ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

Predictive Modelling for Patient Readmission

[1] Aryan Sharma, [2] Adwitiya Bhattacharjee, [3] Dr. Kanipriya M.

[1] Computer Science Engineering, SRM Institute of Science and Technology, Chennai, India
[2] Biotechnology Engineering, SRM Institute of Science and Technology, Chennai, India
[3] Assistant Professor, SRM Institute of Science and Technology, Chennai, India
Corresponding Author Email: [1] hithin2003@gmail.com, [2] rohithkumar2003@gmail.com, [3] kvrforu@gmail.com

***Abstract*— Hospital readmission is one of the significant health- care challenges because it often indicates a gap in patient management and transitions of care. Accurate prediction of readmission likelihood can enhance healthcare delivery, reduce costs, and improve outcomes for patients. This article explores the application of machine learning techniques—Logistic Regression, Decision Tree, and Random Forest—to predict hospital readmis- sions using a more comprehensive dataset of patient records. The dataset was preprocessed by the elimination of missing values and encoding categorical variables into numerical values, along with the removal of irrelevant features. Class imbalance was carried out using Synthetic Minority Over-sampling Technique (SMOTE) to ensure excellent generalization of the model.**

The performances of the models are assessed in accuracy, pre- cision, recall, F1-score, and AUC-ROC. Baseline model: logistic regression. Interpretation into which factors are most impactful on readmission regarding the time spent in the hospital and number of medications. Decision Tree and Random Forest utilize a non-linear relationship towards improvement of the prediction accuracy of the models. The best accuracy among these models is of Random Forest, along with the best trade-off between precision and recall. Meanwhile, Logistic Regression became to be a very interpretable one.

This research puts the interest of using machine learning for fighting against healthcare issues, brought forward as this paper will present a data-driven approach that predicts and mitigates hospital readmission. Those models help in identifying high-risk patients, thereby aiding the healthcare providers in targeted interventions toward optimum patient care and resource utilization.

***Index Terms*— Hospital Readmission Prediction, Machine Learning, Logistic Regression, Decision Tree, Random Forest, SMOTE, Healthcare Data Analytics, Readmission Risk Factors, Predictive Modeling, AUC-ROC, Patient Outcome Prediction, Class Imbalance, Feature Engineering, Precision and Recall, Healthcare Optimization.**

I. INTRODUCTION

Hospital readmission is of major concern to modern health- care systems since it provides one of the main metrics for assessing care quality and efficiency. It often reflects chal- lenges encountered while trying to ensure the transition of patients from the inpatient setting into post-discharge settings. A high readmission rate has serious financial burdens on healthcare facilities but might also signal some deficiencies in managing and continuing patient care. One of the essential steps to mitigate the challenges is therefore to predict hospital readmission: the analysis can enable healthcare providers to determine which patients are most at risk in advance so that intervention can occur in a timely manner.

Here, we use the readmissions task as an application of machine learning methods. The generated dataset from an extremely large repository of hospital records constitutes the patient demographics, clinical, and administrative data. Well, this dataset contains a wide range of features that vary from patient demographics to diagnoses and treatment information. Preprocessing techniques such as handling missing values, encoding categorical data, and feature selection were per- formed so that this dataset was competent enough for machine learning applications. This paper further addresses the class imbalance issue prevalent in health care data and applies the Synthetic Minority Over-sampling Technique, also referred to as SMOTE, which over-samples the minority class to improve the performance of the model.

Three machine learning models—Logistic Regression, De- cision Tree, and Random Forest—were implemented to predict readmission likelihood. These models were evaluated on var- ious metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, providing a comprehensive understanding of their predictive capabilities. Logistic Regression offered interpretability, highlighting the key factors influencing read- missions, while Decision Tree and Random Forest captured non-linear relationships, achieving higher predictive accuracy. This research demonstrates the potential of data-driven approaches in addressing critical healthcare challenges. By leveraging machine learning, healthcare providers can gain actionable insights into the factors driving readmissions and allocate resources more efficiently. The findings of this study underscore the importance of integrating predictive analytics into healthcare workflows to enhance patient outcomes and reduce the financial and operational strain of hospital read-missions.

