

A Review Prediction of Type-2 Diabetes Mellitus for Nephropathy and Neuropathy Using Machine Learning Techniques

¹Tejashree T K, ²Dr. Asha K R

¹Research Scholar, Siddhartha Academy of Higher Education, Tumkur, India

²Associate Professor, Department of Computer Science, Sri Siddhartha Institute of Technology, Tumkur, India

Abstract - Diabetes mellitus (DM) is a chronic malady that is measured to be life-threatening. Diabetes Mellitus type 2 is accounts for the majority of increased risk of Neuropathy and Nephropathy. Utilizing machine learning (ML) predictive model algorithms can aid in the management of diabetes mellitus type 2 (T2DM) by lowering the risk of neuropathy and nephropathy problems in early stages. The increased screening can be avoided the loss of life our study creates predictive model to forecast the diabetes through Gradient Boosting, SVM and Random forest algorithms. The development and work by comparing the performance accomplished using various parameters and algorithms, it originates that by using a selected number of features, we can still build sufficient results.

Keywords: Random Forest, Gradient Boosting, Neuropathy, Machine Learning, Nephropathy, Type-2 Diabetes.

I. INTRODUCTION

Diabetes mellitus is a huge global health burden due to its multifaceted nature, especially Type-2 Diabetes Mellitus (T2DM), and the numerous problems it is connected with. A major contributor to end-stage renal disease, diabetic nephropathy (DN) lowers patient quality of life and puts a heavy burden on healthcare systems. In order to stop the progression of DN and lessen Associated morbidity, early detection and appropriate care are essential.

In recent years, machine learning (ML) techniques have modernized the prediction and management of diabetic complications by integrating diverse clinical and biomarker data. Several studies have highlighted the efficacy of Machine Learning algorithms in predicting Diabetic Neuropathy and other diabetic complications [1-6]. For instance, Nicolucci et al. [1] employed ML approaches to predict T2DM complications, demonstrating the potential of predictive modeling in clinical settings. Sarkhosh et al. [2] and Zhu et al. [3] focused on predicting DN in T2DM patients, utilizing

various biomarkers and methodologies to enhance prediction accuracy.

This literature review aims to synthesize existing research on ML applications for predicting diabetic complications, specifically focusing on DN in T2DM patients. By examining the methodologies, findings, and advancements reported in relevant studies, this review seeks to elucidate the current landscape of ML in diabetic complication prediction. By identifying common trends, challenges, and future directions, this review contributes to the understanding of how ML can enhance early detection and personalized management strategies in diabetes care.

In summary, this paper consolidates and analyzes the existing literature on ML-based predictive modeling for diabetic complications, particularly DN. By exploring the methodologies and outcomes of relevant studies, we aim to provide insights into the potential of ML to transform clinical decision-making and improve outcomes in T2DM management

II. LITERATURE REVIEW

The applications of machine learning to forecast problems in patients with Type-2 Diabetes Mellitus (T2DM), such as neuropathy and nephropathy, is growing. Nicolucci et al. [1] utilized various ML techniques on electronic health records (EHRs) to develop predictive models, finding that random forests and gradient boosting machines (GBM) were particularly effective. Key predictors included HbA1c levels, diabetes duration, and blood pressure. Their findings suggest that ML can significantly aid in early identification and management of high-risk patients, enhancing personalized treatment and improving clinical outcomes. Other reviews have similarly demonstrated the potential of ML in predicting various T2DM-related problems, emphasizing its value in healthcare.

Studies by Nicolucci et al., [1] & Hosseini Sarkhosh et al. demonstrate the efficacy of ML in identifying the difficulties of T2DM. These models, by identifying key predictors like

HbA1c and serum creatinine, can significantly aid in early diagnosis and intervention, thus improving management and outcomes for T2DM patients. Integrating ML into healthcare can personalize treatment plans and optimize resource allocation.

This study focused on using integrated biomarkers for predictive modeling. The research utilized ML techniques to analyze a comprehensive dataset, including biomarkers and clinical parameters, to predict the onset of diabetic nephropathy. The study demonstrated that combining traditional clinical parameters with advanced biomarkers significantly improved predictive accuracy, highlighting the value of integrated approaches in managing diabetic complications.

