Short-Term and Long-Term Readmission Prediction in Uncontrolled Diabetic Patients using Machine Learning Techniques

Monira Mahmoud[1], Mohamed Bader[1] and James McNicholas[1,2]
[1]University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, U.K.
[2]Queen Alexandra Hospital, Portsmouth Hospitals NHS Trust, U.K.

Keywords: Machine Learning, Data Mining, Diabetes, Uncontrolled Diabetes, Readmission.

Abstract: Diabetes is a chronic disease and major health problem which leads to many complications if not managed probably. Hyperglycemia, or raised blood sugar, is a common effect of Uncontrolled diabetes that may leads overtime to serious complications, especially in the nerves and blood vessels. As well as leads to repeated hospital admission. The main purpose of this study is to help clinicians to improve healthcare of uncontrolled diabetic patients through using machine learning as a tool in decision making, consequently this will improve patient care and reduce the readmission which considered a medical quality measurement and cost reduction objective. This study aims to predict the hospital readmission of the uncontrolled diabetic patient who is considered more susceptible to developing life-threatening diabetes complications and based on the Diabetes 130-US hospitals dataset. Several machine learning employed to predict the short term (within 30 days), and both short and long-term readmission (within or after 30 days) of uncontrolled diabetic patient. As expected, the results are in line with other research in the literature. For the first scenario of whole readmission prediction, our model achieved a better accuracy of 64.5 % with SVM and attribute selection and for the second scenario, RF achieved the highest accuracy of 86.38 % which still come in context with other research in the literature.

1 INTRODUCTION

Diabetes Mellitus (DM) is a major public health problem, Worldwide, 415 million adults—or one in every eleven—are projected to have diabetes. By 2040, there will likely be 642 million individuals living with diabetes worldwide. (diabetes UK organization). According to the world health organization, the number of people with diabetes rose from 108 million in 1980 to 422 million in 2014. Also, within the UK, there are 3.5 million diabetics, up from 1.4 million in 2000. Hospital readmission is an episode when a patient who has been discharged from the hospital is readmitted again within a specified period. Indeed, the burdens of inpatient diabetes is huge, growing, and expensive, and readmission can greatly increase these burdens. Hospital readmission is used as a measure of a hospital's ability to provide quality service and patient care. Also, hospital readmission is often used as a benchmark, since a high proportion of readmission is likely to be preventable if the hospital provided adequate care. Thus, the reduction of readmission is a medical quality measurement and cost reduction objective (Battineni et al, 2020). In particular, uncontrolled diabetes implies high blood sugar levels over a prolonged time even if the patient on treatment, it is diagnosed when HbA1c is higher than 6.5. According to Diabetes UK Organization, HbA1c is one of the tests used to diagnose and monitor the diabetic patient, known as glycated hemoglobin, and refers to average blood glucose levels for the last two to three months. For a diabetic patient, an ideal HbA1c level is 48mmol/mol (6.5). Uncontrolled diabetes can result in hyperglycemia, which damages many of the body's systems, particularly the nerves and blood vessels, over time. Nearly every organ in an uncontrolled diabetic patient's body can suffer a toll from diabetes, including, the eyes, kidneys, nerves, heart, blood vessels, gastrointestinal tract, teeth, and gum. Interestingly, on a daily basis, hospitals generate a great deal of data, but that information usually remains as data that is not always converted into knowledge. Through the application of ML techniques, it is possible to uncover hidden relationships or patterns among the data and convert them into knowledge that can be used by healthcare

Mahmoud, M., Bader, M. and McNicholas, J.
Short-Term and Long-Term Readmission Prediction in Uncontrolled Diabetic Patients using Machine Learning Techniques.
DOI: 10.5220/0011926000003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 5: HEALTHINF*, pages 680-688
ISBN: 978-989-758-631-6; ISSN: 2184-4305

professionals to make better decisions. Prediction of readmission could play a role in early intervention for the management of the uncontrolled diabetic patient who is considered a host of complications if not managed properly. Hence, this study aims to apply a set of machine learning techniques to predict uncontrolled patient readmission. Therefore, predicting readmission will ultimately allow hospitals to better calculate and assess the quality of care.

