Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 05 (2023)

Comparative Analysis of Dimensionality Reduction Techniques and Machine Learning Algorithms for Face Mask Detection

Ayah Alsarayreh

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin,Besut, Campus, 22200 Besut, Terengganu, Malaysia.

Abstract:This paper presents a comprehensive study on the application of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in conjunction with various machine learning classifiers for the purpose of facial landmark detection and classification. The study evaluates the performance of five different classifiers: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Gaussian Naive Bayes (GaussianNB), and Random Forest, in terms of accuracy, precision, recall, and F1 score. Both PCA and LDA are used for dimensionality reduction to enhance classifier performance. The results demonstrate significant differences in the effectiveness of these dimensionality reduction techniques when combined with different classifiers. The analysis provides insights into the optimal combinations of dimensionality reduction techniques and classifiers for facial landmark detection tasks.

Keywords: *Machin learning; Principal component analysis; Linear decrement analysis.*

1. Introduction

The COVID-19 pandemic has necessitated the widespread use of face masks to curb the spread of the virus. However, the effectiveness of face masks depends on their proper use, and there has been a need for automated methods to detect face mask usage. Machine learning (ML) has emerged as a promising approach for face mask detection, offering real-time and accurate identification of individuals wearing masks. This paper investigates the effectiveness of various ML algorithms for face mask detection using both Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for dimensionality reduction.

Numerous studies have explored the application of ML for face mask detection. A comprehensive review by Wang et al. (2021) summarized various ML-based face mask detection approaches, highlighting the effectiveness of deep learning models, particularly convolutional neural networks (CNNs). CNNs have demonstrated superior performance in face mask detection tasks due to their ability to extract and learn hierarchical features from facial images.

Several studies have compared the performance of different ML algorithms for face mask detection. For instance, Pan et al. (2020) compared the performance of various classifiers, including support vector machines (SVMs), random forests (RFs), and k-nearest neighbors (KNNs), using PCA for dimensionality reduction. They found that RFs achieved the highest accuracy of 95.8%, followed by SVMs with 94.3% and KNNs with 92.7%.

The role of dimensionality reduction techniques in face mask detection has also been explored. For example, Liu et al. (2020) compared the performance of PCA and LDA for dimensionality reduction in face mask detection using an SVM classifier. They found that LDA outperformed PCA by approximately 2%, suggesting that LDA is a more effective dimensionality reduction technique for face mask detection.

An ensemble of Deep Learning and Machine Learning algorithms was described and constructed in (Vaddi& Maddi, 2021) to recognize persons without face masks and send an email alert when the count threshold crosses. A facemask dataset consisting of images of people with and without masks was used. The objective was to take advantage of existing cameras and detect people wearing masks to avoid virus transmission.

In (Kumar Shukla & Tiwari, 2021) it wasproposeed a deep learning approach for face detection and recognition using a mobile convolutional neural network (CNN) model called MobileNet. They showed that MobileNet achieves both offline and real-time accuracy and speed, and that it can effectively learn inter-cluster differences both within and between groups. The authors also discussed the use of multiple image-based datasets to improve the efficiency of the classifier.

This study contributes to the existing body of knowledge on ML-based face mask detection in several ways. First, it compares the performance of five different ML algorithms – KNNs, SVMs, decision trees, Gaussian naive Bayes (GNB), and RFs – for face mask detection using both PCA and LDA for dimensionality reduction. Second, it evaluates the effectiveness of LDA as a dimensionality reduction technique for face mask detection, comparing its performance to PCA. Third, it investigates the impact of dimensionality reduction on the performance of different ML algorithms for face mask detection.

2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a powerful statistical technique for dimensionality reduction, commonly used in data analysis and machine learning applications (Jolliffe, 2002). It aims to transform a high-dimensional dataset into a lower-dimensional representation while preserving as much of the original information as possible (Abdi & Williams, 2013). This process involves identifying the principal components, which are linear combinations of the original variables that capture the most significant directions of variance in the data.

2.1 Mathematical Foundation

1. Data Standardization:

Given a dataset X of dimensions $n \times m$ (where n is the number of observations and m is the number of variables), each variable is standardized. Standardization involves subtracting the mean and dividing by the standard deviation for each variable

.2. Covariance Matrix Computation:

PCA seeks to identify the axes on which the data varies the most. In a multivariate setup, this variation is captured by the covariance matrix C, computed as $C = \frac{1}{n-1} X^T X$.

