TYPE II DIABETES MELLITUS MICROVASCULAR AND MACROVASCULAR COMPLICATIONS PREDICTION USING MACHINE LEARNING

[1]SHANKU RAJESH, [2]SUSHMA ATHE, [3]MUDISHETTI SRILATHA, [4]B.SINDHUJA
[1]Professor, [2,3]Assistant Professor, [4]UG Student, [1,2,3,4]Department of Pharmacy, Brilliant Grammer School Educational Society Group of Institutions-Integrated Campus, Hyderabad, India.

ABSTRACT
Background: Type 2 diabetes is becoming more and more common at an alarming rate. The consequences of diabetes mellitus pose a serious danger to world health. Type 2 diabetes is the most common form of major organ failure among diabetics and is associated with a number of co morbidities, including diabetic nephropathy, neuropathy, retinopathy, cardiovascular disease, and peripheral vascular disease. Patients are frequently already severely impaired when they are diagnosed with the condition. A risk prediction tool might therefore be useful for implementing early treatment and prevention.
Aim: To construct and internally validate prediction models to estimate the risk of developing complications in patients with type II diabetes mellitus. The main aim of our study is to design and construct a prediction model that predicts the risk of occurrence of diabetes-related complications in the future using a machine learning algorithm.
Methodology: A study was conducted in the outpatient and inpatient department of VIMS hospital, King George hospital, Apollo hospital which included a sample size of 3069 patients having Type II Diabetes Mellitus. Using the data obtained, prediction model was constructed using machine learning algorithm. A combination of random forest with decision tree and random forest with logistic regression was used.
Results: A total of 3069 subjects were included in the model. Most accurate results during internal validation were given by combination of random forest and logistic regression with high sensitivity values for nephropathy (0.91), retinopathy (0.63) and cardiovascular diseases (0.86). While a combination of random forest and decision tree gave best results for neuropathy and peripheral vascular disease with sensitivity of 0.68 and 1 respectively.
Conclusion: Our model can be used to inform diabetic patients about the risk and severity of probable complications, to help health advisors to convince patients to change their lifestyle, and to inform healthcare providers so they can design immediate preventive interventions before a patient loses her/his capability.
Keywords: Diabetes Mellitus, Complications, Machine Learning, Random Forest, Logistic Regression

1. INTRODUCTION

A metabolic illness called diabetes mellitus (DM), which affects 422 million individuals worldwide—most of whom reside in low- and middle-income nations—reports 1.6 million fatalities annually. If uncontrolled, diabetes can result in both acute consequences like diabetic ketoacidosis and hypoglycemia as well as chronic problems like microvascular (diabetic nephropathy, neuropathy, and retinopathy) and macrovascular (heart disease, cerebrovascular, and peripheral artery disease). The primary component of the significant burden of diabetes, among many other associated difficulties connected to managing the condition, is its consequences. They account for over 60% of direct expenditures and approximately 80% to 90% of associated indirect costs. Treatment should begin before the onset of irreversible clinical signs to avoid these problems. This means that predicting them is essential in order to intervene successfully. E.g. early detection and proper treatment of diabetes can prevent up to 90% of blindness, at least 50% of kidney failure and nearly 80% of amputations (AJ & L, 2000)
According to a recent study, a high prevalence of complications related to Type 2 diabetes was seen in the patients with long-standing diabetes (9.9 ± 5.5 years) most of whom were on Insulin with or without oral hypoglycaemic agents. The overall prevalence of both macro vascular and micro vascular

complications was high due to poor glycemic control. Neuropathy (24.6%) was the most common complication seen in these patients followed by cardiovascular (23.6%) renal (21.1%) and eye complications (16.6%).

Machine Learning is the most advanced technique used today for pattern and decision rule extraction from a particular dataset. Despite being a branch of Artificial Intelligence, at its core, Machine Learning depends on statistical techniques..