This study applies machine learning prediction tools for a specific group of diabetic patients (uncontrolled patients), based on UCI diabetes dataset. Moreover, it considers different scenarios for prediction (i.e. short-term or short- and long-term readmission prediction) with feature selection. Six supervised ML technique used for the prediction (RF, NB, KNN, Ada-Boost, SVM, bagging, and NN) of readmission. The study benefits from two scenarios. The first scenario (i.e. using the first subset of data) predicts the readmission event, while the second scenario (using subset data two) predicts of the early readmission (readmission within 30 days). Experiments employed for both sub data sets with and without attribute selection. Results shows that, in the first scenario (all readmission events), SVM achieved the highest accuracy of 64 % and NB achieved the best AUROC area of 0.65. In the second scenario (early readmission only), RF achieved the highest accuracy of 86 % and the best AUROC area of 0.63.

Our goal within the healthcare prospective is to use data mining, data analytical and ML to predict If the uncontrolled diabetic patient will be readmitted at any time point, as a first scenario or If will be readmitted in a short-term time (within 30 days), as a second scenario.

This research aims to develop a model that can accurately predict the readmission of uncontrolled diabetic patient. Also, to provide a better understanding of the readmitted patient characteristics through descriptive analysis. Therefore, using ML for prediction of uncontrolled diabetics readmission will boost early intervention, and consequently lead to better disease management and cost reduction

The paper is structured as follows: Section 2 presents the literature review on machine learning (ML) and diabetes. Section 3 discusses the material and methods. Section 4 provides the results, and Section 5 provides conclusion.

2 LITERATURE SURVEY

A large and growing body of literature has investigated the application of machine learning algorithms in the healthcare domain ((Battineni et al, 2020), (Kumar et al, 2018), (Kohli et al, 2018)), (Ali et al, 2020). In particular, a stream of research examines the accuracy of machine learning algorithms in predicting hospital readmission of diabetic patients. The following section discuss the existing literature related to this paper.

2.1 Diabetes and Machine Learning

Diabetes is linked to micro and macrovascular diseases such as heart disease, kidney failure, eye disease, and amputation, which also leads to a high rate of repeated admission of diabetic patients. Moreover, it is fast becoming a key instrument in complicating other unrelated medical conditions like infections, accidents, and surgery. For instance, the United States (US) health system endures a significant economic burden for diabetes care. This cost reached about 327 billion dollars in 2017 (Kavakiotis et al 2017). Nevertheless, the cost of diabetes is not directly related to the diagnosis and management of diabetes itself but also costs generated by long-term complications and their economic and social consequences (Alamer et al, 2019).

Furthermore, uncontrolled diabetes, if not managed properly, often leads to biochemical imbalances that can cause acute life-threatening events and hospitalizations. Evidently, the uncontrolled diabetic patient is nine times higher risk of admission (Boutayeb et al, 2004), three times more susceptible to developing severe periodontitis (Hu et al ,2019), much greater risk for presenting with later stages of diabetic retinopathy, other rare diabetic ocular complications, including glaucoma, cataract, and dry eye disease (Eldarrat et al ,2011). Extent research links uncontrolled diabetes with substantial mortality and cardiovascular disease burden (Alamer et al, 2019) and increases the risk of peroperative complication (Threatt et al ,2013).

Therefore, predicting readmission will ultimately allow hospitals to better calculate and assess the quality of care they provide to their patients (Navarro-Pérez et al ,2018). The readmission of an individual with uncontrolled diabetes falls into the Potentially Preventable Readmission (PPRs) category. Since ambulatory care (outpatient care) plays an important role in diabetes management, most hospitalizations with uncontrolled diabetes are a direct reflection of the quality of primary health care received outside of hospitals (Kim et al, 2010). Accordingly, the Agency for Healthcare Research and Quality (AHRQ) selected uncontrolled diabetes as a prevention quality indicator (PQI) where hospitalization would be decreased through timely and appropriate ambulatory care ((Pujianto et al ,2019), (Kim et al, 2007)).

Machine learning (ML) is a subclass of artificial intelligence technology, where algorithms process large data sets to detect patterns, learn from them, and execute tasks autonomously without being instructed on exactly how to address the problem. There is ample evidence on the rapid increase in Machine learning applications in disease prediction and diagnosis ((Battineni et al,2020), (Kumar et al,2018), (Kohli et al, 2018) and (Ali et al,2020)). Thus, using machine learning to predict the readmission of diabetic patients will play a role in improving the healthcare system by decreasing the negative consequences related to diabetes readmission.

