2022 IEEE Fourth International Conference on Advances in Electronics, Computers and Communications (ICAECC) | 978-1-6654-0239-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECC54045.2022.9716590

Face Recognition Using Machine Learning Models - Comparative Analysis and impact of dimensionality reduction

1st Yaswanthram P
Department of computer science and Engineering
Amrita School of Engineering,Amrita Vishwa Vidyapeetham
Coimbatore, India
cb.en.u4cse18243@cb.students.amrita.edu

2nd Sabarish B.A
Department of computer science and Engineering
Amrita School of Engineering,Amrita Vishwa Vidyapeetham
Coimbatore, India
ba_sabarish@cb.amrita.edu

***Abstract*—Face Recognition is considered a biometric technique where it is capable of uniquely identifying and verifying a person just by analysing and comparing the facial patterns on the facial contours. Face Recognition has gained significant importance in security aspects and it has been widely used and accepted biometric. It has given greater importance during pandemic situations in terms of cheapest and widely accepted touchless biometrics. This paper studies the impact of dimensionality reduction on the efficiency or accuracy of machine learning algorithms in face recognition. The analysis is carried out over various algorithms include Random Forests, Support Vector Machine, Linear Regression, Logistic Regression, K-Nearest Neighbor. Based on the analysis, Logistic Regression gives better performance in terms of accuracy and time with an accuracy score of 0.97 within a time of 5.74 sec when implemented without principal component analysis whereas with principal component analysis, Logistic Regression achieved an accuracy score of 0.93 within a time of 0. 15sec. There is a huge difference in computation time approximately 20 times, the difference in accuracy is minimal.**

***Index Terms*—Machine Learning, Classifiers, Regression, Face Recognition**

I. Introduction

Face Recognition is a recognition technique where it matches the face templates and verifies the user. Today's technological growth which makes verification and validation process easier, faster and accurate. Face based smart attendance system, traffic control, fraud detection and face unlock are some of the applications developed using face recognition techniques. There are many advantages of using face recognition which includes automation, security, no contact, high accuracy and speed. Methods of face recognition are template based, wholistic, appearance based, statistical approach, Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Linear 2 Discriminant Analysis (LDA), Locality Preserving Projections (LPP), Gabor Wavelet, Independent Component Analysis (ICA), Kernel PCA and Neural Networks [1] [2]. Face recognition can also be possible with Machine learning and cloud computing. Many cloud platforms offer face recognition as a feature, some of them include Google Cloud Vision API, Amazon Recognition. With the help of face recognition API, one can work on cloud platforms also. In this research work, tested some Machine Learning algorithms on Olivetti face dataset and calculated the accuracy values and time taken for each algorithm with both PCA and without PCA. Considering all the values from the result of this research work, made a comparative analysis.

978-1-6654-0239-2/22/$31.00 ©2022 IEEE

II. Problem Statement

The objective of this research is to compare the accuracy and time taken for each of the machine learning algorithms and impact of dimensionality reduction using PCA. Comparative analysis of each algorithm with respect to the computation time is performed to suggest and recommend an algorithm with the best accuracy and reduced dimensions.

III. Literature Review

Researchers pointed out that use of facial recognition using Machine learning algorithms made a tremendous change in today's technology. From the past decades, there have been many research works on the use of machine learning algorithms when implemented on any face dataset. Many research works proved that when these algorithms were implemented on a face dataset, they gave the best accuracy within a time. It also inferred that these models can be used for face recognition applications [1]. In this research work, the researchers developed a vision-based face recognition attendance monitoring system for surveillance using computer vision and deep learning techniques, also stated that artificial neural networks can be used to detect and recognize faces instantly. Researchers used real time surveillance algorithms in their work to eliminate robust and to provide user friendly face recognition attendance systems, achieved an accuracy score of 74% when implementing a real time surveillance algorithm [2]. In another work, the researchers developed a computational model for real time face tracking using a simple rule-based face detection algorithm and Support Vector Machine (SVM) for real time face detection as well as tracking and also pointed out

