A Comparative Analysis of K-Nearest Neighbours & Support Vector Machine for Classification of Iris African Dataset

Falola Peace[1]
[1]Department of Computer Sciences, Precious Cornerstone University, Ibadan, Nigeria.
Adeniyi Emmanuel[1]
[1]Department of Computer Sciences, Precious Cornerstone University, Ibadan, Nigeria.
Adebiyi Marion[2]
[2]Department of Computer Science, Landmark University, Omu-Aran, Nigeria.
Adebiyi Ariyo[2]
[2]Department of Computer Science, Landmark University, Omu-Aran, Nigeria.
Awotunde Joseph[3]
[3]Department of Computer Science, University of Ilorin, Ilorin, Nigeria
Jimoh Rasheed[3]
[3]Department of Computer Science, University of Ilorin, Ilorin, Nigeria
Ogunbona Emmanuel[2]
[2]Department of Computer Science, Landmark University, Omu-Aran, Nigeria

Abstract

Early eye illness detection is a major issue; early eye disease identification is crucial to preventing future complications. Early identification is crucial in several vision-losing disorders such as cataracts, diabetic retinopathy, and diabetes mellitus cataract, which cause blindness in working people at younger ages. This study aims to develop an eye disease detection model. The model was created by using an African iris dataset from Kaggle, PCA, KNN, and SVM (Support Vector Machine). The results decided which algorithms classified myopia or hyperopia best. Evaluation metrics were used to evaluate the performance implementation. The SVM algorithm outperformed the other algorithms, achieving a classification testing accuracy with PCA of 71.6%. The study concluded that the proposed approach can be used to accurately classify eye diseases in African patients and highlights the importance of considering the specific population when developing models for classifying or detecting eye diseases.

Keywords: Machine learning, Classification algorithm, Feature extraction, Eyes, Disease detection.

1.0 Introduction

Early diagnosis of eye disorders is a major cause of visual impairments and blindness worldwide. Only approximately half of diabetics get annual eye exams due to a dearth of ophthalmologists (Ravindran, 2019). Some studies have constraints due to limiting their dataset to certain years due to rapid field advances and the review of deep learning-based algorithms due to their performance on picture categorization. Vision impairment and eye illnesses affect 26.3 million Africans, including 20.4 million with impaired vision and 5.9 million who are blind. Africa has 15.3% of the world's blind people (Outbreak, n.d.). In Nigeria alone, 24.3 million individuals are blind, 5.3 million have moderate to severe blindness, 7.8 million have mild blindness, and 9.9 million are near blind (Map, 2020). This makes Nigeria the continent's leader in eye disease.

This study, therefore, used K-Nearest Neighbor and SVM to improve eye illness identification using an African dataset with the following contributions:

i. The study utilized PCA to reduce the dimensionality of the dataset and improve the performance of the classification algorithms. By reducing the number of features, the model becomes more efficient and less prone to overfitting.

ii. The study compared the performance of two classification algorithms, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). This comparison helped identify the most effective algorithm for classifying eye diseases in the African population.

iii. The study evaluated the proposed model using a dataset of African patients, which is critical as many eye disease detection models have been developed using datasets from other regions. By using an African dataset, the paper ensures that the model is appropriate for the population it is intended to serve.

2.0 Methods

Below indicates the breakdown of this study’s methodology:

Step 1: Collected project-related data. The dataset gotten was from Kaggle

Step 2: Minimized picture dataset dimensionality with PCA preprocessing.

Step 3: Classified the dataset using K-Nearest Neighbor to simplify its performance.

Step 4: Used Support Vector Machine to classify and predict dataset performance.

Step 5: Compared PCA and non-PCA results i.e. comparing the findings from the data using Principal Component Analysis (PCA) utilizing Machine Learning Algorithms with those without preprocessing. Assessed the accuracy, specificity, sensitivity, precision, and computational time of both results.

3.0 Result and discussion

This study used data from Kaggle; here's the link:

https://www.kaggle.com/code/akshat0007/diabetic-retinopathy-detection-and-classification/notebook

3.1 Result analysis for SVM

Figure 1 shows the result gotten after training the SVM classifier.

```
Evaluating the Classifier. Please Wait...
                precision    recall  f1-score   support

         No_DR       0.90      0.96      0.93       361
          Mild       0.41      0.39      0.40        74
      Moderate       0.63      0.69      0.66       200
        Severe       0.25      0.13      0.17        39
Proliferate_DR       0.35      0.22      0.27        59

      accuracy                           0.73       733
     macro avg       0.51      0.48      0.49       733
  weighted avg       0.70      0.73      0.71       733

[[348   3   8   0   2]
 [  8  29  29   4   4]
 [ 20  26 138   5  11]
 [  5   6  16   5   7]
 [  4   7  29   6  13]]
acc: [0.93178718 0.88130969 0.80354707 0.93315143 0.90450205]
sensitivity: [0.96398892 0.39189189 0.69       0.12820513 0.22033898]
specificity: [0.90053763 0.93626707 0.84615385 0.97838617 0.96439169]
Matthews Correlation Coefficient: 0.5782
```

