Advanced Science and Technology Letters
Vol.145 (NGCIT 2017), pp.44-49
http://dx.doi.org/10.14257/astl.2017.145.09

A Deep Learning Approach to Identify Diabetes

Sushant Ramesh, Ronnie D. Caytiles* and N.Ch.S.N Iyengar**

School of Computer Science and Engineering VIT University, Vellore-632014
*Multimedia Engineering department, Hannam University, Daejeon, Korea
**Dept. of Information Technology, Sreenidhi Institute of Science and Technology, Hyderabad, India
sushant.ramesh2013@vit.ac.in, rdcaytiles@gmail.com, srimannarayanach@sreenidhi.edu.in

Abstract. The primary objective of this paper is to predict onset of using deep learning and also as to predict the risk factor and severity of diabetics. The methods are implemented on in a conditional data set of diabetes. The model involves deep learning, in the form of a deep neural network through which we are able to apply predictive analytics on said diabetes data set and obtain optimal results. At the end, a comparative study is done between the implementation of this model on type 1 diabetes mellitus, Pima Indians diabetes and the Rough set theory model. The results add value to additional reports because the number of studies done on diabetes using a deep learning model is few to none. This will help to predict diabetes with much more precision as shown by the results obtained.

Keywords: Deep Learning, Diabetes Mellitus, Restricted Boltzmann Machine.

1 Introduction

Diabetes mellitus, commonly known as diabetes, is group of metabolic diseases which marks high blood sugar levels over a prolonged period. As per the survey done by International Diabetes Federation estimated in 2040 to rise to 642 million diabetic individuals. Type 1 diabetes is a disorder in which the pancreas can no longer produce insulin. According to the WHO (World Health Organization) the number of children having type 1 diabetes is very high as mentioned in the motivation. Hence it can be said that Diabetes is a serious chronic disease. Doctors usually take patients' blood samples and check the sugar concentration in their blood in order to diagnose them with diabetes. This is a highly time- consuming process and there are many other features which need to be reviewed while attempting to detect whether a patient is diabetic or not. These other factors are: insulin, body mass index, blood pressure and age. Even family history can play a vital role in a person's diabetic status. If a patient's ancestors show a presence of diabetes, then there is a high chance of the patient exhibiting symptoms of diabetes as well. Presently, there is no non-intrusive technique to detect if a person has type 1 diabetes and how severe the diabetes will

ISSN: 2287-1233 ASTL

affect him/her. Hence the need arises for some sort of efficient application to predict the onset of type 1 diabetes in patients for a quicker diagnosis which can lead to quicker treatment.

Deep learning involves working on a model quite similar to that of the human brain. It has the capability to decode complex problems much like that of the human brain. Deep neural networks can handle massive sets of data (as done in this work), grasp complex tasks and handle data with small manual inputs. It is a learning method with a deep architecture and algorithms which can learn features that have no labels i.e. unsupervised data. This drives the idea that a heavy amount of unsupervised learning is required. The ability of deep neural networks to follow the same decision process as that of the human brain, along with the methodical learning algorithms that can ensure this ability, make it fast in terms of processing speed and it can also handle data of high dimensions better than simple machine learning algorithms. RStudio and Tensor Flow used in this work for investigation.

2 Related Work

There is so much literature work available and the most important work needed is referred here with. Asma Shaheen Khan et.al., [1], Tawfik Saeed Zeki, et.al., [2] Rahman Ali, et.al., [3] studied different approaches for predicting of diabetics. T.P. Kamble et.al., [6] proposed a system where Deep learning based Restricted Boltzmann machine approach is used to detect whether patient is diabetic or not as Restricted Boltzmann machine is popular for classification and recognition purpose. To detect either patient is having type 1 or type 2 diabetes decision tree technique used. Riccardo Miotto et.al., [6] presented a novel unsupervised deep feature learning method to derive a general-purpose patient representation from EHR data that facilitates clinical predictive modelling. Jack W. Smith, et.al., [7] tested the ability of an early neural network model, ADAP, to forecast the onset of diabetes mellitus in a high risk population of Pima Indians.

