AN AUTOMATED RESEARCH FOR ANALYSING AND VISUALISING THE DATA USING MACHINE LEARNING ANALYTICS

Mr. A. Shravan Kumar [1] Mr. G. Raghupal Reddy [2]
Assistant Professor [1] Programmer[2]
JNTUH College of Engineering Hyderabad

Received: 21.06.2019 **Revised: 22.07.2019** **Accepted: 25.08.2019**

ABSTRACT
Advances in data analytics have remained as a transformative force in various fields, their effect has shaped and converted technology-dependent dynamics that historically sustain technology fields. Health care is one such area where it is also at an emerging level. Artificial intelligence trends have realized their ability with the necessary datasets that supported, subsequently machine learning algorithms have a data ravening existence. The importance of recent progress in data analytics and their use in the health sector, with particular focus on predictive and simulation applications, have been discussed in the paper.
Key words: Healthcare**,** AI, Analytics, Visualization, Prediction, Machine Learning.

INTRODUCTION
Most of the present-day medical records are locked and they might not be digitized causing Incomplete database networks. This would be problematical as ingenious tools are widely used across various sectors; therefore, the multivariate sources of medical data need to be integrated in a proper manner and absolute structure that saves processing and enabling secured access. Data regarding patients, doctors, services and the places are actually divided across a number of systems, without the involvement of an external or a unified body, it is very challenging to establish data management partnerships across all these entities. The problem for scientists is the presence of incomplete, incompatible medical data sets and no prediction-based analysis mechanism.

The introduction of a healthcare data management system would address this issue by preserving a consolidated, stable patient data database by offering specialized analytical and display resources. The position of doctors also must be valued and their guidance plays an important role in the execution's success.

Neural networks are computing structures designed based on how a human brain neuron operate. At the same time, nonlinear connections between inputs and outputs were modeled and processed. Complex pattern detection challenges are the subject of these technologies' commercial implementations. Numerous experiments were conducted identifying health gains by the usage of neural networks in diagnostic and prognostic clinical functions, vital attention and fundamental examination of survival.

Neural networks are flexible, modified by repetitive iterations and initial preparation. More knowledge about the data is provided by successive iterations. In certain instances, the learning model focuses on weighing the input sources. Each node is weighed according to the value of each of its ancestors when processing.

As the iterated data educates the input nodes, which activates the layers of hidden node units and this in turn results in the output node units, flow information patterns into the neural network. This is considered as a network architecture with feedback. Not every node device fires as per the node receives entries in the left (previous layer) node units and the weights of the links they are connected which multiply the entries. All nodes are then added to all the inputs from the previous level and if the number exceeds a certain threshold amount, the node will activate all the nodes to which it has been linked (on its right). When a neural network is properly equipped, a performance pattern may be predicted when new patterns are introduced. [1]

LITERATURE REVIEW

A. Predictive Models for Early Detection of Diseases.

In several fields of medicine, predictive models were used as an aid for decision-making. This involves breast cancer estimation, type II diabetes, lung illness and the number of chronically ill adult deaths. Intensive treatments and attention (e.g., assessment and counselling) are referred to a high-risk patient to avoid future illnesses according to statistical models.

These models will decrease mortality and increase the quality of life of patients with high risk, controlling costs and complications for patients with low risks. Predictive models have especially become vital resources for helping to make medical decisions which can support health workers and patients alike. Precise prediction of mental health conditions will potentially contribute to more prompt detection and care of mental health diseases of physicians that recognize high-risk patients at an early level.

The prediction of disease may be considered a categorization issue in which several well-established classifications in this field are actively used. Researchers tried to predict the magnitude of depression to better individualize care.

In 18, gender, ICD-9 codes, conditions of diseases and medicinal ingredients, and average number of visits are the characteristics used for supervised learning. A LASSO model for logistic regression is used to forecast possible depression on feature vectors. The model is better at identifying low-risk patients with a specificity of 90%, and 25% sensitivity 12 months before diagnosis and 50% sensitivity were seen in patients with higher risk of depression.

