5th African International Conference on Industrial Engineering and Operations Management, Johannesburg/Pretoria, South Africa, April 23 - 25, 2024
Publisher: IEOM Society International, USA
Published: April 23, 2024 DOI: 10.46254/AF05.20240252

Predictive Model for Hospital Readmission of Diabetic Patients

Luyanda Mpofu, Belinda Ndlovu and Sibusisiwe Dube
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
luyandampofu2@gmail.com; belinda.ndlovu@nust.ac.zw;
sibusisiwe.dube@nust.ac.zw

Martin Muduva
Department of Information and Marketing Science
Midlands State University
Gweru, Zimbabwe
mmuduvah@gmail.com

Fungai Jacqueline
ICT and Electronics Department
School of Engineering Science and Technology
Chinhoyi University of Technology
Chinhoyi, Zimbabwe
fungaijacquelinekiwa@gmail.com

Kudakwashe Maguraushe
Department of Information and Communication Technology
Mangosuthu University of Technology
Durban, South Africa
maguraushe.kuda@mut.ac.za

Abstract

Diabetes is recognized as one of the world's most prevalent health problems. As diabetic patients grew, so did the percentage of diabetic hospital readmissions. Early readmissions can impact patient well-being, operational efficiency, and financial burden. This study uses machine learning approaches to predict hospital readmissions among diabetes patients. Data was collected from 130 US hospitals. CRISP-DM is used for analysis. Logistic regression (LR) and random forest (RF) classifiers were implemented. The classifier performance was compared. Random Forest outperformed the other model, with an accuracy of 0.89. The model was chosen to enable practical deployments. Researchers used a web-based interface to get data and receive real-time predictions. The results showed that the predictive model used alongside an interface creates a clear and understandable prediction platform. However, the research might involve various datasets and Deep Learning to improve models and findings, in future studies. Furthermore, the model could explore the integration of machine learning interpretability approaches to increase transparency and promote better comprehension of the model's predictions by healthcare practitioners.

Keywords:
prediction model, readmission, diabetes; machine learning

1. Introduction

Diabetes is a chronic illness characterized either by insufficient insulin production by the pancreas or inefficient insulin utilization by the body to control blood sugar levels (Soh et al. 2020). It is classified into two categories- Type 1 the former and Type 2 the latter. Type 2 diabetes is common among adults and the percentage of people having it is rising at an alarming rate. In Zimbabwe diabetes is estimated to have the highest prevalence rate among African countries, with forecasts suggesting that by 2035, there may be more than a million two hundred thousand (1 200 000) diabetic patients in Zimbabwe (Mureyi et al. 2022). The rate of diabetes admissions has risen recently, accounting for 3.09% of all deaths in Zimbabwe. According to Chopera et al. (2021), diabetes clinics in Zimbabwe's main city welcome thirty to forty new patients each month. Patients with diabetes are more likely to be readmitted to the hospital since their condition is chronic and recurrent. The number of people with diabetes will continue to rise and surpass that of all other global areas since Africa has numerous obstacles to overcome, such as scarce resources for social and health care (Mutunhu et al. 2022). There may be a clear correlation between the growth in diabetes and the increase in readmissions from hospitals among inpatients.

It is possible to utilize machine learning algorithms to predict such situations. When a patient leaves the hospital and returns within a predetermined time frame, it is referred to as a hospital readmission (Sharma et al. 2019). According to Kuguyo et al. (2023), the annual cost of treatment for each patient with diabetes in Zimbabwe is projected to be around one thousand three hundred dollars (US$1,300). In addition to taxing medical resources, readmissions reveal weaknesses in the healthcare system, including patient management and care coordination. According to recent research (Cui et al. 2018), (Hammodeuh et al. 2018), (Goudjerkan and Jayabalan, 2019), some researchers have employed machine learning algorithms to predict diabetes readmissions. However, predicting readmissions and identifying risk factors do provide enough tools that health stakeholders can utilize. Predictive modeling in healthcare can be complicated, counterintuitive, and frequently difficult to explain without more exploration, these qualities contribute to the opaque nature of predictive solutions in the healthcare field which limits their acceptance and usefulness in the clinical setting (Yang 2022).

The analysis will be based on risk factors such as a patient's demographics, admission details, diagnosis, and medical data. This research, therefore, seeks to leverage machine learning algorithms to add more usability to the model, increase the utility of its output, and provide a system usable and understood by stakeholders.A model is not very useful until the end user can access its results (Chumbar 2019). The researcher acknowledges the work done in several studies however, the link to the healthcare providers is missing. This project seeks to expand on this aspect by creating and deploying a predictive model with an easy-to-use interface that identifies high-risk patients for Zimbabwean healthcare personnel. Healthcare professionals will be informed of high-risk patients in real time for prompt interventions.

1.1 Objectives

To train a model.