3 Framework

The deep learning framework used in this work is TensorFlow that run on multiple CPUs and GPUs (with optional CUDA extensions for general-purpose computing on graphics processing units). TensorFlow is available on 64-bit Linux, macOS, and mobile computing platforms including Android and iOS.

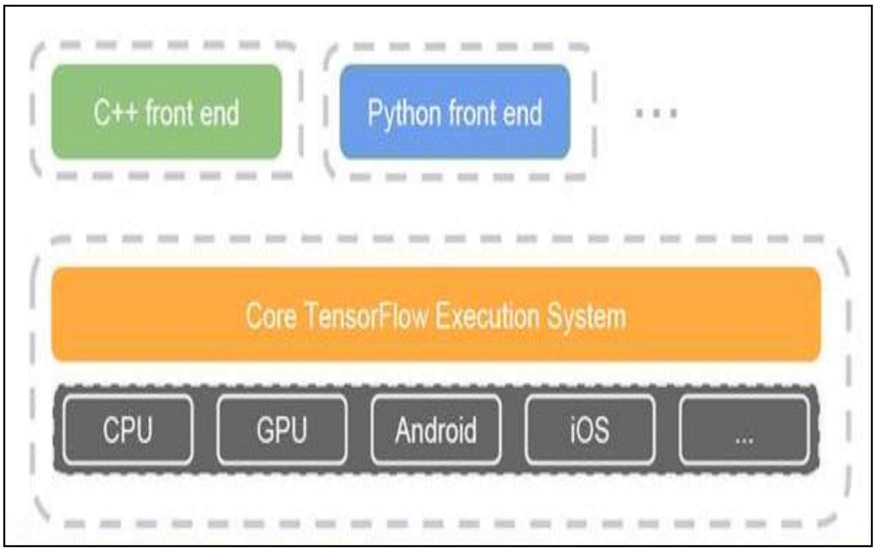

Fig. 1. Deep Learning Framework – TensorFlow

The deep learning model chosen for this endeavour is a Recurrent Deep Neural Network (RNN). The most widely used neural network is a feed-forward neural network (MLP). However, that has not been chosen for this workt specifically because even when both neural networks are well trained, a RNN uses more information than a MLP. Also, while a MLP can approximate any function to an arbitrary precision, the accuracy obtained by a RNN is much higher due to the presence of a layer that considers inputs from different time points i.e. the recurrent formation of the neural network.

4 Methodology Adopted

The data sets considered for this project were obtained from the online Machine Learning UCI repository. The links of the data sets are given.

T1DM [4]: https://archive.ics.uci.edu/ml/datasets/diabetes

Pima Indians Diabetes [5]: https://archive.ics.uci.edu/ml/datasets/pima+indians+diabetes

The proposed system follows the steps as shown in Figure 2, which is the process flow diagram.

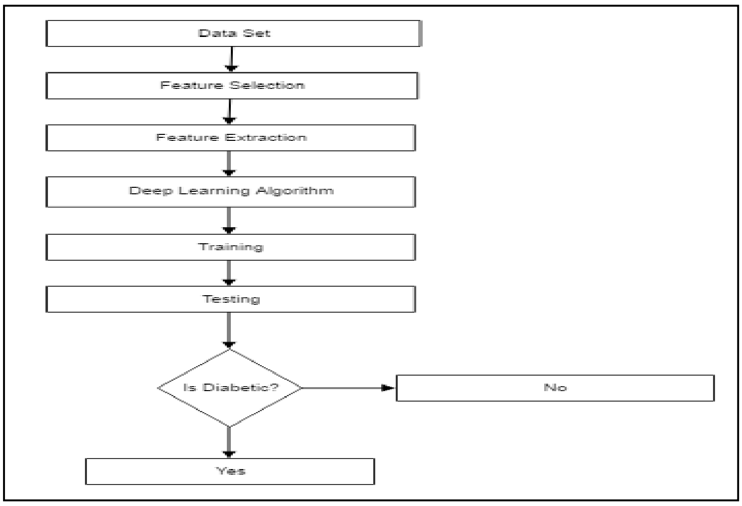

Fig. 2. Process flow diagram for the deep learning model

In the data pre-processing step, the features are selected from the file on the basis of their uniqueness. The attributes, in descending order of their importance are Glucose, BMI, Age, Pregnancies, Diabetes Pedigree Function, Blood Pressure, Skin Thickness and Insulin. The graphical representation of this order as calculated on RStudio using the Random Forest Algorithm is presented in Figures 3