II. LITERATURE REVIEW

Hospital readmission has been one of the concerning areas in health care studies for a long time. Research on readmission has covered myriad approaches. The application of machine learning, however, introduces for the first time new avenues for resolving intricacies in predicting

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

readmission, with an upgraded accuracy and greater efficiency compared to statis- tical methods. Researches have initiated from the application of statistical models, notably logistic regression and survival analysis, focused largely on factors contributing to readmis- sion. Hasan et al. (2010) used multivariate regression to study the interaction of patient demographics and comorbidities as determinants of risk for readmission, thus delineating the need for more complex models beyond the linear ones [1]. Krumholz et al. (2009) built a risk predictive model for heart failure readmissions through conventional statistical analyses but pointed out the emerging need for adaptive models to portray nonlinear patterns [2].

Many strides have been achieved through the revolution in the field of machine learning for readmission prediction. LeCun et al. illustrated the ability of neural networks for learning and recognition of intricate patterns in healthcare data, which led to new opportunities for applying usage in readmission prediction models [3]. Kansagara et al. reviewed 30 prediction models and found that machine learning methods are on the rise, almost gaining even more importance com- pared to statistical techniques because they allow flexibility and scalability [4]. According to Ghassemi et al. (2014), ensemble learning methods such as Random Forest and Gra- dient Boosting Machines are applicable in predicting 30-day hospital readmissions with higher accuracy than the standalone models [5]. In addition, Xiao et al. (2018) used Support Vector Machines for EHR analysis due to their strength in detecting risk patients [6].

The performance of predictive models relies much on feature engineering. Choi et al. (2016) had applied deep learning techniques to EHRs and explained the significance of temporal data, such as medication history and lab results in readmission prediction [7]. Zhou et al. (2019) considered the impacts of the social determinants - including housing stability and availability of a caregiver, for instance - on readmission rates and pleaded for inclusion of non-clinical factors in the predictive models [8]. One more serious challenge that is faced in readmission predictions is missing data handling. Lipton et al. (2016) researched on RNNs with imputations with the purpose of filling gaps in health care data to make better models [9]. Estiri et al. extended the study on imputation strategy, which showed an imputation strategy to be highly critical for data integrity and reliability of the model itself [10].

Class imbalance has been an issue in readmission predic- tion since the number of readmitted patients is most often extremely low compared to non-readmitted cases. A method to balance datasets that was first introduced by Chawla et al. (2002), and which has been applied in most literature since then rather than healthcare, is the Synthetic Minority Over- sampling Technique [11]. Johnson et al. (2016) used SMOTE on hospital readmission data, and their performance was highly improved [12].

The comparative studies demonstrate the strengths and weaknesses of the various machine learning models. In another study by Rajkomar et al. (2018), the authors compared Logistic Regression, Random Forest, and Naive Bayes for readmission predictors. They pointed out that ensemble methods typically perform better than the simpler models [13]. At the same time, they highlighted the interpretability advantages of Logistic Re- gression, especially where, in health applications, importance of features plays a very key role. Deep learning models, such as LSTMs, and CNNs had held a great promise for healthcare predictive tasks. Miotto et al. applied an LSTM to work with time-series EHR data that could have brought better accuracy in terms of predicting the readmission [14]. Nguyen et al. applied CNN to clinical notes towards the pattern discovery task to enhance the quality of prediction [15].

Feature selection is critical to reduce the complexity of the models and for making a model more intuitive. Guyon et al. provided techniques for feature selection, wherein these resulted in improved accuracy for the models [16]. Suresh et al. utilized mutual information-based feature selection to find very informative features for readmission prediction while being balanced between accuracy and simplicity [17]. Machine learning in the sphere of health care presents fair and biased issues on the ethical front. Obermeyer et al., (2019) analyzed the biases that actually exist in many predictive models and provided strategies to address disparities in care delivery [18]. These authors have emphasized the need for open algorithms and diverse data sets for equitable healthcare.