To predict whether a patient is likely to be readmitted or not.

To alert authorized users when a high-risk patient is identified.

2. Literature Review

Much research has been conducted on diabetes prediction using various machine-learning methods. Relatively few research studies addressed the problem of hospital readmission likelihood prediction in diabetes. Most researchers focused on comparing different machine learning techniques for addressing this problem. Unfortunately, only a few research papers have attempted to address this issue in the literature; most research studies are concerned with forecasting the likelihood of the diseases themselves. So far, attempts have been made to increase hospital readmission predictability. However, due to data quality and volume limitations, only a few models have been proven to be accurate and generalizable enough for readmission prediction (Goudjerkan and Jayabalan 2019). The Neural Network and Multilayer Perceptron models were successful in predicting readmission for diabetic patients with an accuracy of 95%.

Deng and Lauba (2023) built three machine learning models; logistic regression with an elastic net penalty, random forest, and XGBoost to forecast diabetic patients' hospital readmission.

Only a few features were identified to have importance in predicting readmission. Zaky et al. (2021) study used a model that predicted readmission within thirty days (30) for the same chronic based on data normalization analysis. The evidence concluded that Z-score normalization is better than min-max normalization when predicting diabetes readmissions on some clinical data. The scholars highlighted that the model would have a direct effect on healthcare costs and the hospital's efficiency and reputation. Thapa et al. (2022) resorted to Natural Language Processing (NLP) algorithms using clinical admission notes gathered at various time intervals to predict readmissions. NLP approaches generate predictive signals and make it easier to model readmission risk. The findings can help healthcare practitioners with realistic treatment interventions and discharge planning. Lu and Uddin (2022) employed a stacking-based model to tackle this problem. This model was explainable and interpretable to help medical experts and stakeholders understand and visualize the risk factors of readmissions.

Soh et al. (2020) utilized patient-related factors to predict readmissions. The scholars managed to identify that age, gender, race, and comorbidities including heart failure and renal disease, insulin therapy, and insurance status can be used to note whether diabetic patients are likely to be readmitted or not. In addition, Rodriguez-Gutierrez et al. (2019) concluded numerous factors, including the patient's economic status, clinical features, patient-level characteristics, and hospital characteristics were found to have an impact on the odds of an unplanned all-cause readmission. Their findings were consistent with those reported by Robbins et al. (2019); the relationship among these factors can be used to predict readmission of diabetes cases. Hammoudeh et al. (2018) constructed a convolutional neural network to address the issue at hand. They wanted to tell the difference between patients who returned to the hospital and those who did not. They reported an 80% accuracy rate.

3. Method

CRISP-DM was the chosen methodology. The stages comprise business understanding, data understanding, data preparation, modeling, evaluation, and deployment. We consider two traditional ML models. The models under consideration are random forest and logistic regression. We implemented all these models using the Scikit-learn library in Python programming language.

Business Understanding

Hospitals frequently managed diabetes arbitrarily, which raised expenses because of readmissions. Better patient care and lower costs can result from fewer readmissions. Therefore, this study's goal utilize data to combat the problem of diabetic patients' hospital readmissions. To do this, the study developed a machine-learning algorithm that predicts and recognizes whether patients are readmitted or not. Additionally, it comprehended the main causes of these readmissions. Through this effort, hospital managers and other healthcare stakeholders can efficiently distribute resources and cut expenditures. Medical personnel also learn more about patient care and possible areas for development. Better care, fewer problems, and overall better health outcomes are thus given to patients. The accuracy metric gauged the model's ability to make satisfactory predictions.

Data Understanding

The purpose of data understanding was to become acquainted with the dataset, detect issues with data quality, and unearth insights. The diabetic_data.csv file contained the dataset. Aspects of the data included race, patient number, age, gender, kind of admission, length of stay in the hospital, admitting physician's medical specialty, number of lab tests completed, etc. Figure 1 shows the data summary.

[6]:

