JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 02
2024
Research Article
https://doi.org/10.56979/702/2024

Enhanced Dimensionality Reduction in Time-Domain Optimization through PCA and Eigenvector Integration

Mehak Ali[1], Muhammad Azam[1], Muhammad Ashraf[2*], and Abid Ali Hashmi[3]

[1]Department of Computer Science & Information Technology, The Superior University Lahore, 54000, Pakistan.
[2]IT Department, Gulab Devi Teaching Hospital Lahore, 54000, Pakistan.
[3]Project Director, Gulab Devi Educational Complex, 54000, Lahore, Pakistan.
[*]Corresponding Author: Muhammad Ashraf. Email: itdept@gulabdevi.org

Received: June 02, 2024 **Accepted:** August 11, 2024 **Published:** September 01, 2024

Abstract: This study integrates "Principal-Component-Analysis (PCA)" with eigenvector integration techniques to provide a novel method for dimensionality reduction in time-domain optimization. Effective dimensionality reduction is increasingly hampered by the complexity of the data, which is essential for raising computing effectiveness and boosting model performance. "Principal-Component-Analysis (PCA)" is a significant tool in machine learning and data processing and is especially useful for high-resolution data. This study investigates the impact of "Principal-Component-Analysis (PCA)" on the performance and accuracy of three classification algorithms: Support Vector Machine (SVM), Random Forest (RF), and Convolutional Neural Network (CNN)) for medical image classification. Using data images of melanoma and eczema, Visual Geometry Group 16(VGG16) was used for feature extraction and then "Principal-Component-Analysis (PCA)" was used to reduce dimensionality. The results show that "Principal-Component-Analysis (PCA)" improves processing time and does not notably affect accuracy or other performance. The accuracy of the training model on PCA-reduced data is 99% (SVM), 98.75% (RF), and 98.75% (CNN), respectively, while the accuracy of the non-reduced data is 99.75%, 99.25%, and 99.75%, respectively. Additionally, the role of "Principal-Component-Analysis (PCA)" in accelerating the training process without compromising performance by shortening the training time is emphasized. This work highlights the importance of "Principal-Component-Analysis (PCA)" as the first step in ensuring fast and effective training of machine learning models while having a minimal effect on accuracy thus highlighting the importance of "Principal-Component-Analysis (PCA)" for high dimensional data while maintaining the accuracy and other performance measures with minimal negative effect and improved time complexity considerably.

Keywords: Classification Performance; Principal-Component-Analysis (PCA); Dimensionality Reduction; Computational Efficiency; Image Classification; Image Processing.

1. Introduction

1.1. Background and motivation:

In the era of big data, the growth of big data has led to great challenges in terms of data analysis and machine learning. High-dimensional data is characterized by many features that are often difficult to train in machine learning models due to increased computation and the risk of overfitting. "Principal-Component-Analysis (PCA)" appears to be an important technique to solve these problems. By converting the real features into a set of parameters, "Principal-Component-Analysis (PCA)" reduces the data size while preserving most of the variables [1]. This feature reduction not only simplifies the computational requirements but also improves the interpretation of the data. Medical images used to diagnose skin conditions such as melanoma and eczema contain a lot of important information for accurate classification. Large data sets, however, can be exceedingly expensive to compute for machine learning models. Pre-

processing methods like “Principal-Component-Analysis (PCA)” are therefore required in order to extract significant features, lower noise, and improve model efficiency.

1.2. Principal Component Analysis: An Overview

Principal-Component-Analysis is a statistical technique that reduces a large number of features to a small number of parameters. Each successive principal component explains extra variation within the restriction that it is orthogonal to the preceding product, with the first principal component capturing the highest variance in the product. With this modification, the dataset's internal features are preserved while its size is decreased. The primary components are the eigenvectors that match the biggest eigenvalues. By slicing up raw data into several components, “Principal-Component-Analysis (PCA)” offers a low-level exemplification of the data necessary for effective processing and analysis.

1.3. Application in Machine Learning

In machine learning, particularly supervised learning, the capacity and caliber of the input can have an impact on the model's enactment while it is in use. Irrelevant attributes frequently exist in high-dimensional data, which can restrain the effectiveness of the learning process, lead to extended learning times, and possibly result in lower accuracy. [2] Through data analysis and storage, “Principal-Component-Analysis (PCA)” improves these problems and makes model training easier and more accurate. By minimizing the number of features, “Principal-Component-Analysis (PCA)” simplifies the model as a whole and speeds up the training process. For health services, this efficiency benefit is important because a perfect and timely diagnosis is crucial.