Some studies have tested the possibility of applying these predictive models to real-world domains. Goldstein et al. 2017 implemented a readmission prediction model in a hospital which resulted in a 15% decline in readmission through appro- priate interventions [19]. Futoma et al. 2015 similarly showed the prowess of real-time analytics in preventing unnecessary healthcare expenditure and enhancing patient outcomes in real time [20].

III. METHODOLOGY

The methodology of this study describes the orderly pro- cesses followed in predicting hospital readmissions with the use of machine learning techniques. It combines data preclu- sion, feature engineering, modeling, and analysis to improve precision and dependability of the forecasts made.

A. Dataset Description

For this research purpose, publicly available hospital data sources containing more than 100,000 patient records have been used. These sources comprise personal identification characteristics, including age, sex, and race, clinical character- istics covering information on diagnosis codes, total number of drugs prescribed and days spent in the hospital, as well as auxiliary data such as the type of admission, disposition of the patient at discharge, and source of admission. The dependent variable “Readmitted” divides

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

patients into three categories; those who were readmitted within 30 days (¡30), those who were readmitted after 30 days (¿30), and patients who were not readmitted (NO)). This dataset serves as an excellent starting point to investigate the clinical factors associated with readmissions and the development of different predictive models.

B. Data Preprocessing

Preparation of the dataset involved efficient data preprocess- ing. There were several steps carried out:

1) *Missing and Invalid Values Handling:* What? unknown/invalid was standardized as NaN. Rows having some critical missing values in demographic or diagnostic columns were simply deleted; other missing values were simply im- puted when needed.
2) *Feature Selection:* Irrelevant and redundant columns, such as weight, payer_code, and medical_specialty, were re- moved. Diagnostic codes diag_2 and diag_3 were excluded in favor of focusing on diag_1, the primary diagnosis.
3) *Categorical Encoding:* Categorical variables such as admission _ type and discharge _ disposition were transformed into meaningful categories to enhance interpretability. For instance, admission types were grouped into categories like Emergency, Elective, and Newborn, and discharge dispositions were classified as Home or Other.
4) *Class Balancing:* The target variable was not balanced by the underlying data in the form of mostly non-readmitted patients and least numbers of readmitted ones. SMOTE tech- nique was used to balance the classes so as the learning in the model would not be biased.

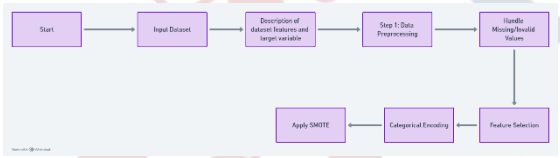

Fig. 1. Work flow for Data preprocessing

C. Feature Engineering

Feature engineering has been used to enhance the quality of the dataset and add predictivity as well as interpretability to it.

1) *Age Grouping:* The age feature, which was originally represented as ranges like [0-10) and [40-50), had been binned into categories like [0-40], [40-50], and [50-60]. This aggregation simplified the data while retaining the value of information.
2) *Diagnosis Categorization:* Diagnostic codes have been mapped into high-level categories such as Circulatory, Res- piratory, and Digestive, thereby reducing dimensionality but retaining clinical significance.
3) *Removal of Features with Low Influence on Outcome Variable:* Drug-related features with high variability were removed. These included repaglinide and tolbutamide. This allowed the consideration of more impactful predictors such as time_in_hospital, num_ medications and number_diagnoses.

D. Model Development

Three machine learning models were used: logistic regression, random forest, and decision tree, to predict the hospital readmission.

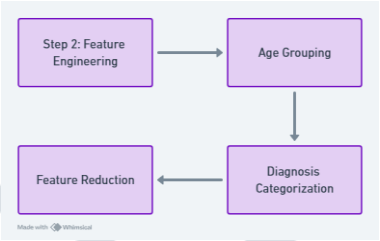

Fig. 2. Flowchart for Feature Engineering

1) *Logistic Regression:* This is a base model because of its simplicity and interpretability. Through logistic regression, feature coefficients show the likelihood that a readmission would occur through the features used.
2) *Decision Tree:* The decision tree accounted for the interaction between features and outcome variables as non- linear, and its intuitive structure yielded decision rules defining conditions that result in readmission.
3) *Random Forest:* The decision tree accounted for the interaction between features and outcome variables as non- linear, and its intuitive structure yielded decision rules defining conditions that result in readmission.