Hossein Sarkhosh et al. [2]: The study published in the Journal of Endocrinological Investigation, focused on predicting diabetic kidney disease. Using machine learning techniques like Random Forest and Gradient Boosting, the researchers created a risk score based on clinical data such as blood pressure, blood glucose levels, and kidney function tests. High accuracy and robustness were shown by the random forests model, which was verified on a different dataset. The study emphasized the significance of machine learning in developing risk scores that can be applied in clinical settings for early diagnosis and enhanced patient management

Shin et al. [3] investigated machine learning models to predict diabetic sensorimotor polyneuropathy in their paper. The study investigated several machine learning algorithms, such as decision trees, random forests, and neural networks, using clinical data from nerve conduction tests, patient demographics, and test results. The outcomes demonstrated the predictive power of neural networks and random forests in polyneuropathy, underscoring the necessity of utilizing state-of-the-art machine learning techniques for the early identification and management of diabetes complications.

Usharani and Shanthini [4] conducted a study focused on predicting the severity levels of neuropathy in T2DM patients using ML techniques. Study likely found that ML models could effectively predict the severity levels of neuropathy in T2DM patients. Key features contributing to prediction accuracy may include glycemic control, duration of diabetes, peripheral nerve function tests, and other clinical parameters. The study may have highlighted the potential of ML-based approaches in stratifying patients based on neuropathy severity, aiding in personalized treatment strategies.

Saini et al. [5] investigated the intertwined relationships between diabetic retinopathy, nephropathy, and neuropathy. Their study highlights shared pathophysiological Mechanisms

and emphasizes the importance of integrated management strategies. Understanding these correlations is crucial for enhancing holistic care and improving outcomes in diabetic patients.

Cao et al.[6] used two different cross-sectional studies with adult Chinese participants to examine the relationship between amino acids associated with the urea cycle and the risk of diabetic nephropathy. According to their research, urea cycle metabolites might be important in the pathogenesis of diabetic nephropathy. That these biomarkers might be used to forecast or track the course of the illness. The molecular mechanisms of diabetic nephropathy are clarified by this research, which also suggests future directions for customized treatment strategies that target metabolic pathways.

III. METHODOLOGY

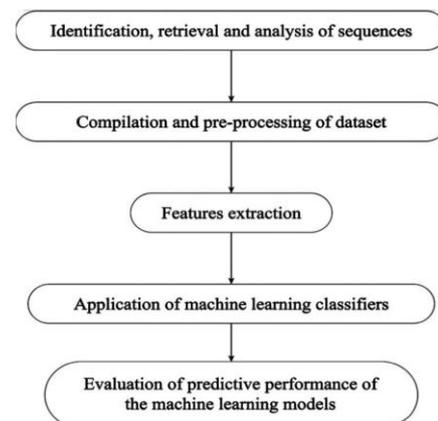


Figure 1: Prediction process of DMT2

IV. METHODOLOGY OVERVIEW

The study follows a structured methodology to predict complications in Type 2 Diabetes Mellitus (T2DM) patients, focusing on diabetic neuropathy and nephropathy. The steps are as follows:

Identification, Retrieval, and Analysis of Sequences: This step involves identifying relevant patient data and retrieving it from electronic health records (EHRs). The data is then analyzed to ensure it meets the study's criteria and relevance.

- 1. Compilation and Pre-processing of Dataset:** A comprehensive dataset is created from the raw data. Preprocessing prepares the dataset for feature extraction and the use of machine learning algorithms by fixing missing values, standardizing data, and encoding categorical variables.
- 2. Features Extraction:** Important features related to neuropathy and nephropathy outcomes are extracted. This involves using statistical methods and domain

knowledge to identify the most relevant predictors, such as HbA1c levels, diabetes duration, and blood pressure.

- Application of Machine Learning Classifiers:** The pre-processed data is subjected to several machine learning techniques, including support vector machines (SVM) and others. These classifiers are designed to forecast the possibility of problems in individuals with type 2 diabetes.

- Evaluation of Predictive Performance of the Machine Learning Techniques:** Assessing Predictive Performance of the Machine Learning Models: Metrics such as Accuracy-A1, Precision-P1, Recall-R1, and F1(score) are used to assess the effectiveness of the machine learning models. Cross-validation techniques ensure the models' robustness and generalizability. Furthermore, hyperparameter tweaking optimizes and enhances model performance.