In the context of diabetes, ML methods have been used to detect, predict, and diagnose i.e. bio-marker Prediction and Diagnosis in DM (Farajollahi etal ,2021), Diabetic Complications (Dagliati et al,2018), Drugs and Therapies (Donsa et al, 2015), Genetic Background and Environment (Urban et al, 2018), and Health Care Management which includes the readmission prediction (Sharma et al ,2019). An example, Chaki et. al (2020) surveyed 107 papers that addressed the application of machine learning and artificial intelligence techniques in DM detection, diagnosis, and self-management. Likewise, Dagliati et. al (Dagliati et al., 2018) provides empirical evidence on the importance of ML in predicting the complications of diabetes.

2.2 Related Work

Several studies used the UCI diabetes dataset for the purpose of diabetic patient readmission prediction. However, the results are mixed due to the variation in the data prepossessing and the used ML algorisms. Bhuvan et. al studies both short-term and long-term readmission as two scenarios (Bhuvan et. al,2016). The first scenario considered the short-term readmission versus all readmission cases. The second scenario combined all the readmission cases versus non-readmitted cases. They found that RF was optimal for this task, compared to NB, Ada-Boost, and NN. Moreover, they employed an ablation study to identify risk factors and association rule mining to identify the association across critical risk factors. They found that the number of inpatient visits, discharge disposition, and admission type are the most important for identifying the high risk patient.

Proposing an ensemble model and cluster analysis, Pham et al (Pham et al,2019), investigate the whole readmission events. the final ensemble model was created using the five best models, which were chosen from a pool of 15 models. The final ensemble reaches a 56 % sensitivity while maintaining a 63.5 % accuracy. Using cluster analysis, they identified four unique patient groupings. Their results suggest that patients who have had previous in-patient visits or who received a large amount of treatment during their most recent visit were shown to be more likely to be readmitted.

Addressing short-term readmission, Al-Ars et al (Al-Ars et al, 2022), Farajollahi et al (Farajollahi et al,2021), Sharma et al (Sharma et al ,2019) and Neto et al (Neto et al,2021) explored the accuracy of alternative predictors and the attributes selections for predicting rea mission of diabetic patients.

Sharma et al (Sharma et al ,2019) investigate the prediction of short-term readmission using RF, LR, XGBoost, Adaboost and DT. They concluded that random forest achieved t highest accuracy of 94. They also pointed out the most important 10 attributes which contribute mostly to the hospital readmission of a diabetes patient in case of using RF an DT algorisms, However, the handling out of the prediction attribute not defined.

Furthermore, Al-Ars et al (Al-Ars et al, 2022) studies prediction of the short-term readmission based on the measurement of HbA1c and the primary diagnosis using LR, NB, J8 and comparing the results with and without using t discretization step. They found that the discretization o numerical attributes step improves the performance of N into 93.51.

Applying principal component analysis (PCA) for feature selection, Farajollahi et al (Farajollahi et al,2021) identified three scenarios of attribute selection. In This paper, they employed RF, DT, XGBoost, KNN, AdaBoost, and Deep learning to predict the short-term readmission and found that dee learning achieved the highest accuracy of 86.8%. However, the handling out of the prediction attribute not defined. the study showed that a machine learning model's effectiveness depends on the choice of the prediction model, the numb of selected features, and the number" k" for k-fold validation.

Furthermore, using six different scenarios based on attribute selection, Neto et al (Neto et al,2021) considered the short-term readmission, using RF, J48, NB, IBK, and MLP algorithms. Comparing alternative scenarios, they documented that the best performance is for the RF with an accuracy of 0,898 in the case of the scenarios with the highest number of attributes.

3 MATERIALS AND METHODS

3.1 Materials

This study is based on a dataset obtained from the UCI machine learning repository about diabetic patients (Dua and Graff,2019). The data set contains about 100,000 instances and it includes 55 features from 130 hospitals in the United States for10 years (1999-2008). the attributes describing the diabetic encounters, including demographics, diagnoses, diabetic medications, number of visits in the year preceding the encounter, and payer information.

3.2 Methodology

This research will follow The CRoss Industry Standard Process for Data Mining (CRISP-DM) methodology. The CRISP-DM steps will be described in details next. For the data preparation phase of this study, Excel used for data preparation and WEKA for the Modelling and Evaluation. Excel's usability and the number of classifiers available by WEKA made it the ideal tool for this analysis.