Figure 1 Results obtained after training the SVM

3.1.1 ROC curve for SVM

This is used for creating visualization with readable labels. Where *0* represents *'NO_DR'*, *1* represents *'Mild'*, *2* represents *'Moderate'*, *3* represents *'Severe'*, and *4* represents *'Proliferate_DR'.* Figure 2 shows the ROC curve for SVM.

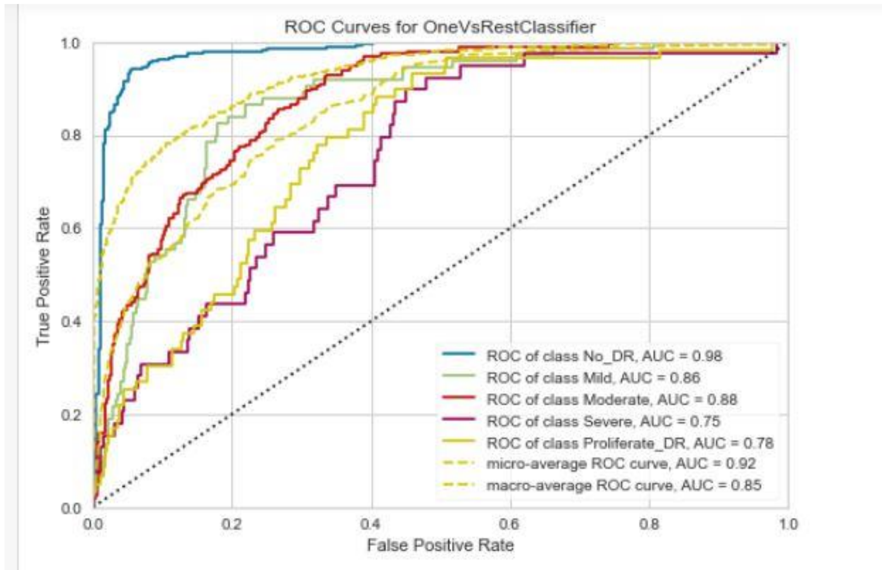

Figure 2 ROC curve for SVM

This shows the performance of the classification model at all classification thresholds. For example, ROC of class No_DR, AUC = 0.98, which means the SVM's accuracy in detecting images that do not have DR is 98%.

3.2 Result analysis for KNN

Figure 3 shows the results of the confusion matrix for KNN. Figure 5 shows the result gotten after training the KNN classifier.

Evaluating the Classifier. Please Wait...

	precision	recall	f1-score	support
No_DR	0.87	0.68	0.76	361
Mild	0.32	0.34	0.33	74
Moderate	0.48	0.65	0.55	200
Severe	0.16	0.13	0.14	39
Proliferate_DR	0.29	0.36	0.32	59
accuracy			0.58	733
macro avg	0.42	0.43	0.42	733
weighted avg	0.62	0.58	0.59	733

Figure 3 Result after evaluation of the KNN classifier with multiclass

3.2.1 ROC curve for KNN

This is used for creating visualization with readable labels. Where *0* represents *'NO_DR'*, *1* represents *'Mild'*, *2* represents *'Moderate'*, *3* represents *'Severe'*, and *4* represents *'Proliferate_DR'.* Figure 4 shows the ROC curve for KNN.

