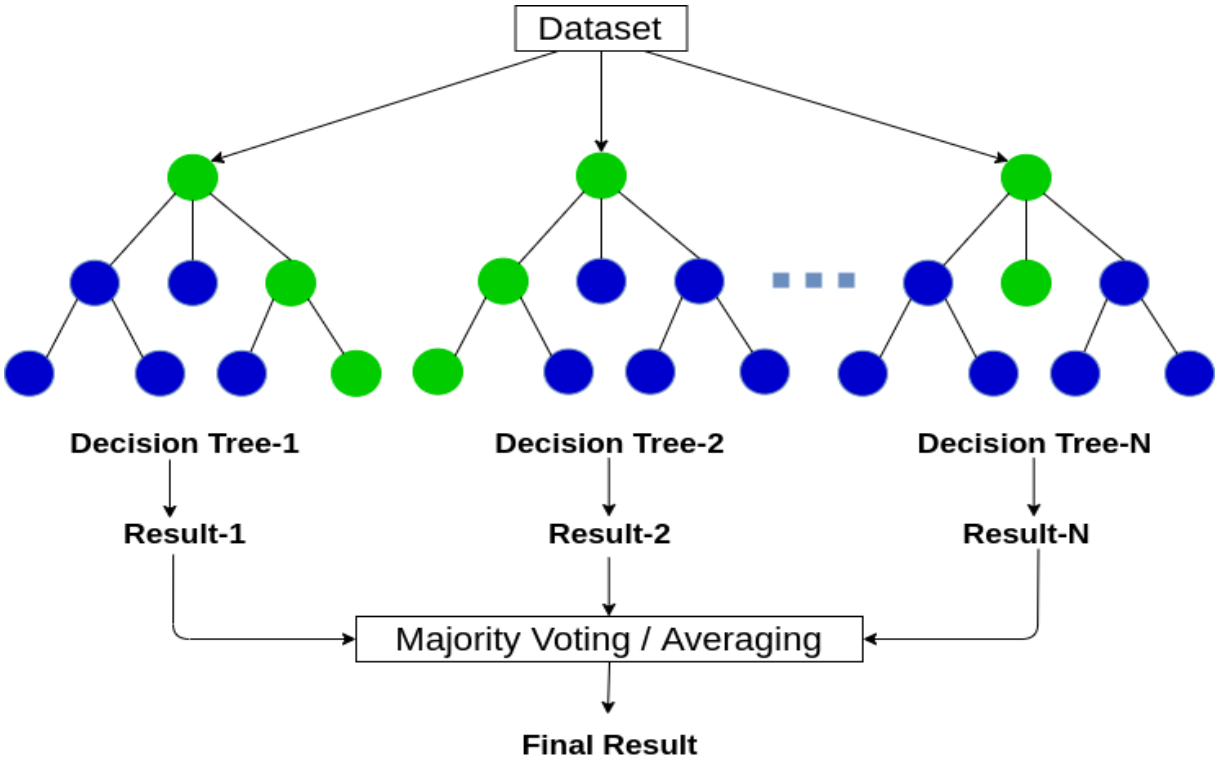


Figure 3: Random Forest Algorithm.

3.5 Naive Baye's Classifiers (NB)

Naive Baye's Classifier is a supervised classification algorithm driven by 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 [25] [12]:

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

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 [12]. 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 [24] :

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

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, $\hat{y}$.

3.6 Artificial neural networks (ANN)

In the field of machine learning and cognitive science, Artificial Neural Networks (ANN) are a family of models which are inspired by the biological neural networks that constitute the human brain [18]. Indeed, from a simple node called formal neuron carrying a value between 0 and 1, called activation, we can create a network composed of neurons organized in three different layers: the input layer, the hidden layers and the output layer shown in the following figure 2. 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 [12], is considered as the first simple and complete neural network [25]. 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.

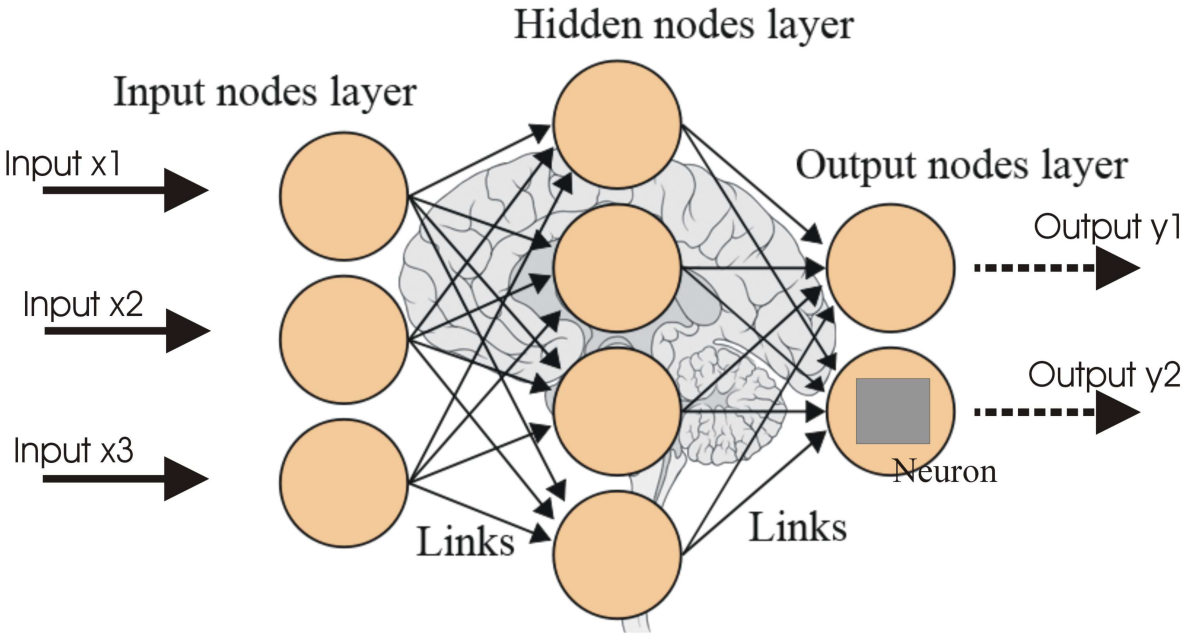


Figure 4: Artificial Neural Network Structure.