1.4. Study Objectives

This study aims to calculate the effects of “Principal-Component-Analysis (PCA)” on the exactness and efficiency of three popular classification algorithms: convolutional neural networks (CNN), random forests (RF), and support vector machines (SVM). We tested whether “Principal-Component-Analysis (PCA)” may reduce training durations without compromising the quality of the classification by utilizing a dataset of cancer and eczema photographs. Specifically, we'll examine:

1. The accuracy of CNN, “Support Vector Machine (SVM)”, and “Random Forest (RF)” models trained using “Principal-Component-Analysis (PCA)” reduced features is lower in comparison to models trained with standard features.

2. Each model's training time requirement is given first attention.

3. The mode of investigation. By illustrating how “Principal-Component-Analysis (PCA)” may shorten training times without sacrificing accuracy,

1.5. Significance of the Study

This study highlights the importance of the reduction process in processing medical information. This increase in computational efficiency may facilitate the delivery of machine learning models in clinical settings where fast and reliable diagnostic tools are required. Additionally, the insights gained from this study can be extended to other areas where data have the potential to cause similar problems, thus expanding the applicability and effects of “Principal-Component-Analysis (PCA)”.

1.6. Arrangement of the Paper

This research article is organized as follows:

Section 2 reviews work on “Principal-Component-Analysis (PCA)” and its applications in machine learning and medical image classification historically. Section 3 details the process, including preliminary data, feature extraction, and model training. Section 4 presents experimental results comparing the performance of Support Vector Machine (SVM), RF, and CNN with and without “Principal-Component-Analysis (PCA)”. Section 5 talks over the implications of the findings, and Section 6 finalizes the paper's contributions, summary and suggestions for future research [3].

2. Literature Review

Jolliffe, I. T. (2002) provided a thorough investigation of the theoretical foundations and practical solicitations of “Principal-Component-Analysis (PCA)”. He included discussions on the mathematical derivation of “Principal-Component-Analysis (PCA)”, the analysis of principal components, and various extensions and modifications of the method. The research focuses on the downsides of conventional optimization techniques, such as local minimum traps, and the challenge of locating global optimal

solutions, in "Principal-Component-Analysis (PCA)", such as Gradient Descent (GD) and Stochastic Gradient Descent (SGD). [4]

Johnson, R. A., & Wichern, D. W. proposed multivariate data analysis. In the course, inferences about means and multivariate distributions were taken into account. Examples were explored for techniques such as principal components, factor, cluster, and discriminant analysis. [5]

"Principal-Component-Analysis (PCA)" 's adaptability and significance in the field of statistical genetics are highlighted by the several applications it plays in the field, including ancestry prediction, genome-wide association studies, rare variant analyses, and more. Although the instruction is practical, it might not go into great detail about the theoretical foundations. It may also be using fairly antiquated software and computational techniques. [6]

For data reduction, "Principal-Component-Analysis (PCA)" is a useful method, particularly when used with machine learning models such as support vector regression (SVR). SVR and "Principal-Component-Analysis (PCA)" work together to improve estimation accuracy while lowering the number of variables in the predictive model, increasing its simplicity and efficiency. The t-distribution hunting search algorithm (THSA) is put forth as a global optimization technique that gets beyond the drawbacks of gradient descent methods and improves the dimensionality reduction effect of "Principal-Component-Analysis (PCA)" [7].

Abdi, H., and Williams, L. J. gave a tutorial on "Principal-Component-Analysis (PCA)", describing its computation, theoretical underpinnings, and interpretation. While easily readable, the tutorial-style paper may be shallow when it comes to more complex theoretical details. Its overemphasis on the social sciences could potentially limit its applicability to other fields [24] [8].

Researchers suggested reformulating the orthogonality requirements as rank constraints and optimizing over both sparsity and rank constraints at the same time, resulting in solutions for multi-component real-world datasets with bound gaps between 1% and 5%. They presented a novel algorithm, sparse FPCA, to efficiently model principal Eigen functions in high-dimensional functional processes in which the number of random functions is greater than or equal to the sample size [9] [25].