All the models were trained on the balanced dataset and then fine-tuned to optimize the performance of that very dataset. The distinct models were then tested on a separate test set to check their generalizability.

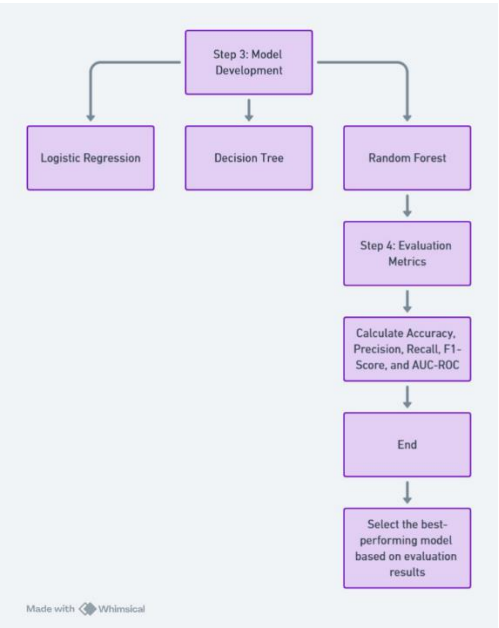

Fig. 3. Model Development and Evaluation

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

E. Evaluation Metrics

To this end, the evaluation would be based on a set of widely accepted metrics for the evaluation of performance in the machine learning models developed within this study. This would allow for an all-rounded and effective assessment of the strengths and weaknesses of the developed models for put- together effective comparison toward helping and informing proper decision-making.

1) *Accuracy:* Accuracy represents one of the most basic evaluation metrics, evaluating the proportion of all correct predictions from the model. Though such a metric captures the high-level performance of the model, accuracy alone in such imbalanced datasets, like the one used in this study, may not prove very reliable. This is because it may become biased from the dominant class and leads to misleading the actual strength of the model in identifying minority cases, such as early readmissions from the hospital.
2) *Precision:* Precision is the number of true positives predicted by the model divided by the sum of true positive and false positive. It is very effective for scenarios where false positives incurred higher costs or resulted in serious conse- quences. For instance, for a hospital readmission scenario, high precision would indicate the exactness of how well the model identifies which of the patients are likely to be readmitted to the hospital. High recall would ensure no patient, who might return, would not be left without resources being made available to him or her.
3) *Recall (Sensitivity):* Recall measures the number of actual positive cases (patients who were readmitted) against which the model correctly identifies. High recall is highly essential in health care applications as it reduces the oppor- tunities of missing potentially high-risk patients. In this case, recall was basic so that most at-risk patients were flagged for potential intervention.
4) *F1 Score:* A hybrid measure of precision and recall to give one measure is achieved by computing their harmonic mean. This makes it rather suitable for imbalanced datasets where precision or recall on its own may not give enough representation of how well the model is actually performing. So, an F1-Score guarantees balancing of the model; there has to be some trade-off between precision (aversion to false positives) and recall (true positive capture).
5) *AUC ROC:* The Area Under the Receiver Operating Characteristic Curve, AUC-ROC, measures the ability of the model to correctly classify instances between classes for various thresholds. The ROC curve is a plot of True Positive Rate (TPR) vs False Positive Rate (FPR) at different threshold levels. The AUC value ranges between 0 to 1; the best performance corresponds to higher values close to 1. AUC- ROC was a good measure of discriminative power in this study, showing how good the model was at separating patients likely to be readmitted from those that were not.