A. Business Understanding

As a measure of a hospital's ability to provide quality service and care, readmissions are often used as a benchmark since most readmissions can be prevented if patients receive adequate treatment. In addition to being a quality indicator of healthcare systems, readmissions are also a financial burden, about 3.3 million readmissions were reported in the United States after 30 days, according to the Agency for Healthcare Research and Quality (AHRQ). The burden of inpatient diabetes is huge, growing, and expensive, and readmission can greatly increase this burden. Nevertheless, reducing readmission rates for diabetics could significantly reduce medical costs while improving care outcomes. The reduction of readmission is a medical quality measurement and cost reduction objective (Battineni et al,2020). As well as the uncontrolled diabetic patent is considered a host of diabetes complications which is considered a cost burden as well. As a result, predicting cases of uncontrolled diabetes patients who are likely to have hospital readmission is the project's commercial goal in order to help decrease the readmission rate.

This graph shows the summary of this research methodology.

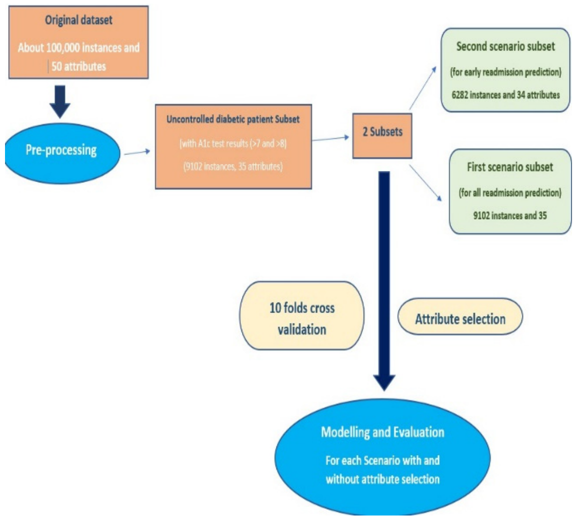

Figure 1: Methodology.

B. Data Understanding

This study is based on a data set obtained from the UCI machine learning repository about diabetic patients (Dua and Graff,2019). The dataset contains about 100,000 instances and it includes 50 features from 130 hospitals in the United States for 10 years (1999-2008). the attributes describing the diabetic encounters, including demographics, diagnoses, diabetic medications, number of visits in the year preceding the encounter, and payer information. The full list of the features and their description is provided in Table1 (Strack et al, 2014).

C. Data Preparation

To ensure that the data is suitable to be used in the various models, the following data prepossessing methods are applied. Figure 2 shows a summary of the prepossessing steps.

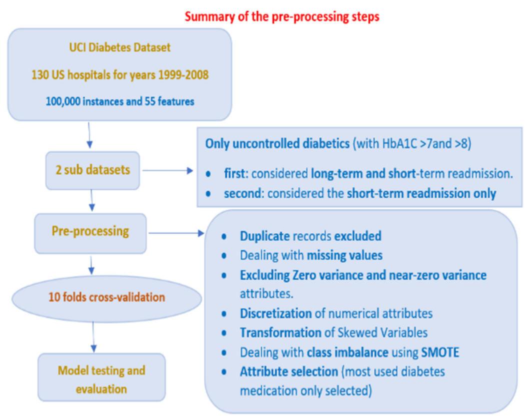

Figure 2: Summary of the prepossessing steps.

Table 1: Data description.