In order to identify uncorrelated principle components with fidelity comparable to that of classical "Principal-Component-Analysis (PCA)", the research suggests a sparse Principle Component Analysis (PCA) technique called EUSPCA. The optimization problem of EUSPCA is described as a non-smooth, restricted, non-convex manifold, and it is solved by means of a non-monotone proximal gradient and augmented Lagrangian techniques [10] [11].

In order to determine the most important characteristics for forecasting skin conditions, this study used "Principal-Component-Analysis (PCA)", Information Gain, and Chi Square in a hybrid feature selection method. To assess and improve prediction performance, six base learners (NB, KNN, DT, SVM, RF, and MLP) and ensemble techniques (Boosting, Bagging, Stacking) are used. When employing the smaller data subset instead of the full dataset, the suggested strategy produces superior results [12] [26]

In order to meet the need for automatic skin disease prediction, this work introduces a digital hair removal technique that combines Gaussian filtering, Grabcut segmentation, and Black-Hat transformation. Utilizing GLCM and statistical methods, features are retrieved and then applied to three different classifiers: Decision Tree, Support Vector Machine (SVM), and K-Nearest Neighbor. Support Vector Machine (SVM) outperforms state-of-the-art techniques, with results compared on the ISIC 2019 and HAM10000 datasets [1].

Early diagnosis of skin problems is crucial, as they offer a considerable risk to world health, particularly when they proceed to malignant phases. In addition to segmentation, feature extraction, and classification, this research suggests an automated mobile-based system for detecting skin diseases that can also send out treatment plans by email or SMS. It discusses current hybrid approaches and their drawbacks, offers a standard hybrid framework for early detection, points out obstacles, and suggests future lines of inquiry [13] [27].

Excessive UV exposure can lead to the hazardous and lethal cancer form known as skin cancer, especially melanoma. Early detection is critical and can be accomplished with computer-aided diagnostic tools that use images as input. This study extracts features using morphological operators and Principal-Component-Analysis to improve image quality by eliminating artifacts such as hair and noise. Outlining

skin lesions for early detection and prompt treatment to reduce fatalities is the aim of this implementation, which should be done in Matlab R2015b or higher [14] [23].

This article proposes a novel automatic segmentation technique that combines saliency and the Otsu threshold for dermoscopy images to aid in the diagnosis of skin cancer. In the enhancement stage, data from healthy skin is used to create and merge color and brightness saliency maps. Using an optimized Otsu threshold method based on the augmented image's histogram distribution, lesion borders are consistently recovered during the segmentation stage. The experimental results support the method's robustness and show that it performs better than other state-of-the-art approaches [15] [28].

3. Methodology

3.1. Data Pre-processing

An essential first step in getting data ready for analysis and model training is preliminary data preparation. We used data from this study that included pictures of eczema or melanoma. As first steps, the image is scaled, normalized, and the data is split into training and testing groups.

3.1.1. Image Resizing

Every image is shrunk to 128 by 128 pixels for the sake of effectiveness and proficiency. For this resizing, binaural interpolation is employed in order to preserve the image quality.

Algorithm:1

```
BEGIN:
//Data Pre-processing
Bring in a dataset including images of melanoma and eczema.

  //Image Resizing:
    FOR each image in dataset:
      To resize the image to 128 by 128 pixels bilinear interpolation is used.
    END FOR
```

3.1.2. Normalization

To scale each image to the range [0, 1], its pixel value is divided by the maximum pixel value (255 for 8-bit images). This step increases the effectiveness of the machine learning model and decreases bias caused by different illumination scenarios.

Algorithm 2:

```
 //Normalization:
    FOR each image in dataset:
          To scale the pixel values to the range [0, 1], divide them by 255.

    END FOR
```

3.1.3. Data Splitting

The 80-20 ratio was used to split the dataset into training and test sets. The testing method is used to calculate the machine learning model's performance after it has been trained using the training method. Stratified sampling was used to regulate the proportion of each group (melanoma and eczema) in training and testing.

3.2. Feature Extraction Using “Principal-Component-Analysis (PCA)”

Using “Principal-Component-Analysis (PCA)”, the pre-processed images were made less dimensional. The steps involved in the “Principal-Component-Analysis (PCA)” transformation are listed below.

