Article

A Comparative Study of Machine Learning Classifiers Performance with Feature Extraction for Face Recognition

Eimad Abusham *

Article

A Comparative Study of Machine Learning Classifiers Performance with Feature Extraction for Face Recognition

Eimad Abusham [1,*]

[1] Faculty of Computing and Information Technology, Sohar University, Sohar 311, Oman

* Correspondence: eabusham@su.edu.om

Abstract: It is crucial to select the right machine learning classifier for image classification and face recognition. This study examines the effectiveness of four different face recognition classifiers - Support Vector Machines (SVM), Random Forest, K-Nearest Neighbors (KNN), and Neural Networks. An analysis of the Large Faces in the Wild (LFW) dataset was carried out using Principal Component Analysis (PCA). Classifiers are rigorously trained and evaluated based on the extracted features. Comparison of classifier performance is an insightful way to figure out their strengths and weaknesses. Having a visual representation of the classifier's performance gives a complete understanding of its capabilities. Through the selection of the most appropriate classifier, study results contribute to advancements in image classification, recognition, and biometric identification. The comparison study demonstrated that the Neural Network classifier was exceptionally accurate and proficient in recognizing faces from the LFW dataset when used in conjunction with PCA for feature extraction. According to the comparative analysis, the Neural Network classifier proved exceptionally accurate and proficient at identifying faces from the LFW dataset when combined with PCA for feature extraction.

Keywords: machine learning; face recognition; image classification; feature extraction

1. Introduction

The potential of facial recognition has been harnessed in a variety of applications, including security systems, biometric authentication, and surveillance. Real-world scenarios, however, can present challenges such as noise, occlusion, lighting variations, and distortions, which can compromise accurate recognition. Choosing an appropriate classifier algorithm is crucial to ensuring accurate and reliable face recognition results. Novel descriptors and classifier techniques have led to advancements in face recognition systems over the years. Principal Component Analysis (PCA) emerges as a vital strategy for dimensionality reduction in face recognition due to its profound importance. PCA contributes to the efficiency and robustness of face recognition systems by effectively compressing facial data while preserving essential discriminative information. With biometrics gaining prominence in security, authentication, and personal identification applications, proper features and classifiers play a key role in enhancing accuracy and reliability. Feature extraction for hyperspectral images is examined in terms of its role in enhancing classification accuracy. Ensemble learning techniques will also be applied to bolster accuracy and robustness in individual identification from images, particularly in the context of facial recognition. The purpose of this study is to explore classifier algorithms that can be used for accurate recognition, while emphasizing PCA as a way to reduce dimensionality to increase efficiency. Furthermore, ensemble learning contributes to improved accuracy and resilience in image-based identification and recognition.

2. Related Work

The importance of face recognition continues to expand in a variety of domains, such as security, biometrics, and human-computer interaction, and this study plays a significant role in improving the effectiveness of these systems. The study by Poon et al [1] examines the problem of face recognition under image distortions. The study seeks to contribute valuable insight into robust face recognition application selection in real-world scenarios by evaluating and comparing various PCA-based algorithms. In spite of challenging conditions, the findings of this research may guide the development of more accurate and reliable face recognition systems. According to Karanwal [2], it is essential to evaluate and compare different face recognition classifier algorithms. A major goal of his investigation is to understand the efficacy of advanced descriptors within real-world contexts, based on the exploration of advanced descriptors. By analyzing each classifier, insights can be gained into its capabilities and limits. In their study [3], Malakar et al. emphasize the application of Principal Component Analysis (PCA) to face recognition. In this study, we look at the practical implementation of PCA in order to gain a better understanding of its effectiveness, advantages, and limitations. PCA is examined in detail within the context of face recognition systems in this research, with a view to shed light on PCA's operational aspects. As a result, it hopes to contribute to an improved understanding of PCA's impact on improving recognition precision, computational efficiency, and system robustness.

Research conducted by [4] aims to provide a comprehensive overview of the application, difficulties, and prospective progressions associated with deploying deep learning techniques for biometric identification to enhance accuracy and dependability. Incorporating insights from a wide range of academic publications, this study contributes to a better understanding of how deep learning will influence biometric identification in the future. Image analysis features extraction has been revolutionized by deep learning techniques over the past few years. Due to their capability to automatically learn hierarchical features from data, convolutional neural networks (CNNs) have gained attention. According to [5], CNN-based feature extraction can capture intricate spatial-spectral patterns, which improves classification accuracy through hyperspectral image analysis.

