AN INTERPRETABLE MACHINE LEARNING FRAMEWORK FOR THE DETECTION OF DIABETIC RETINOPATHY

[1]Pravin S. Rahate, [2]Dr. Nikhat Raza

PhD Scholar, Department of CSE, MPU, Bhopal, Madhya Pradesh, India[1], Associate Professor Department of CSE MPU, Bhopal, Madhya Pradesh, India[2]

psr.mpu@gmail.com[1], nikhat.raza@gmail.com[2]

ABSTRACT

Continuous movements in Artificial Intelligence (AI) and the addition of computational assets and abilities have set out the opportunity to cultivate Deep Learning (DL) applications for the detection and classification of diabetic retinopathy (DR). It offers most advantageous results. But it is tracked down that the model's general precision is not adequate all alone to permit clinicians to settle on a machine learning (ML) model. Clinicians see reasonableness as a method for legitimizing their clinical dynamic with regards to a model's choice. Subsequently, there is a need of hour to plan ML models working with the justification measure. This article proposes a model agnostic method on the top of ML model to provide explainabliity and interpretability for underlying model for the detection and classification of diabetic retinopathy (DR). It has a great advantage of flexibility by portioning the explainability from ML models. This framework will provide best results to enhance model interpretability, making clinical decisions more robust, bridging the gap between ML solution & human explanations and make better acceptance of ML / AI in sensitive & critical domains where value of human life is of an enormous concern such as healthcare.

Keywords: Machine Learning (ML), Deep Learning (DL), Interpretable Machine Learning, Diabetic Retinopathy (DR)

INTRODUCTION

There is a global epidemic of diabetes largely due to increasing urbanization with associated dietary and lifestyle changes, which includes a reduction in physical activity. About 382 million people live with diabetes (8.3% of the world's adult population in 2013) and by 2035 this will have increased by 55% to 592 million. Available estimates show that there are 65.1 million people in India with diabetes and that this number will rise to 109 million by 2035**[1].** Diabetes increases the risk of a range of systemic and eye diseases, but the main cause of blindness associated with diabetes is diabetic retinopathy (DR). It usually affects both eyes and can lead to vision loss if not detected early and treated.

Early detection and treatment are necessary in order to stoppage or avoid vision deterioration and vision loss. To that end, many Machine Learning (ML) models have been proposed by the research community for the detection and classification of diabetic retinopathy. Ongoing progressions in Artificial Intelligence (AI) and the increment of computational resources and capabilitieshave set out the freedom to foster Deep Learning (DL) applications for precise DR detection and classification. These models often perform the best.

Normally ML models are prepared with respect to high exactness yet as of late there is appeal for understanding the manner in which a particular model works and basic explanations behind the produced decisions. Accurate models work well but are not explainable as they are complicated. Translating ML models effectively to clinical practice requires building up clinicians' confidence in the framework. Clinicians view explainability as a means of justifying their clinical decision-making in the context of a model’s decision.Clinical field is the stream where explainability is pre imperative to guarantee scientific value of outcome.AI should not become a powerful deity which we follow blindly without understanding its reasoning but we should not forget about its beneficial insights it can have.

Clinicians wanted better explanations from the ML model to help them interpret the system's confidence, to better understand the reasoning chain of the system, and to make different diagnoses in order to help them make an assessment of the reliability of the system's decisions **[2].** It is found that the model's overall accuracy was not sufficient on its own to allow clinicians to make an informed decision, clinicians wanted to know the subset of features driving a prediction to allow them to compare the model decision to their clinical judgment.Any assistive tool acting on predictions in medical field look for understanding and rationalizing the predictions.Medical practitioners looks-for better clarifications to assist them with interpreting the framework's certainty, to check that the clinical problem fit the model idea, to more readily comprehend the thinking chain of the framework, and to make various determinations to assist them with making an appraisal of the dependability of the framework's choices. Consequently, there is a need of hour to design ML models facilitating the understanding and rationalization process.

LITERATURE REVIEW

The purpose of a literature review is to gain an understanding of the existing research and debates relevant to an area of study, Identify the relationship of works in context of its contribution to the topic and Identify gaps in research & conflicts in previous studies.