B. Data Representation of Electronic Health Data

Electronic health data are heterogeneous and cannot be represented easily in a coherent room. The foundation for further progress in analytics and modelling is therefore an appropriate representation of such results, Poor depiction. Data without essential details can adversely affect predictive data type. Frequency (or absence) are usually used to reflect the categorical characteristics of an Example, where a binary variable codes presence or absence. Diabetes diagnosis is expected in data, medicines and based on past diagnoses orders for procedure and laboratory examinations from the medical histories of patients. A vector function is built on the basis of each patient data fitness. A characteristic value is developed to add all incidents that arise in the same function the default time window for frequency faces categorical. In the early predictive model ICD-9 codes are translated to depression diagnosis binary characteristics where 1 shows the function is present in the medical record of the patient, and 0 else. However, the characteristics derived from the aforementioned works temporary case orders were overlooked and may provide valuable data for the early diagnosis of a disease. So, we suggest generating a vector representation function for each feature Patient of frequency knowledge both widely used and time orders of health cases to boost the early-stage accuracy of identification using statistical model's vectors feature.

METHODOLOGY

The analytics in the current proof of concept are predictor in nature, enabling the graphical formation of neural networks and utilizing the appropriate model to forecast new data values.

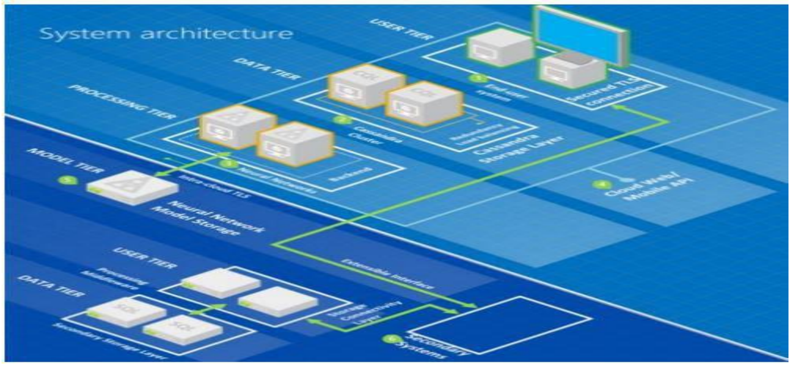

Fig. 1. System flow architecture

THE USER TIER

It manages the data sets and provides researchers with a mechanism through which they can run views/analytics on their choice of data sets. It is a data level interface that secures and integral storage of data sets. Cassandra, a grouped database architecture, supports functionality such as strong scalability, fault tolerance and multi-datacenter replication [4] in cloud storage models in this case. Connection protection is important to avoid data leakage when transmitting in and out of cloud storage.

PROCESSING TIER

The real analysis is conducted throughout the layer. The data are retrieved from the above layer for training and the model is integrated in the data collection depending on the user level parameters. It stores the model to the final level, called the model stage, and for predictive queries previously trained models are collected from that layer. The weights of an existing model are often loaded to facilitate more training on an older model. Another option may be done.

MODEL TIER

The model stage concerns the informative storage of the neural network artefacts and the masses of the neurons involved. It is well aligned with the processing level since the neural network architectures have been lacking universal norms.

SECONDARY SYSTEMS

The framework also provides the interface of data and knowledge import/export capacity for third-party applications. However, this has so far been limited to a hypothetical level since there are currently no such open structures.

Predictive and mathematical operations on the databases, which researchers as well as other health professionals may use, may be conducted for high-level visualization purposes. Geotagged health records, for example, which are regularly aggregated by area may represent all current patterns, such as the influx of a certain infectious disease, contaminants from different sources, etc. This will encourage the authorities to take disciplinary decisions in due course, or even precautionary steps when they arise. The gains may be enhanced by thorough alignment and broad scale framework deployment of secondary monitoring structures. [3] The report was released.

RESULTS & EVALUATION

Dataset

The initial owners of the dataset are, Diabetes and Digestive and Renal Diseases National Institute. Based on the diagnostic measures, the goal is to predict whether a patient is diabetic or not while putting several limitations on the key database. The dataset contains all reports of women aged 21 years or older who have Pima Indian ancestry [9].

