CALIFORNIA STATE SCIENCE FAIR
2016 PROJECT SUMMARY

Name(s)

Rishab Gargeya; Manan Shah

Project Number

36626

Project Title

Automated Diagnosis of Diabetic Retinopathy in Fundus Images: A Novel Synthesis of Biological and Data Driven Approaches

Abstract

Objectives/Goals

Diabetic retinopathy (DR) is one of the leading causes of preventable blindness globally. Today, the disease has the potential to affect 382 million diabetic patients worldwide. Current screening for DR is based on manual grading of severity, leading to undiagnosed and untreated disease where screening specialists are not easily available. Less than half of DR patients are aware of their condition; the disease remains asymptomatic until later stages. The goal of this work is to develop a computer-aided algorithm for the automated gradation and detection of retinopathy severity extracted from color retinal scans. This automated algorithm may be a cost-free tool for early detection of DR, decreasing the workload on retinal specialists and allowing any diabetic patient to obtain a timely and accurate diagnosis.

Methods/Materials

We trained and tested our model on 88,702 fundus images in which disease acuteness was assessed by a retina specialist on a scale of 0 through 4. Image quality within the dataset ranged from well detailed to very noisy. Our model utilized the Python and MATLAB programming languages for image analysis and processing. Our methods for data-driven analysis include convolutional neural networks (CNNs) and intensive regression models. Biological feature extraction methods include mathematical morphology and Gabor filter banks.

Results

The quadratic weighted kappa between predicted and actual classes was 0.75 and the model's AUROC was 0.92, indicating excellent agreement between predictions and the gold standard. Our model performed as well as the average inter-rater agreement between optometrists and ophthalmologists in clinical settings (who report an average 0.75 kappa value), indicating its widespread viability. Furthermore, 80% of misclassified instances were within one class of their true value, representing, on average, small deviation from expected results.

Conclusions/Discussion

We present the first ever data-driven DR detection model based on automated feature learning. We generate applicable features in real time through deep learning with CNNs and utilize manually engineered features representing properties of prognostic biological structures. Our algorithm accounts for varying image quality and ethnicities, indicating robustness and widespread clinical applicability. We hope to extend our model to the diagnosis of other retinal pathologies and deploy an application for public use.

Summary Statement

We developed a novel algorithm that accurately diagnoses diabetic retinopathy from a simple retinal scan, using morphological image processing and data-driven feature engineering to obtain ground-breaking results on large, varied datasets.

Help Received

Our project was conducted at Stanford University's Department of Radiology under the supervision and guidance of Dr. Daniel Rubin, Dr. Luis de Sisternes, and Dr. Theodore Leng.