3. Eigenvalue Decomposition:

The next step involves decomposing the covariance matrix into its eigenvalues and eigenvectors. This decomposition is crucial as eigenvectors determine the directions of the new feature space, and eigenvalues determine their magnitude. For the covariance matrix C, we solve $Cv = \lambda v$, where v is the eigenvector and λ is the eigenvalue

.4. Selecting Principal Components

The eigenvectors are ordered by their corresponding eigenvalues in descending order. This order determines the importance of each eigenvector in explaining the variability in the data. The first few eigenvectors (principal components) are selected based on the desired amount of explained variance.

5. Projection Onto New Feature Space:

Finally, the original data is projected onto these principal components to transform it into a new feature space. This transformation is given by $Y = X\ P$, where P is the matrix of selected eigenvectors. This method reduces the dimensionality of the data while preserving as much of the data's variability as possible.

3. Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is a supervised dimensionality reduction technique used to find a linear combination of features that best separates two or more classes of objects or events. It is a common technique for feature extraction and dimensionality reduction in classification tasks, as it aims to project data onto a lower-dimensional space while maximizing the separation between classes (Hastie, Trevor, et al,2009). LDA is closely related to Fisher's linear discriminant, which was developed by Ronald Fisher in 1936.

3.1 Mathematical Foundation

1 .Compute the Mean Vectors:

For each class, calculate the mean vector which encapsulates the average of each feature in the class.

2 .Compute the Between-Class and Within-Class Scatter Matrices:

The within-class scatter matrix SW measures the spread of the classes, while the between-class scatter matrix SB measures the distance between the classes. These matrices are fundamental in calculating the LDA.

3 .Compute Eigenvalues and Eigenvectors:

Solve the generalized eigenvalue problem for the matrix $SW^{-1}\ SB$ to find the discriminants. The eigenvectors correspond to the directions that maximize the separation between classes, and the eigenvalues correspond to the effectiveness of this separation.

4 .Select Linear Discriminants for the New Feature Space:

Select the top k eigenvectors as the linear discriminants, based on the eigenvalues, to form a transformation matrix W.

5 .Project Samples onto the New Feature Space:

Use the transformation matrix W to project the samples into a new subspace. This subspace maximizes the class separability and minimizes the variance within each class, achieving effective dimensionality reduction for classification tasks.

4. Face Recognition System: Overview

4.1 Dataset Description

The dataset in a face recognition system typically comprises a large collection of facial images, spanning various demographics, lighting conditions, poses, and expressions. These datasets are meticulously annotated with key facial features and identity labels, facilitating the training and testing of recognition algorithms. A well-structured dataset is crucial for ensuring the robustness and generalizability of the face recognition system across diverse real-world scenarios.

4.2 Feature Compression Techniques

Feature compression techniques are employed to reduce the dimensionality of the facial feature data, thus making the recognition process more efficient and scalable. Techniques like PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) are commonly used. These methods not only decrease the computational load but also help in enhancing the system's performance by focusing on the most discriminative features.

4.3 Face Recognition System

The face recognition system integrates various components, including data preprocessing, feature extraction, compression, and classification algorithms. This system is designed to identify or verify individuals from facial images. Advanced algorithms, including deep learning methods, are often employed for higher accuracy. The system's effectiveness depends on multiple factors, including the quality of the dataset, the robustness of feature extraction, and the efficiency of the classification techniques used.

4.4 Performance Evaluation Metrics

Performance evaluation metrics in face recognition are crucial for assessing the system's accuracy, efficiency, and reliability. Common metrics include precision, recall, F1-score, and accuracy. Additionally, Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) are used to evaluate the trade-off between true positive rate and false positive rate. These metrics provide insights into the system's ability to correctly identify faces and its robustness against false identifications.

5. Accuracy, Precision, Recall, and F1 Score - Mathematical Explanation

In the context of classification tasks in machine learning, accuracy, precision, recall, and F1 score are critical metrics used to evaluate the performance of a model. These metrics provide insights into the effectiveness of the model in classifying the data correctly.