Diabetes Mellitus is one such example of health condition. Diabetes is a genetics-dependent disease, which is divided into two broad categories based on occurrence of the disease; Type-1 patients are those who have had Diabetes from birth and it is likely that their pancreatic functions never developed properly in the first place while Type-2 patients are those who develop Diabetes over time as their Pancreas stops working properly owing to old age, excessive intake of glucose, etc.(Rahman et al., 2021)

1.1 FUNDAMENTALS OF MACHINE LEARNING

Machine Learning is a branch of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. In other words, Machine Learning is teaching machines how to learn using Statistics and Probability. Based on the different ways of teaching a machine to learn, Machine Learning divides the types of learning to a few major categories.

1. Supervised Learning uses a dataset as training examples with each example consisting of certain label/s which identify it. Using learning algorithms, the machine is taught to correctly identify new data given to it based on the previous dataset.
2. Unsupervised Learning does not give a definite output like supervised learning. Rather, it aims to find structures, patterns and trends in the given data.
3. Reinforcement Learning is a learning method that interacts with its environment by producing actions and discovers errors or rewards based on the action. This method allows the machine to determine the ideal behaviour in a specific context by itself.

1.2 MACHINE LEARNING MODELS:

Random Forest: Random forest is a nonlinear tree-based integrated learning model. Random forest establishes a forest in a random way. The forest is composed of many decision trees, and there is no correlation between each decision tree. After the random forest model is obtained, each decision tree in the random forest is judged when the new sample enters. For the classification problem, the voting method is used, and the maximum number of votes is the final model output.

***Decision Tree*:** The decision tree is a basic classification method. The decision tree consists of nodes and directed edges. A decision tree contains a root node, an internal node, and a leaf node, in which the internal node represents a feature and the leaf node represents a class. First, the feature is filtered according to the information gain of the feature. Then, each node is divided into sub nodes according to the feature value. The root node contains the sample set. The path from the root node to each leaf node corresponds to a decision sequence.

***Logistic Regression*:** Logistic regression, also known as logarithmic probability regression, is a classification model and suitable for the fitting of numerical binary output data. After the input data are linearly weighted, a sigmoid function is used to process the input data to obtain the output probability result, and then, the probability result is transformed into binary output by a symbol function. The parameters of the input model are obtained by maximum likelihood estimation, which distinguishes it from conventional logistic regressions.(Ye et al., 2020)

2. METHODOLOGY

Study site: The study was conducted in the inpatient and outpatient wards of Department of Endocrinology at Visakha Institute of Medical Sciences (VIMS HOSPITAL), King George Hospital and Apollo Hospital in Visakhapatnam, India.

Study design: The study is a cross-sectional study as multiple outcomes and exposures can be studied at same point in time.

Study duration: The study was conducted from 1st February 2021- 30th July 2021.
Ethical approval: Ethical approval was obtained from the Institutional Ethics Committee.
Sampling techniques: Subjects were selected using simple randomization technique because it creates samples that are highly representative of the population.
Sample size determination: Sample size is 3069

Inclusion criteria	Exclusion criteria
Patients who are diagnosed with type 2 diabetes mellitus having complications; both male and female were enrolled	Pregnant and lactating women
Patients age >30 as there is a high chance of onset of complications in those patients	Patients with juvenile diabetes

Proposed model:
In this paper, several methods were used on a dataset consisting of 3069 T2DM patients to determine the risk of Nephropathy, Neuropathy, Retinopathy, Cardiovascular disorders, Peripheral vascular disorders. Figure 1, below represents the experimental work flow of the model in this project.

2.1. Data collection:
Data was collected from the patients using patient interview and laboratory reports. Data including family history of diabetes mellitus, age, and gender, age of disease onset, body mass index, blood pressure, diabetes duration, blood glucose level, social history, medication use, medication adherence, whether any complications like retinopathy, nephropathy, neuropathy, cardiovascular were obtained.

2.2. Data cleaning:
The obtained data was used for training (80%) and testing (20%) of the prediction model which was constructed using machine learning algorithms. The data was then cleaned by dropping instances with significant amount of missing values.

Study Instruments/Questionnaires:
Modified Morisky Medication Adherence scale for complication (MMMA-4) will be used to measure medication compliance.