For the prediction of diabetes different number of classification techniques has been implemented such as Radial Basis Function (RBF), Multi-Layer Perceptron (MLP), Multi-Level Counter Propagation Network (MLCPN), Rotation Forest Algorithm, Bagging, Back Propagation Network (BPN), Generalized Discriminate Analysis (GDA) and Least Square Support Vector Machine (LS-SVM). In any case, some various kinds of characterization methods which have been acquainted yet due with some inappropriate harmony between datasets, their presentation become most exceedingly terrible in light of the fact that these have been addressed by larger part of classes consequently; some balancing classification technique like RHS-Boost can be applied on them **[3].**

This method is applied on misbalancing datasets for discovering the best precision and forecast. The following calculation has been carried out by Dwivedi and Chouhan. These calculations are Logistic relapse, ANN and Naive Bayes and its outcomes delight the exactness of 78%, 77% and 75% individually. At long last, the proposed calculation, Bagging has been carried out on a similar data set and creating the exactness of 81.77%. The outcome obviously shows that Bagging calculation yield most noteworthy exactness than others contrasted with others **[4]**.

Whereas Pramila M. Chawan et al.**[5]** explored the potential usage of the CNN in retinal image classification. Due to the tedious manual methods by medical personnel, a computerized framework can diminish the work engaged with diagnosing huge amounts of retinal pictures altogether.

Later a Deep Neural Network (DNN) is used for the automated identification and grading system of diabetic retinopathy with a high identification sensitivity of 97.5% and a specificity of 97.7%. On the other hand, the grading model achieved a sensitivity of 98.1% and a specificity of 98.9%**[6].**

On the other hand, Gadekallu et al.**[7]** projected the combination of principal component analysis (PCA) and deep neural network (DNN) for the early detection of diabetic retinopathy. The authors used a diabetes retinopathy dataset available at the UCI machine learning repository and He applied the Z-score technique for the normalization of dataset. Then, PCA was applied to extract the most significant features which was then given as an input to a DNN model for classification and the experimental results showed that the proposed DNN model achieved training accuracies between 72%-82% and testing accuracies between 68%-79%.

A diabetic retinopathy dataset can be normalized using the min-max method and with the help of an ensemble-based ML models Reddy et al. **[8]** detects diabetic retinopathy achieved detection accuracy above 80%. In thiswork, multiple classifiers including RF, DT, Adaboost, K-NN, and Logistic Regression (LR) were considered.

In spite of the promising outcomes shown by ML models for retinopathy detection, further opportunities still exist. One such opportunity is to consider different deep learning models such as convolutional neural networks (CNN) and recurrent neural networks (RNN) and combine them to extend the capabilities of traditional CNN models from the binary to multi-label image classification.Such designs can possibly further improve the performance of deep learning models for early detection of retinopathy**[9].**

The premium in applying profound learning in recognizing diabetic retinopathy has expanded during the previous years and as a few DL frameworks develop and become incorporated into the clinical practice, they will empower the clinicians to treat the patients in need all the more adequately and proficiently. The article **[10]** presents the present status of exploration with respect to the utilization of profound learning in diagnosing diabetic retinopathy. Though deep learning has paved the way for more accurate detection of DR, authors centers upgrades are as yet important with respect to execution, interpretability, dependability and trustworthiness from ophthalmologists.

A. Adadiet al.**[11]** highlights the need of Explainable Artificial Intelligence (XAI), elucidates basic terms related to the arena, XAI application domains, explainability strategies and explainability evaluation.It has then been presumed that impressive exertion will be needed in the future to handle the difficulties and open issues with XAI.

There are many methods in terms of explaining any black-box model as - local interpretable model-agnostic explanations (LIME), Shapley Additive explanations (SHAP), contrastive explanations method (CEM), Permutation importance (PIMP), partial dependence plot (PDP) and Accumulated Local Effect (ALE) plots. The LIME and SHAP strategies are the most far reaching and prevailing across the most comprehensive and dominant methods for visualising feature interactions and feature importance, while Friedman's PDPs, albeit a lot more seasoned and not as complex.The LIME and SHAP methods are not only model-agnostic, but they have been exhibited to be pertinent to any type of data**[12].**

ExplAIn **[13]** is a novel eXplanatory Artificial Intelligence (XAI) framework for multilabel image classification. This framework not only labels images but also classifies each pixel within images. This framework was applied to the diagnosis of Diabetic Retinopathy (DR) using Color Fundus Photography (CFP). In this paper, ExplAIn has been applied to a well-known classification problem (DR severity classification). It permitted us to assess the pertinence of the distinguished local patterns. Nonetheless, ExplAIn would be significantly more helpful for an absolutely new characterization issue as -disease progression prediction, diagnosis of a new disease, etc. it would permit information securing and rapiddistinguishment of new helpful markers in picturescorrespondingly.The proposed framework expects healthcare AI systems to gain the trust of clinicians and patients more easily by augmenting the new explainability feature without loss of classification performance.

