Research Article

Computer Aided Diagnosis Model of Glaucoma with Eye Tracking Data

[1]Pramodini Punde,, [2]Mukti Jadhav, [3]Gangadevi Bedke,

[1]Dept of CS and IT, Dr B.A.M.University, Aurangabad
[2]Dept of CS, Shri Shivaji Arts and Science Chikhli
[3]Dept of CS and IT, Dr B.A.M.University, Aurangabad

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 23 May 2021

ABSTRACT: Computer Aided Diagnostics (CAD) can be considered as a 'second opinion', it is an important application of machine learning to help health professionals in making diagnostic decisions. CAD is a broad concept to assist medical professionals in decision making process, that has been applied to a variety of medical data integrated with image processing, computer vision, mathematics, physics, statistics and machine learning. Use of machine learning in computer aided diagnostics for Glaucoma is based on the fact that there are multiple tests available for the diagnosis, medical data are often not easily interpretable and the interpretation can depend very much on the skill of the doctor, specialization of medical fields that highly impact the accuracy in the diagnosis. Machine learning has models for understanding complex processes, statistical power to build more accurate predictive models. Increasing confidence in such diagnostics would decrease the number of patients or the complications associated with the disease. The only way to save vision due to glaucoma is to early detect its onset. In the proposed work we have used dataset collected from an eye tracker for 55 participants (glaucoma patients and normal subjects) to develop a model for prediction of Glaucoma with logistic regression.

Introduction

While traditional diagnosis remains a key component of clinical workup, many diseases can nowadays benefit from earlier and more refined characterization in order to make the most of increasingly advanced, often personalized therapies and multimodality treatment regimens. Glaucoma also has a major workload of services. It is one of the main causes of irreversible blindness. Early detection is the key to save vision loss due to Glaucoma.

A complete eye exam includes five common tests to detect glaucoma. As effective therapy can inhibit the progress of glaucoma, early diagnosis is one of the main goals in the treatment of this disease. It is strongly believed that the thinning of the RNFL correlates highly with, or even precedes, visual field loss in glaucoma. Once glaucoma is diagnosed, patients need lifelong treatment and monitoring within eye hospitals to control the visual damage. Thus, there is a major workload of eye services for people detected with glaucoma. Eye tracking is a sensor technology that enables a device to know exactly where a person is looking, his eyes are focused. It also determines user's attention, focus, drowsiness, consciousness or other mental state. Here, we propose that how eye tracker data may help us in diagnosis when the user is doing routine tasks.

In the modern multidisciplinary clinical context, such advances tend to rely on the collation and analysis of much varied information and data obtained through at various stages of a patient's workup. The extraction, quantity and variety of such data and their subsequent synthesis is often well beyond human capabilities and thus requires the use of advanced computed analysis. This is where the concept of Computer-Aided Diagnosis and related approaches arose from and, while those were initially developed for extracting and combining features derived from images they have subsequently and more recently been broadened to encompass all types of clinical data and biomarkers, with the aim of providing an actionable understanding of disease to assist with its detailed characterization and optimize interventional strategies. Medical imaging is an indispensable tool for modern healthcare. Machine leaning plays an essential role in the medical imaging field, with applications including medical image analysis, computer-aided diagnosis, organ/lesion segmentation, image fusion, image-guided therapy, and image annotation and image retrieval.

Computer Aided Diagnostics

Computer Aided Diagnostics is fundamentally based on a very complex computational intelligence. X-ray or any other types of images are scanned for suspicious structures. The main goal of CAD systems is to identify abnormal signs at an earliest that a human professional fails to find [1]. Computer-Aided Diagnosis (CAD) demonstrates the practical applications with interdisciplinary approach for eg. in engineering and medical schools, applied-science, medical students, medical engineers, researchers in industry, academia, and health science, radiologists, cardiologists, surgeons, and healthcare professionals.

Computational intelligence (CI) combines elements of learning, adaptation, evolution and fuzzy logic that are all closely related to machine learning, to allow one to create, in some sense, intelligent applications. CI techniques typically rely on heuristic algorithms in neural networks, fuzzy systems and evolutionary computation. Each of these algorithms has advantages and disadvantages, and many computer-aided, review and diagnosis CAD applications have used one of these algorithms. There are three major learning paradigms in CI, each corresponding to a particular abstract learning task: supervised learning, unsupervised learning and reinforcement learning. Most training algorithms can be viewed as a straightforward application of optimization theory and statistical estimation [5].