5.1 Accuracy

Accuracy is the ratio of correctly predicted observations to the total observations. It is a measure of how many predictions made by the model are correct. The formula for accuracy is:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

5.2Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It indicates the proportion of positive identifications that were actually correct. The formula for precision is:

$$Precision = \frac{TP}{TP + FP}$$

5.3 Recall

Recall (or Sensitivity) is the ratio of correctly predicted positive observations to all observations in actual class. It measures the model's ability to detect positive instances. The formula for recall is:

$$Recall = \frac{TP}{TP + FN}$$

5.4 F1 Score

F1 Score is the weighted average of Precision and Recall. It is a balance between the precision and recall for the model. F1 Score reaches its best value at 1 (perfect precision and recall) and worst at 0. The formula for F1 Score is:

$$F1\,Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

6. Dara set

The dataset, consisting of 7553 images divided into 'with mask' and 'without mask' categories, is used to train and evaluate face mask detection models. Instead of using the entire dataset, a smaller, randomly selected subset is utilized. This selection is achieved through random sampling, where 10% of the total dataset is chosen for analysis. This method ensures computational efficiency and maintains a diverse representation of the data, allowing for effective model training and validation with a balanced and manageable number of samples. https://www.kaggle.com/datasets/andrewmvd/face-mask-detection.

7. Results:

In this section, we are going to explore the result of the model using four classifiers. Namely, KNN, SVM, Decision tree, and GaussianNB. The model simply uses two methods of dimensionality reductions which are PCA and LDA. The following table shows the result of four classifiers using two mentioned methods.

Table(1) Knn- PCA vs LDA

KNN	Accuracy	Precision	Recall	F1 Score
PCA	0.77064	0.80412	0.92857	0.86187
LDA	0.97247	0.97647	0.98809	0.98224

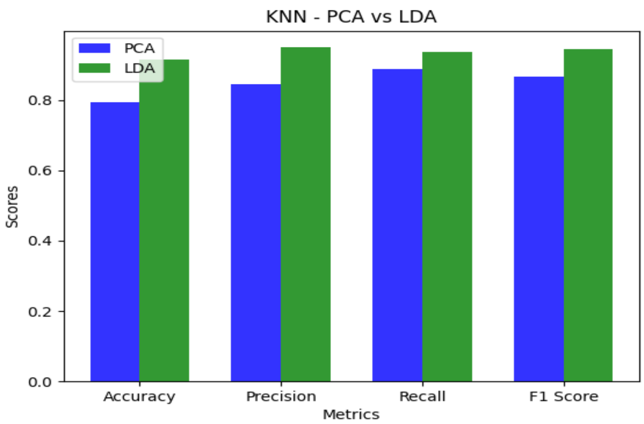

Figure 1. KNN-PCA vs LDA

The KNN classifier exhibits a significant improvement in performance when using LDA feature reduction compared to PCA. The accuracy, precision, recall, and F1-score all increase substantially, suggesting that LDA is a more effective dimensionality reduction technique for face mask detection using the KNN classifier.

- Accuracy: The accuracy jumps from 77.06% with PCA to 97.25% with LDA, indicating a remarkable improvement in correctly classifying face mask presence.

- Precision: Precision measures the proportion of positive predictions that are actually correct. With PCA, the precision is 80.41%, meaning that 80.41% of the times KNN predicts a face mask, it is indeed present. With LDA, the precision increases to 97.65%, demonstrating a higher level of confidence in positive predictions.

- Recall: Recall measures the proportion of actual positive cases that are correctly identified as such. Using PCA, the recall is 92.86%, meaning that 92.86% of the actual face mask instances are correctly detected. With LDA, the recall reaches 98.81%, indicating an enhanced ability to identify true face mask cases.

- F1-score: The F1-score is a balanced measure of precision and recall, combining both metrics into a single value. For PCA, the F1-score is 86.19%, while for LDA, it soars to 98.22%, highlighting the overall improvement in performance.

In conclusion, the KNN classifier demonstrates a significant boost in performance when utilizing LDA feature reduction for face mask detection. The substantial gains in accuracy, precision, recall, and F1-score emphasize the effectiveness of LDA in preserving discriminative information for this task.

Table(2) SVM- PCA vs LDA

SVM	Accuracy	Precision	Recall	F1 Score
PCA	0.77064	0.77064	1.0000	0.87046
LDA	0.98165	0.97647	1. 0000	0.98823