These metrics provided a comprehensive evaluation of each model's strengths and weaknesses, allowing for an informed comparison. give more content on this

IV. RESULTS AND DISCUSSIONS

The accuracy, precision, recall, F1-score, and a confusion matrix are utilised for the evaluation of the three models: Logistic Regression, Decision Tree, and Random Forest. These results will show the relative strengths and weaknesses of each model in respect to predicting the probability of hospital readmissions. Summary of findings:

A. Logistic Regression Results

The baseline model applied in this study is the Logistic Regression model, which is a linear and interpretable clas- sifier. While it is simple yet effective most of the times in classification tasks, it suffered from the inherent complexity and imbalance of the hospital readmission dataset.

At 36% accuracy, this model was capable enough to classify some cases, though not good enough to be practically useful in such a critical domain as healthcare. For the majority class (>30), precision and recall were remarkably high at 36% and 100%, respectively. That indicates the model was identifying the patients robustly who had readmission after more than 30 days. It did show a sharp performance decline for minority classes (<30 and NO), with F1-scores of 0.0. The imbalance therefore confirms the model had a major bias toward the majority class.

Logistic Regression is a model inherently inclined toward data imbalances because it tries to maximize global accuracy and is thus a linear model. For this problem, with a difference of more than 30-fold in cases between classes (>30, <30, and NO), the model is biased toward fitting the majority class (>30). This achieved excellent recall for the >30 class, but at the expense of virtually ignoring the minority classes, hence the zero F1-scores in those categories. Such a result makes the necessity of handling class imbalance during pre-processing evident. Techniques such as SMOTE are used to counterbalance class imbalance. This study employed such an approach. This poor performance of Logistic Regression here underscores a limitation of using it as a baseline model for imbalanced multi-class datasets. While being interpretable is a great feature, the failure to predict accurately in all classes overshadows that. However, in its present form, the model may not be practical without highly significant improvements, like feature engineering or ensemble methods, for high-risk healthcare applications wherein the cost of misclassification may be very high.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

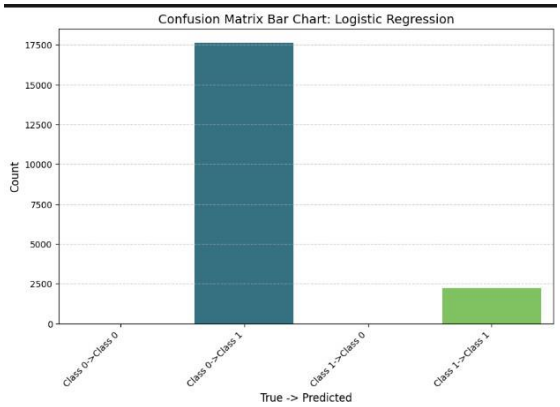

Fig. 4. Confustion Matrix Logistic Regression

B. Decision Tree Results

In this case, the Decision Tree model has shown tremendous advantages over Logistic Regression based on its good ability to capture non-linear patterns and interactions in the data. The model attained 47% accuracy and performed well over the classes, although problems at predicting the minority classes were clearly self-evident.

For the > 30 class (readmissions with in 30 days), the model achieved a precision of 17%, recall of 19%, and an F1-score of 18%. These metrics were not much better than in Logistic Regression but would highlight how challenging it was to predict this minority class, mainly because of its un-derrepresentation in the dataset. Even though SMOTE helped alleviate the imbalance somewhat, it was still required that more measures should be taken to increase sensitivity for this class.

The >30 class (readmissions after 30 days) performed well, yielding a precision of 41%, recall of 42%, and an F1-score of 41%. These suggest that the model was able to capture the main patterns associated with the patients in the majority class, but further gains in precision could reduce false positives in this class.

For the NO class (non-readmitted patients), the model achieved its best performance, with a precision of **59%**, recall of **56%**, and an F1-score of **57%**. This demonstrates the Decision Tree's capability to identify non-readmitted patients effectively, likely because of the class's larger representation in the dataset.

Overall, the macro-average F1-score of 39% suggests better generalization compared with Logistic Regression, but the model still could not balance performance across all classes. The Decision Tree was less biased toward the majority class as well as compared with logistic regression, but it captured more complex patterns that exist in the data. At the same time, however, it is still biased to the larger classes: (> 30 and NO). The tree-based structure of the Decision Tree makes it inherently interpretable. This provides a clear view of the decision-making process, allowing for an understanding of the most influential features, such as time_in_hospital, num_medications, and number_diagnoses. These features played a significant role in improving predictions and offer actionable insights for healthcare professionals aiming to reduce patient readmission rates.

