

ANALYSIS OF DEEP LEARNING ALGORITHMS FOR DETECTING DIABETES

¹B. Narasimha Swamy, ²K. Michael Sadgun Rao, ³Mohammed Moulana

¹Prasad V Potluri Siddhartha Institute of Technology, Vijayawada. swamy_bn@pvpsiddhartha.ac.in

²Lakireddy Bali Reddy College of Engineering (Autonomous), Mylavaram. michael.kona@gmail.com

³Koneru Lakshmaiah Education Foundation, Guntur, AP

Received: 16 March 2020 Revised and Accepted: 16 June 2020

ABSTRACT

In Bangladesh as well as around the world, diabetes is currently one of the most prevalent and serious diseases. It is not only dangerous to the blood, but causes various kinds of conditions, such as blindness, renal disease, kidney disease, heart disease etc., which cause many deaths each year. It is therefore critical that a program is built that can accurately diagnose patients with diabetes with medical data. By five and ten times cross validation of its attributes, we propose a strategy for the diagnosis of diabetes using a deep neural network. Blindness is a leading cause of diabetic retinopathy (DR) in the entire world. Previous detection and timely processing of DR is desirable to reduce vision loss incidence and progression. Deep learning methods (DL) currently provide better efficiency in the detection of DR from images from retinal fundus. Consequently, with a meta-analysis of related research, we conducted a systematic review to assess the efficiency of DL algorithms to detect DR. This research offers a diabetes prediction method that minimizes the issue of overfitting using the drop-out technique. Deep neural learning is used where both layers are totally related and drop-out layers are followed. The results of the proposed neural grid have shown that other state-of-the-art methods have been surpassed and Pima Indian Diabetes Daten Set 's greatest output is reported.

KEYWORDS: diabetic retinopathy (DR), Deep learning methods (DL)

1. INTRODUCTION

Diabetes is among the world's elderly's most prevalent diseases. According to the International Diabetes Federation, 451 million people around the world were diabetic in 2017. It is estimated that this number will rise to 693 million inhabitants in the next 26 years[1]. Diabetes is known to be a chronic disorder associated with an irregular human condition, where blood glucose levels are erratic due to certain pancreatic dysfunctions that lead to little to no insulin output at all causing insulin-resistant type 1 diabetes to insulin-proof cells, causing type 2 diabetes [2,3]. Nevertheless, scientists agree that both genetic and environmental factors play a significant role in diabetes, but the key reasons for diabetes remain unclear. This can be treated by diagnosis and medication while it is incurable. Individuals with diabetes are vulnerable to other side effects including heart disease and nerve damage. This can avoid complications and help to reduce the risk of serious health issues through early diagnosis and treatment of diabetes. Several researchers in bioinformatics have tried to cope with this disease and have tried to build systems and resources to prevent diabetes. They either generated models with various algorithm types such as classification or algorithms associated with it. The most popular algorithms were Decision Trees, SVM help, and Linear Regression [4–6].

Among people of different sexes, races and age, diabetes is a common physiological health issue. The word diabetic is used if a person can not split glucose because of lack of insulin. The human organ called the pancreas produces the hormone called insulin that controls sugar levels in the human bloodstream. It is a very important enzyme. It punishes the human body to use sugar as an energy source; without adequate insulin, the energy required by body cells can't meet, resulting in too large sugar levels in the blood and various problems. Diabetes is not a curable illness. Fortunately, it is a treatable condition. An estimated mortality of nearly 200,000 Americans is caused by diabetes and associated complications[1, 2]. Predicting and treating diseases correctly has become important in the fields of medical prediction in modern health care. The entire diabetes evaluation procedure, which is always suggested by the doctor, is fully manual.