The survey paper by Sagi [6] provides a comprehensive overview of ensemble learning methods in a variety of domains, including facial recognition. In the survey, bagging and boosting are the two main categories of ensemble methods. Boosting methods, like AdaBoost, iteratively adjust weights to focus on misclassified instances, while bagging methods, such as Random Forests, resample training sets for diversity.

Ensemble learning is successful in facial recognition because it mitigates overfitting, bias, and noise issues that can affect individual classifiers. Ensembles improve generalization, reliability, and robustness by aggregating predictions from multiple classifiers. Studies, such as those cited in Sagi's survey, have shown that ensemble methods are highly effective in achieving high recognition accuracy rates, particularly in challenging scenarios like occlusions, changing lighting conditions, and changing poses. In their article [6], Kim et al. emphasize the importance of optimizing and tuning SVMs for face recognition. SVM's discriminative power and generalization ability depend heavily on the choice of hyperparameters, kernel functions, and regularization parameters. It is necessary to develop innovative solutions to overcome the challenges associated with SVM-based face recognition. Based on the study from [6], tailored feature extraction, kernel methods, and parameter optimization are essential to harnessing the potential of SVM for robust and accurate face recognition. In real-world scenarios, SVM-based face recognition could deliver reliable performance as researchers refine these approaches.

Kim's study [7] demonstrated the potential of Random Forest classifiers beyond traditional face recognition tasks by using them to recognize facial expressions. Based on its ensemble-based nature, the algorithm makes robust and accurate predictions by combining multiple decision trees. It is important to note that different illumination, poses, and expressions contribute to intra-class variations in face recognition. These challenges can be mitigated by using Random Forest classifiers that capture diverse feature patterns and adapt to data variations. Feature selection and extraction are highly efficient, allowing for reliable recognition even in high-dimensional feature spaces.

Random Forests and Convolutional Neural Networks (CNNs) work synergistically in detecting facial expressions [7,8]. Random Forests complement CNNs' feature extraction capabilities by providing an ensemble framework for decision-making, which enhances classification accuracy. [9] documented how CNN-based approaches are effective at capturing the intricate details of faces based on raw pixel data, which can be transformed into high-level features. To distinguish edges, textures, and complex facial structures, CNNs utilize convolutional layers and pooling layers to reduce dimensionality. A neural network-based facial identification system is further enhanced by transfer learning and pre-trained models [9]. It is possible to accelerate training and improve accuracy even with limited data by leveraging knowledge gained from one task and applying it to another. VGG, ResNet, and MobileNet are pre-trained CNN architectures that make it easy to build robust facial recognition systems.

A new variation of K-NN classifier, NS-k-NN, based on neuromorphic sets, is introduced by [10]. Considering uncertainty and indeterminacy in real-world data allows this algorithm to adapt and improve face recognition accuracy. Data points with similar features may belong to the same class according to K-NN. Facial features are expected to display patterns indicative of individual identity, and this notion aligns well with facial recognition. In an effort to enhance K-NN's ability to differentiate between individuals, researchers have explored various distance metrics and weighting strategies. Although K-NN is simple, it faces challenges when it comes to face recognition. K-NN might have difficulty accounting for intra-class variations due to changes in lighting conditions, facial expressions, and poses. To address these issues, weighted K-NN and distance normalization have been proposed, demonstrating the algorithm's adaptability.

In Minaee et al. [11], deep learning methods for biometric identification are explored exhaustively. There are various types of deep learning architectures applied to biometric modalities such as face, fingerprint, iris, and voice, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). Deep learning-based biometric identification relies heavily on transfer learning, as discussed in the study [11]. Researchers optimize pre-trained models to perform specific biometric tasks with limited data by leveraging pre-trained models and fine-tuning them.

By exploring information simultaneously in space, scale, and orientation, Li et al. [12] present a unique approach to classifier selection. A key finding of the paper is that different aspects of facial data should be combined in order to improve recognition accuracy. The multidimensional exploration of facial features is in harmony with their inherent complexity, which manifests in a variety of spatial, scale, and orientation variations. Ensemble methods and individual classifier optimization are both used in classifier selection strategies. The results of multiple classifiers are combined in ensemble techniques, such as bagging and boosting. In order to achieve robustness against diverse challenges, such as variations in lighting conditions, poses, and expressions, researchers have integrated multiple classifiers. As discussed in the paper [13], PCA can be used to construct informative facial subspaces based on feature selection. Enhancing recognition accuracy is achieved by keeping important facial characteristics and removing irrelevant noise. By transforming high-dimensional data into meaningful representations at a lower level, this technique can improve the accuracy and efficiency of recognition systems.

