Artificial Neural Network Model for Identifying Early Readmission of Diabetic Patients

Aswathi Anand P, Maya L Pai

Abstract: ***Early re-admission of patients increases the cost of health care and it highly influences the reputation of the hospital. Finding readmission in primary stage, allows the hospitals to give special care for those patients, and then can reduce the rate of readmission. In this work develop a new model using deep learning. It is the comparison method between machine learning and deep learning. Usually Logistic regression is used for all kind of prediction. But according to this data artificial neural network model in deep learning give promising result than logistic regression.***

Index Terms: Logistic Regression, Artificial neural network, Multilayer perception.

I. INTRODUCTION

In the United States, almost 30.2 million people suffering diabetes decease. The figure represents between 27.9 and 32.7 percent of the total population. Diabetes is the large problem in healthcare world. It is lifestyle decease. Diabetes is considered as the dangerous and chronic diseases due to increase in sugar level in blood. Lots of complications may occur if diabetes remains unidentified and untreated. The major identifying process is visiting a patient to a diagnostic center and consultation with a doctor. Prevent diabetes by modifying lifestyles and intervention of new drugs has convincingly shown that these measures can.

Diabetic patient's hospital readmission is an important problem in health care industry. BeataStrack et.al (2014) studied about this topic and they developed new model for predicting hospital readmission using logistic regression. They mainly focus on Hba1c result for their analysis [1].

Several publications are available related to machine learning and neural network Sushant Ramesh et.al (2017) says diabetic prediction using deep learning, with help of prima Indian diabetic data set. Implementation using tensor flow frame work. Calculations done by the help of Random Forest Algorithm. Along with they are using deep neural network model for predicting the diabetic patients [3]. Various published papers studied about patients hospital admissions. Ahmad Hammoudeh et.al (2018) uses cnn model for classification purposes. The frame work they are used are Tensor flow and scikit learn for neural network implementation. Combination of convolutional neural network model and data engineering give promising result for medical fields (4).

Bhuvan M S et.al (2016) studied about the medical data analysis through associate rule mining and feature importance mining along with they studied about classification algorithms related to these data sets. Based on the study finding the risk factors effects the readmission based on the diabetic data (2).

I. DATA AND METHADOLOGY

A. Data

The dataset taken from UCI Machine Learning Repository, it is a big dataset with over more than one lack one hundred rows and 55 column that is attributes, extracted from patient's health records from United States over 150 hospitals from 1999-2008, each row considered as an encounter record of each patient. Encounter Record includes encounter_id, admission_type_id_description, discharge_disposition_id description, admission source id_description, time_in_hospital , num_lab_procedures, and diagnosis result in 3 stages, result of A1C Test and also 23 diabetic medicines. 23 medicines include common diabetic medicines like metformin, glimepirideetc. those are basic medicine. In case of serious diabetic cases doctors may provide combinations of medicines most important part of this data set is some features contains subdivisions. In case of Discharge description id and admission source id descriptions and etc.

Medical records of each patient included more than 50 risk factors and a one column contains class label represents whether a patient was re-admitted or not. Label contains 3 classes readmission within thirty day period, after thirty days or patient was never readmitted. From the data the readmission rate is as follows 35 percentage of patients were readmitted within one month54 percentage patients were not readmitted.

Diagnosis is the method of a patient is diabetic or not. HbA1c test is the best method for diabetic detection. It is Hemoglobin A1c test it explains the average blood sugar level of a patient over past two to three months. It is also known as HbA1c Glycated Hemoglobin test. Normal level of HbA1c test is 4% to 5.6% and 5.7% to 6.4% says that those people have high chance of getting a diabetic patient. 6.5% or more shows they have diabetics. The higher A1c result shows the risk having complicated related to diabetic patients.

Table 1 shows missing values in the data. Data contains one lack one thousand and fifty six encounter. Race, medical specialty and payer code contains high percentage of missing values. Removing these attributes from the data

Revised Manuscript Received on April 07 , 2019.

Aswathi Anand P, Department of Computer Science & IT , Amrita School of Arts and Science Kochi, Amrita VishvaVidyaPeetham, India

Dr Maya L Pai, Department of Computer Science & IT , Amrita School of Arts and Science Kochi, Amrita VishvaVidyaPeetham, India.

in the preprocessing stage. Payer code is the one of the feature in the data it is not an important.