Feature name	Type	Description and Values	% missing
Encounter ID	Numeric	Unique identifier of an encounter	0%
Patient number	Numeric	Unique identifier of a patient	0%
Race	Nominal	Values: Caucasian, Asian, African American, Hispanic, and other	2%
Gender	Nominal	Values: male, female, and unknown/invalid	0%
Age	Nominal	Grouped in 10-year intervals: 0, 10), 10, 20), …, 90, 100)	0%
Weight	Numeric	Weight in pounds.	97%
Admission type	Nominal	Integer identifier corresponding to 9 distinct values, for example, emergency, urgent, elective, newborn, and not available	0%
Discharge disposition	Nominal	Integer identifier corresponding to 29 distinct values, for example, discharged to home, expired, and not available	0%
Admission source	Nominal	Integer identifier corresponding to 21 distinct values, for example, physician referral, emergency room, and transfer from a hospital	0%
Time in hospital	Numeric	Integer number of days between admission and discharge	0%
Payer code	Nominal	Integer identifier corresponding to 23 distinct values, for example,Blue Cross/Blue Shield, Medicare, and self-pay	52%
Medical specialty	Nominal	Integer identifier of a specialty of the admitting physician, corresponding to 84 distinct values, for example, cardiology, internal medicine, family/general practice, and surgeon	53%
Number of lab procedures	Numeric	Number of lab tests performed during the encounter	0%
Number of procedures	Numeric	Number of procedures (other than lab tests) performed during the encounter	0%
Number of medications	Numeric	Number of distinct generic names administered during the encounter	0%
Number of outpatient visits	Numeric	Number of outpatient visits of the patient in the year preceding the encounter	0%
Number of emergency visits	Numeric	Number of emergency visits of the patient in the year preceding the encounter	0%
Number of inpatient visits	Numeric	Number of inpatient visits of the patient in the year preceding the encounter	0%
Diagnosis 1	Nominal	The primary diagnosis (coded as first three digits of ICD9); 848 distinct values	0%
Diagnosis 2	Nominal	Secondary diagnosis (coded as first three digits of ICD9); 923 distinct values	0%
Diagnosis 3	Nominal	Additional secondary diagnosis (coded as first three digits of ICD9); 954 distinct values	1%
Number of diagnoses	Numeric	Number of diagnoses entered to the system	0%
Glucose serum test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>200,” “>300,” “normal,” and “none” if not measured	0%
A1c test result	Nominal	Indicates the range of the result or if the test was not taken. Values: “>8” if the result was greater than 8%, “>7” if the result was greater than 7% but less than 8%, “normal” if the result was less than 7%, and “none” if not measured.	0%
Change of medications	Nominal	Indicates if there was a change in diabetic medications (either dosage or generic name). Values: “change” and “no change”	0%
Diabetes medications	Nominal	Indicates if there was any diabetic medication prescribed.Values: “yes” and “no”	0%
24 features for medications	Nominal	For the generic names: metformin, repaglinide, nateglinide, chlorpropamide, glimepiride, acetohexamide, glipizide, glyburide, tolbutamide, pioglitazone, rosiglitazone, acarbose, miglitol, troglitazone, tolazamide, examide, sitagliptin, insulin, glyburide-metformin, glipizide-metformin, glimepiride-pioglitazone, metformin-rosiglitazone, and metformin-pioglitazone, the feature indicates whether the drug was prescribed or there was a change in the dosage. Values: “up” if the dosage was increased during the encounter “down” if the dosage was decreased, “steady” if the dosage did not change, and “no” if the drug was not prescribed	0%
Readmitted	Nominal	Days to inpatient readmission. Values: “<30” if the patient was readmitted in less than 30 days, “>30” if the patient was readmitted in more than 30 days, and “No” for no record of readmission.	0%

•Missing Data:

weight attribute (97 % missing) was considered to be too sparse and it was not included in further analysis. Furthermore, the payer code attribute is considered irrelevant to the outcome as well as it has a high percentage of missing values so it is excluded too. “medical specialty” refers to the specialty of attending physician which has some missing data so we fill "Missing" in the missing place as this is an important feature for analysis.

•Zero Variance Attributes:

Troglitazone, acetohexamide,citoglipton,glimepirie pioglitazone,metformin pioglitazone, and examide were excluded as no patients on these drugs.

•Near Zero Variance Attributes:

metformin rosiglitazone, glipizide-metformin, tolazamide, tolbutamide, chlorpropamide, and miglitol were excluded as there are only very few cases with steady doses (less than 10 instances).

•Transformation of Skewed Variables:

Age attribute is categorised into 3 distinct groups based on trends proposed by Beata Strack et al (Strack et al,2014). Admission Type id, admission source, and discharge disposition id attribute are categorised with similar categories merged.

•Discretization:

The three diagnosis results are given in icd-9 coding discretized into 9 groups. As well as discretization applied to the numerical attributes (time in hospital, number medications, number lab procedures, number procedures, number outpatient, number emergency and number inpatient) discretized in to 5 pins, using unsupervised splitting technique based on a specified number of bins.

•Class Imbalance:

SMOTE is used to balance the prediction variable classes. For the purpose of uncontrolled diabetic patient readmission prediction.

At the end ,2 sub data sets were extracted from the original one. The first subset for the prediction of all readmission cases (within 30 days or after 30 days counted as yes). The second subset for the prediction of the early readmission cases (excluding all readmission after 30 days).