In his study, Almabdy [14] presents a comprehensive overview of deep learning techniques used in biometric systems, demonstrating their ability to handle various biometric modalities including fingerprints, faces, iris, and voice. The learning of intricate patterns and features from raw data has been demonstrated by deep learning methods such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). It is possible to capture complex relationships within biometric data using deep learning techniques. CNNs, for instance, excel at extracting features from images using hierarchical structures like edges, textures, and facial structures. A RNN, on the other hand, is capable of modeling sequential data, making it suitable for various modalities such as voice and signature recognition. Using various biometric modalities such as fingerprint, face, iris, and voice, Dasgupta et al. [15] present a comprehensive overview of deep learning techniques applied to biometric systems. Learning intricate patterns and

features from raw data is one of the most effective deep learning methods. These methods include generative adversarial networks (GANs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs).

The deep learning method excels at capturing complex relationships among biometric data. CNNs, for instance, are an excellent application for image-based biometrics because they extract features such as edges, texture, and structure of the face hierarchically. On the other hand, RNNs have the capability of modeling sequential data, so they can be used to recognize voice and signatures, for example.

There are a number of performance metrics that can be used to evaluate face recognition algorithms, including accuracy, precision, recall, F1-scores, and receiver operating characteristic (ROC) curves. Based on the dataset, accuracy is measured by the proportion of faces correctly recognized to the total number of faces. An algorithm's precision and recall are indicators of its ability to minimize false positives and false negatives, while an algorithm's F1-score offers a balance between them.

ROC curves can be used to illustrate the trade-off between true positive and false positive rates for different decision thresholds [16]. ROC curves summarize algorithm performance as measured by the Area Under the Curve (AUC). In their study, Ayesha et al. [17] emphasize the importance of evaluating and comparing dimensionality reduction techniques to determine the most suitable approach for a particular facial recognition task. Various aspects of computational efficiency, discriminatory information preservation, and data distribution flexibility are examined in comparative studies. A wide range of supervised learning algorithms have been explored for detecting faces, including Support Vector Machines (SVM), Decision Trees, Random Forests, and Convolutional Neural Networks (CNNs). The deep learning process of CNNs is in comparison with SVM, which provides interpretable results, Decision Trees, which provide ensemble-based classification, and Random Forests, which offer ensemble-based classification.

According to [18], the complexity of the model, training data, and feature extraction influence the performance of supervised learning algorithms for face detection. A CNN's accuracy and adaptability are enhanced by the fact that it learns features automatically from raw data, while traditional algorithms require manually engineered features. Furthermore, the study [18] points out that challenges related to face detection should be addressed, such as occlusions, poses, and lighting conditions. Due to their ability to capture hierarchical features, deep learning algorithms, in particular CNNs, have demonstrated remarkable resilience to such challenges. Using deep learning for medical image processing is described in a comprehensive manner by Razzak et al. [19]. As demonstrated in this research, deep learning techniques, such as convolutional and recurrent neural networks, provide a powerful method for detecting disease, classifying, and segmenting medical images.

It has been demonstrated that deep learning is highly accurate at identifying subtle patterns in medical images that are often indiscernible to the human eye. Specifically, CNNs are excellent at detecting lesions in radiological images and classifying cells in histopathology slides. Medical diagnoses have become more accurate and efficient due to the ability to autonomously learn features from raw data. Through the integration of multi-modal behavioral biometrics, Bailey et al. [20] investigate user identification and authentication. In order to construct a robust and reliable identification system, multiple behavioral traits, such as keystroke dynamics, handwriting patterns, and voice characteristics, must be harnessed in combination. In addition to bagging, boosting, and stacking, ensemble techniques encompass a vast array of approaches. Through these techniques, individual models are aggregated to reduce bias, variance, and instability. Ensemble techniques improve identification robustness and mitigate misclassification risk by combining predictions from diverse models.

Multimodal behavioral biometrics are essential for constructing ensemble identification systems, as described in the paper [20]. A behavioral trait may vary due to mood, environment, or health factors. When multiple traits are combined, not only is the identification more unique but it also reduces the risk of false rejections or acceptances. The study by Wang [21] offers a comprehensive

examination of the interactions between pattern recognition, machine intelligence, and biometrics. To construct efficient and effective identification systems, it is imperative to understand and leverage patterns within biometric data. An essential part of biometrics is pattern recognition, which identifies regularities and recurrent structures. It is possible to automate the extraction and recognition of these patterns from complex biometric data by using machine learning techniques, such as neural networks, support vector machines, and decision trees.