Table 1. Attributes containing missing values

Attribute	Missing values
Race	2.23%
Weight	96.85%
Payer Code	40.5%
Medical Specialty	50.1%
Diag_1	.02%
Diag_2	.35%
Diag_3	1%

Gender having less no of missing values. Data contains three diagnosis result. If more than three features are missing in one encounter it is considered as bad data, so remove an encounter if 3 diagnosis are empty.

Complication	**ICD-9-codes**
Diabetes and specific complications included	250
Microvascular	
Ophthalmologic complications	
Cataract	366
Partially sighted and blindness	369
Neurological complications	
Peripheral autonomic neuropathy	337
Myasthenia syndrome	358
Peripheral vascular disorders	443
Gangrene	785
Amputation	895-897
Skin and strengthening complications	
Chronic Neurophatical ulcer	707
Cellulites	682
Diseases of the genitourinary	
Nephropathy, nephrotic syndrome	581
Nephritis	583
Chronical kidney failure	585
Proteinuria	791
Urinary infections	599
Macrovascular	
Hypertension	401-404
Coronary heart disease	410-414
Stroke	430-438
Congestive Heart Failure	428-429

Table 2: ICD9 code and complications

ICD9 codes are like medical coding. Used for representing disease in medical industry. Codes include values from 001 to 999. Each code represents each disease. 250.00 to 250.99 represents icd9 code of diabetic mellitus. Table 2 represents icd9 code and corresponding disease.

B. System Architecture

Architecture of proposed model shown in fig 1. Data contains lots of missing values, unimportant data and noises so preprocessing is the important step. In preprocessing step data normalize using probability distribution. Data split in to test set and train set. Here 80% data used for training and 20% data used for testing. Model building is using mlp neural network and at last comparing the model accuracy using existing and proposed method. For data validation tenfoldcross validation methods are used.

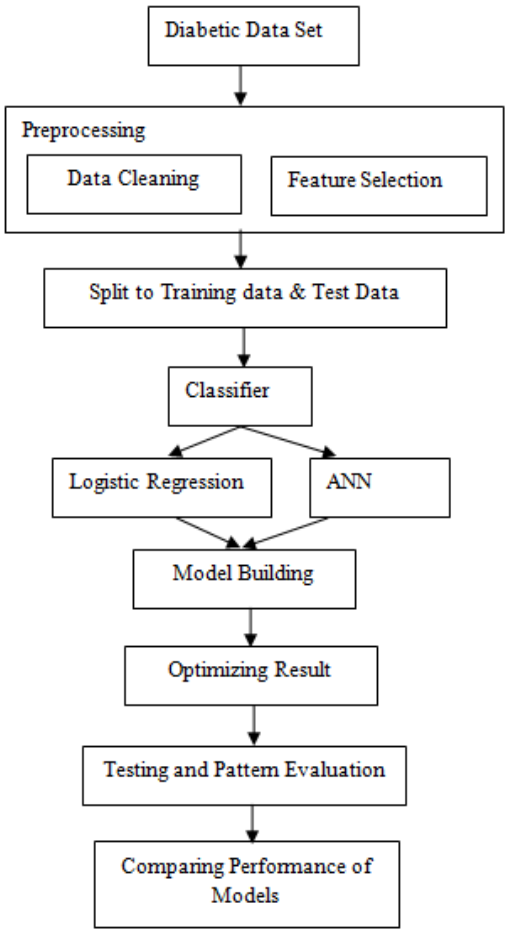

Fig. 1, System Architecture

C. Logistic Regression

Logistic regression is popularly known as machine learning classification method for prediction. Logistic regression contains a dependent variable it is also known as target variable. Independent variable is also known as Predictor variable. Regression analysis is a very important tool for analyzing and modeling data.

$$Y = b_0 + b_1 \times X + \text{E} \quad (1)$$

Equation 1 shows the basic equation of the linear regression. All other equations are driven from this basic equation. It is popularly known as the equation of the straight line. Adding sigmoid function to the linear regression we can develop new equation for the logistic regression.

$$0 \leq h_\theta(x) \leq 1 \quad (2)$$

$$h\theta(x) = g(\theta^T x) \quad (3)$$

$$g(z) = \frac{1}{1+e^{-z}} \quad (4)$$

Decision Boundary

$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2) \quad (5)$$

Predict (y=1) if -3 $+x_1 + x_2 \geq 0$ (6)

Logistic regression Cost Function

$$j(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h\theta(x^i), (y^i)) \quad (7)$$

$$Cost(h\theta(x), y) = \begin{Bmatrix} -\log(h\theta(x)) & if\ y = 1 \\ -\log(1 - h\theta(x)) & if\ y = 0 \end{Bmatrix} \quad (8)$$

D. MLP- Artificial neural networks (ANN)

Artificial neural networks (ANN) is computing systems. It developed by getting the idea from the structure of biological neural networks that that considered as human brain.Neural network itself is never considered as a perfect algorithm, but actually it is a framework helps many different Machine Learning algorithms to process together to run and solve complex data inputs. ANN contains artificial neurons. Artificial neurons means set of connected nodes or units, which mimics the neurons in a human brain. Each and every connection same as the synapses in brain, can transmits signals from one neurons to another neuron. When artificial neuron receives a signal it will process the signal and act accordingly.