Table 2.1 : MMMA-4 scale

Have you ever forgotten to take your diabetes medications?	Yes/ No
At times, are you not careful about taking your diabetes medicine?	Yes/ No
When you feel better, do you sometimes stop taking your diabetes medicine?	Yes/ No
At times, if you feel worse when you take your diabetes medicine, do you stop taking it?	Yes / No

Adherence scores in modified medication adherence scale (Morisky et al., 1986)
No= 0; yes=1
Total score 0 = high adherence
Total score of 1 to 2 = medium adherence
Total score of 3 to 4 = low adherence

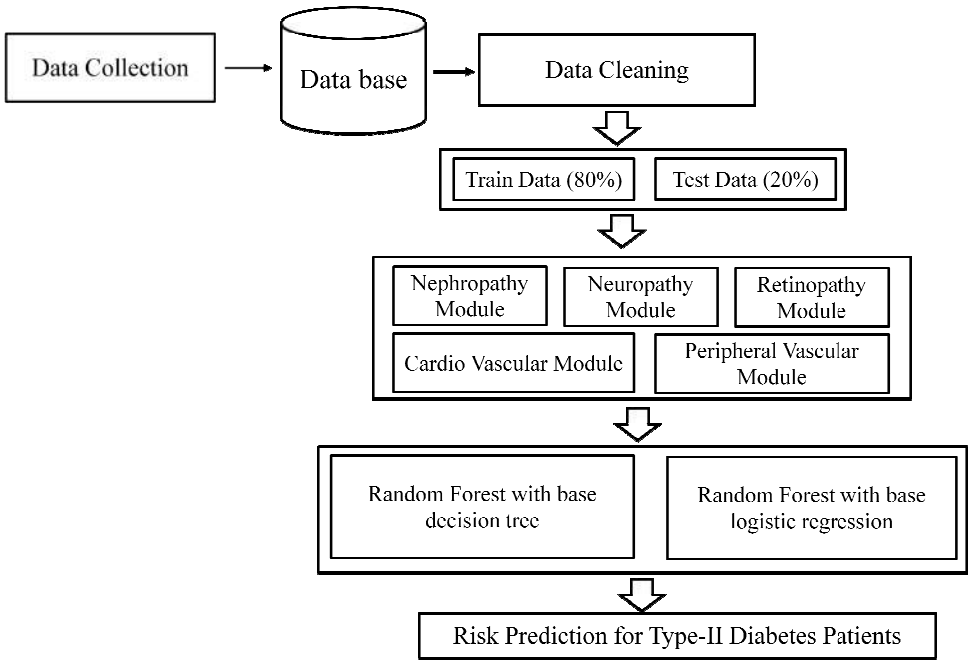

Figure 2.1: System Architecture

2.3. Data base and data analysis:

The in-detailed process is as follows: after cleaning the data, the mean/mode imputation was used to fill the missing values. However, the whole process was done manually and the results of mean/mode imputation were too poor, hence they were discarded. Instead, Random Forest algorithms were used to impute the missing values. The categorical variables were then converted to numerical ones by introducing dummy variables. Since all the variables have different units and ranges, they were all brought to the same scale by apply a feature scaling technique called Standardization. The problem at hand can be defined as a multi-label classification problem since a patient can have either of the complications or all. Hence, it was divided into five separate binary classifications where risk of Nephropathy, Neuropathy, Retinopathy, Cardiovascular disease and Peripheral vascular disorders were predicted separately.

3. Dataset:

The dataset used in the proposed model was found in a loyalty free dataset sharing platform. It is from an open-label, central registration, multicenter, simple randomization. This study was conducted at Visakha Institute of Medical Science (VIMS hospital), King George Hospital, and Apollo Hospital in Visakhapatnam, India.

It has most of the variables needed to implement the proposed Diabetes Complications Prediction Model. The original dataset has 3069 instances and 57 variables. In order to make this dataset useful for the project, some unnecessary variables were removed at the beginning of data cleaning phase due to their lack of relation to the project.

3.1. Numeric features:

This group of the dataset contains 21 features and most of these features have multiple instances. The features are gender, age, weight and height (BMI), HbA1c, fasting blood glucose (FBG), post prandial blood glucose (PPBG), blood pressure (systolic and diastolic), family history, age of onset, diabetic duration, smoking status, physical activity, medication use, medication adherence, presence of complications like nephropathy, neuropathy, retinopathy, cardiovascular disorder, peripheral vascular disease.