In recent years, the idea of deep learning has evolved rapidly and has become overwhelming. Deep learning methods, including medical prognostics and optical character analysis, are used in a variety of diversified areas[5]. This paper uses profound computing methods, including a deep neural learning network, to suggest a

high accuracy model for the prediction of diabetes. This is accomplished using a regularization layer called Dropout to resolve the overfitting problem created by the use of strongly connected columns.

Diagnoses in medical research are one of the main and daunting activities. Information from patients including plasma glucose, diastolic blood and triceps thickness of skin folds, serum insulin, body weight, age etc. are taken to predict diabetes conditions. The patient then consults a physician. The doctor decides on these factors based on his / her experience and expertise. The decision-making cycle is lengthy and often takes several weeks or months which makes it very difficult for the doctor to work[6]. A large number of medical datasets, which are useful for research in different fields of medical science, are now available. Therefore it is difficult to manage the large data by a human being or even impossible. Therefore the conventional modalities are efficient computer-based approaches. Computing system efficiency is enhanced and time and money is saved.

The theory of deep learning evolves rapidly and works like a human mind. The data is expressed in many levels and the invariance problem can be solved efficiently[7]. Deep learning methods in the area of medical prognostics are used in a variety of ways. Many research papers show that the results of deeper learning techniques are improved, the classification error rate is reduced and the noise is more robust than other approaches[8]. It can handle the vast amount of data and can quickly decipher a complex problem. Some computer teaching [9, 10] and bio-inspired computing technologies have recently been implemented in many medical prognostics, and also in profound learning techniques. We used a deep neural network that is a common method in machine learning recently to predict diabetes mellitus. However, the dataset is well educated before diabetes prediction, so that test data in almost all cases produces an accurate result. Past work on the prediction of diabetes using machine learning techniques does not have enough accuracy. Nevertheless, the accuracy rate in our method is much better than that in the outcome portion.

While a variety of modalities have been seen, no approach can deliver correctly or consistently. The article outlines an approach to diabetes prediction from the inputs. Deep Neural Network was used for the correct outcome diagnosis of diabetes.

2. RELATED WORKS USING MACHINE LEARNING

In the medical sector for preventing diseases, ML algorithms are well-known. To order to achieve the best and reliable results, many researchers have used ML techniques to predict diabetes. Multistake SVM, J48, K-Nearest Neighbors and the Random Forest classification systems were used by Kandhasamy and Balamurali[4]. The classification was achieved with a UCI server dataset. The classifier results have been measured based on precision, sensitivity and specificity values. The grading has been performed in two situations where the datasets are pre-processed and 5 times cross-validated without pre-processing. The authors did not clarify what phase the data set had taken before processing, they just noted that the noise had been removed. The Decision Tree J48 classifier was stated to be 73,82% with the highest accuracy without pre-processing, while the KNN (k=1) and Random Forest classifiers displayed the maximum accuracy of 100% following data pre-processing.

Yuvaraj and Sripreetha have been used for the prediction of diabetes using three different ML algorithms, Random Wood, Tree of Decision and Naïve Bayes. Pima Indian Diabetes (PID) set of data was used during preprocessing. The authors did not define how the data were pre-processed but rather discussed the information gain system used to extract the relevant features. Eight of the 13 main features were included. The data set was then split into 70% for training and 30% for research. The data showed a overall accuracy of 94 percent for the random forest algorithm.

In addition , a new integrated improved model for diabetes prediction was introduced by Tafa et al. Tafa and Naïve Bayes. The model was tested using a data collection from three locations in Kosovo. The data collection includes 8 characteristics and 402 patients with Type 2 diabetes in 80 beds. Several characteristics of daily diet and physical activity as well as family history of diabetes were not discussed before. Whether or not the data were pre processed was not specified by the authors. The data set was divided into 50 percent for each training and test set for the validation test. The combined algorithms proposed increased the accuracy of the forecast to 97.6%. This value was compared with 95.52 and 94.52 per cent respectively for SVM and Naïve Bayes.