their results after successful implementation of algorithms [3]. While some researchers proposed a biometric authentication model for facial recognition. In this work, the researchers used PCA for dimensionality reduction and they implemented KNN classification algorithm for face recognition on MIT face dataset. After successful 3 implementation, got a very good accuracy score of 98.66% [4]. In this particular research work, the researchers made a comparative analysis on various image processing algorithms for facial recognition. The researchers implemented these prominent algorithms in OpenCV and completed their evaluation of various databases. After successful implementation, calculated the accuracy values and time taken for each algorithm used in research work [5]. Multi Resolution Sparse Representation based Classification (MRSRC) is designed to improve the performance and execution time of facial recognition. They also proved that MRSRC gave better performance as well as time complexity when compared with existing algorithms. Researchers carried out their complete work on LFW and ORL databases [6]. Researchers explained an approach for illumination invariant face detection using viola jones algorithm by training their system for long hours to identify a person's face even in illumination conditions and discussed their experimental results in their research work [7]. Driver face recognition system using thermal images is developed which can identify whether the driver is drunker or sober. It is developed using Gaussian Mixture model and used Fisher Linear discriminant for dimensionality reduction with the use of capillary junction points on the human face, classified by temperature of blood [8].Some research carried out in a face recognition system using machine learning algorithms. In this work, the researchers used linear discriminant analysis, multilayer perceptron, SVM and naïve bayes approach and achieved an accuracy score of 97% and 100% when used PCA and linear discriminant analysis [9,10,11]. A face net for clustering and face recognition is developed and trained a CNN on the system and implemented it on four datasets. As a result, got an accuracy score of 99.63%when implemented on LFW dataset and accuracy of 95.12% when implemented on YouTube face database. In another machine learning research work, the researchers proposed a probabilistic face embedding (PFEs) method which will improve the performance of face recognition algorithms. In this work, face images are stored in a latent semantic space and converted the embeddings into PFEs.

IV. Implementation

A. Dataset description

In this research work, Olivetti face dataset was used which was taken from Kaggle which has set of face images taken in between April 1992 and April 1994 at AT & T laboratories, Cambridge.This dataset consists of 10 different face images of each 40 distinct subjects and for some other subjects, the face images are taken at different variations at different times followed by light variation, facial expressions with open and closed eyes, smile and without smile faces.

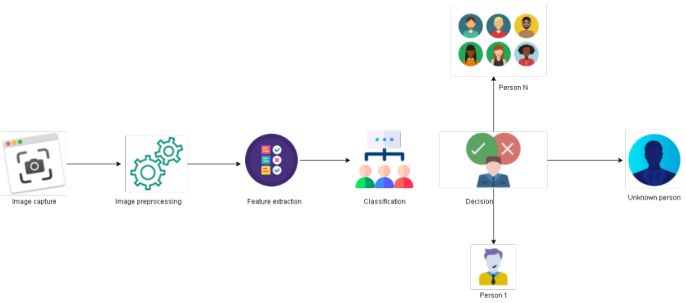

Fig. 1. System architecture of the implementation.

B. Implementation

In this research work, the implementation undergoes many steps of process and the primary step of this research work is to gather a universal face dataset which is most suitable for the work. After checking all the requirements of the work, used Olivetti Universal face dataset which was publicly available on Kaggle. By using various regression and classification techniques, calculated the accuracy values and time taken for each algorithm. and also compared all the actual and predicted values for each Algorithm. In this work, five algorithms are implemented on Olivetti face dataset with PCA and without PCA.The five algorithms are Random Forest, SVM (Support Vector Machine), Linear Regression, KNN (K-Nearest Neighbor) and Logistic Regression. In this work, the implementation requires high dimensionality data. So, in order to eliminate overfitting and redundancy transform data from high dimensional space to low dimensional space using PCA (Principal Component Analysis) which is the linear dimension reduction technique used for dimensionality reduction and it also helps in data compression, reduce storage space and even computation time.

Figure.1. shows the system architecture of overall implementation. Face recognition can be defined as a process that begins with face detection, then moves on to image preprocessing, feature extraction, classification and lastly face recognition. Pre-processing suppresses unwanted distortions to improve specific picture characteristics necessary for further processing, and feature extraction can be done to reduce the number of features in the dataset and then it classifies and recognize the person's face.

V. Results

A. Random Forest

Random Forest is a type of ensemble learning algorithm used for classification and regression methods. It builds decision Trees at training time during classification and regression. [10] It was used for missing values and it gives higher accuracy even for large datasets. Random forest is also called as random decision forest. When random forest was implemented on a face dataset, got an accuracy of 0.92 in a time of 2.44 sec without PCA Whereas with PCA, got an accuracy score of 0.88 in a time of 0.511 sec. As a result, Random Forest is considered as a good algorithm for face recognition.