A wide range of aspects of face recognition and biometric identification are addressed in this collection of studies. In addition to examining robust face recognition methods in the face of image distortions, these research works also explore the practical application of Principal Component Analysis (PCA) during face recognition. A powerful strategy for improving biometric identification accuracy and efficiency can be found in the integration of deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). Facial recognition challenges such as overfitting and variability can be mitigated using ensemble learning methods, as demonstrated by Sagi's survey. To advance the accuracy and reliability of biometric identification systems, it is essential to select appropriate classifier algorithms, leverage dimensionality reduction techniques, and implement ensemble methods.

3. Proposed Framework

This study proposes a framework to reduce dimensionality and select the best face recognition classifier using a combination of dimensionality reduction and diverse classification algorithms. Figure 1 shows the sequential steps in the system, beginning with loading the dataset, testing classifiers, identifying the best classifier, and plotting the performance results. A description of the proposed algorithm can be found in Table 1.
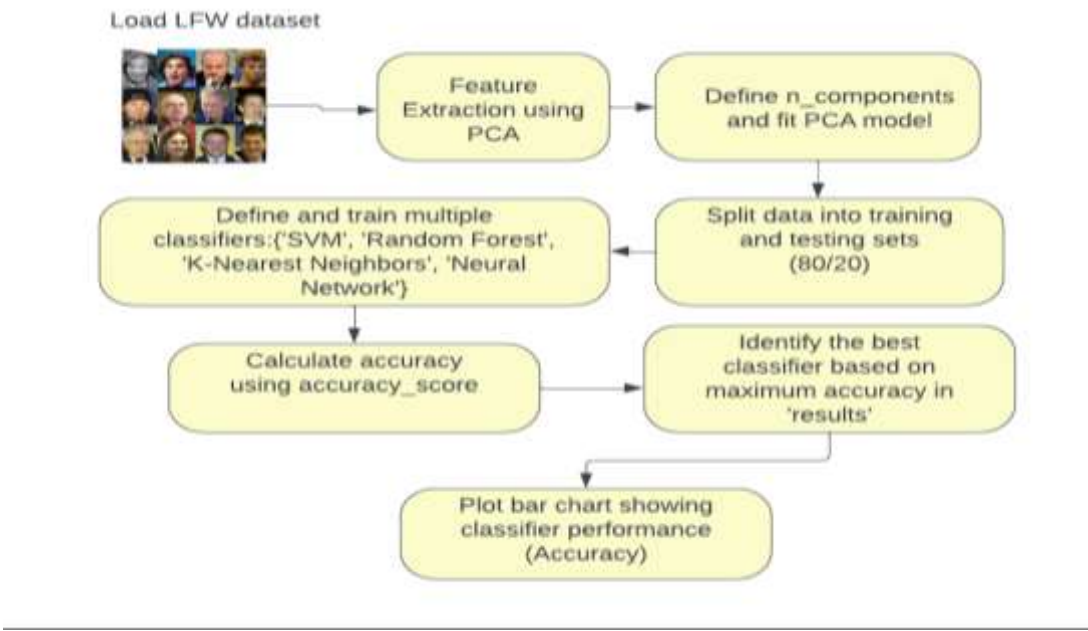

Figure 1. Visual Framework for Face Recognition.

An Algorithm 1 describing the steps involved in comparing and analyzing the performance of multiple classifiers for face recognition based on PCA is presented here.

Algorithm 1: Comparative Analysis of Classifier Algorithms
Input: LFW dataset images (X), corresponding labels (y)
Output: Best performing classifier with its accuracy
1:    Upload images (X) and labels (Y) from the LFW dataset.

2:      PCA features are extracted.

Determine the number of principal components (n_components).

Perform PCA on the images (X) using n_components and whiten=True.

Reduce the image dimensions by transforming it to a pixelated representation (X_pca).

3.   Divide the data into testing and training sets.

4.   Create multiple classifiers and train them:

Initialize SVM, Random Forest, K-Nearest Neighbor, and Neural Network classifier objects.

5.   Create a dictionary 'results' to store accuracy data.

6.   For each of the classifiers in 'classifiers':

a. Use X_train and Y_train to train the classifier.

b. Apply the trained classifier to predict X_test labels.

c. Calculate accuracy by comparing y_test with y_predicted labels (y_pred).

d. Keeping accuracy in a data table called 'results' with classifier names as keys.

7.   Compare the performance of the classifiers:

8.   Display the accuracy of the best classifier.

3.1. Dimensional reduction

By using Principal Component Analysis (PCA), the method extracts features from high-dimensional image data and reduces its dimensionality while preserving important variances. According to Algorithm 2, the graph plots the cumulative explained variance against the number of PCA components. By plotting how much variance is explained by each principal component, 150 was determined as the optimal number of components when reducing dimensionality.