A rapid advancement of XAI is evident, and the recent study by Linardatos et al. **[12]** has identified four main areas of focus: methods for explaining complex black-box models, methods for creating white-box models, methods that promote fairness and restrict the existence of discrimination, and methods for analyzing sensitivity of model predictions. The authors observed a noteworthy amount of work on explaining complex black-box models, especially on neural networks, probably due to the fact that there is great potential in terms of complex analyses and performance. Further, white-box models are labeled as more challenging to create and, accordingly, they give the impression to lost their popularity among developers.

While there are many studies on XAI in medicine, there is a limited number that focus on the context of decision support system for medical diagnosis. In this assessment Antoniadi, A.M et al.**[2]** focused on the use of XAI in decision support system and were able to measure some realized benefits as well as identify future needs in this area. However, in spite of some user studies reporting positive views on this, especially in light of explainability, there is still uncertainty around their use in practice. Thus, research focusing on all stages of medical assistance system development is required to establish more definitely how explainability can be put into useful practice.

PROPOSED SYSTEM

3.1 Analysis

Machine Learning (ML) algorithms suffer from opacity that it is difficult to get insight into their internal mechanism of work.Likewise Deep learning models are notorious for their un-interpretability because of complex approach towards extracting and combining features. Which further compound the issue, on the grounds that entrusting significant choices to a framework that can't account for itself presents clear risks. Therefore, medical experts struggle with the gap between what is output by an ML-based solution and human explanations. It demands for a field focusing on the understanding and interpretation of the behaviour of ML models which explains outputs in a transparent and easy to interpret manner.

3.2 PROBLEM STATEMENT

To develop a ML based explainable framework for the detection of diabetic retinopathy.

3.3 Proposed framework

It is essentially apparent that models which are interpretable and understandable are most favored by clinicians. Additionally, it is advantageous to researchers to validate and improve their work. We proposed an explainable framework for the detection of diabetic retinopathy using ML. The outline for the proposed framework is illustrated in figure 3.3.1.

The first step in the proposed framework is the collection of initial data sets. There are various data sets such as – UCI Machine Learning Repository, Kaggle data set, Public Use Microdata Areas (PUMA) , Optical Coherence Tomography Image Database (OCTID) and EyePACS available for the detection of diabetic retinopathy. For the initial experimentation we will start with UCI Machine Learning Repository. Then the preprocessing of data will be conducted which comprises of cleaning of data, data curation and the removal of redundant data. After that data is divided into training and testing data set. This will be the input to a black box ML model for the detection of DR. There are two reasons for the selection of black box ML model for detection of DR. The first one is that ML models works very well and have better results for the detection as well as grading of DR. Consequently there is no more room for improvement. Secondly separating the explanations from the ML model have greatest advantage of the flexibility. CNN, RNN and Radial Basis Function (RBF) seem the most promising methods. They facilitate extraction of features and classification process effectively. In order to interpret the model feature importance is a basic and often free approach. There is a need to find out preferable attribute as per the experiment to be performed which helps us to make decision system more robust & accurate for the prediction of DR**[14].**

In case of DR local explanation for individual patient is very vital. It helps a lot for the detection, grading as well as progression analysis of DR. The feature level explanations by learning an interpretable model that endeavors to surmised the conduct of the first model is the need of hour. With the purpose of augmentation to the ML model to recalibrate the trust of an individual patient we propose model agnostic method for the understanding of feature importance and interpretation of model. LIME provides a generic framework to discover black boxes provides the why behind AI generated predictions. So if we want to understand how a single prediction is made for a given observation then LIME is the method providing insights into the model and makes the model further trustworthy.

CONCLUSION

Ongoing advances in computerized reasoning (AI) have prompted its broad modern reception, with AI frameworks showing superhuman execution in a critical number of errands. Nonetheless, this surge in performance has regularly been accomplished through expanded model complexity, transforming such frameworks into "discovery" approaches and causing vulnerability with respect to the manner in which they work and, at last, the way that they come to choices. This vagueness has made it untrustworthy for ML models to be confirmed in sensitive yet basic area, where their worth could be massive, like medical care. Thus, logical premium in the field of Explainable Artificial Intelligence (XAI), a field that is worried about the advancement of new techniques that clarify and interpret AI models, has been hugely reignited over late years. To bridge the gap understanding behind these

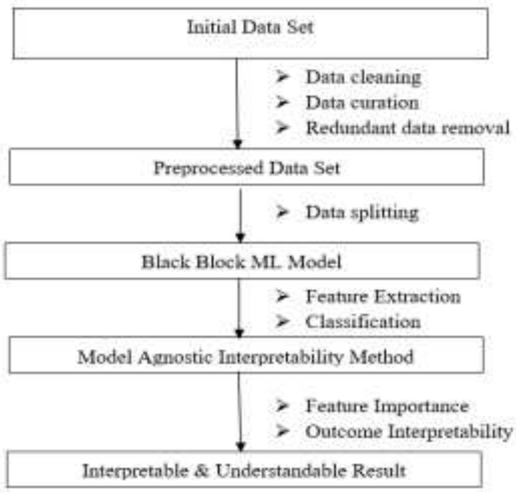

figure 3.3.1 A ML based explainable framework for the detection of diabetic retinopathy.