IV METHODS

4.1 Face recognition databases

For the implementation of a facial recognition system and its evaluation, face images databases are needed. Among those available, some have aroused a growing interest among many researchers. The GRIMACE database contains 360 pictures of 18 different persons. All the images have a resolution of 200×180 pixels and same lighting conditions [27]. The ORL database (Olivetti Research Laboratory) has been gathered from 1992 to 1994 in AT & T laboratory of Cambridge University in England [13] [27]. It includes 400 pictures of 40 distinct persons, with a resolution of 112×92 pixels and contains slight variations in illumination, in pose, and facial details [25]. The YALE image dataset is designed at the CVC (Center for Computational Vision and Control) of Yale University in a fully controlled environment [20]. It is composed of 165 gray-scale frontal images of 15 persons under varying illumination. All the images of these databases are taken in different facial expression situations.

4.2 Image processing and feature extraction techniques

This step of face characterization by pre-processing and extraction is very important because it directly influences the performance of the systems. For this task there are several efficient techniques among which we will use four.

The wavelet transform (DWT) which is used to compress the images by eliminating the redundancies which mobilize a great quantity of computing resources. Principal component analysis (PCA) is generally considered as the main approach to reduce the dimension of dataset by extracting variables; it was, introduced by M. A. Turk and M. P. Pentlanden in 1991 [1] [5]. Linear Discriminant Analysis (LDA), which appeared with the work of Belhumeur in 1997 at Yale University in the USA [5]. It is a good dimension reduction technique always used before a classification algorithm. The GABOR filter was introduced by Dennis Gabor in 1946, then improved in 1980 by John G. Daugman [10]. 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. To perform the experiments and evaluate the performance of the face recognition systems, an experimental methodology in different steps is used. In the first step, we prepare the data on the ORL, YALE and GRIMACE image databases by dividing them into two distinct sets: the training set and the testing set. The first one will be used to train the model and the second to evaluate it. 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, classification is done through the following algorithms: SVM, KNN, RF, LR, NB, and ANN. At this level the model is well trained, saved and able to make the prediction. We will use the testing set which will follow the same process of preprocessing and feature extraction. The model created, takes as parameters these vectors and predicts classes from them. The diagram below explains the experiment graphically.

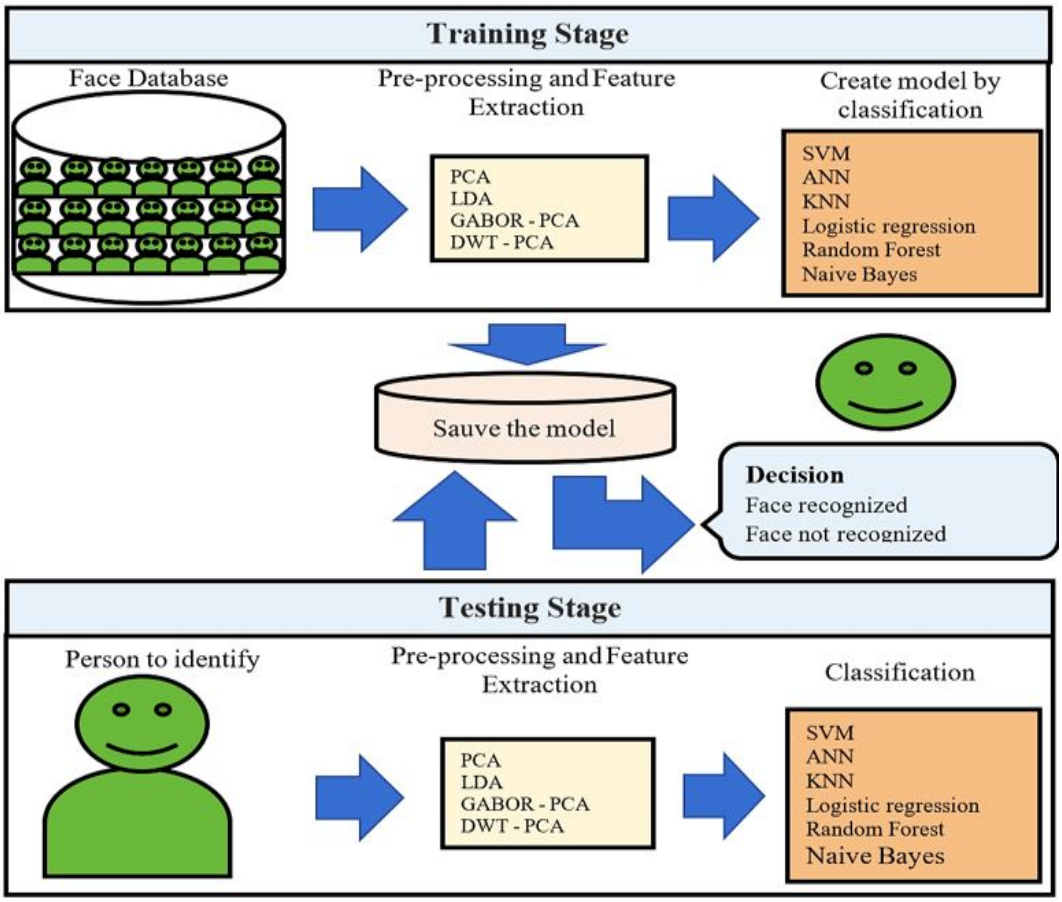


Figure 5: Diagram of the experimental methodology.

