A SURVEY ON DIFFERENT ALGORITHMS FOR DIABETIC PREDICTION

B. Prabha[1], P. Krishna Kumar[2]
[1]PG Scholar, [2]Professor
[1,2]Department Of CSE, VV College Of Engineering, Tisaiyanvilai-628656, India.

ABSTRACT

The large-scale parallel computation that became available on the new generation of GPUs and on cloud-based services can be exploited for use in Healthcare Data analysis. Furthermore, computation workstations suited for Deep Learning are usually equipped with multiple GPUs allowing for workload distribution among multiple GPUs for larger data sets while exploiting parallelism in each GPU. In this paper we utilize distributed and parallel computation techniques to efficiently analyze healthcare data using Deep Learning techniques. We demonstrate the scalability and computational benefits of this approach with a case study of longitudinal assessment of approximately 150,000 type 2 diabetic patients. Diabetes Mellitus (T2DM) is the fourth case of mortality worldwide with rising prevalence. T2DM leads to adverse events such as acute myocardial infarction, major amputations and avoidable hospitalizations. This study aims to establish a relation between laboratory and medical assessment variables with the occurrence of the aforementioned adverse events and its prediction using Machine Learning Techniques.

Index Terms - Deep Learning, GPU Computing, Healthcare.

I. INTRODUCTION

Deep Learning has been quickly adapted to many application domains for data-driven discoveries. Different from both Big Data processing systems such as Hadoop and Spark, and statistical analysis tools such as R, deep learning is a relatively new methodology that emerged from Data Mining that is used to derive insights from data. However, successful application of deep learning to a practical problem requires significant knowledge about the data, the algorithm, programming efforts and computational resources. Usually, the development of these techniques requires exhaustive computing making the development and scalability of machine learning models impractical. One option to overcome this difficulty is Tensor flow, which allows for numerical computation using data flow graphs, with an architecture that makes computations on both CPU and GPU feasible. The developed models are cloud computing ready. Cloud computing provides an on-demand service of IT resources including computing capacities and database storage.

. Worldwide there are 415 million people living with diabetes and around 193 million people have undiagnosed diabetes. T2DM accounts for more than 90% of patients with diabetes and leads to microvascular and macro-vascular complications that cause profound psychological and physical distress to both patients and caregivers, resulting in a huge burden on the health-care system . The range of complications arising from diabetes, due to the damaging nature of glucose molecules on the micro- and macro-vascular system includes: cardiovascular disease, coronary heart disease, blindness, nephropathy, peripheral neural disease, amputations, depression and erectile dysfunction.

Despite increasing knowledge regarding risk factors for type 2 diabetes and evidence for successful prevention programs, the incidence and prevalence of the disease continues to rise globally. In a population-based study in Spain, a prevalence of 13.8% was observed on a population over the age of 18, about half of those diabetes cases were not diagnosed . The prevalence of diagnosed diabetes in the Basque Country has been shown to be lower than the present at national level in Spain. The registered prevalence in the Basque Country was 9.12% in 2011, with an increasing pattern since 2008.

Patients with T2DM have a higher total mortality rate, as well as risk of microvascular and macro-vascular events. These complications reduce the quality of life of patients with T2DM, and lead to an increase in healthcare costs. Studies show that glycemic control is the most important feature to prevent the organ damage and other complications of T2DM. The decrease of blood glucose levels in patients with diabetes decreases the mortality and morbidity rates significantly. Additionally, it has been well established the association between blood pressure and lipid control with disease duration and the development and progression of diabetes related complications.

Several T2DM related complications have been studied through different classical machine learning, deep learning and data mining techniques. Identification of the risks factors associated with these complications is of great value to the clinical management of individuals with diabetes. Due to the high level of disability and incremental costs of the disease, it is necessary to investigate the main factors involved in the genesis of complications.

II. TYPES OF DIABETICS

2.1.Type 1 diabetes

Type 1 diabetes is an autoimmune disease that causes the insulin producing beta cells in the pancreas to be destroyed, preventing the body from being able to produce enough insulin to adequately regulate blood glucose levels. Type 1 diabetes may sometimes be referred to

as juvenile diabetes, however, this term is generally regarded as outdated as, whilst it is commonly diagnosed in children, the condition can develop at any age. Insulin dependent diabetes is another term that may sometimes be used to describe type 1 diabetes. Because type 1 diabetes causes the loss of insulin production, it therefore requires regular insulin administration either by injection or by insulin pump.