3.2. Binary categorical feature:

This set of data includes 19 binary features - Sex, social history (smoking), History or presence of complications like nephropathy, neuropathy, retinopathy, Cardiovascular and peripheral vascular diseases. Medication use, physical activity and medication adherence were included. All these features contain binary values of either 0 or 1. It is mentionable that some numeric variables change values based on the gender

(sex) of the instances. History of Smoking (social history) is taken to consideration due to the fact that affects heart, eye, kidney neurons and vessels.

4. RESULTS

Among 3069 cases with diabetes related complications 1566 participants were male and 1503 participants were female. Individuals in age groups of Young adults (25-35) were 346, adults (36-45) were 589, young middle age(46-55) were 639, middle age(56-65) were 633 and senior adult(>65) were 861 in number.

The percentage of population highly adherent to medications are 32.2 %, moderately adherent to medications are 32.2% and not adherent to medications are 35.2%. This states that the percentage of population who are highly adherent were equal to the percentage of population who are moderately adherence. Compared to highly adherent and moderately adherent population non- adherent populations were slightly high.

The percentage of population with the presence of complication related to diabetes mellitus i.e., nephropathy, neuropathy, retinopathy, cardiovascular, peripheral vascular diseases were 67.6%, 16.22%, 50.4%, 22.4% and 1.36% respectively.

After implementation of different Machine Learning models, the next step is to find out how the models performed. This is done by running the models on the test dataset which was set aside earlier. To determine and compare the performance of the different algorithms, several performance metrics were used.

Performance matrix

The performance of Machine Learning algorithms is evaluated using several performance metrics. Performance metrics relating to classifications are discussed here since the paper only deals with classification problems. For Nephropathy, if the target variable (risk of Nephropathy) is 1 then it is a positive instance, meaning the patient has Kidney complications. And if the target variable is 0, then it a negative instance, meaning the patient does not have Kidney complications. Similarly for cardiovascular disease, if the target variable (risk of cardiovascular disease) is 1 then it is a positive instance, meaning the patient has Heart complications. And if the target variable is 0, then it a negative instance, meaning the patient does not have Heart complications. The performance matrix is same, for neuropathy, retinopathy, and peripheral vascular complications.

Confusion Matrix

Confusion Matrix is the easiest way to determine the performance of a classification model by comparing how many positive instances were correctly/incorrectly classified and how many negative instances were correctly/incorrectly classified. In a Confusion Matrix, the rows represent the actual labels and the columns represent the predicted labels. Table 1 shows the Confusion Matrix.

Predicted Negative	Predicted Positive	
Actual Negative	TN	FP
Actual Positive	FN	TP

Table 4.1: Confusion Matrix

The Confusion Matrix has four values- True Negative, False Positive, False Negative and True Positive. The blocks in confusion matrix represent correctly predicted negative in True Negative, falsely predicted positive in False Positive, wrong prediction of negative in False Negative and correctly predicted positive in True Positive respectively. These values are later used to find the Accuracy, Precision, Recall, Specificity and F1 Score to evaluate the performance of each algorithm.

True Positives (TP): True positives are the instances where both the predicted class and actual class is True (1), i.e., when a patient actually has complications and is also classified by the model to have complications.

True Negatives (TN): True negatives are the instances where both the predicted class and actual class is False (0), i.e., when a patient does not have complications and is also classified by the model as not having complications.

False Negatives (FN): False negatives are the instances where the predicted class is False (0) but actual class is True (1), i.e., when a patient is classified by the model as not having complications even though in reality, they do.

False Positives (FP): False positives are the instances where the predicted class is True (1) while actual class is False (0), i.e., when a patient is classified by the model as having complications even though in reality, they do not.