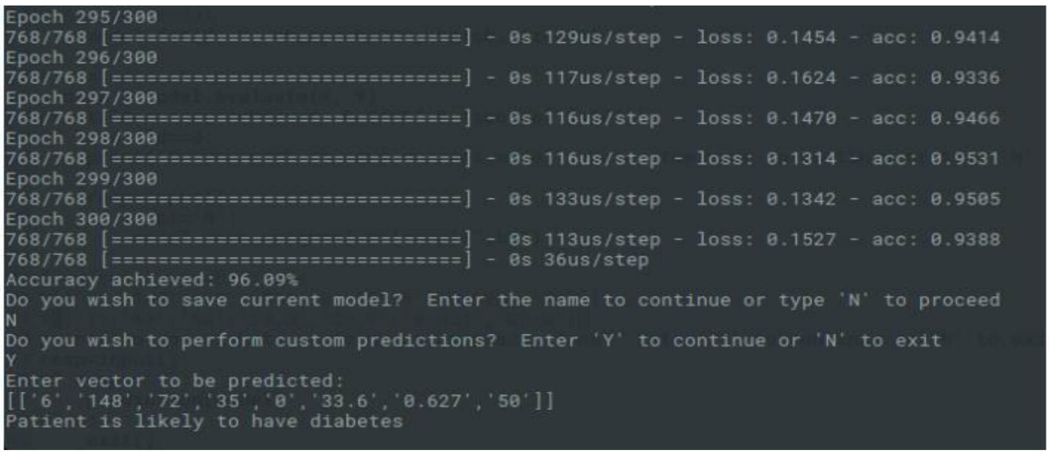

Fig. 4. Console Output

Fig. 4. Is shows a terminal platform. The neural network of the experimental systems is 300 epochs long and 96.09 percent accuracy. The vector containing the 8 input parameters is the input provided for the running programme [6, 148, 72, 35, 0, 33.6, 0.627, 50]. The result is "Yes."

Layers			**Epochs**	**Accuracy (%)**
First	**Second**	**Third**		
12	8	1	100	80.54
12	8	1	200	83.02
12	8	1	300	84.37
24	16	1	100	85.29
24	16	1	200	86.07
24	16	1	300	87.50
48	32	1	100	88.11
48	32	1	200	93.28
48	32	1	300	96.09

Fig. 5. is indeed a table that depicts the results of changing the node number present in the neural network in 3layers. As seen above, the exactness of the predictions produced by the increased times (a series of passes) and number of nodes in the layers are defined.

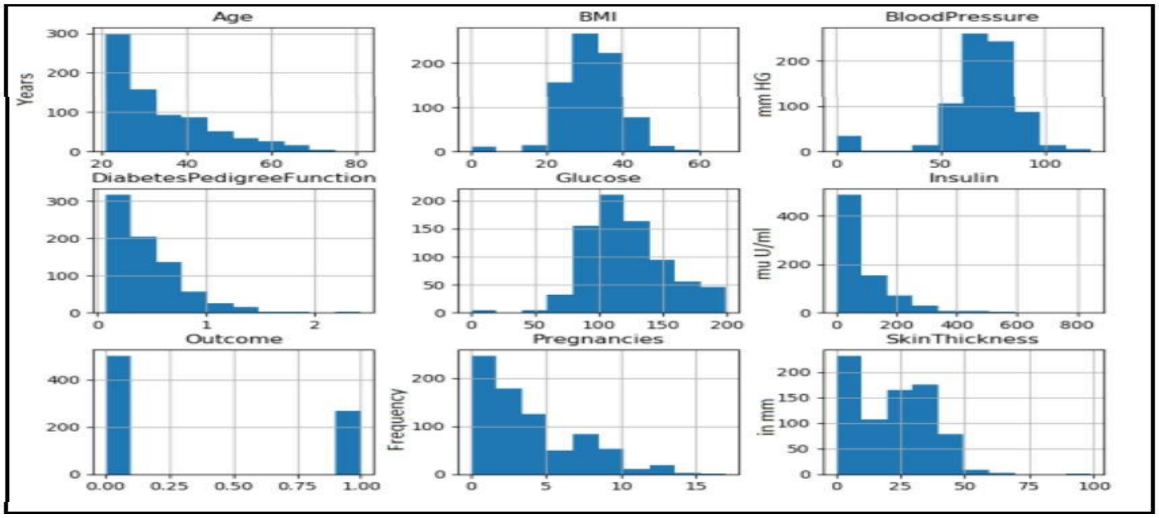

Fig. 6. Visualization 1: Histograms

The continuous data from the dataset in the ordered columns is shown in Fig. 6. In - group, the Y-axis reflects the frequency of objects. The properties and meaning distribution of these nine parameters is shown in the nine histograms in Fig. 6. The consumer can view the data set quickly to appreciate how the values and the meaning of the data set appear.