Deepti and Dilip furthermore have used Diabetes detection classifications from Decision Tree, SVM and Naive Bayes. The goal was to decide the classifier as accurately as possible. This analysis used the Pima Indian dataset. Cross-validation takes place in 10 times the partitioning of the dataset. Data preprocessing was not discussed by the scientists. Evaluated results using accuracy, duration, repetition and F-measurement steps. The Naive Bayes, which reached 76.30 percent, achieved the highest precision.

Six separate classifiers were used by Mercaldo et al. J48, HoeffdingTree, JRip, BayesNet and RandomForest ranked in the categories of multilayer perceptron. This analysis was also based on the Pima Indian dataset. However, two algorithms, GreedyStepwise and BestFirst, were employed by the authors to classify the discriminatory attributes that lead to increased classification efficiency. There are four characteristics, i.e. body weight index, concentration of glucose in plasma, pedigree function in diabetes and age. The dataset is backed by a 10 fold-cross validation. The distinction was based on the importance of accuracy, the alert and the F-measurement. The tests showed an accuracy of 0.757, 0.762, and 0.759 for F-measurement using the Hoeffding Tree algorithm. Compared to the other, this result is the best.

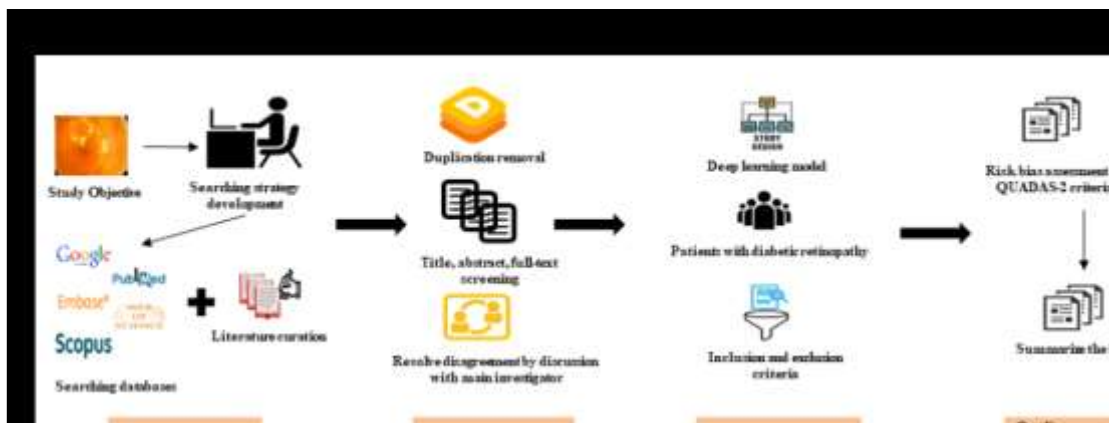
Negi and Jaiswal have targeted SVM in predicting diabetes in addition to the other research. As a composite data collection, Pima Indians and Diabetes 130-Americans were used. The aim of this analysis was to determine if the findings are accurate, as other scientists only use a single dataset. The information kit consists of 102,538 samples and 49 characteristics, 64,419 positive and 38,115 negative. The authors did not list the attributes used in this report. The dataset is pre-processed by eliminating missing and non-numeric values, and the data is finally normalized by 0 to 1. The data sets are pre-processed. Before implementation of the SVM model, various feature selection approaches were used. The LIBSVM package selected four attributes in the Fselect script and selected nine and 20 attributes in the Wrapper and Ranker methods (from the Weka tool). The authors employed a 10-fold cross validation technology for the validation process. The diabetes prediction could be more accurate by using a combined data package with a precision of 72%.

In addition, Olaniyi and Adnan have been using the Multilayer Neural Feed Network. For algorithm training the backpropagation algorithm was used. The goal was to increase the accuracy of diabetes prediction. The Pima Indian Diabetes software has been used. In order to achieve numerical consistency the authors standardized the data set prior to classification processing. The corresponding amplitude of each sample attribute has been extracted to return all data set values from 0 to 1. Subsequently, the data collection is split up into 500 test samples and 268 test samples. The accuracy obtained was 82 percent, a high precision figure.