This ended with 35 attributes in the first subset (long- term and short-term readmission) and 9102 instances. A 6282 instances and 34 attributes in the second subset (short- term readmission only).

The following charts shows the distribution of the prediction variable (readmission) in the data set and the two sub sets. Chart 1 and 2 shows the distribution of the readmission through the subsets. Chart 3 shows the distribution of the excluded data set for uncontrolled diabetics.

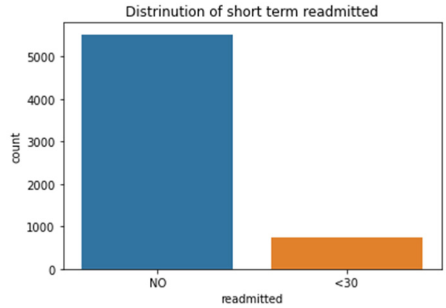

Chart 1: Distribution of short-term readmission.

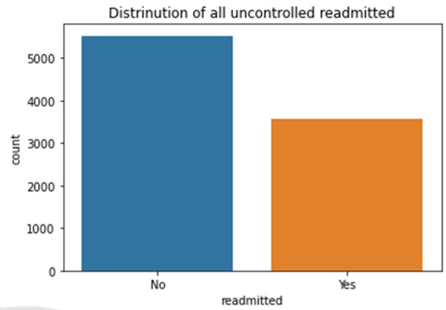

Chart 2: Distribution of uncontrolled diabetic patient.

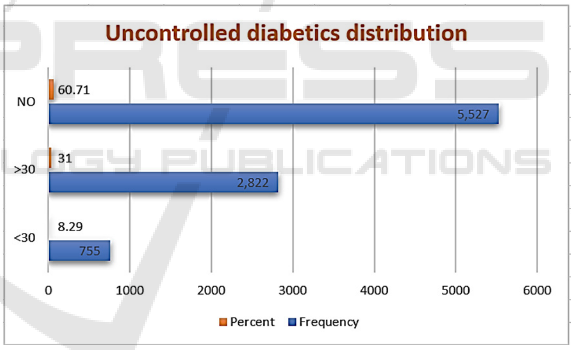

Chart 3: distribution of the excluded data set for uncontrolled diabetics.

D. Modelling

WEKA was the tool that has been chosen for this step because of its variety of classification methods, this study used a tree-based Algorithm (RF), a Bayesian learning algorithms (NB), a function algorithm (SVM, NN), a meta algorithm (Ad boost), and a lazy algorithm (KNN). After choosing the algorithms, Sampling been done with

10 folds cross validation while 30 % used for the test set and 70 % used for the training set. Cross validation has been used to give the model an opportunity to be trained on multiple (10) train test splits as well as it reduces over fitting. Also, all these

algorithms used through the filtered classifier algorism in WEKA to apply the over sampling for the training set only not on both training and test sets.

Finally, two prediction scenarios have been developed to compare the results. The first scenario for the prediction of all readmission cases (within 30 days or after 30 days). The second scenario for the prediction of the early readmission cases (excluding all readmission after 30 days). the output of this step were 2 sub data sets one for the early readmission prediction and the second for the readmission prediction.

E. Evaluation

The basic performance parameters this study considers are the model accuracy and AUROC (Area Under curve for the ROC). While AUROC is the measure of the ability of a classifier to distinguish between classes, the accuracy is the fraction of predictions our model got right.

4 RESULTS

For the first scenario, SVM achieved the highest accuracy of 64.2 % and NB achieved the best AUROC area of 0.65. For the second scenario, RF achieved the highest accuracy of 86.38 % and the best AUROC area of 0.63. Table 2 summarizes the results:

Table 2: Results summary.

The first scenario (considering all readmitted cases)						
	RF	NB	SVM	AdaBoost	KNN	Neural Network
35 attributes						
AUROC	0.64	**0.65**	0.60	0.62	0.57	0.60
Accuracy	63.03 %	61.7 %	**64.2 %**	62.3 %	57.7 %	59.5 %
25 attributes (all medication excluded except insulin and metformin)						
AUROC	0.64	**0.65**	0.60	0.63	0.55	0.59
Accuracy	62.8 %	61.6 %	**64.5 %**	62.9 %	57.3 %	58.7%
The second scenario (short term readmission within 30 days)						
	RF	NB	SVM	AdaBoost	KNN	Neural Network
34 attributes						
AUROC	**0.63**	0.60	0.55	0.58	0.56	0.56
Accuracy	**86.3 %**	74.4 %	78.3 %	70.5 %	68.6 %	82.3 %
25 attributes (all medication excluded except insulin and metformin)						
AUROC	**0.63**	0.61	0.55	0.59	0.58	0.56
Accuracy	**85.6 %**	72 %	76.1 %	68.6 %	70.4 %	79.4 %

Also, as noticed from the results, the second scenario shows a much better accuracy of 86 %, but the first scenario shows a little better AUROC (.65). This figure compares the AUROC and accuracy for each algorithm in both scenarios.