Algorithm 2: PCA-Based Feature Extraction

Input: X (image data), n_components

Output: X_pca (transformed feature matrix)

1.   Compute PCA with n_components for image data X.

2.   Calculate the transformation matrix by fitting the PCA model.

3.   Get X_pca by transforming image data X using the transformation matrix.
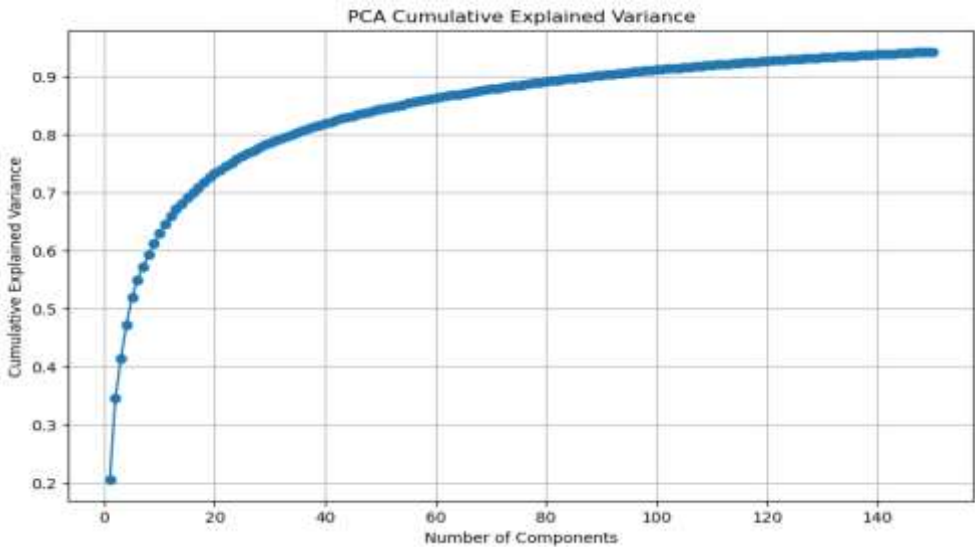
4.   Then return X_pca.

Figure 2. Plotting the variance of the data by each additional principal component will show how much variance can be explained.

3.2. Algorithms

The proposed algorithm defines a diverse set of machine learning classifiers and subsequently trains them to recognize faces. Several algorithmic approaches can be utilized for these classifiers, each with its own strengths and weaknesses. In the dictionary 'classifiers', there are four key-value pairs, each representing the name of the classifier, such as Support Vector Machine (SVM) Algorithm 3 shows how this works, Random Forest as shown in Algorithm 4, K-Nearest Neighbors (KNN) based on Algorithm 5, and Neural Network as shown by on Algorithm 6. Classifier objects are instantiated from the respective classes with default parameters. With this selection of classifiers, different approaches to face recognition can be comprehensively evaluated. As these classifiers are trained on PCA-transformed training data, the subsequent analysis aims to identify their performance and suitability for a particular task by gauging their accuracy.

Algorithm 3: Support Vector Machine (SVM) [22]

Input: X_train_pca, y_train (training labels), X_test_pca

Output: y_pred (predicted labels), accuracy

1. Set the default parameters for the SVM classifier.
2. Use X_train_pca and Y_train to train the SVM classifier.
3. Apply the trained SVM classifier to predict the labels for X_test_pca.
4. Calculate accuracy using accuracy_score by comparing y_test and y_pred.
5. Produce a classification report based on the y_test and y_pred data.
6. Return classification_report, y_pred, and accuracy.

Algorithm 4: Random Forest Classifier [23]

Input: X_train_pca, y_train (training labels), X_test_pca

Output: y_pred (predicted labels), accuracy

1. Set the default parameters for the Random Forest classifier.
2. Use the X_train_pca and Y_train functions to train the Random Forest classi-fier.
3. Utilize the Random Forest classifier to predict labels for X_test_pca.
4. Calculate accuracy using accuracy_score and comparing y_test and y_pred.
5. Create classification report for y_test and y_pred using classifica-tion_report.
4. Get y_pred, accuracy, and classification_report.

Algorithm 5: K-Nearest Neighbors (KNN) [24]

Input: X_train_pca, y_train (training labels), X_test_pca

Output: y_pred (predicted labels), accuracy

1. Set the default parameters for the Random Forest classifier.
2. Use the X_train_pca and Y_train functions to train the Random Forest classi-fier.
3. Utilize the Random Forest classifier to predict labels for X_test_pca.
4. Calculate accuracy using accuracy_score and comparing y_test and y_pred.
5. Create classification report for y_test and y_pred using classifica-tion_report.
4. Get y_pred, accuracy, and classification_report.