Multilayer perceptron (MLP) is a single hidden layer. It is important and classical type of artificial neural network model. MLP contains 3 layers of nodes: first input layer, hidden layer and the last layer is output layer. In this each node is considered as neuron that uses a nonlinear activation function. If layers in MLP are connected, each nodes in each layer is connected with specific weight ωij. It uses back propagation algorithm. MLP algorithms mainly used to solve the classification prediction problem and also classification regression problem.

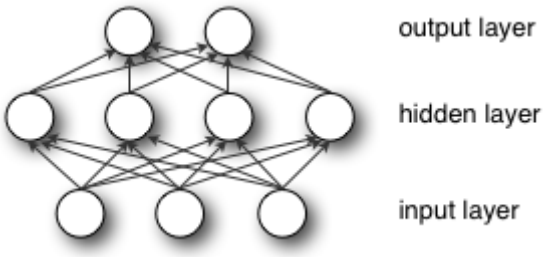

Fig. 2 Layers of neural network

Hidden layer of MLP is a function

$$f: R^D \rightarrow R^L \quad (9)$$

D is size of input, x & L are size of the output vector and $f(x)$ is the matrix notation:

$$f(x) = G(b^{(2)} + W^{(2)}\left(s\left(b^{(1)} + W^{(1)}x\right)\right)) \quad (10)$$

$b^1\ and\ b^2$ are bias vectors and $W^1 and\ W^2$ are weight.

G is the Activation Function

$$\varepsilon(n) = \frac{1}{2}\sum_j e_j^2(n) \quad (11)$$

Using Gradient Boosting

$$\Delta W_{ji}(n) = -n\frac{\partial\,\varepsilon(n)}{\partial v_j(n)} y_i(n) \quad (12)$$

y_i is the previous neurons output and v_j local induced field

$$-\frac{\partial\varepsilon(n)}{\partial v_j(n)} = e_j(n)\emptyset'\left(v_j(n)\right) \quad (13)$$

$\emptyset'$ is activation functions derivate.

$$\frac{-\partial\varepsilon(n)}{-\partial v_j(n)} = \emptyset'\left(v_j(n)\right)\sum_k -\frac{\partial\varepsilon(n)}{\partial v_k(n)} w_{kj}(n) \quad (14)$$

E. MODEL BUILDING

Machine Learning can deals with huge amount of data. But Deep Learning is very efficient for deals with those data. Build a new model using deep learning. Deep learning contains lots of algorithm for classification here using Multilayer perceptron Artificial neural network model for developing new model.

II. RESULT

Logistic Regression Result Contains Dataset details Accuracy and cross validation score from the analysis

Table 2 Result of Logistic Regression Model

Original dataset shape Counter	({0: 56476, 1: 5199})
New dataset shape Counter	({0: 56476, 1: 56476})
No. Observations	90361
Cross Validation Score:	60.70%
Dev Set score:	60.59%
Accuracy	61.62%
Current function value:	0.665612
Iterations:	35

Table 3 Result of MLP Neural Network model.
Table contains data set shape, Accuracy, Deviation score and f1 score of confusion matrix.

Table 3: Result of MLP Artificial neural network Model

Original dataset shape Counter	({0: 56476, 1: 5199})
New dataset shape Counter	({0: 56476, 1: 56476})
No. Observations	90361
Dev Set score:	81.63%
Accuracy	82.86%
F1 score	82%

Table 4 :Result Comparison between Logistic Regression and MLP Neural network

Model	Accuracy	Precision	Recall
Logistic Regression	61.62	62	55
MLP Neural Network	**82.86**	**82**	**80**

Table 4 shows the comparison between two methods. It compares Accuracy and Precision recall scores.