Accuracy (ACC): Accuracy determines the number of correct predictions over the total number of predictions made by the model. Even though it is widely used, it is not a very good measure of performance especially when the dataset is imbalanced like in this case. The formula for Accuracy is:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

Precision:

Precision is a measure of the proportion of patients that actually had complications among those classified to have complications by the system. The formula for Precision is:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Sensitivity/Recall:

Recall or sensitivity is a measure of the proportion of patients that were predicted to have complications among those patients that actually had the complications. The formula is:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Specificity:

Specificity is the opposite of Recall. It is a measure of the number of patients who were classified as not having complications among those who actually did not have the complications. The formula is:

$$\text{Specificity} = \frac{TN}{TN+FP}$$

F1 Score:

F1 Score is the harmonic mean of the Recall and Precision that is used to test for Accuracy. The formula is:

$$F1 = 2 \times \frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}$$

Variables	N= 3069 (%)	Variables	N= 3069 (%)
Sex		**Smoking status**	
male	1566 (51.02)	non smoker	847 (27.59)
female	1503 (48.97)	occasional /EX smoker	752 (24.50)
Age		smoker	741 (24.11)
young adult	346 (11.27)	mean SBP	139.31
adult	589 (19.19)	Mean DBP	84.66
young middle age	639 (20.82)	Family history	1510 (49.2)
middle age	633 (20.62)	mean diabetic duration	12.16
senior adult	861 (28.05)	Mean age	54.9
mean age	54.9	**Complications**	
BMI		nephropathy	498 (16.22)
underweight	155 (5.05)	neuropathy	2066 (67.31)
normal weight	515 (16.78)	retinopathy	1547 (50.40)
over weight	264 (8.60)	CVD	688 (22.41)
obese 1	737 (24.01)	PVD	42 (1.30)
obese 2	1388 (45.22)	Nephro& Neuro	323 (10.52)
mean BMI	28.82	Nephro&Retino	256 (8.34)
Physical activity		Nephro& CVD	94 (3.06)
not active	1510 (49.20)	Nephro& PVD	4 (0.13)
sufficiently active	1558 (50.76)	Nephro, Neuro, Retino &CVD	24 (0.0078)
Medication adherence		Neuro, Retino, CVD & PVD	4 (0.0013)
not adherent	1083 (35.28)		
medium adherence	994 (32.38)		
highly adherence	991 (32.29)		

Table 4.2: Patient Characteristics

ALGORITHM NAME	MODEL NAME	ACCURACY	SENSITIVITY	SPECIFICITY	PRECISION	F 1 SCORE
RANDOM FOREST (DECISION TREE)	**NEPHO**	0.71	0.16	0.98	0. 76	0.26
	NEURO	0.68	0.68	0.69	0.81	0.74
	RETINO	0.81	0.25	0.92	0.40	0.31
	CVD	0.95	0.97	0.94	0.94	0.95
	PVD	1	1	1	1	1

Table 4.3: Performance Evaluation Using Random Forest with Decision Tree

ALGORITHM NAME	MODEL NAME	ACCURACY	SENSITIVITY	SPECIFICITY	PRECISION	F 1 SCORE
RANDOM FOREST (LOGESTIC REGRESSION)	NEPHO	0.64	0.91	0.84	0.67	0.77
	NEURO	0.83	0.20	0.99	0.50	0.29
	RETINO	0.61	0.63	0.60	0.60	0.6
	CVD	0.81	0.86	0.99	1	0.92
	PVD	0.99	0.50	1	1	0.67

Table 4.4: Performance Evaluation Using Random Forest with Logistic Regression

Receiver operating characteristics (ROC) Curve

ROC or Receiver Operating Characteristics is a graphical plot of sensitivity against (1-Specificity) or in other words, a comparison of true positive rate (TPR) and false positive rate (FPR). It is used to visualize a classifier‘s performance at different thresholds to determine the best threshold point for the classifier. The below graph showed the ROC curve of random forest with decision tree and random forest with logistic regression. Orange color represents ROC of logistic regression and blue color represents ROC of decision tree.