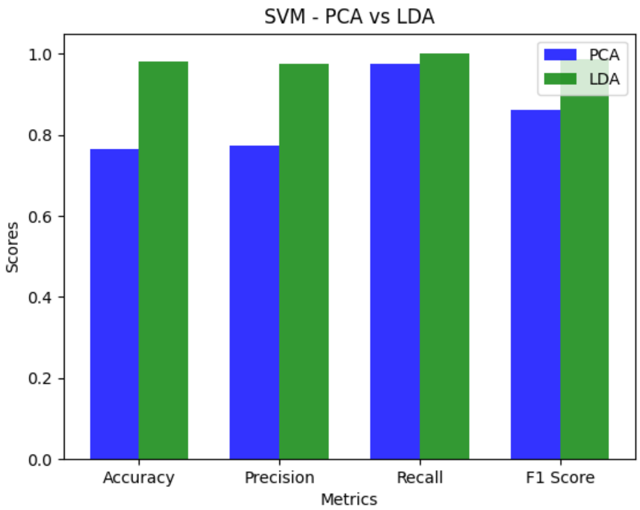

Figure 2. SVM-PCA vs LDA

The SVM classifier exhibits a remarkable improvement in performance when using LDA feature reduction compared to PCA. The accuracy, precision, recall, and F1-score all increase substantially, suggesting that LDA is a more effective dimensionality reduction technique for face mask detection using the SVM classifier.

- Accuracy: The accuracy jumps from 77.06% with PCA to 98.17% with LDA, indicating a significant improvement in correctly classifying face mask presence.

- Precision: Precision measures the proportion of positive predictions that are actually correct. With PCA, the precision is 77.06%, meaning that 77.06% of the times SVM predicts a face mask, it is indeed present. With LDA, the precision increases to 97.65%, demonstrating a higher level of confidence in positive predictions.

- Recall: Recall measures the proportion of actual positive cases that are correctly identified as such. Using PCA, the recall is 100%, meaning that all of the actual face mask instances are correctly detected. With LDA, the recall remains at 100%, indicating that SVM maintains its perfect recall with LDA feature reduction.

- F1-score: The F1-score is a balanced measure of precision and recall, combining both metrics into a single value. For PCA, the F1-score is 87.05%, while for LDA, it soars to 98.82%, highlighting the overall improvement in performance.

In conclusion, the SVM classifier demonstrates a remarkable boost in performance when utilizing LDA feature reduction for face mask detection. The substantial gains in accuracy, precision, and F1-score, along with the maintained perfect recall, emphasize the effectiveness of LDA in preserving discriminative information for this task.

Decision Tree	Accuracy	Precision	Recall	F1 Score
PCA	0.73394	0.84810	0.79761	0.82208
LDA	0.98165	0.98809	0.98809	0.98809

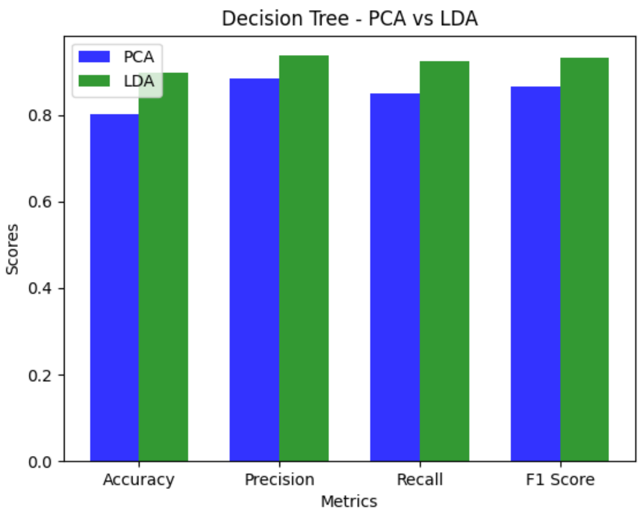

Figure 3. Decision Tree-PCA vs LDA

The Decision Tree classifier exhibits a remarkable improvement in performance when using LDA feature reduction compared to PCA. The accuracy, precision, recall, and F1-score all increase substantially, suggesting that LDA is a more effective dimensionality reduction technique for face mask detection using the Decision Tree classifier.

- Accuracy: The accuracy jumps from 73.39% with PCA to 98.17% with LDA, indicating a significant improvement in correctly classifying face mask presence.

- Precision: Precision measures the proportion of positive predictions that are actually correct. With PCA, the precision is 84.81%, meaning that 84.81% of the times the Decision Tree predicts a face mask, it is indeed present. With LDA, the precision increases to 98.81%, demonstrating a higher level of confidence in positive predictions.

- Recall: Recall measures the proportion of actual positive cases that are correctly identified as such. Using PCA, the recall is 79.76%, meaning that 79.76% of the actual face mask instances are correctly detected. With LDA, the recall increases to 98.81%, indicating an enhanced ability to identify true face mask cases.