V PERFORMANCE MEASURES

Three performance measures will be simultaneously used to evaluate the quality of the 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 [25]. 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:

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

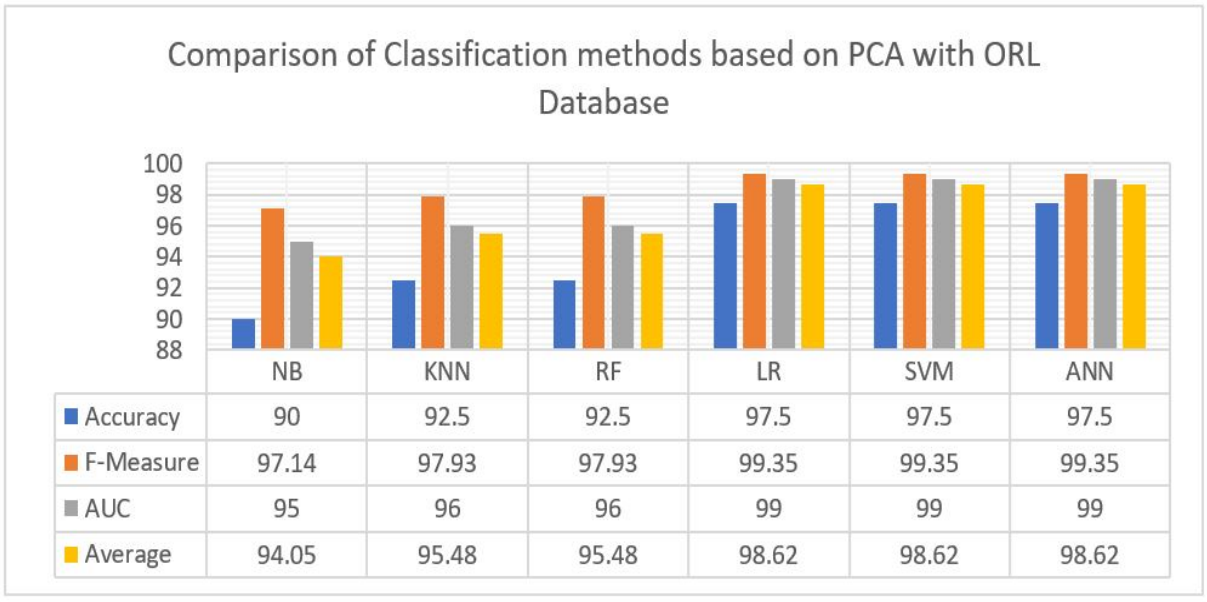
The ROC curve plots graphically the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings [25]. In machine learning for model comparison, researchers focus, most often, on the use of the Area Under the Curve (AUC); we therefore, use the same.

VI RESULTS

Each system is evaluated on the three databases. Its performance is the average value of the accuracy rate, the f-measure and the area under the ROC curve (AUC), defined as:

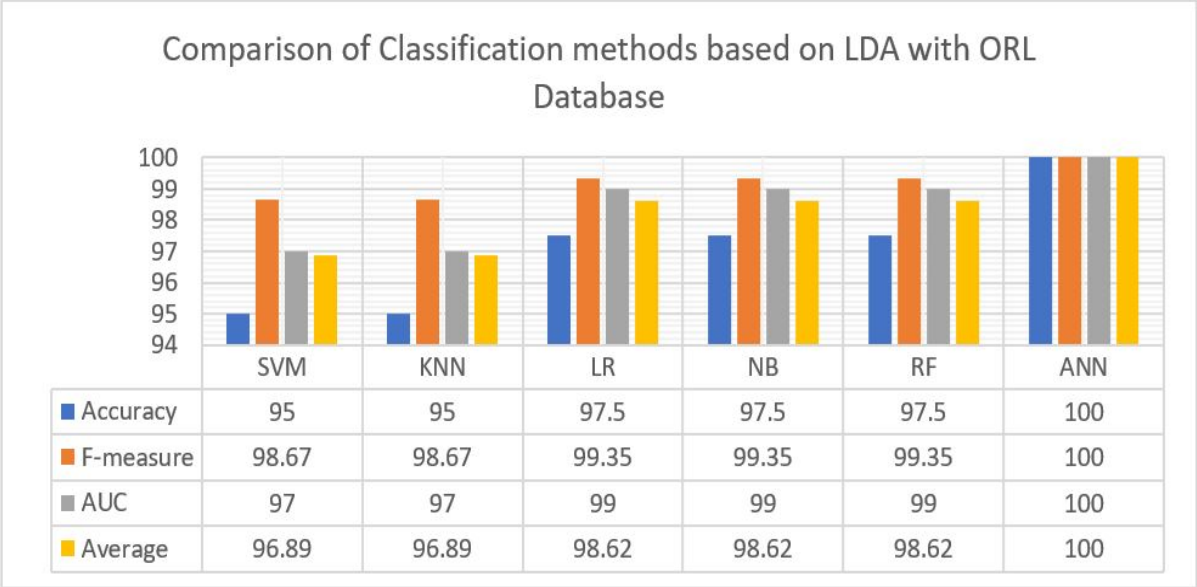
$$Performance = \frac{Accuracy + F\_measure + AUC}{3} \tag{15}$$