Algorithm 5: Neural Network [25]

Input: X_train_pca, y_train (training labels), X_test_pca

Output: y_pred (predicted labels), accuracy

1. Set the default parameters for the MLP classifier.
2. Use X_train_pca and Y_train to train the MLP classifier.
3. Use the trained MLP classifier to predict labels for X_test_pca.
4. Calculate the accuracy by comparing y_test to y_pred using accuracy_score.
5. Create detailed classification reports based on y_test and y_pred data.
6. Calculate the y_pred, the accuracy, and the classification report.

4. Experiments and Results

4.1. LFW Database

There are thousands of faces in the LFW dataset [26], representing a diverse range of ages, genders, ethnicities, and facial expressions. As a result of this diversity, this dataset is ideal for evaluating algorithms across a wide range of demographics and environmental conditions. This dataset contains thousands of images taken under a wide variety of lighting, pose, and occlusion conditions. Figure 3 illustrates the challenges presented by this diversity for algorithms.

Figure 3. Random Samples from LFW dataset.

4.2. Result

Based on PCA-based feature extraction techniques along with the Support Vector Machine (SVM) classifier, the results of the face recognition experiment are as follows in Table 1. Depending on how difficult it was to distinguish individuals, precision, recall, and F1-score varies for each class.

With an overall accuracy of 81%, the SVM classifier is capable of recognizing faces from the dataset when combined with PCA for feature extraction. It is challenging to distinguish some classes accurately than others, however, as performance varies across classes.

Table 1. Support Vector Machine (SVM) performance matrix.

SVM Accuracy: 0.81				
	precision	recall	f1-score	support
0	1.00	0.36	0.53	11
1	0.82	0.85	0.83	47
2	1.00	0.55	0.71	22

3	0.75	0.97	0.85	119
4	1.00	0.74	0.85	19
5	1.00	0.38	0.56	13
6	1.00	0.70	0.83	27
accuracy			0.81	258
macro avg	0.94	0.65	0.74	258
weighted avg	0.85	0.81	0.80	258

Faces identified by the Random Forest classifier were accurate based on the dataset. Table 2 shows that precision, recall, and F1-score varied for each class due to the classifier's performance in distinguishing individuals.

Table 2. Random Forest performance matrix.

Random Forest Accuracy: 0.58				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	11
1	0.79	0.49	0.61	47
2	0.60	0.14	0.22	22
3	0.54	0.97	0.69	119
4	0.50	0.11	0.17	19
5	0.00	0.00	0.00	13
6	1.00	0.19	0.31	27
accuracy			0.58	258
macro avg	0.49	0.27	0.29	258
weighted avg	0.59	0.58	0.50	258

As evidenced by the 58% accuracy rate, the Random Forest classifier was unable to recognize faces from the dataset satisfactorily, when combined with PCA for feature extraction. It is evident that certain individuals are difficult to distinguish accurately across different classes. As a result of this analysis, the classification approach can be improved in terms of accuracy and efficiency by improving its strengths and limitations.

To identify faces from the dataset, the K-NN classifier is applied. Table 3 evaluates the classifier's performance in recognizing individuals by measuring precision, recall, and F1-score for each class.

Table 3. K-Nearest Neighbors performance matrix.

K-Nearest Neighbors Accuracy: 0.69				
	precision	recall	f1-score	support
0	0.33	0.36	0.35	11
1	0.67	0.70	0.69	47
2	0.67	0.55	0.60	22
3	0.70	0.92	0.80	119
4	0.75	0.32	0.44	19

	precision	recall	f1-score	support
5	1.00	0.31	0.47	13
6	0.75	0.33	0.46	27
accuracy			0.69	258
macro avg	0.70	0.50	0.54	258
weighted avg	0.70	0.69	0.66	258

The overall accuracy of 69% indicates that the K-NN classifier, combined with PCA to extract features, performed reasonably well in recognizing faces in the dataset. However, classifier performance varies across classes, suggesting that some individuals are hard to recognize. As a result of this analysis, the classification approach's strengths and limitations are made clear, providing a basis for further refining its accuracy and effectiveness.

Neural networks successfully identified faces in the dataset with impressive results. In Table 4, precision, recall, and F1-scores were calculated for each class to evaluate the performance of the classifier.

Table 4. Neural Network performance matrix.