2.2.Type 2 diabetes

Type 2 diabetes mellitus is a metabolic disorder that results in hyperglycemia (high blood glucose levels) due to the body. Being ineffective at using the insulin it has produced; also known as insulin resistance and/or Being unable to produce enough insulin

Type 2 diabetes is characterized by the body being unable to metabolise glucose (a simple sugar). This leads to high levels of blood glucose which over time may damage the organs of the body. From this, it can be understood that for someone with diabetes something that is food for ordinary people can become a sort of metabolic poison. This is why people with diabetes are advised to avoid sources of dietary sugar. The good news is for very many people with type 2 diabetes this is all they have to do to stay well. If you can keep your blood sugar lower by avoiding dietary sugar, likely you will never need long-term medication. Type 2 diabetes was formerly known as non-insulin-dependent or adult-onset diabetes due to its occurrence mainly in people over 40. However, type 2 diabetes is now becoming more common in young adults, teens and children and accounts for roughly 90% of all diabetes cases worldwide.

2.3.Gestational diabetes

Gestational diabetes occurs when you have hyperglycemia (high blood glucose levels) during pregnancy. Gestational diabetes usually develops in the third trimester (between 24 and 28 weeks) and typically disappears after the baby is born. Women who develop gestational diabetes during pregnancy are more likely to develop type 2 diabetes later on in life. Finding out you have gestational diabetes can be very frightening as not only do you have to deal with all the emotions (the ups and the downs) and the questions that come with being pregnant, but also the uncertainty of this new-found condition. Fortunately, as with all types of diabetes, there are many well-informed health professionals to help answer your questions and to guide you through this very important time in your life. The more you know, the easier it is to accept and make the necessary changes for a successful and happy pregnancy.

2.4.Diabetes LADA

LADA stands for Latent Autoimmune Diabetes of Adulthood. LADA is a form of type 1 diabetes that develops later into adulthood. LADA tends to develop more slowly than type 1 diabetes in childhood and, because LADA can sometimes appear similar to type 2 diabetes,

doctors may mistakenly diagnose LADA as type 2 diabetes. The definition provided by Prof. David Leslie, Principle Investigator of Action LADA, is that in Europe: LADA is defined as initially non-insulin requiring diabetes diagnosed in people aged 30-50 years with antibodies to GAD - glutamic acid decarboxylase.

2.5.Double diabetes

Double diabetes is when someone with type 1 diabetes develops insulin resistance, the key feature of type 2 diabetes. Someone with double diabetes will always have type 1 diabetes present but the effects of insulin resistance can be reduced somewhat. The most common reason for developing insulin resistance is obesity and whilst type 1 diabetes is not itself brought on by obesity. People with type 1 diabetes are able to become obese and suffer from insulin resistance as much as anyone else

III. COMPARING VARIOUS ALGORITHMS FOR DIABETIC PREDICTION

Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem [2] by assuming the independence of predictors [2]. It calculates the posterior probability of being one class given the predictors. The Naïve Bayes algorithm is naturally incremental and can keep the model up-to-date by constantly updating the prior probability according to the incoming training instances.

Logistic regression[3] was conducted to assess whether the three variables significantly estimate the risk of having increased arterial stiffness among diabetic patients a. Three independent predictors for logistic regression

selected are age, HbA1c level and auc-PPG. Output of the model is a risk level, 1 for high risk and 0 for low risk.

Logistic regression analysis was conducted using the SPSS software program (SPSS 16). The first step in this analysis is to obtain the information regarding the intercorrelations among of the predictor variables (multicollinearity) as multicollinearity could lead to misleading and/or inaccurate result. This is done by tabulating the Pearson correlation coefficients of pairs of variable

To conduct this study four years of medical records from all patients diagnosed with Type 2 diabetes in a complete population were used. The information contained on the database was cleaned and homogenized in order to apply different machine learning techniques. The classical machine learning algorithms implemented techniques were Logistic Regression, Linear Discriminant Analysis[5], Quadratic Discriminant Analysis and Support Vector Machines[4]; the Deep Learning technique implemented was Recurrent Neural Network.