6.1 Classification methods based on ORL Database



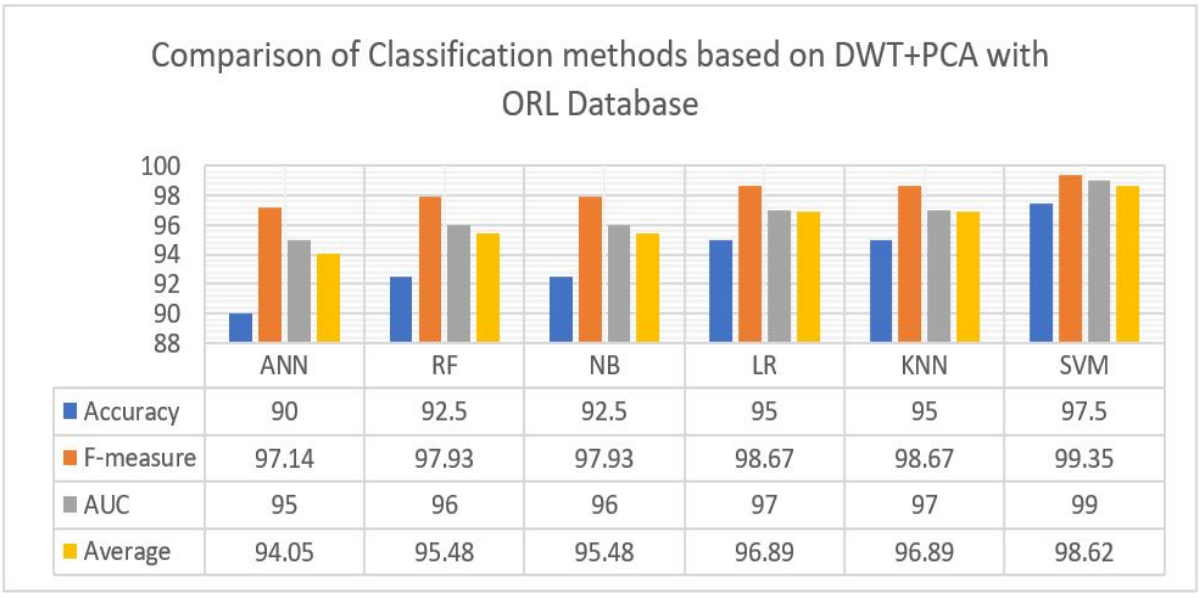
	NB	KNN	RF	LR	SVM	ANN
Accuracy	90	92.5	92.5	97.5	97.5	97.5
F-Measure	97.14	97.93	97.93	99.35	99.35	99.35
AUC	95	96	96	99	99	99
Average	94.05	95.48	95.48	98.62	98.62	98.62

Figure 6: Performance of some classifiers on ORL database using PCA.



	SVM	KNN	LR	NB	RF	ANN
Accuracy	95	95	97.5	97.5	97.5	100
F-measure	98.67	98.67	99.35	99.35	99.35	100
AUC	97	97	99	99	99	100
Average	96.89	96.89	98.62	98.62	98.62	100

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



	ANN	RF	NB	LR	KNN	SVM
Accuracy	90	92.5	92.5	95	95	97.5
F-measure	97.14	97.93	97.93	98.67	98.67	99.35
AUC	95	96	96	97	97	99
Average	94.05	95.48	95.48	96.89	96.89	98.62

Figure 8: Performance of some classifiers on ORL database using DWT and PCA.

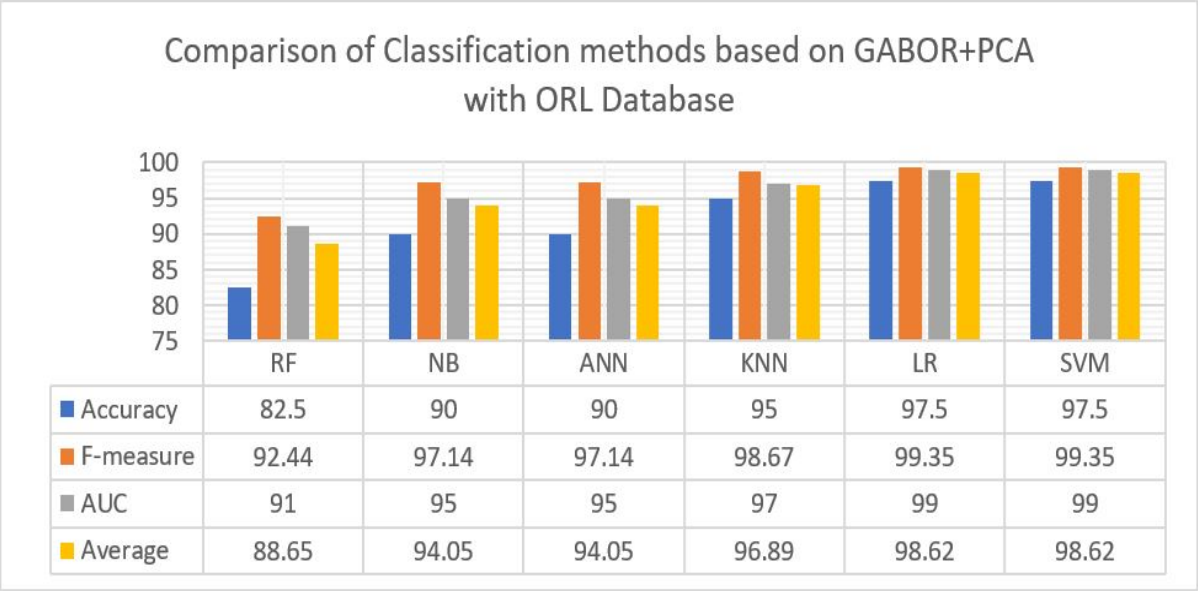


Figure 9: Performance of some classifiers on ORL database using GABOR and PCA.