B. SVM (Support Vector Machine)

Support Vector Machine is also one type of supervised Machine learning algorithms that can be used for classification as well as Regression methods but it was mostly used for classification methods. [6] It was very effective during high dimensional spaces. SVM is not recommended for large datasets as it takes high training time. When SVM was implemented on a face dataset, got an accuracy score of 0.96 in a time of 107.68 sec without PCA whereas with PCA, got an accuracy score of 0.94in a time of 4.77 sec. As a result, SVM can also be considered as a good algorithm for face recognition.

C. Linear Regression

Linear regression aims to model the connection between two variables by fitting a linear equation to observed data. One variable is an explanatory variable, while the other is a dependent variable. . [4] Learning a linear regression model entails estimating the values of the coefficients used in the representation using the data that we have available. When Linear Regression was implemented on a face dataset, got an accuracy score of 0.61 within a time of 0.203 sec without PCA Whereas with PCA, got an accuracy score of 0.52 within a time of 0.004 sec. The accuracy scores resulted in this work are not good hence it was not recommended.

D. KNN (K-Nearest Neighbor)

KNN (K-Nearest Neighbor) is a type of supervised learning algorithm in machine learning. [3] The K-NN approach may be used for both regression and classification, though it is more commonly utilized for classification tasks. During the training phase, the KNN algorithm simply saves the dataset, and when new data is received, it classifies it into a category that is quite similar to the new data. When K-NN was implemented on a face dataset, got an accuracy score of 0.86 within a time of 0.3 sec without PCA, whereas with PCA, got an accuracy score of 0.88 within a time of 0.209 sec. As a result, K-NN also gave a good accuracy when implemented hence, KNN is also recommended for classification problems.

E. Logistic Regression

Logistic regression is a classification technique that uses supervised learning to estimate the likelihood of a target variable. [2]. It is one of the most basic ML techniques that may be used to solve a variety of classification issues such as spam detection, diabetes prediction, cancer diagnosis, and so on. Logistic regression, often known as logit regression or logit model, is a type of regression analysis. When logistic regression implemented on the face dataset, shown an accuracy score of 0.97 Within a time of 5.74 sec without PCA whereas with PCA, shown an accuracy score of 0.93 within a time of 0.15 sec. To conclude, logistic regression gave the best accuracy among all other algorithms when compared.
Figure.2 shows the comparison of actual values with predicted values when Logistic regression was implemented without PCA.Figure.3 shows the comparison of actual values with predicted values when Logistic regression was implemented with PCA.

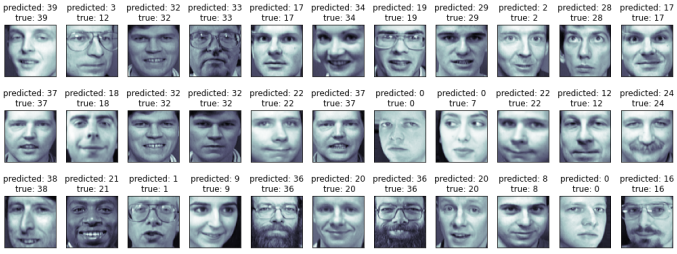

Fig. 2. Actual values and Predicted values when Logistic Regression implemented without PCA.

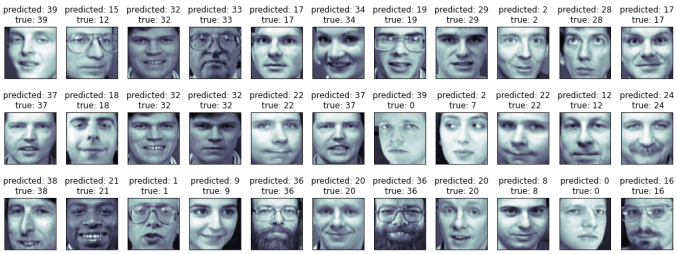

Fig. 3. Actual values and Predicted values when Logistic Regression implemented with PCA..

Figure.4. shows the graphical representation of accuracy values of all models when implemented with and without PCA.The x-axis represents the machine learning models used in the work and y-axis represents the accuracy percentage. The blue colour line indicates the data values of specified models when implemented without PCA whereas the red colour line indicates the data values of specified models with PCA. When compared, the graph is identical and shape of the graph is parabola for both PCA and without PCA however, Logistic Regression has better accuracy score when implemented without PCA and SVM has better accuracy score when implemented with PCA.Linear Regression has the lowest