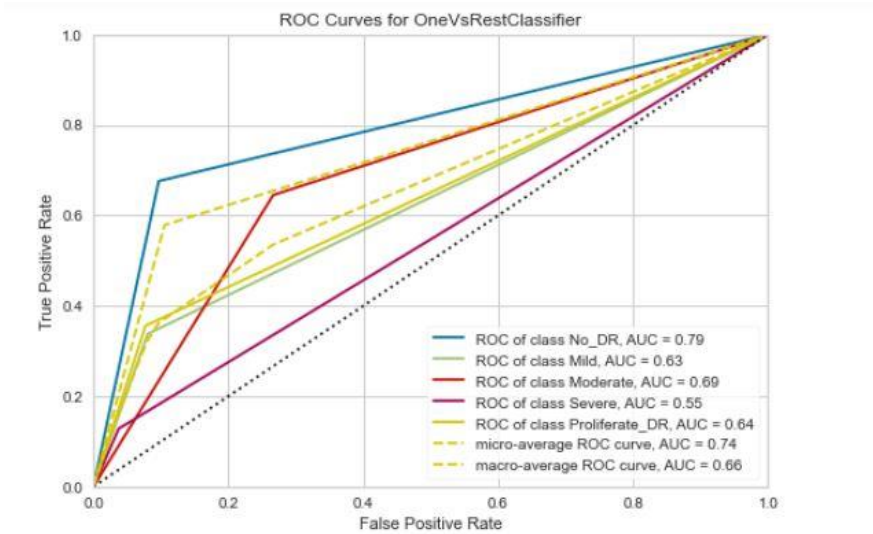

Figure 4 ROC curve for KNN

This shows the performance of the classification model at all classification thresholds. For example, ROC of class No_DR, AUC = 0.79, which means the KNN's accuracy in detecting images that do not have DR is 79%.

3.3 Comparison of models with PCA

This is used to compare the models' performance. Figure 5 shows the comparison of machine models with PCA learning.

	ML Model	Train Accuracy with PCA	Test Accuracy with PCA
0	Support Vector Machines	0.975	0.716
1	K-Nearest Neighbors	0.994	0.578

Figure 5 Comparison of Machine Learning Models of the Train and Test Accuracies with PCA

Hence, from the above comparison, it is clear that the **SVM** outperformed KNN for the classification of eye diseases using an African dataset.

4.0 Conclusion

The study limits the dataset to only that of Africans. This study developed a classification model for analyzing a small number of eye diseases; myopia and hyperopia. Future research of this study is to extend the dataset, employ new algorithms and hybridize them instead of employing PCA or other feature extraction methods.

References

Akcay, S., Atapour-Abarghouei, A., & Breckon, T. P. (2019). Skip-GANomaly: Skip Connected and Adversarially Trained Encoder-Decoder Anomaly Detection. *Proceedings of the International Joint Conference on Neural Networks*, *2019-July*(1), 894–896. https://doi.org/10.1109/IJCNN.2019.8851808

Gargeya, R., & Leng, T. (2017). Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*, *124*(7), 962–969. https://doi.org/10.1016/j.ophtha.2017.02.008

Gollapalli M., Alansari A., Alkhorasani H., Alsubaii M., Sakloua R., Alzahrani R., Al-Hariri M., Alfares M., AlKhafaji D., R Al Argan R., Albaker W. (2022). A novel stacking ensemble for detecting three types of diabetes mellitus using a Saudi Arabian dataset: Pre-diabetes, T1DM, and T2DM. *Computers in Biology and Medicine,* vol 147, https://doi.org/10.1016/j.compbiomed.2022.105757.

Hemanth, D. J., Deperlioglu, O., & Kose, U. (2020). An enhanced diabetic retinopathy detection and classification approach using deep convolutional neural network. *Neural Computing and Applications*, *32*(3), 707–721. https://doi.org/10.1007/s00521-018-03974-0

Map, C. (2020). Country Map & Estimates of Vision Loss The 10 countries with the highest number of persons with vision loss, pp 2–7.

Mesa F., Ospina-Ospina R., Devia-Narvaez D. (2023). Comparison of Support Vector Machines and Classification and Regression Tree Classifiers on the Iris Data Set. Journal of Southwest Jiaotong University, vol 58, no 2. DOI: https://doi.org/10.35741/issn.0258-2724.58.2.59

Nuzzi, R., Boscia, G., Marolo, P., & Ricardi, F. (2021). The Impact of Artificial Intelligence and Deep Learning in Eye Diseases: A Review. *Frontiers in Medicine*, *8*, 1–24. https://doi.org/10.3389/fmed.2021.710329

Outbreak, D. (n.d.). *Eye health*, pp 20–23.

Ravindran, S. (2019). How artificial intelligence is helping to prevent blindness. *Nature*, *April*, 1–11.https://doi.org/10.1038/d41586-019-01111-y

Sarki, R., Ahmed, K., Wang, H., & Zhang, Y. (2020). Automatic Detection of Diabetic Eye Disease through Deep Learning Using Fundus Images: A Survey. *IEEE Access*, *8*, 151133–151149. https://doi.org/10.1109/ACCESS.2020.3015258

WHO. (2021). Blindness and vision impairment: Refractive errors. *World Health Organization*, *11*(October 2013), 6–8. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment%0Ahttps://www.who.int/news-room/q-a-detail/blindness-and-vision-impairment-refractive-errors