These techniques were selected given the nature of the employed data; the medical records contained in the database relate the diabetes patient control information over time and the

Table 1 Comparison table

ALGORITHM	CONTENT	ADVANTAGES	DISADVANTAGES
Neural Network Model[1]	The network architecture has three layers; the activation function is sigmoid in the two hidden layers. Neurons in different layers are totally connected and feed forward. The output layer has one neuron with linear activation function.	It can handle large amount of data sets and also has the ability to implicitly detect complex nonlinear relationships	It include its "black box" nature, greater computational burden, proneness to overfitting.
Naïve Bayes[2]	Naïve Bayes is a probabilistic classification algorithm based on Bayes' theorem by assuming the independence of predictors . It calculates the posterior probability of being one class given the predictors.	Naïve Bayes, is used to build candidate models.	These algorithms used to build the model are not good at handling highly skewed data.
Logistic regression[3]	Logistic regression was conducted to assess whether the three variables significantly estimate the risk of having increased arterial stiffness among diabetic patients	It is more robust: the independent variables don't have to be normally distributed, or have equal variance in each group	It is high reliance on a proper presentation of your data.
SVM[4]	Support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis.	SVM is an effective tool in high-dimensional spaces, which is particularly applicable to sentiment analysis .	There is no direct probabilistic interpretation for group membership.
LDA[5]	LDA is a simple model in both preparation and application. There is some interesting statistics behind how the model is setup and how the prediction equation is derived	Linear Discriminant Analysis (LDA) and its quadratic brother (QDA) were chosen due to the literature pointing as good feature selectors	The class containing more records is iteratively employed during training, allowing for an optimization of data usage.

IV. CONCLUSION

In this Study, T2DM has been a typical machine learning benchmark problem for many years. It has been dealt with using various machine learning algorithms. The problem an diagnosis system for predicting the diabetes based on different algorithm is presented. And it explains about the different ML approaches and their applications in diabetes used to analyze the dataset in different biomedical database.

REFERENCES

[1] A. Géron, (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. " O'Reilly Media, Inc.".

[2] Guo-li DU, et al. "Metabolic Risk Factors of Type 2 Diabetes Mellitus and Correlated Glycemic Control/Complications: A Cross-Sectional Study between Rural and Urban Uygur Residents in Xinjiang Uygur Autonomous Region," PloS one, vol. 11, 2016.

[3] Christian, BOMMER, et al. "The global economic burden of diabetes in adults aged 20–79 years: a cost-of-illness study," The lancet Diabetes & endocrinology, vol. 5, pp. 423-430, 2017.

[4] Alessandra M., et al. MANTOVANI, "Relationship between amputation and risk factors in individuals with diabetes mellitus: a study with Brazilian patients," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 11, pp. 47-50, 2017.

[5] Arturo Corbatón ANCHUELO, Rafael Cuervo PINTO, and Manuel Serrano. RÍOS, "La diabetes mellitus tipo 2 como enfermedad cardiovascular," Revista Española de Cardiología Suplementos, vol. 7, pp. 9A-22A, 2007.

[6] Jonathan E. SHAW, Richard A. SICREE, and Paul Z. ZIMMET, "Global estimates of the prevalence of diabetes for 2010 and 2030," Diabetes research and clinical practice, vol. 87, pp. 4-14, 2010.

[7] Colin D. MATHERS and Dejan LONCAR, "Projections of global mortality and burden of disease from 2002 to 2030," PLoS medicine, vol. 3, p. 442, 2006.

[8] Federico, SORIGUER, et al. "Prevalence of diabetes mellitus and impaired glucose regulation in Spain: the Diabetes Study," Diabetologia, vol. 55, pp. 88-93, 2012.

[9] J. M., ARTEAGOITIA, et al. "Incidence, prevalence and coronary heart disease risk level in known Type 2 diabetes: a sentinel practice network study in the Basque Country, Spain," Diabetologia, vol. 46, pp. 899-909, 2003.

[10] Edurne, ALONSO-MORÁN, et al. "Prevalence and quality of care indicators of type 2 diabetes in the population of the Basque Country (Spain)," Avances en Diabetología, vol. 31, pp. 72-79, 2015.

Author's Biography

B. Prabha received the B.E degree in Computer Science and Engineering from Dr. Sivanthi Aditinar College Of Engineering, Tiruchendur in 2018. She is pursuing M.E degree in Computer Science and Engineering from VV college of Engineering, Tisaiyanvilai. Her area of interest includes Data mining, Cloud computing, and Bigdata analysis.

P. Krishna Kumar received the B.E., degree in Computer Science and Engineering from Bharathidasan University, India in 1992, received M.E., degree in Computer Science and Engineering from Anna University, India in 2007 and Ph.D degree in Information and Communication Engineering from Anna University, India in 2017.He has more than 9 years of industrial experience in the field of Software and 16 years of teaching experience. His area of interest include Network Security, Data mining, Cloud computing, Medical Image Processing and Bigdata analysis. He has more than 10 International Journal Publication in his credit.