A Data-Driven Approach to Hospital Readmissions in Diabetic Patients

Samridh Vikhas S
Department of Mathematics – SAS
Vellore Institute of Technology, Vellore
Vellore, India
samridhvikhas.s2020@vitstudent.ac.in

Subhashree S
Department of Mathematics – SAS
Vellore Institute of Technology, Vellore
Vellore, India
subhashree.s2020@vitstudent.ac.in

Dr. B S R V Prasad
Department of Mathematics – SAS
Vellore Institute of Technology, Vellore
Vellore, India
srvprasad.bh@vit.ac.in

Abstract - Hospital Readmission is a major problem for patients with diabetes and the Hospitals providing care. For the patients, readmission might be costly and disruptive to their lives, whereas the hospitals have problems with operational costs and logistic issues. In this present work, an exhaustive investigation to produce a forward-thinking predictive model was carried out. The primary goal of the presented methodology is to predict the individuals who have been diagnosed with diabetes and are at an increased risk of readmission. This work used a combination of statistical tools, including Logistic Regression, Decision Trees, and Random Forests. The proposed model uses a comprehensive approach that efficiently identifies patients requiring further support to prevent readmission. The model considers various factors, including the patient's medical history, clinical data, and social determinants of health. Our sophisticated model has been put through training on a comprehensive dataset of diabetic patients who have been treated for their condition in a hospital setting. The proposed model uses a comprehensive approach that efficiently identifies patients requiring further support to prevent readmission. Thus, the work finds its applications in the Healthcare Service Provider sector, helping doctors in decision-making and cutting down unnecessary hospital readmissions.

Keywords **– Hospital Readmission, Diabetes, Predictive Modelling, Decision-Making, Machine Learning, Classification Algorithms, Random Forests**

I. INTRODUCTION

The continuous endeavor to enhance patient outcomes and implement cost-effective strategies in the healthcare sector poses an ongoing and persistent challenge. The issue at hand is notably apparent in the management of individuals with diabetes, as the repeated need for hospital readmissions poses a significant financial burden on patients and presents a logistical challenge for healthcare institutions. Given the circumstances, we undertook a comprehensive inquiry to formulate a sophisticated prognostic framework. The main aim of our study is to identify individuals who have been diagnosed with diabetes and are at a heightened risk of being readmitted to the hospital. The utilization of an advanced model[4], which has undergone training using an extensive dataset consisting of diabetic patients within a hospital environment enables healthcare providers to effectively discern individuals who require supplementary assistance to mitigate the likelihood of readmission [11]. This paper signifies the culmination of our endeavors to utilize machine learning techniques and healthcare data to enhance the healthcare ecosystem's overall health and efficiency.

II. METHODOLOGY

The study adopts a meticulously devised methodological framework to address the pressing concern of diminishing hospital readmissions among individuals diagnosed with diabetes, primarily aiming to enhance patient outcomes and optimize healthcare practices.

The dataset utilized in model development spans a decade, from 1999 to 2008, encompassing data from 130 United States hospitals and integrated delivery networks. This dataset comprises records of individuals diagnosed with diabetes who underwent laboratory tests, received pharmacological treatments, and experienced hospital stays lasting up to 14 days. The data preprocessing phase involved a comprehensive analysis, encompassing the handling of missing values, data type transformations, and the normalization of numerical features [2]. Categorical variables were encoded to facilitate subsequent analyses.

An exhaustive set of relevant features was meticulously constructed to augment the model's predictive capabilities. This process involved the creation of patient-centric variables, including demographic data, medical histories, laboratory results, medication profiles, and prior healthcare utilization. Temporal attributes, such as the frequency of past hospitalizations and interactions with healthcare professionals, were also incorporated, given their pivotal role in predicting readmission likelihood.

A range of machine learning models was evaluated for their suitability in predicting hospital readmission, a binary classification challenge. The models employed in this study encompass Decision Trees, Random Forests [6], Naive Bayes, and XGBoost. The selection of models was determined by evaluating their performance across various metrics,

including accuracy, precision, recall, F1-score, and AUC-ROC. This evaluation considered the trade-off between false positives and false negatives.

To advance patient care and healthcare efficiency, this research leverages a decade-spanning dataset and a diverse set of machine-learning techniques to confront the critical challenge of reducing hospital readmissions among diabetic patients.

III. RESULTS & DISCUSSIONS

This study addresses a significant concern within the healthcare sector [7]: mitigating hospital readmissions among individuals diagnosed with diabetes. By implementing an advanced predictive model, we have successfully established a connection between evidence-based findings derived from data analysis and applying these insights in clinical practice. The approach employed in our study integrates extensive data preprocessing, feature engineering, and rigorous model selection techniques to identify patients at a heightened risk of readmission accurately. Significantly, we have addressed the class imbalance issue [6], improving the model's performance. The responsible utilisation of the model in healthcare is underscored by its transparency and commitment to ethical considerations. This technology's deployment and potential integration within healthcare systems provide timely and up-to-date insights. Our collaboration with medical professionals has confirmed the clinical significance of the model. This study signifies a substantial advancement in the effort to decrease the frequency of readmissions, enhance patient outcomes, and optimise healthcare strategies, ultimately improving the quality of diabetes care.

The following table provides a comparative analysis of the performance of different Machine Learning Algorithms in predicting hospital readmissions. The models were evaluated based on metrics such as accuracy, F1-score, and AUC-ROC, utilising the Test dataset [12].