	count	unique	top	freq	mean	std	min	25%	75%	max	missing_values
encounter_id	101766.0	NaN	NaN	NaN	165201645.622978	102640295.983457	12522.0	84961194.0	230270887.5	443867222.0	0.0
patient_nbr	101766.0	NaN	NaN	NaN	54330400.694947	38696359.346534	135.0	23413221.0	87545949.75	189502619.0	0.0
race	101766	6	Caucasian	76099	NaN	NaN	NaN	NaN	NaN	NaN	0
gender	101766	3	Female	54708	NaN	NaN	NaN	NaN	NaN	NaN	0
age	101766	10	[70-80)	26068	NaN	NaN	NaN	NaN	NaN	NaN	0
weight	101766	10	?	98569	NaN	NaN	NaN	NaN	NaN	NaN	0
admission_type_id	101766.0	NaN	NaN	NaN	2.024006	1.445403	1.0	1.0	3.0	8.0	0.0
discharge_disposition_id	101766.0	NaN	NaN	NaN	3.715642	5.280166	1.0	1.0	4.0	28.0	0.0
admission_source_id	101766.0	NaN	NaN	NaN	5.754437	4.064081	1.0	1.0	7.0	25.0	0.0
time_in_hospital	101766.0	NaN	NaN	NaN	4.395987	2.985108	1.0	2.0	6.0	14.0	0.0
payer_code	101766	18	?	40256	NaN	NaN	NaN	NaN	NaN	NaN	0
medical_specialty	101766	73	?	49949	NaN	NaN	NaN	NaN	NaN	NaN	0
num_lab_procedures	101766.0	NaN	NaN	NaN	43.095641	19.674362	1.0	31.0	57.0	132.0	0.0
num_procedures	101766.0	NaN	NaN	NaN	1.33973	1.705807	0.0	0.0	2.0	6.0	0.0

Figure 1. Data Summary

The researchers encountered particular problems with the data quality, such as a significant percentage of missing values in the variable "weight." The following observations are derived from the data summary:

- There are 101,766 records in the collection.
- A significant number of entries are missing from the weight column.
- There is categorical data in columns like age, gender, and race.

Data Preparation

A dataset is prepared by transforming the label to the needed format, identifying missing values in the dataset, eliminating extraneous columns based on individual comprehension of the data set, and imputing missing values (Shankar and Janikadan, 2019). During this stage, the data is processed to make it optimal for modeling. We focused on missing values, filling missing values with mode, and encoding categorical values. As determined during the data understanding phase, some columns have a high percentage of missing data. The 'weight' column is eliminated. The ‘payer code’ and ‘medical_speciality’ are also eliminated due to an absence of data and are not necessary for this study. The only rows with valid values in the 'gender' column were included, while rows with 'unknown/invalid' genders were removed. The authors replaced missing values in the columns with the column's mode.

The readmitted column represents the goal variable, indicating whether a patient has been readmitted as well as the time frame. To make things easier, the column is converted to a binary result. "1" indicates that the individual returned within 30 days. Otherwise, return "0". There are 11 357 cases when patients were readmitted again within 30 days. 90 409 cases in which patients did not get readmitted nor readmitted after 30 days. ‘Admission_type_id’, ‘Discahrge_disposition_id', and 'admission_source_id' contain string values, the author then mapped them to numerical values for modeling. After data preparation, the dataset now had 35 columns, depreciated from the original 50 columns. The authors used the Sklearn library to partition the data into training and testing sets. Training features (X_train) included 81,410 records with 44 features apiece. Test features (X_test): 20,353 records, each with 44 features. The training target (y_train) is 81,410 records. Test target (y_test): 20,353 records.

Modelling

This study used two machine learning algorithms: the Random Forest and the Logistic Regression algorithm. The best model was chosen after training on the same dataset.

Evaluation

After the models had been trained, the researchers assessed their performance on the test set. For this research, the authors considered the accuracy, precision, recall, and F1 Score metrics, which provide the number of correctly classified occurrences (Neto et al. 2021). This also allowed for the selection of the optimum performing algorithm by assessing the results of the metrics employed in the evaluation process.

Deployment

A model is not very useful until the end user can access its results (Chumbar 2019). The highest-performing model was chosen to enable practical deployments. The deployment of the readmission model involved the creation of a web-based interface for healthcare practitioners to enter data and receive real-time predictions simply. A Flask framework was used to deploy the model.

4. Data Collection

The study's data came from the UCI machine learning repository and included information on people with diabetes. The dataset used in this study encompasses 100,000 instances, missing values, and 55 attributes from 130 hospitals in the US over ten years (1999–2008). The study's foundation is an already-existing, Health Insurance Portability and Accountability Act (HIPAA) compliant dataset devoid of any personally identifiable data. This research was not deemed to include human subjects or require consent because the datasets used were de-identified.

5. Results and Discussion

This research gets results from the final stages of evaluation and deployment. In the evaluation process, we compared the performance of logistic regression and random forest.

5.1 Numerical Results

Figure 2 shows probable associations that might be evident while examining the individual variables as well as comprehending how variables are associated with the readmission of diabetes patients. Seven variables in total are randomly chosen to compute linear correlation. Features of generic medications, such as "Repaglinide" and "Metformin," are eliminated. The findings indicate that there is no linear link between the two variables when the variables are at 0 and have a substantial negative linear correlation when the variables are at +1. At 24, the correlation coefficient is the highest, and at -.09, the lowest. At -.05, there is a moderate negative linear association between "age" and "gender," with an increase in "emergencies" and a drop in "number of procedures." There is a weak negative linear association between "age" and "number of procedures" at -.02. After the age of 75, fewer treatments are performed as people age. The number of procedures and the length of stay in the hospital have a somewhat positive linear connection (.19). The longer patients stay in the hospital, the more lab procedures there are. The positive linear association between "number of diagnoses" and "age" is somewhat strong, at.24. The longer someone stays in the hospital, the more meds they need. There is no link between "number of diagnoses" and "gender," suggesting that readmissions are not influenced by this relationship. The moderate result suggests that there is an acceptable linear relationship between the variables, with some points being near and some being far from the line. It has demonstrated that some feature sets such as length of hospital stay, number of diagnoses, age, and gender are more significant than others.

