Florida International University

CNT-4153: IoT Applied Machine Learning Project

Diabetic Retinopathy Debrecen Model

By: Karla Estefany Colmenares
4-14-2024

1) Problem:

Diabetic retinopathy is a serious eye condition that can lead to blindness if not detected early. It's common among people with diabetes but catching it in its early stages can prevent vision loss. Currently, doctors manually examine eye images for signs of retinopathy, which takes a lot of time and can vary between doctors. But with the help of computers and machine learning, we can automate this process. By teaching computers to recognize patterns in eye images that indicate retinopathy, we can speed up diagnosis and ensure that more people get the treatment they need before it's too late. This could make a big difference in preventing blindness in diabetic patients.

2) Motivation:

Diabetic retinopathy poses a significant threat to the vision of individuals living with diabetes. Early detection is crucial for preventing irreversible vision loss, but current manual screening methods are time-consuming and subject to variability. By harnessing the power of machine learning, we are trying to automate the detection of diabetic retinopathy from retinal images, streamlining the diagnosis process and ensuring timely intervention. This could have a profound impact on reducing the risk of blindness among diabetic individuals and improving overall healthcare outcomes.

3) Dataset Information:

The dataset contains various features extracted from retinal images, each providing valuable information for predicting the presence of diabetic retinopathy. Here's a breakdown of the features and dataset:

Quality Assessment (Binary): This feature indicates the quality of the retinal image, with a value of 1 representing sufficient quality and 0 indicating poor quality.

Pre-screening Result (Binary): A binary result indicating the presence (1) or absence (0) of severe retinal abnormality detected during pre-screening.

- 3-8. MA Detection Results: These features represent the results of Microaneurysm (MA) detection at different confidence levels (alpha = 0.5 to 1), with each feature value indicating the number of MAs found.
- 9-15. Exudates Detection Results: Similar to MA detection, these features represent the results of exudates detection at different confidence levels, normalized by dividing the number of lesions with the diameter of the Region of Interest (ROI) to compensate for different image sizes.

Euclidean Distance of Centers: This feature provides information about the Euclidean distance between the center of the macula and the center of the optic disc, normalized with the diameter of the ROI.

Diameter of the Optic Disc: Indicates the diameter of the optic disc in the retinal image.

AM/FM-based Classification Result (Binary): The result of AM/FM-based classification, with a binary value of 1 indicating a positive classification and 0 indicating a negative classification.

Class Label (Binary): The target variable indicating the presence (1) or absence (0) of signs of diabetic retinopathy (DR) in the retinal image.

4) Feature Processing and Feature Engineering:

In the feature processing and feature engineering phase of our project, we employed two approaches to optimize our dataset for machine learning model training. We chose a Random Forest classifier to rank the importance of features in our dataset. By training the Random Forest model on our dataset, we were able to extract the feature importances attribute, which provided valuable insights into the relative importance of each feature in predicting the presence of diabetic retinopathy. This step allowed us to identify the most influential features, which were subsequently prioritized for further analysis and model training.

Building upon the insights gained from feature importance ranking, we then performed feature selection to retain only the most informative features while discarding redundant or less relevant ones. We automatically selected the top-ranked features based on their importance scores. These feature selection and dimensionality reduction techniques helped streamline our dataset, simplify model training, and potentially enhance model generalization performance by mitigating the effects of dimensionality and redundancy.

5) Machine Learning Model Development:

In the machine learning model development phase of the project, we focused on training and evaluating models to predict the presence of diabetic retinopathy based on the optimized dataset prepared during the feature processing and feature engineering phase. We began by splitting the dataset into training and testing sets to ensure an unbiased evaluation of model performance. Next, we selected two powerful classification algorithms, Random Forest and Support Vector Machines (SVM), known for their effectiveness in handling complex classification tasks.

For each algorithm, we initiated the model training process using the training set. With Random Forest, we trained an ensemble of decision trees to collectively make predictions on the target variable. Meanwhile, with SVM, we utilized a linear kernel function to separate the classes in the feature space.