Table 1 - Classification Models' Performance Metrics

Model	**Accuracy**	**F1-Score**	**AUC-ROC**
Gaussian Naive Bayes	0.89473	0.93856	0.86512
CART	0.87433	0.91256	0.89965
ID3 Tree	0.88354	0.90567	0.90068
C4.5 Tree	0.91567	0.93488	0.90198
Random Forest Classifier	0.93872	0.96185	0.98653
XGBoost Classifier	0.91269	0.90123	0.88597

Based on the findings presented in Table [1], it is apparent that the Random Forest Classifier[4] has yielded favourable results when applied to predicting hospital readmission for diabetes. The Random Forest model's ROC Curve is shown in Figure 1; from Figure 1, it is evident that the Random Forest model is almost perfect in predicting the Readmission time gap. The chosen model demonstrates differing performance levels across the assessed metrics, excelling in particular aspects.

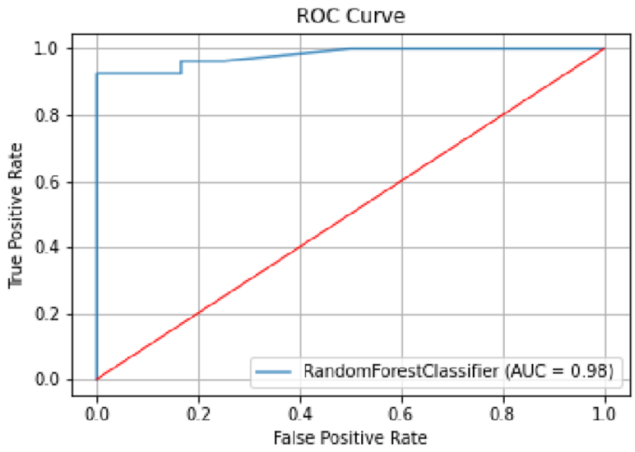

Figure 1 – ROC Curve of the Random Forest Classifier

In summary, this research endeavour signifies a significant advancement in addressing the issue of hospital readmissions, improving patient results, and optimising healthcare methodologies. The findings and knowledge acquired from this study exhibit considerable potential for enhancing the standard of diabetes management. Additional investigation and practical application will play a pivotal role in advancing the quality of healthcare for individuals with diabetes and mitigating the frequency of hospital readmissions.

IV. REFERENCES

[1] Kexin Huang, Jaan Altosaar, Rajesh Ranganath, "ClinicalBERT: Modeling ClinicalNotes and Predicting Hospital Readmission", arXiv:1904.05342v3 [cs.CL] 29 Nov 2020

[2] Daniel J. Rubin, Elizabeth A. Handorf,Sherita Hill Golden,Deborah B. Nelson, Marie E. McDonnell, Huaqing Zhao, "Development and Validation of a Novel Tool to Predict Hospital Readmission Risk among Patients with Diabetes", ENDOCRINE PRACTICE Vol 22 No. 10 October 2016

[3] Daniel J. Rubin, Kelly Donnell-Jackson, Ram Jhingan, Sherita Hill Golden, Anuradha Paranjape, "Early readmission among patients with diabetes: A qualitativeassessment of contributing factors", Journal of Diabetes and Its Complications, 28 June 2014

[4] Avishek Choudhury, Dr. Christopher M. Greene, "Evaluating Patient Readmission Risk: A Predictive AnalyticsApproach", American Journal of Engineering and Applied Sciences, 6 December 2018

[5] Daniel J Rubin, "Hospital Readmission of Patients with Diabetes", Current Diabetes Reports (2018) 18: 21, 13 March 2018

[6] Shaoze Cui, Dujuan Wang, Yanzhang Wang, Pay-Wen Yu, Yaochu Jin, "An improved Support Vector Machinebased diabetic readmission prediction".

Computer Methods and Programs in Biomedicine, 12 October 2018

[7] Sara J Healy, Andrew Lorenz, Dawn Black, Kathleen M Dungan, Cara Harris, “Inpatient Diabetes Education isAssociated With Less Frequent Hospital Readmission Among Patients With Poor Glycemic Control”, Diabetes Care Publish Ahead of Print, July 8 2013

[8] Siyi Tang, Amara Tariq, Jared Dunnmon, Umesh Sharma, Praneetha Elugunti, Daniel Rubin, Bhavik N Patel, Imon Banerjee, “Multimodal spatiotemporal graph neuralnetworks forimproved prediction of 30-day all-cause hospital readmission.”

[9] Ahmed Allam, Mate Nagyy, George Thomaz, Michael Krauthammer, “Neural networksversus Logistic regression for 30days all-cause readmission prediction”, arXiv:1812.09549v1 [cs.LG] 22 Dec 2018

[10] Ahmad Hammoudeh, Ghazi Al-Naymat, Ibrahim Ghannam, Nadim Obie, “Predicting Hospital Readmission among Diabetics using Deep Learning”, The 5th International Symposium on Emerging Information, Communication and Networks(EICN-2018)

[11] Hongsoo Kim, Joseph S. Ross, Gail D. Melkus,Zhonglin Zhao, Kenneth Boockvar, “Scheduled and Unscheduled Hospital Readmissions amongDiabetes Patients”, NIH Public Access, 2010 October

[12] Naina Sinha Gregory, Jane J Seley, Savira Kochhar Dargar, Naveen Galla Linda M Gerber, Jennifer I Lee, “Strategies to Prevent Readmission in High-Risk Patients with Diabetes: The Importance of an Interdisciplinary Approach”, Current Diabetes Reports (2018) 18:54s