3. METHODS

3.1. Research Design:

Present work is performed using the Cochrane Systemic Reviews Handbook (PRISMA), the Recommended Reporting Elements for Systematic Analyses and Meta-Analyses. The published manual was not reviewed. The analysis process is as follows:



5: Flow-chart of the Systematic Review Process

3.2. Search Methods for Identification of Studies:

3.2.1. Electronic Database Search:

We methodically search the most generally utilized web crawlers, for example, EMBASE, PubMed, Scopus, Google Scholar, and Web of Science, to discover what was distributed between 1 January 2000 and 1 March 2019 utilizing the most suitable free content terms ("Retinopathy" OR "Diabetic retinopathy," OR "Referable diabetic retinopathy," OR "Retinal fundus picture," OR "Vision undermining diabetic retinopathy," OR "DR"), as well as "D"

3.2.2. Searching for Other Sources:

We additionally deliberately looked the reference records for the recovered examinations and the significant past audit articles for additional incorporation.

3.3. Inclusion and Exclusion Criteria:

Two scholars (MMI, TNP) self-sufficiently screened the titles and altered arrangements for recuperated articles fundamentally considered in the intentional review. They picked the most appropriate and expected articles from the hidden screening and spared it for full-text review. Contrasts were settled by discussion with another ace (HCY). Finally, all inconsistencies were settled by a discussion with the standard specialist (YCL). To be fused, all the examinations expected to fulfill the going with models: 1) study must be in English and be peer-evaluated, 2) gave a consequence of DL figurings and DR area, 3) expected to give information any of the appraisal matric, for instance, accuracy, AUROC, affectability and expressness, 4) expected to part with from about database and number of pictures 5) expected to part with a from of DR, 6) clearly portrayed DL counts and strategy used in the DR revelation. Distributions, short reports, ordinary procedures for recognizing DR were restricted. All the assessments meeting fuse models at this stage were additionally examined by a comparative two makers to ensure the fittingness of the last examination. All distinction between two makers for picking potential assessments were then settled by the major operator (YCL). Studies giving the most bare essential information regarding counts were put something aside for references.

3.4. Data Extraction:

Two of us (MMI) and (TNP) independently recouped all information from the picked articles reliant on the predefined, standardized show and data combination instrument. They entered the data to Review Manager programming (RevMan-5) and as such checked for exactness. Finally, they accumulated the going with information from the included articles: (a) methods: model used in the assessment, database information, number of pictures, assessment of retinopathy, recognizing confirmation strategy of retinopathy, thought and dismissal norms, camera information, degree of planning and testing dataset, dataset used in getting ready, testing and endorsement process, number of diabetic patients, evaluation estimations, for instance, precision, affectability, expressness and AUROC; (b) individuals: supreme number of individuals, number of DR pictures, mean age, age go, sexual direction, level of sex; (c) mediations: CNN, significant neural framework; (d) result: desire for DR (Referral, vision threatening, and different sorts of DR).

3.5. Quality Assessment:

Any efficient reviews with meta-examination of indicative precision ponders are as often as possible depicted by such heterogeneous revelations that starts from contrasts in the structure and direct of included assessments. MMI and TNP independently used the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) instrument for evaluating the idea of the included examinations. The QUADAS-2 scale involves four key spaces for threat of inclination, for instance, calm assurance, list test, reference standard, and stream and timing and three regions for genuine nature worries, for instance, getting decision, list test, and reference standard. The risk of inclination was requested into three areas (low, high and obfuscated danger tendency).

3.6. Outcome Parameters:

The two fundamental outcome limits of this efficient study and meta-examination were: (1) to survey the display of DL estimations for referable DR distinguishing proof from the electronic fundus pictures (2) to perceive the wellspring of inclination and assortment in decisive precision focuses by QUADAS-2.