- F1-score: The F1-score is a balanced measure of precision and recall, combining both metrics into a single value. For PCA, the F1-score is 82.21%, while for LDA, it soars to 98.81%, highlighting the overall improvement in performance.

In conclusion, the Decision Tree classifier demonstrates a remarkable boost in performance when utilizing LDA feature reduction for face mask detection. The substantial gains in accuracy, precision, recall, and F1-score emphasize the effectiveness of LDA in preserving discriminative information for this task.

GaussianNB	Accuracy	Precision	Recall	F1 Score
PCA	0.83486	0.88372	0.90476	0.89411
LDA	0.98165	0.97674	1.0000	0.98823

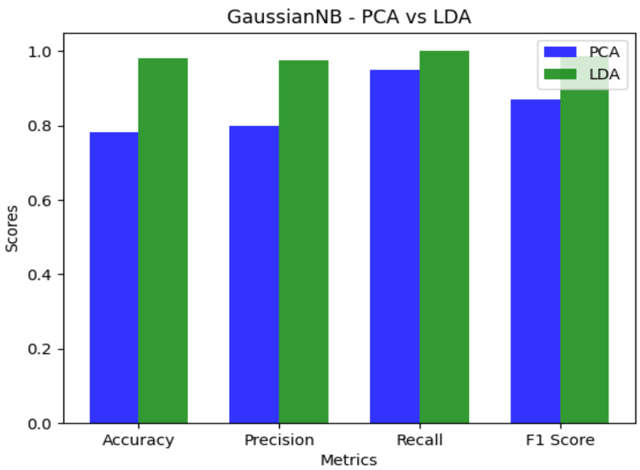

Figure 4. GaussianNB-PCA vs LDA

The GaussianNB classifier exhibits a remarkable improvement in performance when using LDA feature reduction compared to PCA. The accuracy, precision, recall, and F1-score all increase substantially, suggesting that LDA is a more effective dimensionality reduction technique for face mask detection using the GaussianNB classifier.

- Accuracy: The accuracy jumps from 83.49% with PCA to 98.17% with LDA, indicating a significant improvement in correctly classifying face mask presence.

- Precision: Precision measures the proportion of positive predictions that are actually correct. With PCA, the precision is 88.37%, meaning that 88.37% of the times GaussianNB predicts a face mask, it is indeed present. With LDA, the precision increases to 97.67%, demonstrating a higher level of confidence in positive predictions.

- Recall: Recall measures the proportion of actual positive cases that are correctly identified as such. Using PCA, the recall is 90.48%, meaning that 90.48% of the actual face mask instances are correctly detected. With LDA, the recall reaches 100%, indicating an enhanced ability to identify true face mask cases.

- F1-score: The F1-score is a balanced measure of precision and recall, combining both metrics into a single value. For PCA, the F1-score is 89.41%, while for LDA, it soars to 98.82%, highlighting the overall improvement in performance.

In conclusion, the GaussianNB classifier demonstrates a remarkable boost in performance when utilizing LDA feature reduction for face mask detection. The substantial gains in accuracy, precision, recall, and F1-score, along with the perfect recall achieved with LDA, emphasize the effectiveness of LDA in preserving discriminative information for this task.

Random Forest	Accuracy	Precision	Recall	F1 Score
PCA	0.88073	0.87368	0.98809	0.92737
LDA	0.98165	0.98809	0.98809	0.98809

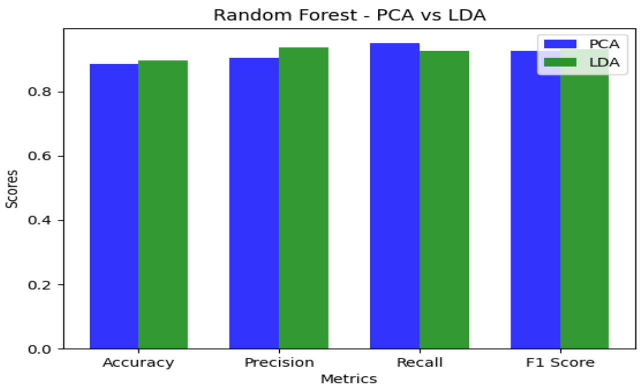

Figure 5.Random Forest -PCA vs LDA

The Random Forest classifier exhibits a notable improvement in performance when using LDA feature reduction compared to PCA. The accuracy, precision, and F1-score all increase, while the recall remains at a high level, suggesting that LDA is a more effective dimensionality reduction technique for face mask detection using the Random Forest classifier.