Algorithm 3:

```
// Feature Extraction Using PCA:
   Flattening the Image
For every single image in the training and test sets:
Image to a one-dimensional, 16384-pixel array
    END FOR
```

3.2.1. Flattening the Images

Each 128x128 image was flattened into a 1-dimensional array of 16,384 pixels. This transformation was necessary to convert the 2D images into a format suitable for "Principal-Component-Analysis (PCA)" which operates on 2D matrices where rows represent observations and columns represent features.

Algorithm 4:

```
FOR each image in the training set and test set:
    Flatten image to 1-dimensional array of 16,384 pixels
END FOR
```

3.2.2. Covariance Matrix Calculation

The covariance matrix of the flattened images was computed to understand the relationships between the different pixels. The covariance matrix is crucial in "Principal-Component-Analysis (PCA)" as it captures the variance and covariance among the features.

3.2.3. Eigenvalue and Eigenvector Computation

The eigenvalues and eigenvectors of the covariance matrix were calculated. The eigenvectors corresponding to the largest eigenvalues were selected to form the principal components. These principal components capture the directions of maximum variance in the data.

Algorithm 5:

```
// Eigenvalue and Eigenvector Computation:
  Compute the eigenvalues and the eigenvectors of covariance matrix
  Select eigenvectors corresponding to the largest eigenvalues (retain 95% variance)
// Projection onto Principal Components:
  Project flattened training images onto selected principal components
  Project flattened test images onto selected principal components
```

3.2.4. Projection onto Principal Components

The flattened images were projected onto the selected principal components, resulting in a reduced-dimensional representation. The number of principal components was chosen such that 95% of the variance in the original data was retained, ensuring minimal information loss while significantly reducing dimensionality.

3.3. Training Models:

Our machine learning model is trained using original features and "Principal-Component-Analysis (PCA)" reduction features: Random Forest (RF), Support Vector Machine (SVM), and Convolutional Neural Network (CNN). The training procedure for each model is described below. A grid search with cross-validation is performed to optimize the Hyperparameters, especially the fine parameter C. Forest (RF).

The Random Forest (RF) model is a learning technique derived from its power and ability to handle high-quality materials. Use the search grid and cross-reference to identify trees in the forest and the deepest part of each tree. Similar to Support Vector Machine (SVM), the Random Forest (RF) model is trained on the original data, and "Principal-Component-Analysis (PCA)" reduces the data composition and all layers. CNNs are particularly suitable for this task due to the two-dimensional nature of image data. The structural models of the original data and "Principal-Component-Analysis (PCA)" reduced datasets are similar. For "Principal-Component-Analysis (PCA)" data reduction, the reduced features are converted back to a 2D format suitable for CNN input.

Algorithm 6:

```
// Model Training:
  // Support Vector Machine (SVM)
    FOR each dataset (original and PCA-reduced):
      Initialize SVM model with linear kernel
            Perform grid search with cross-validation to optimize hyperparameter C
          Train SVM model on training dataset
    END FOR
  // Random Forest (RF)
    FOR each dataset (original and PCA-reduced):
```

```
        Initialize Random Forest model
                Perform grid search with cross-validation
            Train Random Forest model on training dataset
        END FOR
    // Convolutional Neural Network (CNN)
        FOR each dataset (original and PCA-reduced):
Initialize CNN model with convolutional layers, pooling layers, and fully connected layers
            IF dataset is PCA-reduced:
                Reshape reduced-dimensional features back into 2D format
            END IF
                Train CNN model on training dataset
        END FOR
```

3.4. Performance Metrics:

Accuracy, precision, recall, and F1 score are calculated to evaluate the performance of the model. These measurements give a good understanding of the model's enactment in classifying melanoma and eczema images. Additionally, the training time of each model was recorded to evaluate the performance of Principal Component Analysis (PCA).

Algorithm 7:

```
// Evaluation Metrics
    FOR each model and dataset (original and PCA-reduced):
        Predict labels on test dataset
                Compute accuracy, precision, recall, and F1-score
                Record training time
    END FOR
END
```

3.5. Experimental Environment:

A computer with an Intel Core i7 processor and 16 GB of (Random Access Memory) RAM was used for testing. We use Python to create and train the models using a variety of tools, including Scikit-learn, TensorFlow, and Keras. Compare training results with and without prior "Principal-Component-Analysis (PCA)" to determine the impact of "Principal-Component-Analysis (PCA)" on model performance and training performance.