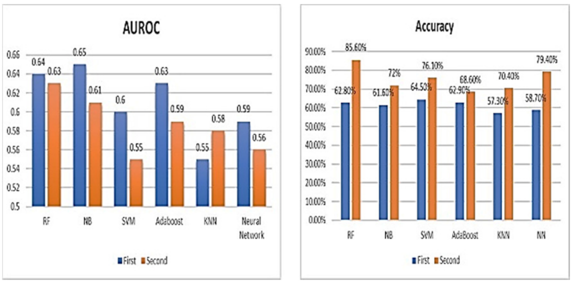

Chart 4: Results graph.

5 DISCUSSION

Although this research targets uncontrolled diabetics and all research in literature targets all diabetic patients. The results comes, as expected, in context with other research in literature, especially the whole readmission predictions as example ,Pham et al ensemble model achieved an accuracy of 63 % accuracy , our model achieved a better accuracy of 64.5 % with SVM and attribute selection .For the second scenario for the short term uncontrolled diabetic readmission prediction, although, Sharma et al RF model achieved 94 % accuracy and Alars et al NB model achieved 93.5 % ,our model is still in context with other research in literature, as example Neto et al RF model achieved 89.8 % and Farajollahi et al achieved 86.8 % using deep learning. the difference in the data sample used in this research (uncontrolled diabetic patient may explain the difference in accuracy with other researchers.

6 CONCLUSION

In this study, several machinebased methods were proposed to predict short-term and long-term uncontrolled diabetic readmission. SMOTE-based data pre-processing is introduced to address the imbalanced data. In addition, comparisons have been done between Random forest, Neural network, KNN, Naïve Bayes, SVM, and Adaboost. The experimental results indicate that in the first scenario, SVM outperforms other methods in the prediction of short-term and long-term readmission with an accuracy of 64 % but NB achieved a better AUROC 0.65 in both cases with and without attribute selection. Also, In the second scenario, the prediction of early readmission with Random forest outperforms other methods with an accuracy of 86,38 % and an AUROC of 0.63 in both experiments with and without attribute selection.

In this study, uncontrolled diabetic patients are targeted; nevertheless, we expect that this early study will pave the way for future research that can improve the accuracy of readmission risk estimates for other health conditions like heart and kidney diseases. Also, an improved data set, including other important features such as age, weight, and laboratory values, could prove valuable and warrant further study.

REFERENCES

Alamer, A. A., Patanwala, A. E., Aldayyen, A. M., & Fazel, M. T. (2019). Validation and comparison of two 30-day re-admission prediction models in patients with diabetes. Endocrine Practice, 25(11), 1151-1157.

Al-Ars, Z. T., & Aldabbagh, A. M. (2021). Predicting the Early Re-admission of Diabetic Patients Using Different Data Mining Techinques. In 2021 Fourth International Conference on Electrical, Computer and Communication Technologies (ICECCT) (pp. 1-8). IEEE.

Ali, F., El-Sappagh, S., Islam, S. R., Kwak, D., Ali, A., Imran, M., & Kwak, K. S. (2020). A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. Information Fusion, 63, 208-222.

Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of machine learning predictive models in the chronic disease diagnosis. Journal of personalized medicine, 10(2), 21.

Bhuvan, M. S., Kumar, A., Zafar, A., & Kishore, V. (2016). Identifying diabetic patients with high risk of readmission. arXiv preprint arXiv:1602.04257.

Boutayeb, A., Twizell, E. H., Achouayb, K., & Chetouani, A. (2004). A mathematical model for the burden of diabetes and its complications. Biomedical engineering online, 3(1), 1-8.

Dagliati, A., Marini, S., Sacchi, L., Cogni, G., Teliti, M., Tibollo, V., & Bellazzi, R. (2018). Machine learning methods to predict diabetes complications. Journal of diabetes science and technology, 12(2), 295-302.