In conclusion, the Decision Tree model outperforms Logis- tic Regression in terms of both accuracy and interpretability. However, the results call for a further refinement in order to deal with the class imbalance issue and make a better performance on minority classes. Suggested visualizations include a confusion matrix that will illustrate class-wise pre- diction performance, a precision-recall chart that will illustrate the respective trade-off among different classes, and feature importance plot to highlight the most impactful predictors.

C. Random Forest

The Random Forest model was correct more often than both Logistic Regression and Decision Tree models: 56% accuracy

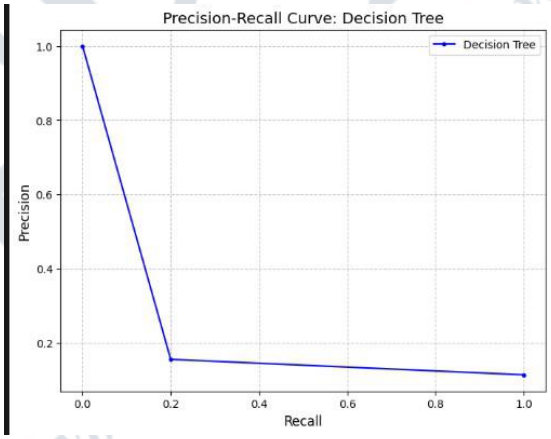

Fig. 5. Precision Recall Decision Tree

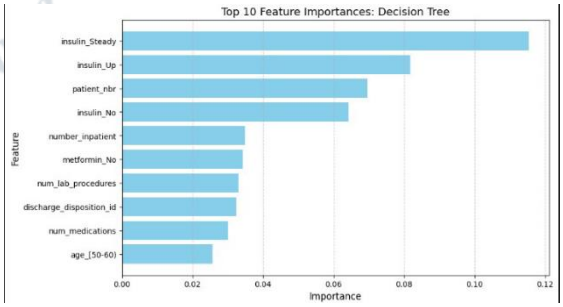

Fig. 6. Feature Importance of Decision Tree

for the Random Forest, making it more sensitive to patient readmission. Because the model consists of an ensemble of de- cision trees, Random Forest significantly improves the quality of predictions by reducing overfitting and capturing complex patterns in data, thus calling for its ability to generalize much better across classes of the dataset.

The precision for the minority class ¡30 (readmission within 30 days) was about 31%, recall 4%, and F1-score 8%. Despite the fact that prediction of this class remained challenging, the Random Forest model showed slight precision advantages over the Decision Tree model, but it

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 12, Issue 03, March 2025

better perceives less obvious patterns. Majority class >30 - readmissions more than 30 days- Precision and recall are at 47% and 42%, respectively, with the F1-score at 45%. This was actually balanced with stable performance. NO Class (not readmitted patients) showed the best performance - precision at 61% recall at 77% and F1-score at 68%. This shows that the model is strong in identifying not readmitted patients.

The Random Forest model carries significant advantages over Logistic Regression and Decision Tree models, and it further robustness and reduced variance significantly. Using the ensemble learning approach for averaging predictions from multiple decision trees effectively managed to capture complex relationships between features and outcomes, leading to excellent generalization across all classes in a performance that no previous model possessed in any strengths.

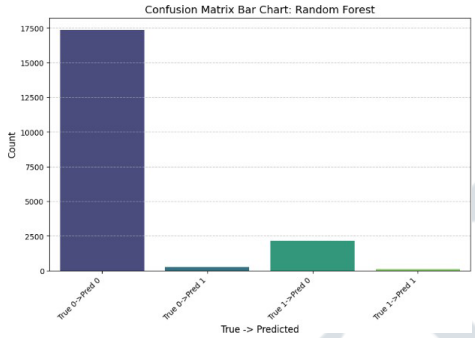

Fig. 7. Confusion Matrix of Random Forest

One of the significant strengths of Random Forest is that it can identify key predictors that may lead to hospital readmission. Established key risk factors for hospital readmission involved time_in_hospital, num_medications, and number_diagnoses. Such findings provide actionable information to healthcare providers in identifying at-risk patients more efficiently; thus, Random Forest gets closer to applied use in real-world clinical practice.