Table 1: Literature Review of Machine Learning Techniques for Predictive Modeling of Type 2 Diabetes Mellitus Complications

Authors	Paper's Title	Methodology	Used algorithm	Conclusion
Nicolucci et al. (2022)	Prediction of T2DM complications (neuropathy, nephropathy) using ML techniques	Data collection: Electronic health records (EHRs)	Random forests and GBM models performed best in prediction	ML models aid in early identification and personalized treatment
Hosseini Sarkhosh et al. (2023)	establishing and assessing a risk score to assist individuals with type-2 diabetes in predicting the onset of diabetic kidney disease	Gathered clinical data using logistic regression, random forests, and SVM, including blood pressure, kidney function, and blood glucose levels	A risk score was created and validated; random forests and SVM demonstrated high effectiveness	The risk score developed using machine learning techniques enhances early diagnosis and management of diabetic kidney disease
Zhu et al. (2024)	Enhancing predictive modeling of diabetic nephropathy through the use of integrated biomarkers	Data from clinical parameters combined with advanced biomarkers	Integrating traditional clinical parameters with advanced biomarkers improves predictive accuracy	Integration of biomarkers enhances predictive modeling for diabetic complications
Shin et al. (2021)	Forecasting diabetic sensorimotor polyneuropathy using machine learning techniques	Data collection: Nerve conduction studies, demographics, and laboratory results	Random forests and neural networks effectively predicted polyneuropathy	ML techniques enable early diagnosis and management of diabetic complications
Usharani & Shanthini (2020)	Prediction of patient severity levels in T2DM neuropathy using ML approaches	Data collection: Likely included demographic information and clinical measurements	ML models effectively predicted severity levels of neuropathy in T2DM patients	ML-based prediction aids in risk stratification and personalized treatment

V. DISCUSSION

5.1 Methodological Approaches in ML Algorithms for DN Prediction

Several important methods and algorithms have been identified by the literature study on machine learning applications for diabetic nephropathy prediction in patients with Type 2 Diabetes Mellitus. Notably, the capacity for prediction and adaptability of Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB) are acknowledged.

5.2 Random Forest Algorithm

- During the training stage, numerous decision trees are constructed using the Random Forest collaborative

learning technique, which then combines the outputs to produce predictions. For problems involving classification, it determines which class is most frequent among the trees, while for tasks involving regression, it calculates the average forecast made by the trees. This method works especially well for handling high-dimensional data since it averages the output of multiple decision trees that were bootstrapped and trained on different subsets of data to reduce overfitting. Among the basic mathematical ideas at play are:

- Tree Construction:** In order to decrease correlation between trees and increase model variety, each tree is developed using a subset of attributes that are randomly chosen at each node split.

$$T = \text{Tree}(X_{\text{subset}}, Y_{\text{subset}}) \dots\dots (1)$$

- **Bootstrap Sampling:** Random samples of the training data are used to train each tree independently, ensuring robustness against variance and improving generalizability.

$$X_{subset} = \text{BootstrapSample}(X) \dots (2)$$

- **Voting Mechanism:** Classification tasks aggregate predictions through majority voting, while regression tasks average predictions across all trees, providing stable and accurate results.

$$Y = \text{MajorityVote}(Y_{trees}) \text{ for classification}$$

$$Y = \frac{1}{N} \sum_{i=0}^N y_i \dots (3)$$

Random Forest's ability to handle complex datasets and provide comprehensions into feature importance makes it a popular choice in DN prediction studies.

5.3 Gradient Boosting Algorithm

Using a collaborative approach called gradient boosting, models are constructed one after the other, with each new model intended to fix the flaws in the prior one. It enhances a differentiable loss function by introducing weak learners—typically decision trees—gradually and step-by-step. This algorithm's mathematical building blocks include:

Gradient Descent: improves the model iteratively by minimizing the gradient of loss functions with respect to the ensemble's predictions.

$$F_{m+1}(x) = F_m(x) + \mu \cdot hm(x) \dots (4)$$

- **Weak Learners:** Typically shallow decision trees are used as weak learners, minimizing computational complexity while maximizing predictive power

$$hm(x) = \text{argmin}_h \sum_{i=0}^N [L(y_i, F_m(x_i) + h(x_i))] \dots (5)$$

- **Boosting Strategy:** Focuses on reducing bias and variance by emphasizing errors made by previous models, resulting in strong predictive performance.

$$residual_i = y_i - F_m(x_i) \dots (6)$$

Gradient Boosting algorithms, such as XGBoost and AdaBoost, have demonstrated superior performance in various DN prediction models, showcasing their effectiveness in clinical applications.