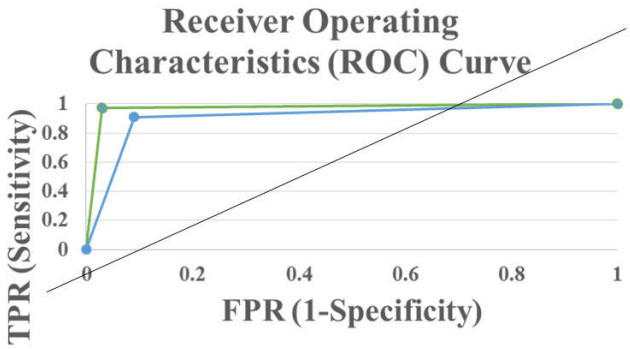

Figure 4.1: ROC of Decision Tree and Logistic Regression

To predict the risk of nephropathy, decision tree gives an accuracy of 0.71, precision of 0.76, specificity of 0.98 and F1 score of 0.26. However, sensitivity is not satisfactory with the value of 0.16. Conversely, logistic regression gave an accuracy of 0.64, sensitivity of 0.91, specificity of 0.84, and precision of 0.67, F1 score of 0.77. If the performance parameters are considered accuracy, specificity, precision are lower in logistic regression when compared to decision tree. Whereas, precision, sensitivity and F1 score were higher in logistic regression when compared with decision tree. So, logistic regression is the method which is predicting the risk of nephropathy accurately.

In the prediction of neuropathy, following are the prediction performance parameters of decision tree and logistic regression with accuracy of 0.68 and 0.83, sensitivity of 0.68 and 0.20, specificity of 0.69 and 0.99, precision of 0.81 and 0.50 and F1 score of 0.74 and 0.29 respectively. Even though, accuracy, and specificity were higher but the sensitivity and precision and F1 score were low in logistic regression in comparison to decision tree, so decision tree is the best algorithm in predicting the risk of neuropathy.

For retinopathy risk prediction accuracy, specificity values are higher with 0.81 and 0.92 but sensitivity, precision and F1 score values are low with 0.25, 0.40, 0.31 respectively in decision tree when compared with logistic regression with accuracy 0.61, sensitivity 0.63, specificity 0.60, precision 0.60 and F1 score 0.61. So, logistic regression is the algorithm which predicts the risk accurately than decision tree.

In cardiovascular risk prediction, decision tree algorithm provided the satisfactory performance parameters with accuracy 0.95, sensitivity 0.97, F1 scores 0.95 when compared with performance parameters of logistic regression with accuracy, sensitivity and F1 score 0.81, 0.86 and 0.92 respectively. So, decision tree is best algorithm in predicting cardiovascular risk.

Accuracy 1, sensitivity 1, specificity 1, precision 1, F1 score 1 of decision tree is higher in decision tree when compared to logistic regression with accuracy 0.99,sensitivity 0.50, specificity 1, precision 1 and F1 score 0.67. So in the prediction of risk for peripheral vascular disease, decision tree is the best algorithm when compared with logistic regression.

DISCUSSION:

A prediction model must provide valid and accurate estimates, and these estimates should be able to inform management and clinical decision-making and subsequently improve outcome and cost-effectiveness of care. A study has indicated that only ~50% of studies are externally validated and ~25% are internally validated(Collins et al., 2011). The model must be accepted and understood by clinicians for the model to be adopted on a wider scale. These requirement soften imply that the models become oversimplified, which could weaken their accuracy. This trend of focusing on simplification rather than performance has been observed in many studies.

Convincing documentation and evidence for all relevant aspects must be provided, which is not always possible in a pragmatic context. Prediction models are there for often based simply on multiple logistics or similar linear regression. The advantage of this approach lies in the transparency of its functionality; however, this advantage comes at the cost of not taking into account that predictors are rarely independent.

In our study, we have mainly focused on the prediction of complications of Diabetes mellitus using predictive models. Several prediction models are used nowadays that are developed using machine learning, to predict different patient outcomes like treatment outcomes and disease outcomes. Using this model several complications can be predicted, prevented, and managed at an early stage by healthcare interventions or lifestyle changes, ultimately improving patient outcomes. This helps healthcare providers to identify disease progression even before the occurrence of signs and symptoms.