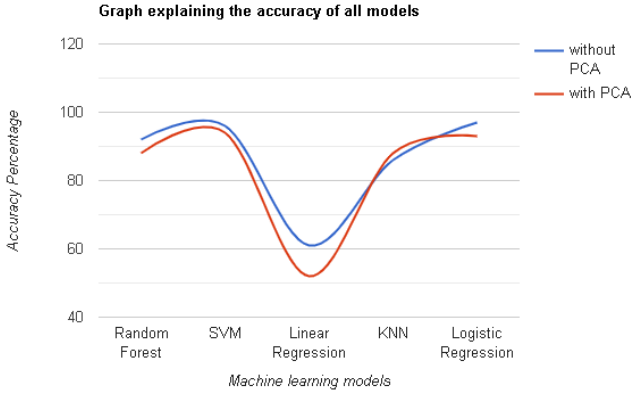

Fig. 4. Graph explaining the accuracy of all models .

TABLE I
MACHINE LEARNING MODELS AND THEIR ACCURACY WITHOUT PCA

Sl. No	Model	Without PCA	
		Accuracy	Time Taken
1	Random Forest	0.92	2.44
2	SVM (Support Vector Machine)	0.96	107.68
3	Linear Regression	0.61	0.203
4	KNN (K-Nearest Neighbor)	0.86	0.3
5	Logistic Regression	0.97	5.74

TABLE II
MACHINE LEARNING MODELS AND THEIR ACCURACY WITH PCA

Sl. No	Model	With PCA	
		Accuracy	Time Taken
1	Random Forest	0.88	0.511
2	SVM (Support Vector Machine)	0.94	4.77
3	Linear Regression	0.52	0.004
4	KNN (K-Nearest Neighbor)	0.88	0.209
5	Logistic Regression	0.93	0.15

accuracy when implemented in both the ways.

VI. CONCLUSION

Various Machine learning algorithms used for facial recognition is studied over facial recognition. Comparative analysis is made on accuracy scores and computational times with respect to dimensionality reduction using PCA. As a result, Logistic Regression can be the best algorithm for face recognition with an accuracy difference of only 4%when implemented without PCA and SVM can be the best algorithm for face recognition when implemented with PCA. Even though SVM achieved an accuracy of 94% and 96% when implemented with PCA and without PCA, the computational time (107.68 sec) is relatively high in without PCA that means almost 26 times higher when compared with the time taken (4.77 sec) by SVM in PCA. However, the difference in accuracy is just 2% and hence they are recommended for future works and both Logistic Regression and SVM shown better performance with respective to the time and accuracy scores.

REFERENCES

[1] J. Harikrishnan, Sudarsan, A., Sadashiv, A., and Remya Ajai A. S., "Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision", in 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN), Vellore, India, 2019.

[2] Dr. Tripty Singh and C. Mary, J., "Development of Accurate Face Recognition Process Flow for Authentication", in Emerging Research in Computing, Information, Communication and Applications ERCICA-2016, NMIT, Bangalore, 2016.

[3] K. Nimmy and Dr. M. Sethumadhavan, "Biometric Authentication via Facial Recognition", 2014.

[4] R. Neha and Nithin S, "Comparative Analysis of Image Processing Algorithms for Face Recognition", in 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018.

[5] J. Madarkar, Sharma, P., and Rimjhim Padam Singh, "Improved Performance and Execution Time of Face Recognition Using MRSRC", in Soft Computing for Problem Solving, Singapore, 2020.

[6] M. Nehru and Dr. Padmavathi S., "Illumination Invariant Face Detection using Viola Jones Algorithm", in 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), 2017.

[7] S. Menon, Swathi, J., Anit, S. K., Nair, A. P., and Sarath S., "Driver face recognition and sober drunk classification using thermal images", Proceedings of the 2019 IEEE International Conference on Communication and Signal Processing, ICCSP 2019. Institute of Electrical and Electronics Engineers Inc., Melmaruvathur; India, pp. 400-404, 2019.

[8] Kirill Smelyakov, Anastasiya Chupryna, Oleksandr Bohomolov, Nikita Hunko, "The Neural Network Models Effectiveness for Face Detection and Face Recognition", Electrical Electronic and Information Sciences (eStream) 2021 IEEE Open Conference of, pp. 1-7, 2021.

[9] Schroff, Florian & Kalenichenko, Dmitry & Philbin, James. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. Proc. CVPR.

[10] Waldemar Wójcik, Konrad Gromaszek and Muhtar Junisbekov (July 6th 2016). Face Recognition: Issues, Methods and Alternative Applications, Face Recognition - Semi Supervised Classification, Subspace Projection and Evaluation Methods, S. Ramakrishnan, IntechOpen, DOI: 10.5772/62950.