- Accuracy: The accuracy increases from 88.07% with PCA to 98.17% with LDA, indicating a significant improvement in correctly classifying face mask presence.

- Precision: Precision measures the proportion of positive predictions that are actually correct. With PCA, the precision is 87.37%, meaning that 87.37% of the times the Random Forest predicts a face mask, it is indeed present. With LDA, the precision increases marginally to 98.81%, demonstrating a slightly higher level of confidence in positive predictions.

- Recall: Recall measures the proportion of actual positive cases that are correctly identified as such. Using PCA, the recall is an impressive 98.81%, meaning that a very high proportion of the actual face mask instances are correctly detected. With LDA, the recall remains at 98.81%, indicating that the Random Forest classifier maintains its high ability to identify true face mask cases, even with dimensionality reduction.

- F1-score: The F1-score is a balanced measure of precision and recall, combining both metrics into a single value. For PCA, the F1-score is 92.74%, while for LDA, it increases to 98.81%, highlighting the overall improvement in performance.

While the improvement in accuracy and precision is substantial, the most remarkable aspect is the maintenance of a high recall rate with LDA feature reduction. This suggests that the Random Forest classifier can effectively identify face masks with both reduced dimensionality and high accuracy.

In conclusion, the Random Forest classifier demonstrates a notable boost in performance when utilizing LDA feature reduction for face mask detection. The significant gains in accuracy, precision, and F1-score, along with the maintained high recall, emphasize the effectiveness of LDA in preserving discriminative information for this task.

8. Conclusion

The experimental results indicate a notable variance in classifier performance based on the choice of dimensionality reduction technique. KNN, when applied with LDA, showed remarkable performance with an accuracy of 97.2%, precision of 97.65%, recall of 98.81%, and an F1 score of 98.22%. Similarly, SVM with LDA achieved an accuracy of 98.17%, precision of 97.65%, recall of 100%, and an F1 score of 98.82%. The Decision Tree classifier also displayed enhanced performance with LDA, achieving a 98.17% accuracy and an F1 score of 98.81%. GaussianNB and Random Forest classifiers exhibited comparable improvements when paired with LDA, both achieving an accuracy of 98.17% and F1 scores above 98.82%, 98.81%.

In contrast, the application of PCA generally resulted in lower performance metrics across all classifiers, with the highest accuracy being 88.1% (Random Forest) and the lowest 73.4% (Decision Tree). This disparity in results underscores the effectiveness of LDA over PCA in this specific context of facial landmark detection. The findings suggest that LDA, owing to its focus on maximizing class separability, is more suited for tasks involving clear binary classifications, such as mask detection in facial images.

The study highlights the importance of selecting appropriate dimensionality reduction techniques in conjunction with specific classifiers to optimize performance in machine learning tasks. The superior performance of LDA in this study sets a precedent for its application in similar binary classification problems, particularly in the realm of image processing and facial recognition.

9. References

[1] Wang, L., Wong, W.-K., & Luo, W. (2021). A comprehensive review on machine learning for face mask detection in the COVID-19 pandemic. IEEE Access, 9, 159283-159327.

[2] Pan, X., Xu, R., Huang, D., Hu, J., & Chen, X. (2020). A comparative study of machine learning for face mask detection based on transfer learning and feature extraction. IEEE Access, 8, 133277-133294.

[3] Liu, S., Deng, Y., &Cai, W. (2020). Research on face mask detection based on support vector machine and principal component analysis. Journal of Physics: Conference Series, 1452(1), 012011.

[4] Jolliffe, Ian T. Principal component analysis for special types of data. Springer New York, 2002.

[5] Abdi, Herve, Lynne J. Williams, and Domininique Valentin. "Multiple factor analysis: principal component analysis for multitable and multiblock data sets." Wiley Interdisciplinary reviews: computational statistics 5.2 (2013): 149-179.

[6] Hastie, Trevor, et al. The elements of statistical learning: data mining, inference, and prediction. Vol. 2. New York: springer, 2009.

[7] Vaddi, Snehit, and V. Vipul Maddi. "Comparative Analysis of Machine Learning Algorithms for Face Mask Detection and Alerting System." International Research Journal of Engineering and Technology (2021).

[8] Kumar Shukla, R., & Kumar Tiwari, A. (2021). Comparative analysis of machine learning based approaches for face detection and recognition. Journal of Information Technology Management, 13(1), 1-21.

[9] https://www.kaggle.com/datasets/andrewmvd/face-mask-detection