Machine Learning

Machine learning (ML) is the study of computer algorithms that improve automatically through experience. It is seen as a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as ‘training data’ to make predictions or decisions without being explicitly programmed to do so. In machine learning, algorithms are 'trained' to find patterns and features in massive amounts of data in order to make decisions and predictions based on new data. The better the algorithm, the more accurate the decisions and predictions will become as it processes more data[5]. Machine learning algorithms are to analyze any dataset to extract data-driven model, prediction rule, or decision rule from the dataset. Various machine learning algorithms are now used to develop high-performance medical image processing systems such as computer-aided detection system which detects clinically significant objects from medical images and computer-aided diagnosis system which quantifies malignancy of manually or automatically detected clinical objects. In this paper, we introduce some applications of machine learning algorithms to the development of medical image processing system.

Machine learning involves computers discovering how they can perform tasks without being explicitly programmed to do so. It involves computers learning from data provided so that they carry out certain tasks. For simple tasks assigned to computers, it is possible to program algorithms telling the machine how to execute all steps required to solve the problem at hand; on the computer's part, no learning is needed. For more advanced tasks, it can be challenging for a human to manually create the needed algorithms. In practice, it can turn out to be more effective to help the machine develop its own algorithm, rather than having human programmers specify every needed step.

Machine learning methods have been applied to classifying images acquired with various imaging modalities, using a variety of features, for various diseases, and with tools such as CAD and radiomics. Developing a machine learning method involves creating a training function for a dataset (the feature vectors, in the case of image classification) and making use of logical inference. When classes (diagnoses or clinical outcomes) for final decision of the model are pre-established, the training process is supervised. When there is no defined class, the process is unsupervised. In the latter case, the algorithm is aimed at the formation of clusters of similar samples (“exams with a similar pattern”), which may or may not be related to a known condition or disease.

Machine learning is basically using data to answer a question. An algorithm is a set of statistical processing steps towards answer. The type of algorithm depends on the type (labeled or unlabeled) and amount of data in the training data set and on the type of problem to be solved[6]. The types of machine learning algorithms differ in their approach, the type of data they input and output, and the type of task or problem that they are intended to solve.

Medical Imaging Intelligence and Analysis provides a comprehensive overview of machine learning research and technology in medical decision-making based on medical images.

Tools that employ artificial intelligence, machine learning, and deep learning can be used in different ways to analyze images. In the field of radiology and diagnostic imaging, such tools have been applied primarily in CAD, content-based image retrieval (CBIR), and radiomics/radio genomics.

Coupled with rapid improvements in computer processing, these AI-based systems are already improving the accuracy and efficiency of diagnosis and treatment across various specializations. The increasing focus of AI in radiology has led to some experts suggesting that someday AI may even replace radiologists. These suggestions raise the question of whether AI-based systems will eventually replace physicians in some specializations or will augment the role of physicians without actually replacing them. To that end this paper researches the role of AI-based systems in performing medical work in specializations including radiology, pathology, ophthalmology, and cardiology. It concludes that AI-based systems will augment physicians and are unlikely to replace the traditional physician–patient relationship.

Statistical methods of automated decision making and modeling have been invented (and reinvented) in numerous fields for more than a century. Important problems in this arena include pattern classification, regression, control, system identification, and prediction.

Machine Learning Model for Glaucoma Detection

There are four basic steps for building a machine learning application in computer aided diagnostics. These are typically performed by data scientists working closely with the medical health professionals for whom the model is being developed.

Step 1: Select and prepare a training data set

Training data is a data set representative of the data the machine learning model will ingest to solve the problem it's designed to solve. In some cases, the training data is labeled data, 'tagged' to call out features and classifications the model will need to identify. Other data is unlabeled, and the model will need to extract those features and assign classifications on its own.
In either case, the training data needs to be properly prepared, randomized, de-duplicated, and checked for imbalances or biases that could impact the training. It should also be divided into two subsets: the training subset, which will be used to train the application, and the evaluation subset, used to test and refine it.
Table 1.The Participants in the dataset