The following table summarizes the average of the three performance measures for different classifiers associated with the different preprocessing and extraction techniques.

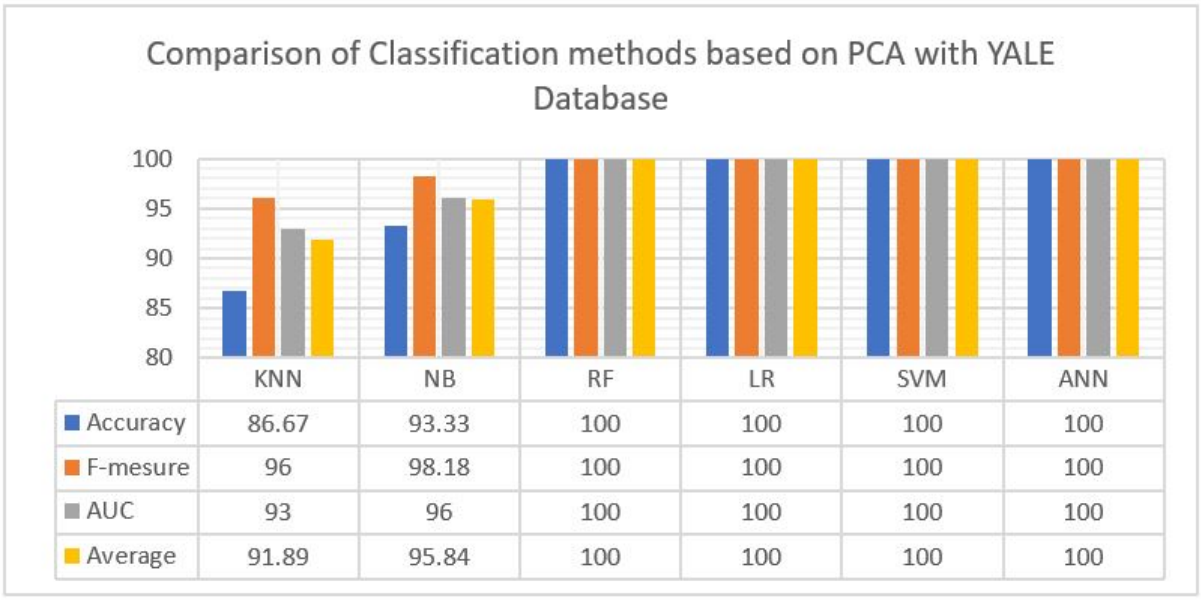
ALGORITHM	PCA	LDA	DWT+PCA	GABOR+PCA	Average (%)
RF	95.48	98.62	95.48	88.65	94.56
NB	94.05	96.89	95.48	94.05	95.12
KNN	95.48	96.89	96.89	96.89	96.54
ANN	98.62	100	94.05	94.05	96.68
LR	98.62	96.89	96.89	98.62	97.76
SVM	98.62	96.89	98.62	98.62	98.19

Table 1: : Comparative Analysis of Face Recognition Approaches on ORL Database using different classifiers with an average value of the F- measure, the accuracy and the AUC.

Support vector machines (SVM) outperform the others, with an overall average performance of 98.19%, as in Figure 2. 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 of 94.56%. From the discussion it is clear that, the SVM classifier has comparatively higher performance than the other classifiers.

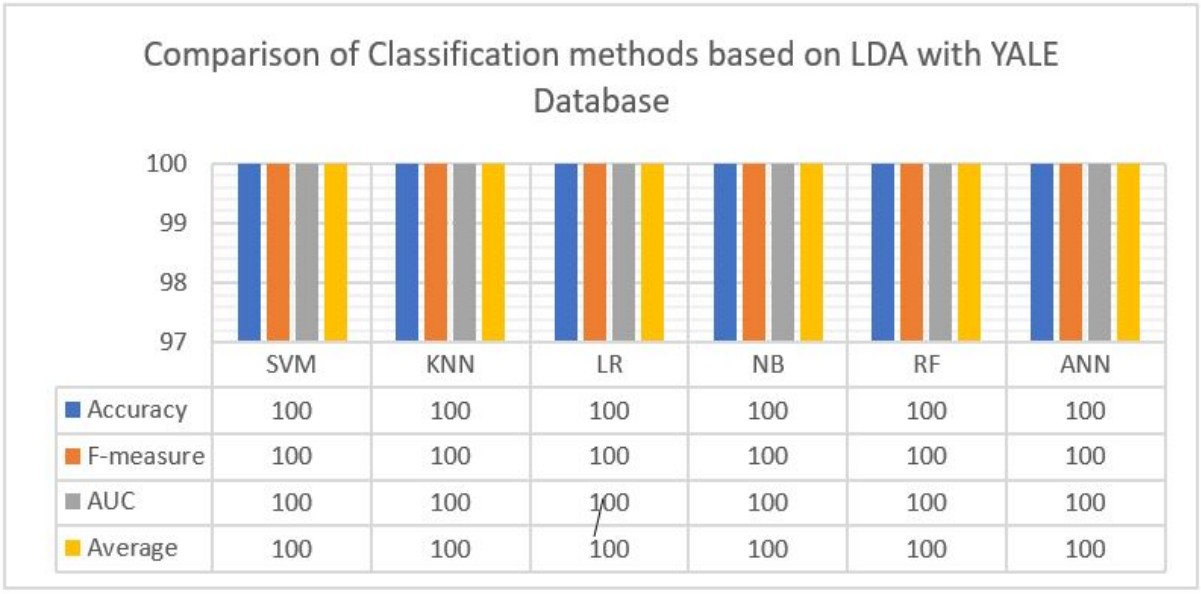
6.2 Classification methods based on YALE Database