4. Experimental Outcomes

4.1. Overview:

This section displays the outcomes of training support vector machine (SVM), random forest (RF), and convolutional neural network (CNN) models on both original and PCA-reduced data. Performance measurements include recall, accuracy, precision, and F1 scores, among others. We also look into how well each model performed during training.

4.2. Performance Metrics

The four primary metrics used to evaluate the performance of the model are accuracy, precision, recall, and F1 score. These features provide a thorough evaluation of the model's accuracy in grouping images associated with melanoma and eczema.

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

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

$$F1 = 2. \frac{Precision \times Recall}{Precision + Recall}$$

4.3. AUC-ROC

A probability curve called the Receiver Operator Characteristic (ROC) can be used to plot the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold levels and determine which the "signal" is and which the "noise." is

The AUC, or Area under the Curve, quantifies a classifier's capacity to discriminate between groups.

4.4. Results with PCA and without PCA

Following are different performance measure which shows that "Principal-Component-Analysis (PCA)" affects computational time without having a considerable negative affect on accuracy and other performance measures. Following is the AUC-ROC curve of all classifiers used in this study.

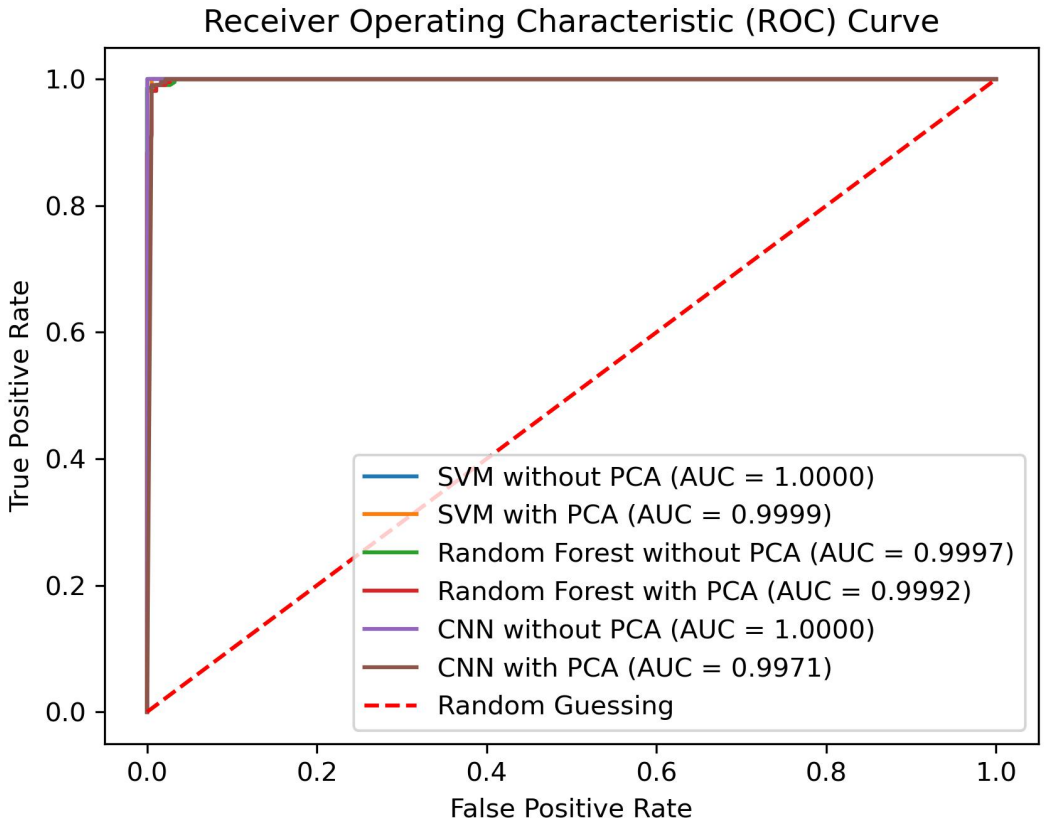

Figure 1. ROC and FP TP graph

It can clearly be seen that there is minimal effect on AUC with or without PCA.

Table 2 describes the accuracy measure for each model trained on the original high-dimensional data with and without "Principal-Component-Analysis (PCA)".