Once trained, we evaluated the performance of each model using various evaluation metrics, including accuracy, precision, recall, F1-score, and ROC curve analysis. These metrics provided insights into the models' ability to correctly classify retinal images as containing signs of diabetic retinopathy or not. By comparing the performance of Random Forest and SVM, we were able to identify the most effective model for our specific prediction task.

6) Prediction/Result:

In the prediction and result analysis phase, we evaluated the performance of two machine learning classifiers, Random Forest and Support Vector Machine (SVM), in predicting the presence of diabetic retinopathy from retinal images.

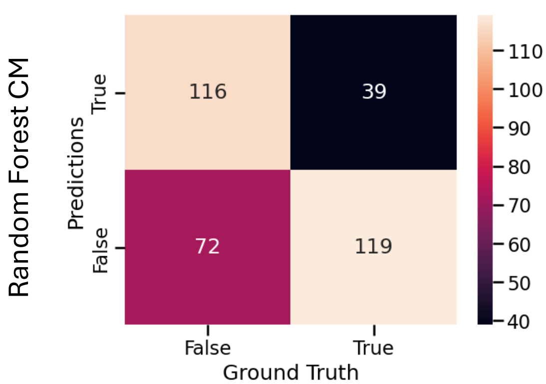
For the Random Forest classifier, we observed an accuracy of 67.92%, precision of 75.32%, recall of 62.30%, F1-score of 68.19%, and an AUC of 68.57%. These metrics indicate that the Random Forest model achieved moderate performance in classifying retinal images. While the model demonstrated relatively high precision, correctly identifying instances of diabetic retinopathy among positive predictions, it exhibited lower recall, indicating a proportionately higher number of false negatives.

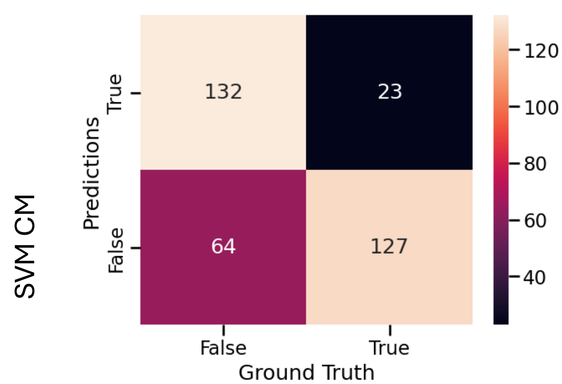
On the other hand, the SVM classifier yielded an accuracy of 74.86%, precision of 84.67%, recall of 66.49%, F1-score of 74.49%, and an AUC of 75.83%. These metrics suggest that the SVM model performed reasonably well in predicting diabetic retinopathy. The model achieved high precision, effectively identifying instances of diabetic retinopathy among positive predictions, and demonstrated a balanced trade-off between precision and recall.

7) Evaluating the result/metrics (Include graphs):

To visually represent the performance of the classifiers, we generated several graphs. We plotted a confusion matrix for each classifier, illustrating the true positive, true negative, false positive, and false negative predictions. These matrices provide insights into the classifiers' abilities to correctly classify instances of diabetic retinopathy and non-diabetic retinopathy.

Overall, both classifiers showed promise in predicting diabetic retinopathy from retinal images, with the SVM classifier exhibiting slightly superior performance compared to the Random Forest classifier. However, further analysis and comparison with additional models may be necessary to identify the most effective approach for this prediction task. These findings underscore the potential of machine learning techniques in assisting clinicians in the early detection and intervention of diabetic retinopathy, ultimately improving patient outcomes, and reducing the risk of vision loss among diabetic individuals.





8) Conclusion:

In conclusion, this model showcases the effectiveness of machine learning in automating the detection of diabetic retinopathy from retinal images. Both Random Forest and Support Vector Machine classifiers demonstrate promise in this task, with SVM showing slightly superior performance. These findings highlight the potential of machine learning to revolutionize diabetic retinopathy screening, paving the way for improved patient care through early detection and intervention.

9) References:

Antal,Balint and Hajdu,Andras. (2014). Diabetic Retinopathy Debrecen. UCI Machine Learning Repository. https://doi.org/10.24432/C5XP4P.