Neural Network Accuracy: 0.85				
	precision	recall	f1-score	support
0	0.55	0.55	0.55	11
1	0.77	0.87	0.82	47
2	0.79	0.86	0.83	22
3	0.92	0.92	0.92	119
4	0.82	0.74	0.78	19
5	1.00	0.69	0.82	13
6	0.84	0.78	0.81	27
accuracy			0.85	258
macro avg	0.81	0.77	0.79	258
weighted avg	0.85	0.85	0.85	258
Best Classifier: Neural Network with Accuracy: 0.85				

The Best Classifier for faces was recognized with 85% accuracy using the Neural Network classifier with the PCA feature extraction. A classifier's performance across different classes can also be determined using precision, recall, and F1-score metrics. Moreover, the classifier was ranked as the best classifier among the tested algorithms, with 85% accuracy. Based on the results of the analysis, it is clear that the neural network is capable of identifying individuals accurately and efficiently, making it a strong candidate for face recognition.

Figure 4 illustrates the overall performance of the neural network classifier, combined with PCA for feature extraction, when it comes to identifying faces in the LFW dataset.
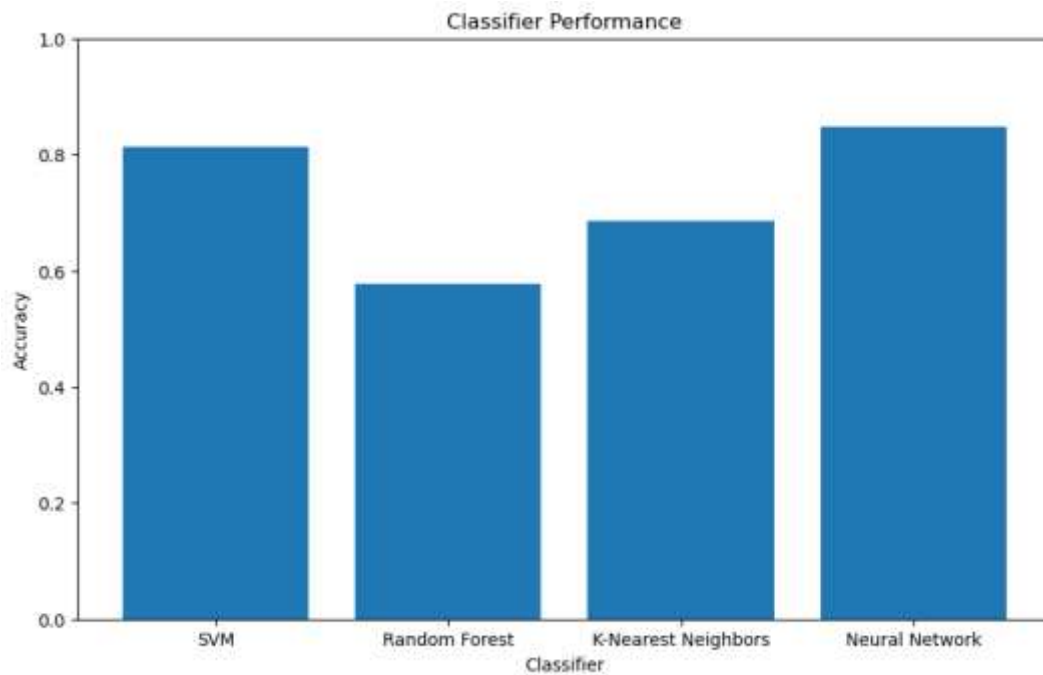
Figure 4. Performance of four machine learning classifiers on LFW dataset.

5. Conclusions

The aim of this study was to evaluate the performance of four machine learning classifiers in the context of face recognition: Support Vector Machines (SVMs), Random Forests, K-Nearest Neighbors (KNNs), and Neural Networks. Using the Large Faces in the Wild (LFW) dataset, Principal Component Analysis (PCA) was used to extract features. The primary goal of this study was to assess this classifier's efficiency and accuracy in recognizing faces, in order to gain insight into its strengths and weaknesses.

In identifying faces from the dataset, the Support Vector Machine (SVM) achieved an accuracy of 81%. There was a difference in precision, recall, and F1-scores among classes, indicating that discriminating between individuals is difficult at different levels of difficulty. The SVM demonstrated a respectable performance, focusing on precision and recall. A Random Forest classifier, however, was able to recognize faces with an accuracy of 58%. In a similar manner to SVM, the performance varied according to the class. Although the Random Forest classifier presented challenges in certain cases, it showed potential. A K-Nearest Neighbors (KNN) classifier achieved 69% accuracy in identifying faces. There were challenges with precision, recall, and F1-scores in some classes, while there were reasonable results in others. It appears that the system recognizes faces moderately well. In comparison with the other algorithms tested, the Neural Network classifier achieved an impressive 85% accuracy. In most classes, it displayed high precision, recall, and F1-score, demonstrating its accuracy. As a result, the Neural Network, a powerful classifier with real-world applications, was found to be the best classifier overall. The comparative analysis revealed that the LFW dataset was recognized exceptionally accurately and proficiently by the Neural Network classifier combined with the PCA feature extraction. By aiding in the selection of suitable classifiers for face recognition tasks, this study contributes to the advancement of image classification and biometric identification. As a result of this analysis, further refinements can be made based on the strengths and limitations it highlights.