According to one study, the use of insulin and duration of diabetes are described as novel interpretable features to assist with clinical decisions in identifying the high-risk populations for diabetic retinopathy. If the duration of diabetes mellitus increases by one year, the odds ratio to have retinopathy is increased by 9.3%. The odds ratio to have retinopathy is increased by 3.561 times for patients who use insulin compared to patients who do not use insulin. (Tsao et al., 2018)

Early and appropriate intervention in diabetes is able to reduce the rate of complications, to

Prolong life expectancy and to reduce the financial cost hence, it is very important to predict the risk of developing complication. The most common predictive models for diabetes complications, however, are not able to deal with all the major complications, but are mainly focused on cardiovascular diseases, diabetic nephropathy and diabetic retinopathy(Lefteris et al., 2012).

In most of these studies only relatively simple statistical approaches, such as additive scores or logistic regression assuming independence between variables, have been applied In our research, we have built a model to include all chronic complications of diabetes mellitus including micro vascular complications(retinopathy, neuropathy and nephropathy) as well as macro vascular complications (coronary heart diseases and peripheral vascular diseases)(Lagani et al., 2013)

Recurrent neural model (RRN) gives an accuracy of 0.7686± 0.012, 0.7466±0.053, 0.7676±0.002, 7966±0.014; sensitivity of 0.8266±0.017,0.7956±0.041,0.7746±0.005, 0.7996±0.007; specificity of 0.6546±0.021, 0.7016±0.049, 0.7536±0.011, and 0.7926±0.018 to predict the risk of nephropathy, neuropathy, PVD and retinopathy respectively for subject with 4 hospitalizations. The number of hospitalizations was an important factor for the prediction accuracy. In the 4-visit scenario for all complications of DM2, RRN was the model that achieved the best prediction accuracy. The results for 2 and 3 visits are omitted because they were consistent with results for 4 hospitalizations. The prediction accuracy of complications decreases over time period according toLjubic et al., 2020

Linear regression (LR) gives an accuracy of 0.777, 0.647, 0.746; sensitivity of 0.820,0.652,0.783; specificity of 0.730, 0.642, 0.707; positive predictive value of 0.771, 0.680, 0.743; negative predictive value of 0.785, 0.613, and 0.750 to predict the risk of retinopathy, nephropathy, and neuropathy respectively. Micro-vascular complications account for a larger number of cases developed after the first visit (20.1% and 79.9% before and after the first visit respectively), as compared to macro-vascular complications (39.4% and 60.6% before and after the first visit respectively) according to Dagliati et al., 2017.

In our study by the combination of random forest with LR gave an accuracy of 0.64, sensitivity of 0.91 and specificity of 0.84 for nephropathy. Our model yielded best results for prediction of PVD with better accuracy, sensitivity and specificity compared to other studies.

Complication	**Best algorithm**
Nephropathy	Logistic regression
Neuropathy	Decision tree
Retinopathy	Logistic regression
Cardiovascular disease	Logistic regression
Peripheral vascular disease	Decision tree

Table 5.1: Complication and best algorithm for it.

CONCLUSION

This model has benefits for diabetic patients and the health workers who are involved in diabetes diagnosis and treatment. It can be used to inform diabetic patients about the risk and severity of probable complications, to help health advisors to convince patients to change their lifestyle, and to inform healthcare providers so they can design immediate preventive interventions before a patient loses her/his capability. The contributions of this research are,

It creates a predictive model to indicate the relationship between individual risk factors and complications. It suggests a method to integrate a number of different models and derive probability tables from them. This model should be further validated externally by using data of patients without complications of diabetes mellitus to predict their risk of developing those in future.

