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

Falola Peace[1], Adeniyi Emmanuel[2], Adebiyi Marion[3], Adebiyi Ariyo[3], Awotunde Joseph[4], Jimoh Rasheed[4], Ogunbona Emmanuel[3]

[1]Department of Computer Science, University of Ibadan, Ibadan, Nigeria
[2]Department of Computer Sciences, Precious Cornerstone University, Ibadan, Nigeria.
[3] Department of Computer Science, Landmark University, Omu-Aran. Nigeria.
[4]Department of Computer Science, University of Ilorin, 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.

Methods

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.

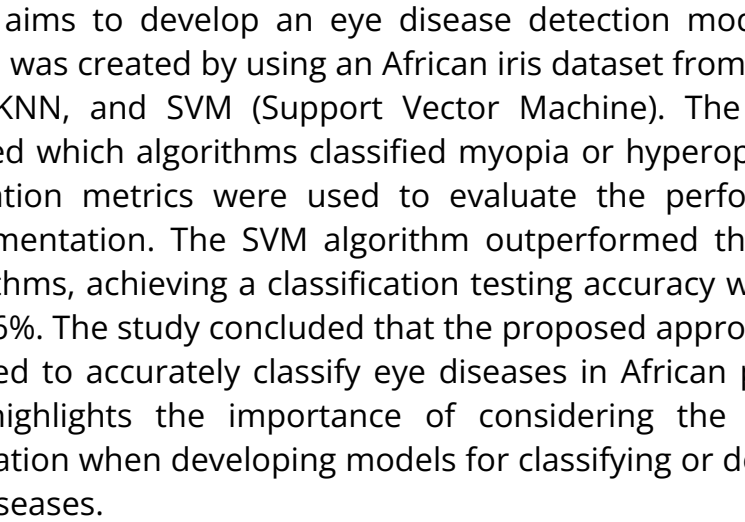
Result Analysis for SVM

```
aluating 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
oliferate_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]]
c: [0.93178718 0.88130969 0.80354707 0.93315143 0.90450205]
nsitivity: [0.96398892 0.39189189 0.69       0.12820513 0.22033898]
ecificity: [0.90053763 0.93626707 0.84615385 0.97838617 0.96439169]
```

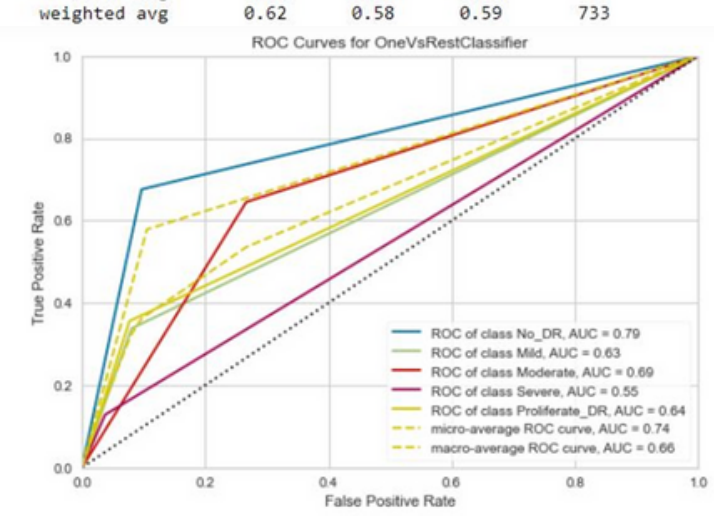


Result Analysis for KNN

```
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
```



	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

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