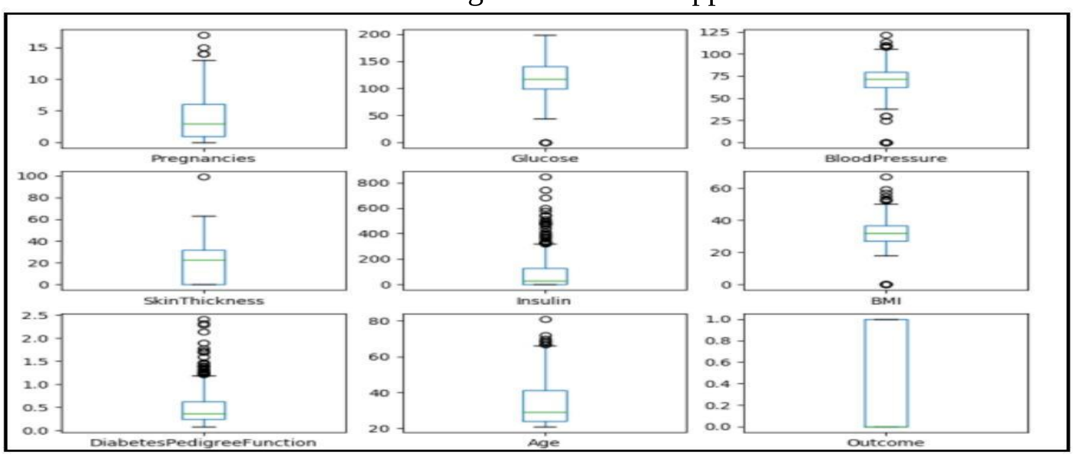

Fig. 7. Box Plot Analysis

Fig. 7. describes the data distribution review. Each of the 9 box tracks shows the place of the vector and spreads details on the symmetry and skews of the details in one vision.

The nine diagrams of the box show the properties and character of the values for the nine parameters mentioned above. For any parameter, the consumer can conveniently see the minimum, limit and range of values.

The consumer can conveniently view outliers using Fig. 6 and Fig. 7.

CONCLUSION

This model achieves its ultimate objective of offering an optimized medium for use of predictive as well as general analytics of medical data sets. It is currently reviewed with many datasets to expand its usage cases further to make sure that it is not linked to a specific dataset.

A framework introduced with this paradigm also has constraints on the supply of genuine, accessible medical data, which imposes limits on the precision of modern AI technologies. Present trends are most likely due to computing power breakthroughs to increase training speed; the reliability of advanced equipment is often related.

In relation to these, moreover, no basic developments, such as their theoretical mathematical structures, have been made within the structure of neural networks themselves. Instead, due to improved source data generation/collection, digitization of current data and other causes, there was sudden inflow of accessible data. In this regard, however, there is a clear potential for progress, especially in the healthcare sector in developed countries.

Any legal grey issues need to be addressed prior to large-scale implementation of such an architecture. Such programmers or suppliers have little certainty over the property of patient data; even though patients have full privileges in their systems. These problems, related to questions about the privacy of patients because of the sensitivity of data, have to be addressed before it can be extended to enable deep predictive analysis in an area or generically.

REFERENCES

[1]. "Early Detection of Diseases Using Electronic Health Records Data and Covariance-regularized Linear Discriminant Analysis" by Jiang Bian, Laura E Barnes, Guanling Chen, Haoyi Xiong, IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 2017

[2]. "A Beginner's Guide to Understanding Convolutional Neural Networks" by Adit Deshpande, DZone, 2017.

[3]. "Impact of Predicting Health Care Utilization Via Web Search Behavior: A Data-Driven Analysis" by Vibhu Agarwal, Liangliang Zhang, Josh Zhu, Shiyuan Fang, Tim Cheng, Chloe Hong, and Nigam H Shah, 2016.

[4]. "The NoSQL Principles and Basic Application of Cassandra Model" by Guoxi Wang; Jianfeng Tang, IEEE, 2012.

[5]. "Back-Propagation vs Particle Swarm Optimization Algorithm: which Algorithm is better to adjust the Synaptic Weights of a Feed-Forward ANN?" by Beatriz A. Garro, Humberto Sossa, Roberto A. Vázquez, Computer Research Center (CIC), International Journal of Artificial Intelligence, 2011.

[6]. "Applications of Artificial Neural Networks in Medical Science" by Jignesh Patel, Ramesh K Goyal, ResearchGate, 2007.

[7]. "Prediction and cross-validation of neural networks versus logistic regression: using hepatic disorders as an example" by M. Duh, A. Walker, K. Kronlund, American journal of Epidemiology, 1998.

[8]. "Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction" by William G. Baxt, MD, Annals of Internal Medicine, 1991.

[9]. "Pima Indians Diabetes Data Set" by National Institute of Diabetes and Digestive and Kidney Diseases. https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes.

[10]. National eHealth Authority (NeHA) Executive Summary.