5.4 Support Vector Machines (SVM)

Support vector machines, supervised learning models, are used for tasks involving regression and classification. SVM's primary goal is to identify the best hyperplane for classifying data into different groups. The closest data points to the hyperplane are the support vectors, and this is achieved by optimizing the margin between them. SVM's mathematical foundation consists:

Table 2: Accuracy comparison of Machine learning Algorithms

	Accuracy (%)	Area Under Curve (AUC)	Precision (%)	Recall (%)	F1 Score
	81.00	0.77	81.00	74.00	0.76
Random Forest	80.00	0.76	77.00	73.00	0.75
Gradient Boosting	83.00	0.82	82.00	78.00	0.80
Support Vector Machine	84.00	0.81	81.00	80.00	0.81

Hyperplane Construction: Finding the hyperplane that optimizes the distance between the classes is the goal.

$$\text{Maximize } \frac{2}{\|w\|}$$

- **Decision Function:** The decision function for classification is based on the sign of the function

$$F(x) = \text{Sign}(w \cdot x + b)$$

- **Lagrange Multipliers:** SVM uses Lagrange multipliers to solve the optimization problem.

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha [y_i (w \cdot x_i + b) - 1]$$

SVM's effectiveness in classification tasks with complex datasets and its ability to construct clear margins between classes make it a valuable tool in predicting DN.

VI. COMPARATIVE ANALYSIS AND APPLICATIONS

Both Random Forest, Gradient Boosting, and SVM algorithms offer distinct advantages in DN prediction models. Random Forest excels in handling large datasets and identifying significant predictors through feature importance analysis, while Gradient Boosting optimizes predictive accuracy by sequentially minimizing errors. SVM is particularly effective in classification tasks, constructing optimal hyperplanes to separate classes.

Choosing between these algorithms often depends on dataset characteristics, computational resources, and the specific clinical context. Understanding the underlying mathematical principles and methodological approaches of these techniques, researchers can leverage their strengths to improve early detection and personalized management strategies in diabetes care.

Performance Metrics of Various Machine Learning Algorithms for Predicting Diabetic Nephropathy

Finally in conclusion, choosing suitable machine learning algorithms, such as Gradient Boosting, Support vector machine, & Random Forest is vital for creating effective predictive models for diabetic neuropathy in Type-2 Diabetes Mellitus patients. By utilizing these advanced techniques, researchers can improve the predictive accuracy and clinical relevance of their models. This, in turn, can lead to better patient outcomes by enabling early detection and personalized treatment strategies.