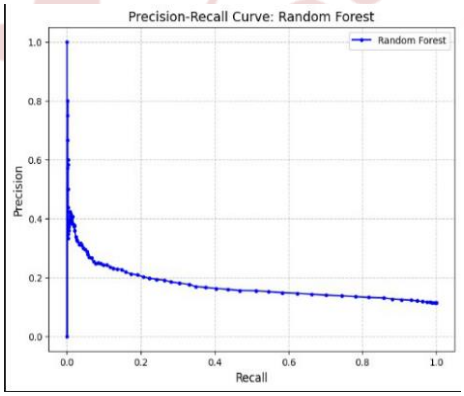

Fig. 8. Precision Recall Chart Random Forest

The model also showed balanced generalization across all classes unlike Logistic Regression and Decision Tree, which showed bias towards the majority class. This observed the model score a macro-average F1-score of 40%, which was consistent for all classes. Still, even though SMOTE is used to balance the classes, it was still very hard to predict the class

>30. Although the precision for the minority class improved with the Decision Tree model, more optimization will ensure that recall for these cases increases.

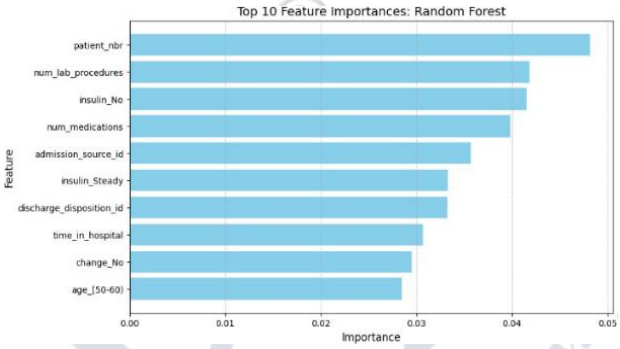

Fig. 9. Feature Imporatnce of Random forest

D. Comparative Analysis

The table below summarizes the performance metrics for all three models: Comparative analysis points out that while

Table I: Comparison of Model Performance Metrics

Metric	Logistic Regression	Decision Tree	Random Forest
Accuracy	36%	47%	56%
Precision	0.13	0.48	0.53
Recall	0.36	0.47	0.56
F1-Score	0.19	0.47	0.53
Macro Avg F1-Score	0.18	0.39	0.40

Logistic Regression provides a baseline, Random Forest stands as the strongest model with higher accuracy and better F1-score. Decision Tree, although not very accurate, is really good in terms of recall and may be used in those scenarios where a high-risk patient has to be diagnosed.

V. CONCLUSION

This study assessed the application of three machine learning models, namely Logistic Regression, Decision Tree, and Random Forest to utilize patient information in predicting readmission back to the hospital. From this experiment, Random Forest came out as the best model with 56% accuracy while still helping identify predictors such as time_in_hospital, num_medications, and number_diagnoses, thus it indicates the promising use of data-driven approaches to inform healthcare decision-making through better efficiency in detecting patients at a risk of being readmitted into the hospital.

However, the results also brought significant challenges, particularly in predicting early readmissions (>30 days).

Even after balancing the dataset using SMOTE, the minority class still didn't get easy to predict, thus bringing complexity in trying to model healthcare outcomes in a good way. The models showed reasonable performance for the majority and non-readmitted classes but was a challenge to bring balanced accuracy across all classes.

Based on the health care requirements, interpretability of Logistic Regression and the strong generalization ability of the Decision Tree and Random Forest models make model selection essential. Thus, the study is structured to demonstrate that machine learning can be used to bring actionability to reduce readmission rates and improve patient care either by priority to interpretability for clinician or reaching high accuracy for operational decision-making.