The comparison between the average performance of the six classification algorithms are represented in Figure 7, Figure 8, Figure 9 and Figure 10. It is clearly observed from the analysis of these figures that all the supervised machine learning techniques namely NB, KNN, LR, ANN, SVM, and RF used to detect and recognize human faces, give significantly good performances.



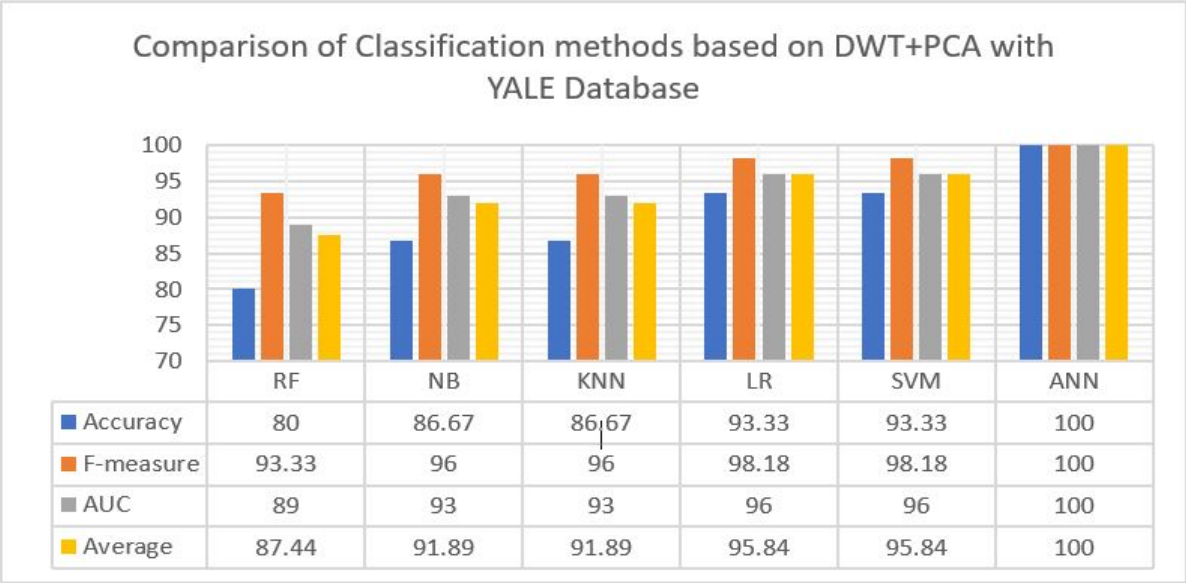
	KNN	NB	RF	LR	SVM	ANN
Accuracy	86.67	93.33	100	100	100	100
F-mesure	96	98.18	100	100	100	100
AUC	93	96	100	100	100	100
Average	91.89	95.84	100	100	100	100

Figure 10: Performance of some classifiers on YALE database using PCA.



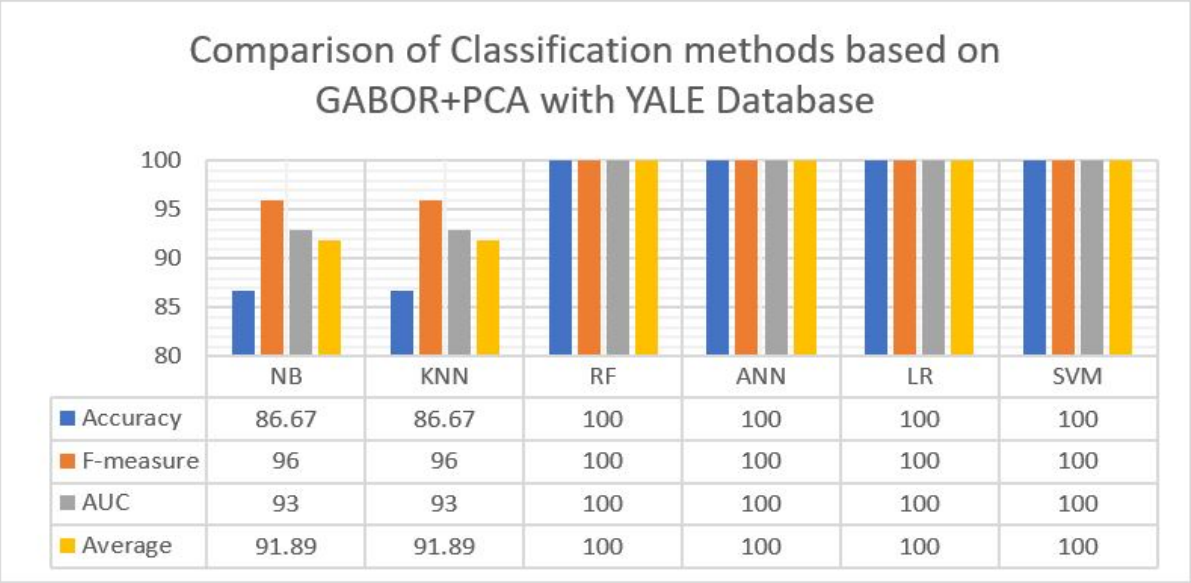
	SVM	KNN	LR	NB	RF	ANN
Accuracy	100	100	100	100	100	100
F-measure	100	100	100	100	100	100
AUC	100	100	100	100	100	100
Average	100	100	100	100	100	100

Figure 11: 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 12: Performance of some classifiers on YALE database using DWT and PCA.