REFERENCES

1. Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., & Tibollo, V. et al. (2017). Machine Learning Methods to Predict Diabetes Complications. *Journal Of Diabetes Science And Technology*, *12*(2), 295-302. https://doi.org/10.1177/1932296817706375
2. Diabetes. Who.int. (2021). Retrieved 2 February 2021, from https://www.who.int/health-topics/diabetes#tab=tab_3
3. Research, E., & Atlas, I. (2017). *IDF Diabetes Atlas*. Idf.org. Retrieved 12 August 2021, from https://www.idf.org/e-library/epidemiology-research/diabetes-atlas.html.
4. (2007). Retrieved 12 August 2021, from https://www.researchgate.net/publication/6336330_Economic_analysis_of_diabetes_Care
5. AJ, B., & L, V. (2000). *The diabetic foot: the scope of the problem*. PubMed. Retrieved 12 August 2021, from https://www.ncbi.nlm.nih.gov/pubmed/11093553
6. Jacques, C., Jones, R., Houts, P., Bauer, L., Dwyer, K., Lynch, J., & Maria Casale, T. (1991). Reported Practice Behaviors for Medical Care of Patients With Diabetes Mellitus by Primary-Care Physicians in Pennsylvania. *Diabetes Care*, *14*(8), 712-717. https://doi.org/10.2337/diacare.14.8.712
7. Harris, M. (1990). Testing for Blood Glucose by Office-Based Physicians in the U.S. *Diabetes Care*, *13*(4), 419-426. https://doi.org/10.2337/diacare.13.4.419
8. Rahman, T., Farzana, S., & Khanom, A. (2021). *Prediction of diabetes induced complications using different machine learning algorithms*. Hdl.handle.net. Retrieved 12 August 2021, from http://hdl.handle.net/10361/10945.

9. Ye, Y., Xiong, Y., Zhou, Q., Wu, J., Li, X., & Xiao, X. (2020). Comparison of Machine Learning Methods and Conventional Logistic Regressions for Predicting Gestational Diabetes Using Routine Clinical Data: A Retrospective Cohort Study. *Journal Of Diabetes Research*, *2020*, 1-10. https://doi.org/10.1155/2020/4168340
10. Makino, M., Yoshimoto, R., Ono, M., Itoko, T., Katsuki, T., & Koseki, A. et al. (2019). Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-48263-5
11. Rodriguez-Romero, V., Bergstrom, R., Decker, B., Lahu, G., Vakilynejad, M., & Bies, R. (2019). Prediction of Nephropathy in Type 2 Diabetes: An Analysis of the ACCORD Trial Applying Machine Learning Techniques. *Clinical And Translational Science*, *12*(5), 519-528. https://doi.org/10.1111/cts.12647
12. Aminian, A., Zajichek, A., Arterburn, D., Wolski, K., Brethauer, S., & Schauer, P. et al. (2020). Predicting 10-Year Risk of End-Organ Complications of Type 2 Diabetes With and Without Metabolic Surgery: A Machine Learning Approach. *Diabetes Care*, *43*(4), 852-859. https://doi.org/10.2337/dc19-2057
13. Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting Diabetes Mellitus With Machine Learning Techniques. *Frontiers In Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00515
14. Huang, G., Huang, K., Lee, T., & Weng, J. (2015). An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. *BMC Bioinformatics*, *16*(Suppl 1), S5. https://doi.org/10.1186/1471-2105-16-s1-s5
15. Tsao, H., Chan, P., & Su, E. (2018). Predicting diabetic retinopathy and identifying interpretable biomedical features using machine learning algorithms. *BMC Bioinformatics*, *19*(S9). https://doi.org/10.1186/s12859-018-2277-0
16. Tanaka, S., Tanaka, S., Iimuro, S., Yamashita, H., Katayama, S., & Akanuma, Y. et al. (2013). Predicting Macro- and Microvascular Complications in Type 2 Diabetes: The Japan Diabetes Complications Study/the Japanese Elderly Diabetes Intervention Trial risk engine. *Diabetes Care*, *36*(5), 1193-1199. https://doi.org/10.2337/dc12-0958
17. Fan, Y., Long, E., Cai, L., Cao, Q., Wu, X., & Tong, R. (2021). Machine Learning Approaches to Predict Risks of Diabetic Complications and Poor Glycemic Control in Nonadherent Type 2 Diabetes. *Frontiers In Pharmacology*, *12*. https://doi.org/10.3389/fphar.2021.665951
18. Ljubic, B., Hai, A., Stanojevic, M., Diaz, W., Polimac, D., Pavlovski, M., & Obradovic, Z. (2020). Predicting complications of diabetes mellitus using advanced machine learning algorithms. *Journal Of The American Medical Informatics Association*, *27*(9), 1343-1351. https://doi.org/10.1093/jamia/ocaa120