	No.	Group 0	Group 1
Male	32	15	17
Female	24	10	14

Step 2: Choose an algorithm to run on the training data set
Again, an algorithm is a set of statistical processing steps. The type of algorithm depends on the type (labeled or unlabeled) and amount of data in the training data set and on the type of problem to be solved. These kinds of algorithms are known as supervised machine learning algorithms.
Classification algorithms are used when the outputs are restricted to a limited set of values, and regression algorithms are used when the outputs may have any numerical value within a range. As an example, for a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email.
Common types of machine learning algorithms for use with labeled data include the following:
Regression algorithms:
Linear and logistic regression are examples of regression algorithms used to understand relationships in data. Linear regression is used to predict the value of a dependent variable based on the value of an independent variable. Logistic regression can be used when the dependent variable is binary in nature. Here we have used Logistic regression on the eye tracking dataset to predict whether a person is having Glaucoma or not based on the features fixation duration, number of fixations, saccades.

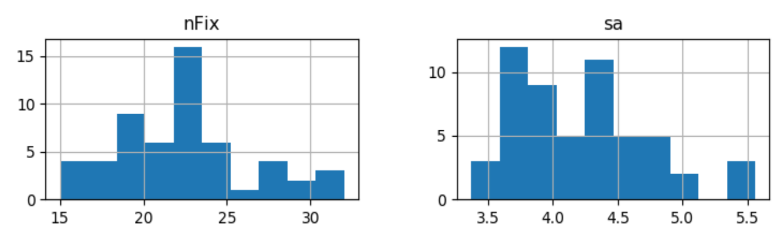

Fig.1 Histogram plot of features Number of fixations and saccades
Model Training and model evaluation
Evaluation of CAD systems are measured by two major factors, such as sensitivity and specificity, they seek for suspicious structure. CAD systems may not be 100% but their hit rate means sensitivity can be up to 98% these days.
Another important step in the machine learning process is validation and performance assessment. Given a set of data, a machine learning classifier must use at least two different subsets to perform algorithm training and predictive model validation.
The performance of each model is to be evaluated on the test set that was held out before training the model (25% of the dataset). The performance on the test set is evaluated exactly as described for the train set above. A confusion matrix is built and all the performance measures are computed as for the train set.
Performance of Logistic Regression:

	Precision	Recall	F1-score
Class 0	0.89	0.89	0.89
Class 1	0.90	0.90	0.90
		Accuracy	0.89

Performance evaluation
Each model is to be evaluated by calculating performance measures using the confusion matrix. Confusion matrix, or contingency table, is used to evaluate the performance of a classification model on a set of data for which the true values are known. The confusion matrix has four categories. True positives (TP) are cases in which the classification

model predicted them to be high responders, and indeed, those cases were high responders, whereas true negatives (TN) correspond to cases correctly labeled as low responders. Finally, false negatives (FN) and false positives (FP) refer to low responders or high responders that were incorrectly labeled. From a confusion matrix, to evaluate classification models, we calculated following performance measures. Accuracy, a measure of how often the classifier is correct, was calculated as (TP + TN)/(total number of observations). Specificity, the proportion of actual negative cases (low responders) that were correctly identified was calculated as TN/(FP + TN), whereas sensitivity (also known as recall or TP rate), the proportion of actual positive cases (high responders) correctly labeled, was calculated as TP/(TP + FN).

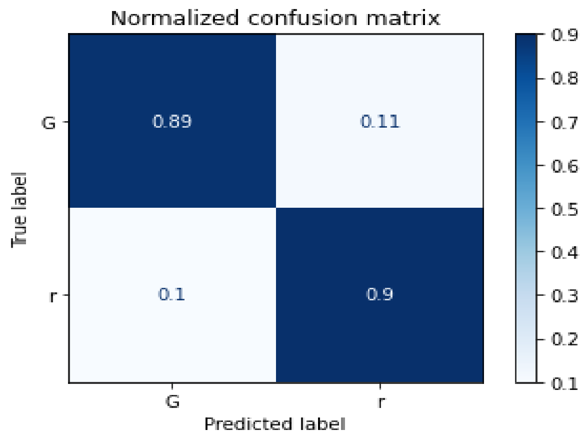

Fig. Confusion Matrix with Logistic Regression

Conclusion

CAD has undergone significant advances due to technological revolution in artificial intelligence, computational intelligence and machine learning. The advances in computational models have provided facilities in the clinical decision making, prognosis and prediction process. This paper discussed the machine learning approach related to clinical data for computer aided diagnostics in Glaucoma with the eye tracker data. To develop a machine learning model involves selecting the feature vectors as a training data, function, rules, algorithm specific data structures, and algorithm. We got 89% accuracy with logistic regression. We believe that Artificial Intelligence, Computational Intelligence, machine learning and CAD will change the way health professional work and change the perspective everyone in the clinical field has on their work. Increasing confidence in such diagnostics would decrease the number of patients or the complications associated with the disease.