	NB	KNN	RF	ANN	LR	SVM
Accuracy	86.67	86.67	100	100	100	100
F-measure	96	96	100	100	100	100
AUC	93	93	100	100	100	100
Average	91.89	91.89	100	100	100	100

Figure 13: Performance of some classifiers on YALE database using GABOR and PCA.

The table of averages of the three performance measures of each classification algorithm associated with the different extractors, is shown below.

ALGORITHM	PCA	LDA	DWT+PCA	GABOR+PCA	Average (%)
RF	91.89	100	91.89	91.89	93.92
NB	95.84	100	91.89	91.89	94.91
KNN	100	100	87.44	100	96.86
ANN	100	100	95.84	100	98.96
LR	100	100	95.84	100	98.96
SVM	100	100	100	100	100

Table 2: : Comparative Analysis of Face Recognition Approaches on YALE Database using different classifiers with an average value of the F- measure, the accuracy and the AUC.

On the YALE database, artificial neural networks ANN outperform all other classification algorithms by achieving a perfect average score of 100%, as shown in Table 2.

6.3 Classification methods based on GRIMACE Database

For all algorithms, the average performance rate is 100%, except the Naive Bayes' Classifier associated to the combination of DWT and PCA which gives 90.59%, as illustrated in the table and histogram below.

ALGORITHM	PCA	LDA	DWT+PCA	GABOR+PCA	Average (%)
RF	100	100	100	100	100
NB	100	100	95.59	100	97.65
KNN	100	100	100	100	100
ANN	100	100	100	100	100
LR	100	100	100	100	100
SVM	100	100	100	100	100

Table 3: : Comparative Analysis of Face Recognition Approaches on GRIMACE Database using different classifiers with an average value of the F- measure, the accuracy and the AUC.

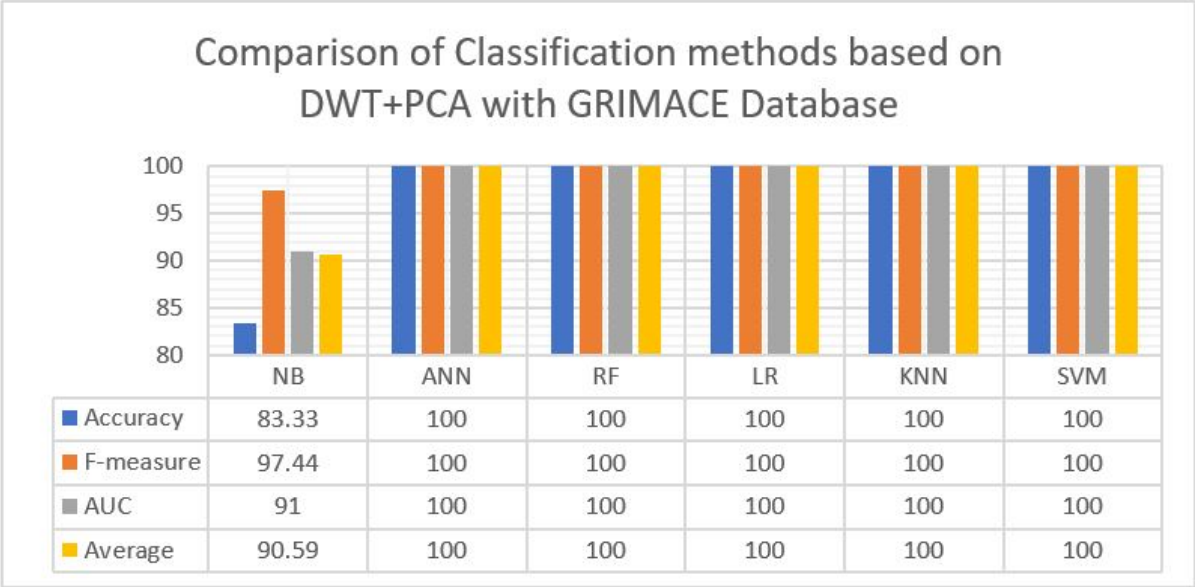


Figure 14: Performance of some classifiers on GRIMACE database using DWT and PCA.

With the Grimace database, the summary in the table 3 below shows the importance of the image acquisition stage. The Grimace database does not bring us any difference from the performance point of view, necessary in establishing a real comparison of the machine learning algorithms.

6.4 Comparison of classifications methods and discussion

For the three databases studied: ORL, YALE and GRIMACE, the final results obtained are shown in the Table 4.

ALGORITHM	ORL	YALE	GRIMACE	Average (%)
NB	95.12	94.91	97.65	95.89
KNN	96.54	93.92	100	96.82
RF	94.56	96.86	100	97.14
ANN	96.68	100	100	98.89
LR	97.76	98.96	100	98.91
SVM	98.19	98.96	100	99.05

Table 4: : Comparative Analysis of Face Recognition Approaches on ORL, YALE, GRIMACE Databases using different classifiers with an average value of the F- measure, the accuracy and the AUC of the ROC curve..