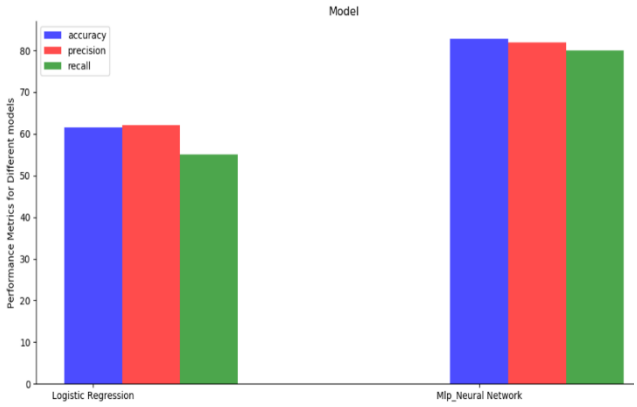

Fig 3: Graphical representation of performance matrix

III. CONCLUSION

Proposed model gives promising result than the existing model. Here we are predicting early readmission of diabetic patients using machine learning and deep learning classification algorithms.Multilayer perceptron Artificial neural network model gives better accuracy and lower error rate than traditional predicting method logistic regression.

REFERENCES

1. Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., &Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. BioMed research international, 2014.
2. Bhuvan, M. S., Kumar, A., Zafar, A., & Kishore, V. (2016). Identifying diabetic patients with high risk of readmission. arXiv preprint arXiv:1602.04257.
3. Hammoudeh, A., Al-Naymat, G., Ghannam, I., &Obied, N. (2018). Predicting Hospital Readmission among Diabetics using Deep Learning. *Procedia Computer Science*, *141*, 484-489.
4. Ramesh, S., Caytiles, R. D., &Iyengar, N. C. S. (2017). A Deep Learning Approach to Identify Diabetes. *Advanced Science and Technology Letter*, *145*, 44-49.
5. Aswathi Anand P &Dr Maya L Pai. (2019) Hospital readmission prediction based on diabetic patient records using classification algorithms (Unpublished)
6. Safavian, S. R., &Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), 660-674.
7. Chan, J. C. W., &Paelinckx, D. (2008). Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery. *Remote Sensing of Environment*, *112*(6), 2999-3011.
8. Pearce, J., & Ferrier, S. (2000). Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological modelling*, *133*(3), 225-245.
9. Mogensen, C. E., & Christensen, C. K. (1984). Predicting diabetic nephropathy in insulin-dependent patients. *New England Journal of Medicine*, *311*(2), 89-93.
10. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., &Kripalani, S. (2011). Risk prediction models for hospital readmission: a systematic review. *Jama*, *306*(15), 1688-1698.
11. Hasan, O., Meltzer, D. O., Shaykevich, S. A., Bell, C. M., Kaboli, P. J., Auerbach, A. D., ... &Schnipper, J. L. (2010). Hospital readmission in general medicine patients: a prediction model. *Journal of general internal medicine*, *25*(3), 211-219.
12. Specht, D. F. (1991). A general regression neural network. *IEEE transactions on neural networks*, *2*(6), 568-576.
13. Frosini, A., Gori, M., &Priami, P. (1996). A neural network-based model for paper currency recognition and verification. *IEEE transactions on neural networks*, *7*(6), 1482-1490.
14. Haykin, S. (1994). *Neural networks* (Vol. 2). New York: Prentice hall.
15. Hsu, K. L., Gupta, H. V., &Sorooshian, S. (1995). Artificial neural network modeling of the rainfall-runoff process. *Water resources research*, *31*(10), 2517-2530.
16. Sarle, W. S. (1994). Neural networks and statistical models.
17. Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. MIT press.
18. Karlik, B., &Olgac, A. V. (2011). Performance analysis of various activation functions in generalized MLP architectures of neural networks. *International Journal of Artificial Intelligence and Expert Systems*, *1*(4), 111-122.
19. Zare, M., Pourghasemi, H. R., Vafakhah, M., & Pradhan, B. (2013). Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arabian Journal of Geosciences*, *6*(8), 2873-2888.
20. https://archive.ics.uci.edu/ml/

AUTHORS PROFILE

Aswathi Anand P: Mphil Scholar, Department of computer science & it, Amrita school of arts and Science Kochi, Amrita VishvaVidyaPeetham, India.Completed Bachelor of Technology from Cochin University of Science and Technology and Master of Engineering from Anna University Chennai.

Dr. Maya L Pai:Asst. Professor (Sr.Gr.), Research Guide, (Chairperson, Board of Studies in Computer Science and IT)., Department of computer science & it, Amrita school of arts and Science Kochi, Amrita VishvaVidyaPeetham, India.ObtainedPh.D, M. Sc. (Applied Mathematics), B. Ed., M.Phil., SET.34 years academic experience, including 8 years in research.