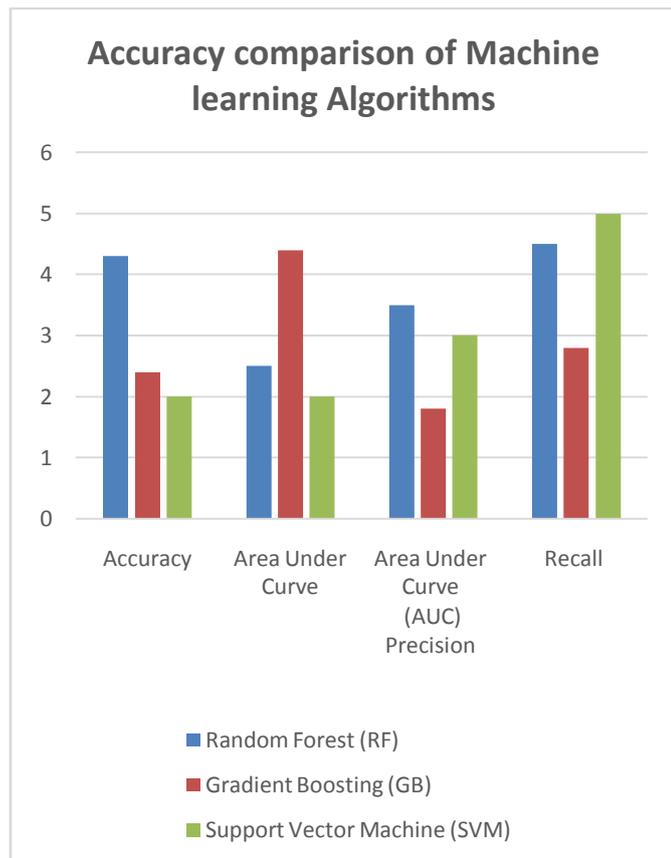


Figure 2: Comparison and prediction of Diabetic nephropathy and neuropathy

VII. CONCLUSION

Finally, utilizing cutting edge machine learning (ML) techniques, our study created a predictive model for diabetic nephropathy in different patients with Type-2 Diabetes

Mellitus. We showcased the ability of machine learning to precisely forecast diabetes (DN) through the use of boosting, support vector machine, and Random Forest algorithms. This allowed for the implementation of early detection and customized treatment plans. Our rigorous technique guaranteed the predictive model's stability and generalizability through substantial data preparation, feature engineering, and iterative training and validation. These machine learning algorithms are useful in clinical practice because of their ability to handle high-dimensional clinical and biomarker data. Our findings highlight the importance of utilizing machine learning approaches to improve patient outcomes and contribute to the growing body of research on predictive analytics for diabetes complications. In the end, this research should open the door to more proactive and individualized treatment approaches by improving these models and exploring their application to other diabetic problems.

REFERENCES

- [1] Nicolucci, A., Romeo, L., Bernardini, M., Vespasiani, M., Rossi, M. C., Petrelli, M., ...& Vespasiani, G. (2022). Prediction of complications of type 2 Diabetes: A Machine learning approach. *Diabetes Research and Clinical Practice*, 190, 110013.
- [2] Hosseini Sarkhosh, S. M., Esteghamati, A., Hemmatabadi, M., & Daraei, M. (2022). Predicting diabetic nephropathy in type 2 diabetic patients using machine learning algorithms. *Journal of Diabetes & Metabolic Disorders*, 21(2), 1433-1441.
- [3] Shin, D. Y., Lee, B., Yoo, W. S., Park, J. W., & Hyun, J. K. (2021). Prediction of diabetic sensorimotor polyneuropathy using machine learning techniques. *Journal of Clinical Medicine*, 10(19), 4576.
- [4] Usharani, R., & Shanthini, A. (2020). Machine learning approaches for predicting patient severity levels in T2dm complications neuropathy. *European Journal of Molecular & Clinical Medicine*, 7(10), 2020.
- [5] Saini, D. C., Kochar, A., & Poonia, R. (2021). Clinical correlation of diabetic retinopathy with nephropathy and neuropathy. *Indian Journal of Ophthalmology*, 69(11), 3364-3368.
- [6] Cao P, Huang B, Hong M, et al. Association of amino acids related to urea cycle with risk of diabetic nephropathy in two independent cross-sectional studies of Chinese adults. *Front Endocrinol*. 2022; 13:983747. doi:10.3389/fendo.2022.983747.
- [7] Zhu, Y., Zhang, Y., Yang, M., Tang, N., Liu, L., Wu, J., & Yang, Y. (2024). Machine Learning-Based Predictive Modeling of Diabetic Nephropathy in Type 2 Diabetes Using Integrated Biomarkers: A Single-Center Retrospective Study. *Diabetes, Metabolic Syndrome and Obesity*, 1987-1997.

- [8] Hosseini Sarkhosh, S. M., Hemmatabadi, M, & Esteghamati, A. (2023). Development and validation of a risk score for diabetic kidney disease prediction in type 2 diabetes patients: a machine learning approach. *Journal of endocrinological investigation*, 46(2), 415-423.
- [9] Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
- [10] Zhang, Z., Zhao, Y., Canes, A., Steinberg, D., & Lyashevskaya, O. (2019). Predictive analytics with gradient boosting in clinical medicine. *Annals of translational medicine*, 7(7).

Citation of this Article:

Tejashree T K, & Dr. Asha K R. (2024). A Review Prediction of Type-2 Diabetes Mellitus for Nephropathy and Neuropathy Using Machine Learning Techniques. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 8(10), 7-12. Article DOI <https://doi.org/10.47001/IRJIET/2024.810002>