In this paper, our study performed a validation of the performance of the machine learning techniques applied in face recognition, as presented in Table 4. The experimental results show that support machine vector model achieved 99.05% average performance. We note that this best result corroborates that of the literature in which SVMs also give the best classification accuracy rate as 98.80%. Also, LR, ANN, RF, KNN and NB produced respectively 98.91%, 98.89%, 97.14%, 96.82% and 95.89% as average performance rates. This ranking is roughly equal to what we found in the literature. In the latter, after SVM, we found that ANN, LR, RF, KNN and NB provided the following results in order: 97.25%, 96.44%, 95.56%, 95.14%, 94.82%. This small difference is probably due to the imbalance in the dataset used but also to the unequal conditions and criteria for comparison in the literature.

VII CONCLUSION

We are in a favorable context for the development of facial recognition systems. The technological evolution, the apparition of powerful machines in calculation and treatment, the advent of artificial intelligence, are the important factors which allowed the researchers to direct their work towards this topic. Thus, many approaches using machine learning algorithms and giving different results have emerged in the recent literature [15] [11] [9] [3] [21]. That is why, this paper found it necessary to search for the best performing method by a comparative study of classification algorithms. Using a specific methodology, many systems are implemented with different supervised machine learning techniques such as: SVM, LR, ANN, RF, KNN, NB. For the evaluation of each technique on the ORL, Yale and Grimace 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. And, the SVM, LR, ANN, RF, KNN, NB classifiers produced respectively the following average performances: 99.05%, 98.91%, 98.89%, 97.14%, 96.82% and 95.89%. SVM take the first place in the ranking. Thus, the future perspectives are the research of the best parameterization of the SVMs, and the research of the best methods of preprocessing and extraction of the features, in order to increase the performances of this classifier.

REFERENCES

[1] M. A. Turk and A. P. Pentland. “Face Recognition Using Eigenfaces”. In: *IEEE, Vision and Modeling Group, The Media Laboratory Massachusetts Institute of Technology* (1991), pages 586–591.

[2] P. N. Belhumeur, J. P. Hespanha, and D. 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.

[3] X. Wang and X. Tang. “Bayesian Face Recognition Using Gabor Features”. In: *Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications* (2003), 70—73.

[4] A. Eleyan and H. Demirel. “PCA and LDA Based Face Recognition Using Feedforward Neural Network Classifier”. In: *Springer-Verlag* 4105 (2006), pages 199–206.

[5] N. MORIZET, F. A. Thomas EA Florence ROSSANT, and A. 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.

[6] B. Z. H. A. Hossein Sahoolizadeh and C. H. Dehghani. “A New Face Recognition Method using PCA, LDA and Neural Network”. In: *International Journal of Computer, Electrical, Automation, Control and Information Engineering* 2.5 (2008), pages 1354–1359.

[7] M. M. K. Mayank Agarwal Nikunj Jain and H. Agrawal. “Face Recognition Using Eigen Faces and Artificial Neural Network”. In: *International Journal of Computer Theory and Engineering* 2.4 (2010), pages 624–629.

[8] H. Rady. “Face Recognition using Principle Component Analysis with Different Distance Classifiers”. In: *International Journal of Computer Science and Network Security (IJCSNS)* 11.10 (2011), pages 135–144.

[9] A. I. Salhi, M. Kardouchi, and N. Belacel. “Fast and efficient face recognition system using random forest and histograms of oriented gradients”. In: *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*. Darmstadt, Germany: IEEE, 2012, pages 1–11.

[10] F. Bellakhdhar, K. Loukil, and M. ABID. “Face recognition approach using Gabor Wavelets, PCA and SVM”. In: *International Journal of Computer Science Issues (IJCSI)* 10.3 (2013), pages 201–207.

[11] Q. Z. Changjun Zhou Lan Wang and X. Wei. “Face recognition based on PCA and logistic regression analysis”. In: *Elsevier* 125.20 (2014), 5916—5919.

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

[13] P. S. Gharamaleki and H. 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.

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

[15] K.S.Maheswari and C. Babu. “A Color Face Recognition Using PCA and KNN Classifier”. In: *International Journal and Magazine of engineering, Technology, Management and research* 2.9 (2015), pages 1110–1116.

[16] V. W. Parate and P. 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.

[17] E. Setiawan and A. Muttaqin. “Implementation of K-Nearest Neighbors Face Recognition on Low-power Processor”. In: *TELKOMNIKA* 13.3 (2015), pages 949–954.

[18] G. Kaur and H. 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.

[19] P. S. Manju Bala and M. S. Meena. “FACE RECOGNITION USING LINEAR DISCRIMINANT ANALYSIS”. In: *International Journal of Electrical and Electronics Research* 4.2 (2016), pages 96–103.

[20] B. Pathya and S. 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.

[21] S. S. H. Ab Waheed Lone and M. A. Ahad. “Face Recognition by SVM Using Local Binary Patterns”. In: *International Journal on Computer Science and Engineering* 9.5 (2017), pages 22–26.

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

[23] Vanlalhruaia, Y. K. Singh, and N. D. 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.

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

[25] C.-A. Azencott. *Introduction au Machine Learning*. Malakoff (France): Dunod, 2018.

[26] F. M. M. Shakir Fattah Kak and P. R. Valente. "Discrete Wavelet Transform with Eigenface to Enhance Face Recognition Rate". In: *Academic Journal of Nawroz University (AJNU)* 7.4 (2018), pages 9–17.

[27] V. S. Putta Sujitha and K. K. K. V. "Scale Invariant Face Recognition with Gabor Wavelets and SVM". In: *International Journal of Recent Technology and Engineering (IJRTE)* 7.5S4 (2019), pages 100–104.