REFERENCES

1. Christopher M. Bishop, ”Pattern Recognition and Machine Learning”, Springer, 2006 ISBN,-10: 0-387-31073-8
2. Santos MK, Ferreira Júnior JR, Wada DT Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine Radiol Bras 2019 387-396
3. Wikipaedia ,Machine_learning,https://en.wikipedia.org/wiki/Machine_learning
4. IBM Cloud Education, Machine_learning, https://www.ibm.com/cloud/learn/machine-learning, July, 2020
5. Bhagirathi Halalli Aziz Makandar, Computer Aided Diagnosis - Medical Image Analysis Techniques, Open access peer-reviewed chapter Intechopen ,2017
6. Heidi Daoxian Zhang Patrick Bernard Heffernan, Combination machine learning algorithms for computer-aided detection, Google Patents
7. Suzuki, Kenji, ”Machine Learning in Computer-Aided Diagnosis of the Thorax and Colon in CT: A Survey”, IEICE transactions on information and systems.2013, 772-783
8. Hadjiiski,L, "Advances in CAD for diagnosis of breast cancer", Curr Opin Obstet Gynecol. Feb. 2006; 18(1) 64-70
9. K. Doi, “Computer-aided diagnosis in medical imaging: Historical review, current status and future potential”, Computerized Medical Imaging and Graphics, 2017, 31(4), 198-211
10. Giger ML, “Computer-aided diagnosis in mammography”, Handbook of Medical Imaging, 2nd ed, SPIE Digital Library, Europe, 2000, 915-1004

11. Cheng HD, " Approaches for automated detection and classification of masses in mammograms", Pattern Recognition, 2006, 646-668
12. Innocent, "Computer Aided Fuzzy Medical Diagnosis", Information Science, 2004, vol. 162, pp. 81-104,
13. Penedo, "Computer-Aided Diagnosis: A Neural-Network-Based Approach to Lung Nodule Detection", IEEE Transactions on Medical Imaging, Dec 1998, vol. 17
14. Kaur, M; Talwar, R., "Review on: blood vessel extraction and eye retinopathy detection". International Journal of Computer Science and Information Technologie, 2014, 5 (6), 7513–7516
15. Golemati, Spyretta, Nikita, Konstantina, "Cardiovascular Computing-Methodologies and Clinical Applications", 2019, Springer
16. Gastounioti, Aimilia Makrodimitris, Stavros, Golemati, Nikita, "A novel computerized tool to stratify risk in carotid atherosclerosis using kinematic features of the arterial wall" IEEE Journal of Biomedical and Health Informatics, 2014, 1137–1145.
17. Golemati, Nikita, Konstantina, "Cardiovascular Computing-Methodologies and Clinical Applications", 2019, Springer
18. Abe, Yoshiyuk, Hanai, Kouzo, Nakano, Makiko, "A Computer-aided Diagnosis (CAD) System in Lung Cancer Screening with Computed Tomography", Anticancer Research, 2005, 483–488
19. Chen, S.; Suzuki, K., "Computerized Detection of Lung Nodules by Means of Virtual Dual-Energy; Radiography", IEEE Transactions on Biomedical Engineering, 2013, 369–378
20. Padma, A.; Sukanesh, R., "Segmentation and Classification of Brain CT Images Using Combined Wavelet Statistical Texture Features", Arabian Journal for Science and Engineering,2014, 767–776.
21. Ahuja, Abhimanyu S. "The impact of artificial intelligence in medicine on the future role of the physician." PeerJ vol. 7 e7702. 4 Oct. 2019, doi:10.7717/peerj.7702
22. Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C. (2010). Machine Learning in Medical Imaging. IEEE signal processing magazine, 27(4), 25–38.