Donsa, K., Spat, S., Beck, P., Pieber, T. R., & Holzinger, A. (2015). Towards personalization of diabetes therapy using computerized decision support and machine learning: some open problems and challenges. In Smart Health (pp. 237-260). Springer, Cham.

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science

Eldarrat, A. H. (2011). Diabetic patients: their knowledge and perception of oral health. Libyan Journal of Medicine, 6(1), 5691.

Farajollahi, B., Mehmannavaz, M., Mehrjoo, H., Moghbeli, F., & Sayadi, M. J. (2021). Diabetes diagnosis using machine learning. Frontiers in Health Informatics, 10(1), 65.

Farajollahi, B., Mehmannavaz, M., Mehrjoo, H., Moghbeli, F., & Sayadi, M. J. (2021). Predicting hospital readmission of diabetic patients using machine learning. Frontiers in Health Informatics, 10(1), 74.

Hu, P., Li, S., Huang, Y. A., & Hu, L. (2019, June). Predicting hospital readmission of diabetics using deep forest. In 2019 IEEE International Conference on Healthcare Informatics (ICHI) (pp. 1-2). IEEE.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and structural biotechnology journal, 15, 104-116.

Kim, H., Ross, J. S., Melkus, G. D., Zhao, Z., & Boockvar, K. (2010). Scheduled and unscheduled hospital readmissions among diabetes patients. The American journal of managed care, 16(10), 760.

Kim, S. (2007). Burden of hospitalizations primarily due to uncontrolled diabetes: implications of inadequate primary health care in the United States. Diabetes care, 30(5), 1281-1282.

Kohli, P. S., & Arora, S. (2018, December). Application of machine learning in disease prediction. In 2018 4th International conference on computing communication and automation (ICCCA) (pp. 1-4). IEEE.

"Kumar, P. M1. Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of machine learning predictive models in the chronic disease diagnosis. Journal of personalized medicine, 10(2), 21.., Lokesh, S., Varatharajan, R., Babu, G. C., & Parthasarathy, P. (2018). Cloud and IoT based disease prediction and diagnosis system for healthcare using Fuzzy neural classifier. Future Generation Computer Systems, 86, 527-534."

Navarro-Pérez, J., Orozco-Beltran, D., Gil-Guillen, V., Pallares, V., Valls, F., Fernandez, A., & Tellez-Plaza, M. (2018). Mortality and cardiovascular disease burden of uncontrolled diabetes in a registry-based cohort: the ESCARVAL-risk study. BMC cardiovascular disorders, 18(1), 1-9.

Neto, C., Senra, F., Leite, J., Rei, N., Rodrigues, R., Ferreira, D., & Machado, J. (2021). Different scenarios for the prediction of hospital readmission of diabetic patients. Journal of Medical Systems, 45(1), 1-9.

Pham, H. N., Chatterjee, A., Narasimhan, B., Lee, C. W., Jha, D. K., Wong, E. Y. F., & Chua, M. C. (2019, July). Predicting hospital readmission patterns of diabetic patients using ensemble model and cluster analysis. In 2019 International Conference on System Science and Engineering (ICSSE) (pp. 273-278). IEEE.

Pujianto, U., Setiawan, A. L., Rosyid, H. A., & Salah, A. M. M. (2019). Comparison of Naïve Bayes Algorithm and Decision Tree C4. 5 for Hospital Readmission Diabetes Patients using HbA1c Measurement. Knowl. Eng. Data Sci., 2(2), 58-71.

Sharma, A., Agrawal, P., Madaan, V., & Goyal, S. (2019, June). Prediction on diabetes patient's hospital readmission rates. In Proceedings of the Third International Conference on Advanced Informatics for Computing Research (pp. 1-5).

Sharma, A., Agrawal, P., Madaan, V., & Goyal, S. (2019, June). Prediction on diabetes patient's hospital readmission rates. In Proceedings of the Third International Conference on Advanced Informatics for Computing Research (pp. 1-5).
Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records. BioMed research international, 2014.
Threatt, J., Williamson, J. F., Huynh, K., Davis, R. M., & Hermayer, K. (2013). Ocular disease, knowledge and technology applications in patients with diabetes. The American journal of the medical sciences, 345(4), 266-270.
Urban, G., Tripathi, P., Alkayali, T., Mittal, M., Jalali, F., Karnes, W., & Baldi, P. (2018). Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology, 155(4), 1069-1078.