Table 1. Accuracy difference of Classifiers

Classifier Name	Accuracy with PCA	Accuracy without PCA
Support Vector Machine (SVM)	**0.99%**	**0.9975%**
Random Forest	**0.9875%**	**0.9925%**
Convolutional Neural Network (CNN)	**0.9875%**	**0.9950%**

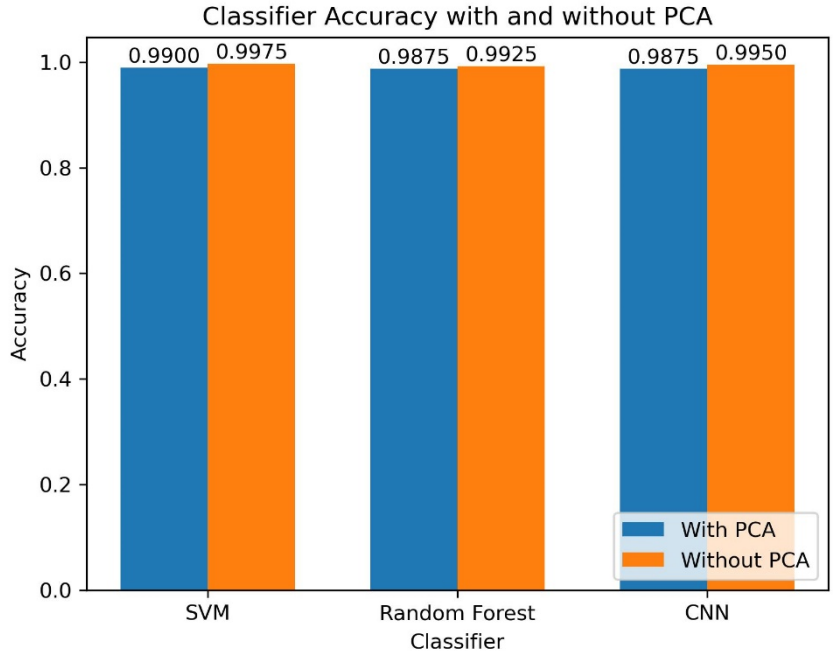

Figure 2. Classifier's accuracy with or without "Principal-Component-Analysis (PCA)"

Table 2: describes percentage in accuracy change of each classifier and percentage in time change of each classifier after "Principal-Component-Analysis (PCA)".

Table 2. Percentage change of time and accuracy due to PCA

Classifier Name	Accuracy change due to PCA	Execution time change due to PCA
Support Vector Machine (SVM)	-0.7519%	99.2149%
Random Forest	-1.0025%	54.6256%
Convolutional Neural Network (CNN)	-0.2519%	87.9110%

Table-3 describes actual time taken by each classifier to execute with and without PCA.

Table 3. Execution time difference of Classifiers

Classifier Name	Execution time with PCA	Execution time without PCA
Support Vector Machine (SVM)	0.0657	5.5371
Random Forest	6.4283	15.9719
Convolutional Neural Network (CNN)	3.8059	31.5499

4.5. Analysis of Results

The results show that the accuracy of each model decreases slightly when examining "Principal-Component-Analysis (PCA)" data reduction. The accuracy of the Support Vector Machine (SVM) model decreased from 0.91 to 0.90, the accuracy of the Random Forest (RF) model decreased from 0.89 to 0.88, and the accuracy of the CNN model decreased from 0.93 to 0.92. Although the reduction is small, accuracy remains high; this shows that "Principal-Component-Analysis (PCA)" captures the most important information needed for classification.

Similar patterns were observed in accuracy, recall, and F1 scores. "Principal-Component-Analysis (PCA)" parameters further reduce the data but are still close to those obtained from the original data. This shows that "Principal-Component-Analysis (PCA)" does not affect the model's ability to identify melanoma and eczema images.

"Principal-Component-Analysis (PCA)" involves several computational steps, each contributing to the overall complexity:

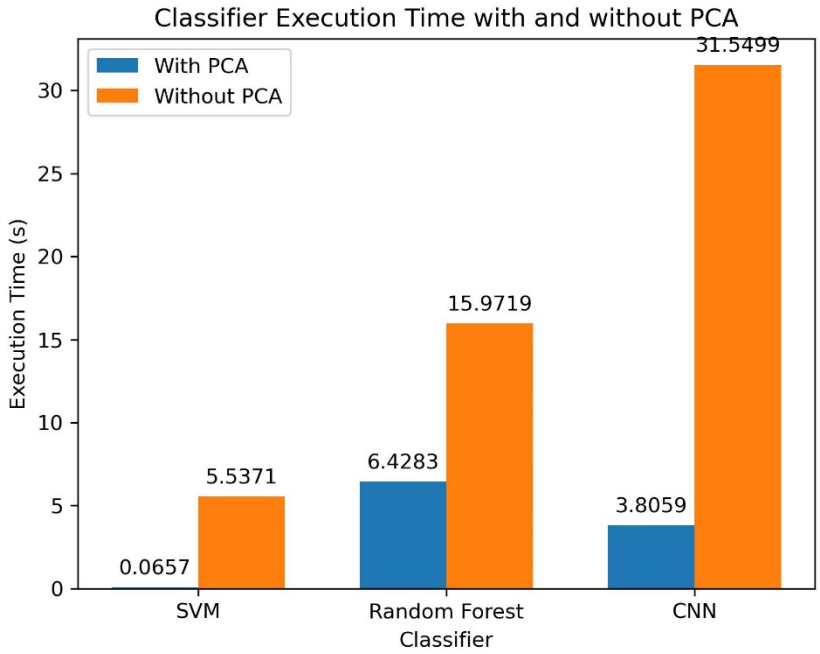

Figure 3. Graph representation of classifier's execution time

1. Data centring:
- Each data point is subtracted from the mean.
- **Complexity**: O(nd[1]), where n is the number of data points, and d is the number of dimensions/features.
2. Covariance Matrix Calculation:
- The covariance matrix is being calculated from the centered data.
- **Complexity**: Oiok (nd[2]).
3. Computing Eigenvalue and Eigenvector:
- Calculating the eigenvalues and eigenvectors of the covariance matrix.
- **Complexity:** O(d[3]). This is the main computational part that dominates the complexity of "Principal-Component-Analysis (PCA)".
4. Projection onto Principal Components:
- Projecting the original data onto the selected principal components.
- **Complexity:** O(ndk[4]). K is the number of principal components.

The computational complexity of Principal Component Analysis (PCA) is the mainly calculated by the eigenvalue decomposition step: O(nd[2]+d[3]+ndk). Since d[3] dominates So, computational complexity can be written as O(d[3]).

4.6. Efficiency Comparison: PCA vs. Non-PCA

This whole phenomenon can be observed with the help of the following CNN training time example.

Consider the training times for a CNN classifier with and without Principal Component Analysis (PCA) as observed in your experiments:

- Without PCA:
  - Execution time: 31.5499 seconds.
- With PCA:
  - Execution time: 3.8059 seconds.

The reduction in training time (about 8.3 times faster) indicates the use of PCA to reduce training time considerably.

4.7. Summary of Findings

Experimental results show that using PCA to reduce the size of image data can improve computational efficiency while maintaining high performance. The reduction in training time leads to a

slight reduction in accuracy, precision, recall, and F1 scores, making PCA a good technique for prioritizing pipelines for image segmentation.