prediction is key aspect. Thus we proposed a framework that appends understanding and interpretability of well established models for the detection of DR. A fundamental analysis keep on the choice of techniques used to introduce clarifications in an instructive, effective and thusly clinically helpful way. Our propose structure puts just another layer of - LIME- a model agnostic method on the top of proven ML back model . This will enhance the explainability and interpretability aspect - on feature level for an individual patient – for the detection of DR using ML model. Optimistically it will heighten the reliability of medical field to make use of of ML / AI in the clinical decision support system.

REFERENCES

1. THE EMERGING EPIDEMIC OF DIABETIC RETINOPATHY IN INDIA- A report of a situation analysis and evaluation of existing programmes for screening and treatment of diabetic retinopathy conducted by South Asia Centre for Disability Inclusive Development & Research, IIPH, PHFI, Hyderabad
2. Antoniadi, A.M.; Du, Y.; Guendouz, Y.; Wei, L.; Mazo, C.; Becker, B.A.; Mooney, C. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical

Decision Support Systems: A Systematic Review. Appl. Sci. 2021, 11, 5088. https://doi.org/10.3390/ app11115088

3. Gong, J., & Kim, H. (2017). RHSBoost: Improving classification performance in imbalance data. Computational Statistics & Data Analysis,Volume 111,July 2017, 1-13.
4. Vandana Rawat, Suryakant; A Classification System for Diabetic Patients with Machine Learning Techniques; International Journal of Mathematical, Engineering and Management Sciences; Vol. 4, No. 3, 729–744, 2019
5. Kirti Satpute, Arsh Pakhare, Gaurav Gaikwad, Kunal Warudkar, Vishal Bhat; Detection Of Diabetic Retinopathy Using Machine Learning;
6. International Research Journal of Engineering and Technology (IRJET); Volume: 08 Issue: 05 May 2021Alyoubi, W.L.; Abulkhair, M.F.; Shalash, W.M. Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning. Sensors 2021, 21, 3704, https://doi.org/10.3390/s21113704
7. Gadekallu TR, Khare N, Bhattacharya S, Singh S, Reddy Maddikunta PK, Ra IH, Alazab M (2020) Early detection of diabetic retinopathy using pca-firefly based deep learning model. Electronics 9(2):274
8. Reddy GT, Bhattacharya S, Siva Ramakrishnan S, Chowdhary CL, Hakak S, Kaluri R, Praveen Kumar Reddy M (2020), An ensemble based machine learning model for diabetic retinopathy classification. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), pp 1–6
9. ohammadNoor Injadat, Abdallah Moubayed, AliBouNassif, AbdallahShami; MachineLearningTowardsIntelligentSystems:Applications,Challenges,andOpportunities; arXiv:2101.03655v1 [cs.LG] 11 Jan 2021
10. Nikos Tsiknakis, Dimitris Theodoropoulos, Georgios Manikis, Emmanouil Ktistakis, Ourania Boutsora, Alexa Berto, Fabio Scarpa, Alberto Scarpa, Dimitrios I. Fotiadis, Kostas Marias; Deep Learning for Diabetic Retinopathy Detection and Classification Based on Fundus Images:A Review; Computers in Biology and Medicine Volume 135, August 2021, 104599
11. A. Adadi, M. Berrada; Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI); IEEE Access;Volume 6, 2018; 52138- 52160
12. Linardatos, P.; Papastefanopoulos, V.; Kotsiantis, S. ExplainableAI:AReviewofMachine Learning Interpretability Methods Entropy2021,23,18.
13. Gwenol´e Quelleca, Hassan Al Hajjb, Mathieu Lamardb, Pierre-Henri Conzec, Pascale Massind, B´eatrice Cochenerb; ExplAIn: Explanatory Artiﬁcial Intelligence for Diabetic Retinopathy Diagnosis; arXiv:2008.05731v2 [cs.CV] 1 Sep 2020
14. Pravin S. Rahate , Dr. Nikhat Raza; A Machine Learning Approach for the Diagnosis of Diabetes: A Review, International Journal of Scientific Research in Computer Science, Engineering and Information Technology 2020 IJSRCSEIT | Volume 6 | Issue 2| ISSN : 2456-3307