Table 1. Attributes arranged by priority with assigned weights for Pima Indians Diabetes

Attribute	Glucose	Insulin	BMI	Pregnancies	D. P.F	B.P	S.T	Age
Assigned Weight	0.8	0. 7	0.7	0.615	0.61	0.59	0.55	0.47

Next, the data set is normalized with respect to these features through min-max normalization in order to get an input vector range between 0 and 1, and to avoid computation complexity. And next the dataset is divided into training dataset and test dataset such as 80% of training dataset and 20% of test dataset. Upper bound (UB) =1 and lower bound (LB) =0.

5 Experimental Setup and Simulation Results

The deep neural network was trained and tested in TensorFlow as shown in Figure 9. On constructing and training a deep neural network i.e. an RBM, and simulating the data sets [4][5], the following results were obtained:

RESULTS FOR PIMA INDIANS DIABETES:

Initial Result obtained after applying logistic regression on data generated from restricted Boltzmann machine:-

Correctly Classified Instances: 622 (81%)
Incorrectly Classified Instances: 146 (19%)
Kappa statistic: 0.5297
Mean absolute error: 0.211
Root mean squared error: 0.304
Relative absolute error: 80.1856 %
Root relative squared error: 91.4408 %
Coverage of cases (0.95 level): 97.15 %
Mean rel. region size (0.95 level): 89.35 %
Total Number of Instances: 768

Table 2. Showing the correctly and incorrectly classified instances using RBM.

		Predicted Classes	
		0 (Diabetic)	1 (Non-Diabetic)
Actual Classes	0 (Diabetic)	421	79
	1 (Non-Diabetic)	67	201

6 Comparative Study

Validation with diabetes prediction using Rough Set Theory[3]:

Table 3. Comparative study (validation) with respect to Rough Set Theory.

	Deep Learning (Type 1 Diabetes Mellitus)	Deep Learning (Pima Indians' Diabetes)	Rough Set Theory
Precision for Non Diabetic i.e. 0	0.8064	0.75	0.7977
Precision for Diabetic i.e. 1	0.7777	0.842	0.7781
Recall for Non Diabetic i.e. 0	0.9259	0.9066	0.9133
Recall for Diabetic i.e. 1	0.53846	0.55138	0.54174
Error Rate	0.20	0.145	0.185

7 Conclusion and Future Work

The deep neural network was successfully trained, tested and implemented on the data sets[4][5]. The results obtained were above satisfactory and can be further improved by increasing the size of the data set by adding to it the information gathered by incoming patients in a hospital or in a network of hospitals.

References

1. Shaheen, Asma. Intelligent decision support system in diabetic ehealth care. Diss. Blekinge Institute of Technology, 2009.
2. Zeki, Tawfik Saeed, et al. "An expert system for diabetes diagnosis." American Academic & Scholarly Research Journal 4.5 (2012): 1.
3. Ali, Rahman, et al. "H2RM: A hybrid rough set reasoning model for prediction and management of diabetes mellitus." Sensors 15.7 (2015): 15921-15951.
4. Michael Kahn (1994). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Washington University.
5. National Institute of Diabetes and Digestive and Kidney Diseases (1990). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. John Hopkins University.
6. Miotto, Riccardo, et al. "Deep patient: An unsupervised representation to predict the future of patients from the electronic health records." Scientific reports 6 (2016).
7. Smith, Jack W., et al. "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus." Proceedings of the Annual Symposium on Computer Application in Medical Care. American Medical Informatics Association, 1988.