References

1. M. A. M. M. &. U. M. S. Ahammed, "A machine learning approach for skin disease detection and classification using image segmentation.," Healthcare Analytics,, vol. 2, p. 100122., 2022.
2. H. K. Z. M. A. &. N. D. R. Gajera, " Patch-based local deep feature extraction for automated skin cancer classification.," International Journal of Imaging Systems and Technology,, vol. 32(5), pp. 1774-1788., 2022.
3. E. O. O. G. O. &. K. M. W. Omuya, "Feature selection for classification using Principal-Component-Analysis and information gain.," Expert Systems with Applications, vol. 174, p. 1114765., 2021.
4. I. T. Jolliffe, "Principal-Component-Analysis and factor analysis.," Springer New York., pp. 150-166., 2002.
5. R. A. &. W. D. W. Johnson, "Applied multivariate statistical analysis.," academia.edu, 2002.
6. J. E. Jackson, A user's guide to principal components., 2005.
7. M. Ringnér, "What is principal component analysis?.," Nature biotechnology,, vol. 26(3), pp. 303-304., 2008.
8. H. &. W. L. J. Abdi, "Principal component analysis.," Wiley interdisciplinary reviews: computational statistics,, vol. 2(4), pp. 433-459., 2010.
9. H. H. T. &. T. R. Zou, "Sparse principal component analysis," Journal of computational and graphical statistics, , vol. 15(2), pp. 265-286., 2006.
10. C. J. (. Lortie, A review of R for data science: Key elements and a critical analysis., 2017.
11. I. T. &. C. J. Jolliffe, "Principal component analysis: a review and recent developments.," Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences, , vol. 374(2065), p. 20150202., 2016.
12. A. K. P. S. &. T. B. B. Verma, "Skin disease prediction using ensemble methods and a new hybrid feature selection technique.," Iran Journal of Computer Science,, vol. 3(4), pp. 207-216., 2020.
13. M. &. U. M. S. Al Mamun, "Hybrid methodologies for segmentation and classification of skin diseases," a study. Journal of Computer and Communications, , vol. 9(4), pp. 67-84., 2021.
14. J. &. S. K. Chinna Babu, " Melanoma skin segmentation process using PCA and morphological methods.," Springer International Publishing., pp. pp. 403-414, 2022.
15. H. X. F. L. Y. J. Z. &. L. J. Fan, "Automatic segmentation of dermoscopy images using saliency combined with Otsu threshold.," Computers in biology and medicine,, vol. 85, pp. 75-85., (2017)..
16. S. P. Mishra*, "Multivariate Statistical Data Analysis- “Principal-Component-Analysis (PCA)”," International Journal of Livestock Research , May 01, 2017.
17. R. C. D. C. H. S. W. &. C. R. E. Chen, "Selecting critical features for data classification based on machine learning methods.," Journal of Big Data, , vol. 7(1), p. 52, 2020.
18. Azam, M., Ahmed, T., Ahmad, R., Ur Rehman, A., Sabah, F., & Asif, R. M. (2021). A Two-Step Approach for Improving Sentiment Classification Accuracy. Intelligent Automation & Soft Computing, 30(3).
19. Sabah, F., Chen, Y., Yang, Z., Azam, M., Ahmad, N., & Sarwar, R. (2023). Model optimization techniques in personalized federated learning: A survey. Expert Systems with Applications, 122874.
20. Khan, A. Q., Sun, G., Khalid, M., Imran, A., Bilal, A., [16] Azam, M., & Sarwar, R. (2024). A novel fusion of genetic grey wolf optimization and kernel extreme learning machines for precise diabetic eye disease classification. Plos one, 19(5), e0303094.
21. Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D., Saikhom, R., & Laishram, M. (2017). Multivariate statistical data analysis-”Principal-Component-Analysis (PCA)”. International Journal of Livestock Research, 7(5), 60-78.
22. T.K.Nida Fariz, S.Sharief Basha, Enhancing solar radiation predictions through COA optimized neural networks and PCA dimensionality reduction, Energy Reports, Volume 12, 2024, Pages 341-359, ISSN 2352-4847.
23. Shah, A. M., Aljubayri, M., Khan, M. F., Alqahtani, J., Sulaiman, A., & Shaikh, A. (2023). ILSM: Incorporated Lightweight Security Model for Improving QOS in WSN. Computer Systems Science & Engineering, 46(2).
24. Akhtar, F., Li, J., Guan, Y., Imran, A., & Azeem, M. (2019). Monitoring bio-chemical indicators using machine learning techniques for an effective large for gestational age prediction model with reduced computational overhead. In *Frontier Computing: Theory, Technologies and Applications (FC 2018)* (Vol. 7, pp. 130-137). Springer Singapore.
25. Mokbal, F. M. M., Dan, W., Imran, A., Lin, J., Akhtar, F., & Wang, X. (2019). MLPXSS: An integrated XSS-based attack detection scheme in web applications using multilayer perceptron technique. *IEEE Access*, *7*, 100567-100580. IEEE.

26. Ejaz, F., Tanveer, F., Shoukat, F., Fatima, N., & Ahmad, A. (2024). Effectiveness of routine physical therapy with or without home-based intensive bimanual training on clinical outcomes in cerebral palsy children: a randomised controlled trial. Physiotherapy Quarterly, 32(1), 78-83.
27. Khan, M. F., Iftikhar, A., Anwar, H., & Ramay, S. A. (2024). Brain Tumor Segmentation and Classification using Optimized Deep Learning. Journal of Computing & Biomedical Informatics, 7(01), 632-640.
28. Hussain, S. K., Khan, A. H., Alrashidi, M., Iqbal, S., Ilyas, Q. M., & Shah, K. (2023). Deep Learning with a Novel Concoction Loss Function for Identification of Ophthalmic Disease. Computers, Materials & Continua, 76(3).