Data Availability Statement: The data presented in this study are publicly available and accessible via the sklearn webpage.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Poon, Bruce, M. Ashraful Amin, and Hong Yan. "Performance evaluation and comparison of PCA Based human face recognition methods for distorted images." International Journal of Machine Learning and Cybernetics, 2011,2, 245-259.
2. Karanwal S. A comparative study of 14 state of art descriptors for face recognition. Multimedia Tools and Applications. 2021 Mar;80(8):12195-234.
3. Malakar, S., Chiracharit, W., Chamnongthai, K. and Charoenpong, Masked face recognition using principal component analysis and deep learning. In 18th International conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON),785-788, May, 2018.
4. Sundararajan K, Woodard DL. Deep learning for biometrics: A survey. ACM Computing Surveys (CSUR). 2018,51(3):1-34.
5. Liu, Bing, Xuchu Yu, Pengqiang Zhang, Anzhu Yu, Qiongying Fu, and Xiangpo Wei. "Supervised deep feature extraction for hyperspectral image classification." IEEE Transactions on Geoscience and Remote Sensing 56,2017, 4,1909-1921.
6. Sagi O, Rokach L. Ensemble learning: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2018,8,4,1249.
7. Kim SK, Park YJ, Toh KA, Lee S. SVM-based feature extraction for face recognition. Pattern Recognition. 2010,43(8),2871-81.
8. Wang Y, Li Y, Song Y, Rong X. Facial expression recognition based on random forest and convolutional neural network. Information. 2019,10(12),375.
9. Almabdy S, Elrefaei L. Deep convolutional neural network-based approaches for face recognition. Applied Sciences. 2019 Oct 17;9(20):4397.
10. Akbulut Y, Sengur A, Guo Y, Smarandache F. NS-k-NN: Neutrosophic set-based k-nearest neighbors classifier. Symmetry. 2017,9(9):179.
11. Minaee S, Abdolrashidi A, Su H, Bennamoun M, Zhang D. Biometrics recognition using deep learning: A survey. Artificial Intelligence Review. 2023,1-49.
12. Lei Z, Liao S, Pietikäinen M, Li SZ. Face recognition by exploring information jointly in space, scale and orientation. IEEE transactions on image processing. 2010,20(1),247-56.
13. Song F, Guo Z, Mei D. Feature selection using principal component analysis. In2010 international conference on system science, engineering design and manufacturing informatization 2010 Nov 12,1, 27-30. IEEE.
14. Almabdy SM, Elrefaei LA. An overview of deep learning techniques for biometric systems. Artificial Intelligence for Sustainable Development: Theory, Practice and Future Applications. 2021:127-70.
15. Dasgupta D, Roy A, Nag A. Advances in user authentication. Cham, Switzerland: Springer International Publishing; 2017 Aug 22.
16. Phillips PJ, Wechsler H, Huang J, Rauss PJ. The FERET database and evaluation procedure for face-recognition algorithms. Image and vision computing. 1998 Apr 27;16(5):295-306.
17. Ayesha S, Hanif MK, Talib R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. Information Fusion. 2020,1;59:44-58.
18. Singhal N, Ganganwar V, Yadav M, Chauhan A, Jakhar M, Sharma K. Comparative study of machine learning and deep learning algorithm for face recognition. Jordanian Journal of Computers and Information Technology. 2021,1;7(3).
19. Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. Classification in BioApps: Automation of Decision Making. 2018,323-50.
20. Bailey KO, Okolica JS, Peterson GL. User identification and authentication using multi-modal behavioral biometrics. Computers & Security. 2014 Jun 1;43:77-89.
21. Wang PS, editor. Pattern recognition, machine intelligence and biometrics. Springer Berlin Heidelberg; 2011 Dec 27.
22. Wang L, editor. Support vector machines: theory and applications. Springer Science & Business Media; 2005.
23. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing. 2016,114:24-31.
24. Peterson LE. K-nearest neighbor. Scholarpedia. 2009 Feb 21;4(2):1883. Peterson LE. K-nearest neighbor. Scholarpedia. 2009,4(2),1883.
25. Pinkus A. Approximation theory of the MLP model in neural networks. Acta numerica. 1999,8,143-95.
26. Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. InWorkshop on faces in'Real-Life'Images: detection, alignment, and recognition 2008 Oct.