3.7. Statistical Analysis:

We used Meta-Disk (Version: 1.4) for quantifiable assessment, for instance, a pooled check of AUROC, affectability, identity and illustrative possibilities extent. It was used to a) summarize data from each individual assessment, b) evaluate the homogeneity of included examinations graphically and quantifiably, c) figure the pooled measure. Expressive possibilities extent (DOR) was moreover prepared to know, the amount increasingly vital the odds of having DR are for the people with a useful test result than for the people with a contrary test result. Likelihood extents are resolved to impart the sum continuously visit the in regards to finding is among the individuals with DR than among the individuals without DR.

4. CONCLUSION

The discoveries of our investigation show that DL can assume a significant job in distinguishing referable DR with high affectability, particularity, and repeatability. The use of a DL-based computerized framework may change the way DR is analyzed later on. Robotized devices may improve the nature of DR screening, openness to social insurance as

well as to lessen the expense of screening. Diabetes is an interminable infirmity that must be forestalled before troubles individuals. Each year in the world, diabetes causes a lot of death. Therefore it is critical for treatment to recognize diabetes at its underlying level. This research has modified a deep neural network in order to predict diabetes. This test was combined with a deep neural network procedure with the characteristics of certain clinical components based diabetes. For five-overshoot cross-approval, which is relative high than specific methods used to predict diabetes mellitus, we found the accuracy to be 98.35 percent. For health personnel and general public, the suggested context should be clear.

5. REFERENCES

1. Cho, N.; Shaw, J.; Karuranga, S.; Huang, Y.; Fernandes, J.D.R.; Ohlrogge, A.; Malanda, B. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. *Diabetes Res. Clin. Pr.* **2018**, *138*, 271–281.
2. Sanz, J.A.; Galar, M.; Jurio, A.; Brugos, A.; Pagola, M.; Bustince, H. Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system. *Appl. Soft Comput.* **2014**, *20*, 103–111.
3. Varma, K.V.; Rao, A.A.; Lakshmi, T.S.M.; Rao, P.N. A computational intelligence approach for a better diagnosis of diabetic patients. *Comput. Electr. Eng.* **2014**, *40*, 1758–1765.
4. Kandhasamy, J.P.; Balamurali, S. Performance Analysis of Classifier Models to Predict Diabetes Mellitus. *Procedia Comput. Sci.* **2015**, *47*, 45–51.
5. M. Sailaja and Dr. C.V.P.R. Prasad “Finding *shortest path for community structures with Deep Learning*” in International Journal of Pure and Applied Mathematics” ISSN: 1314-3395, Volume-118, No-14.
6. Iyer, A.; Jeyalatha, S.; Sumbaly, R. Diagnosis of Diabetes Using Classification Mining Techniques. *Int. J. Data Min. Knowl. Manag. Process.* **2015**, *5*, 1–14.
7. Razavian, N.; Blecker, S.; Schmidt, A.M.; Smith-McLallen, A.; Nigam, S.; Sontag, D. Population-Level Prediction of Type 2 Diabetes from Claims Data and Analysis of Risk Factors. *Big Data* **2015**, *3*, 277–287.
8. Ashiquzzaman, A.; Kawsar Tushar, A.; Rashedul Islam, M.D.; Shon, D.; Kichang, L.M.; Jeong-Ho, P.; Dong-Sun, L.; Jongmyon, K. Reduction of overfitting in diabetes prediction using deep learning neural network. In *IT Convergence and Security*; Lecture Notes in Electrical Engineering; Springer: Singapore, 2017; Volume 449.
9. Swapna, G.; Soman, K.P.; Vinayakumar, R. Automated detection of diabetes using CNN and CNN-LSTM network and heart rate signals. *Procedia Comput. Sci.* **2018**, *132*, 1253–1262.