VI. FUTURE DIRECTIONS

The findings from this study open up several avenues for future improvement to optimize predictive performance and real-world utility:

A. Integration of Relevant Variables

Future models can consider incorporating relevant social determinants of health, such as socioeconomic status, availability of healthcare services, and discharge planning support, to better capture the complex nature of readmissions.

Adding time-series data, such as trends in vital signs or medication changes during hospital stays, could contribute significantly to improving models' ability to predict early readmissions.

B. More Advanced Machine Learning Algorithms

Boosting algorithms, such as Gradient Boosting Machines (GBM) or XGBoost, might be employed to address the performance gap in predicting the minority class (< 30). These algorithms are particularly effective at handling imbalanced datasets.

Ensemble learning methods could also result in higher accuracy and robustness by leveraging the strengths of multiple algorithms.

C. Real-Time Predictive Systems

The models can be deployed within real-time healthcare systems, enabling early identification of at-risk patients during hospitalization.

To facilitate this, optimization for computational efficiency without sacrificing accuracy will be necessary. Integration with electronic health records (EHRs) could allow for seamless prediction and intervention within normal clinical workflows.

D. Patient-Specific Recommendations

By utilizing feature importance from models such as Random Forest, targeted interventions can be developed to address specific risk factors—for example, prolonged hospital stays or complex medication regimens.

Predictive insights could also guide the allocation of resources, such as follow-up calls or home visits, to lower the risk of readmissions.

E. Model Bias

Future research could focus on addressing bias in data and models to ensure consistent performance across different demographic groups. Techniques such as fairness-aware machine learning could be applied to mitigate biases and promote equitable outcomes.

F. Validation with External Datasets

External validation using diverse datasets is essential to assess the stability and generalizability of the models for various patient populations. This step is critical for transitioning these methods from research into real-world applications.

- Incorporation of external datasets will test the scalability of these models for broader populations.
- This transition is vital to bridge the gap between academic research and practical implementation in healthcare sys-tems.

REFERENCES

[1] Hasan, O., et al. (2010). Factors influencing readmission risk. Journal of Hospital Medicine.

[2] Krumholz, H.M., et al. (2009). Predicting heart failure readmissions. Circulation.

[3] LeCun, Y., et al. (2015). Deep learning in healthcare. Nature.

[4] Kansagara, D., et al. (2011). Predicting risk for hospital readmission: a systematic review. JAMA.

[5] Ghassemi, M., et al. (2014). Ensemble learning for hospital readmis- sions. Healthcare Informatics Research.

[6] Xiao, C., et al. (2018). Support Vector Machines in healthcare. IEEE Transactions on Knowledge and Data Engineering.

[7] Choi, E., et al. (2016). Temporal EHR analysis with deep learning. Scientific Reports.

[8] Zhou, X., et al. (2019). Social determinants and readmissions. BMJ Open.

[9] Lipton, Z.C., et al. (2016). Handling missing data in EHRs. ICLR.

[10] Estiri, H., et al. (2020). Data imputation strategies in predictive model- ing. Journal of Biomedical Informatics.

[11] Chawla, N.V., et al. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research.

[12] Johnson, A., et al. (2016). SMOTE in healthcare datasets. Critical Care Medicine.

[13] Rajkomar, A., et al. (2018). Comparative analysis of machine learning models in healthcare. PLOS ONE.
[14] Miotto, R., et al. (2016). LSTMs for healthcare predictions. Journal of Machine Learning Research.
[15] Nguyen, A., et al. (2020). CNNs for clinical note analysis. JAMA Network Open.
[16] Guyon, I., et al. (2003). Feature selection techniques. Machine Learning Journal.
[17] Suresh, H., et al. (2017). Mutual information for feature selection. NIPS.
[18] Obermeyer, Z., et al. (2019). Bias in healthcare algorithms. Science.
[19] Goldstein, B.A., et al. (2017). Real-world hospital readmission models. Health Affairs.
[20] Futoma, J., et al. (2015). Real-time predictive analytics in healthcare. Journal of Healthcare Informatics.