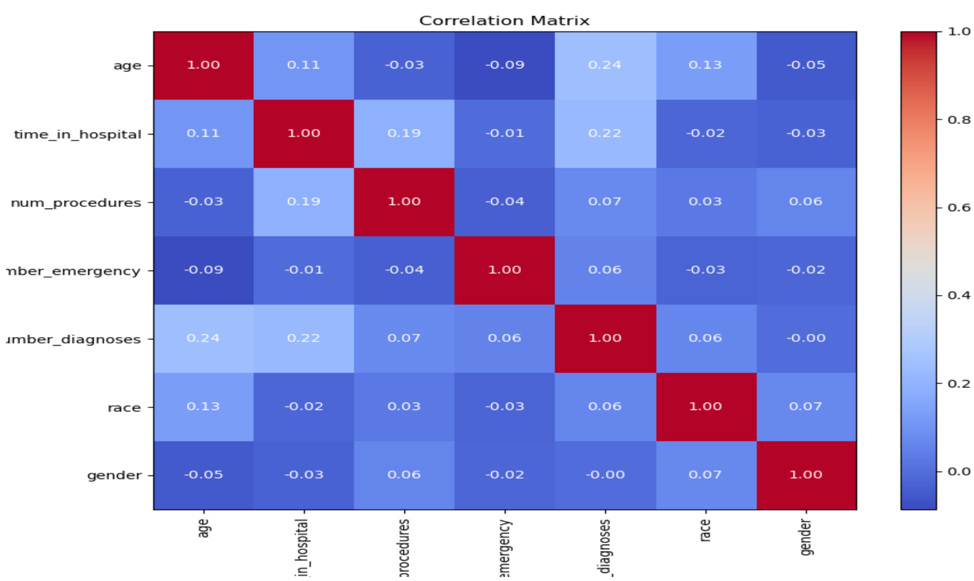

Figure 2. Correlation Matrix

Table 1 depicts the metrics performance for each model. Based on the primary metrics: accuracy, Recall, Precision and F1, a comparison of models is conducted. Those parameters are defined in terms of the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) as indicated in equations (1). When the model correctly predicts the positive category, the result is a true positive (TP); similarly, when the model correctly predicts the negative category, the negative result is true (TN). If the model incorrectly predicts the positive category, the result is a false positive (FP); similarly, when the model incorrectly predicts the negative category, the negative result is false (FN). The following formulas are employed in this study to calculate F1, Accuracy, Recall, and Precision.

Accuracy = (TP+TN)/ (TP+FP+FN+TN). (1)

$$Recall = (TP)/\ (TP+FN) \quad (2)$$

$$Precision = (TP)/\ (TP+FP) \quad (3)$$

$$F1 = 2.\ (Precision.\ Recall)/\ (Precision + Recall) \quad (4)$$

Logistic Regression had lower performance compared to random forest with accuracy (88.83%), recall (72%), precision (66%) and F1 Score (67%). RF achieved the highest accuracy of 88.97%, recall (89%), precision (87%) and F1 Score (85). The outcome of this comparison verifies that the RF is consistent with the published works. The obtained scores are a sign that the model can reasonably predict the likelihood of readmission, which makes it a viable method to predict readmission rates while protecting the privacy and confidentiality of the patient's private medical data. The model excels in capturing complex, non-linear relationships hence the high accuracy.

Table 1. Comparison of results

Algorithm	Accuracy	Precision	Recall	F1 Score
Logistic Regression	88.83	66	72	67
Random Forest	88.97	87	89	85

We then calculated the Random Forest confusion matrix to see the completed prediction tasks more clearly. Figure 3 displays the impressive confusion matrix for the model. The confusion matrix findings demonstrate that 18098 true negatives (TN) were properly predicted by the model. In addition to two false positives (FP) which were mistakenly labelled as "Readmission" whereas they were in fact "No Readmission," it shows the number of cases that the model correctly identified as "No Readmission." 2253 indicates true positives (TP), while 0 false negatives (FN) were mistakenly labeled as "No Readmission" while their true designation was actually "Readmission." The number of cases that the model accurately categorized as "Readmission" is indicated. All things considered; this shows the Random Forest technique has done an impressive task of categorizing cases of diabetic hospital readmission.

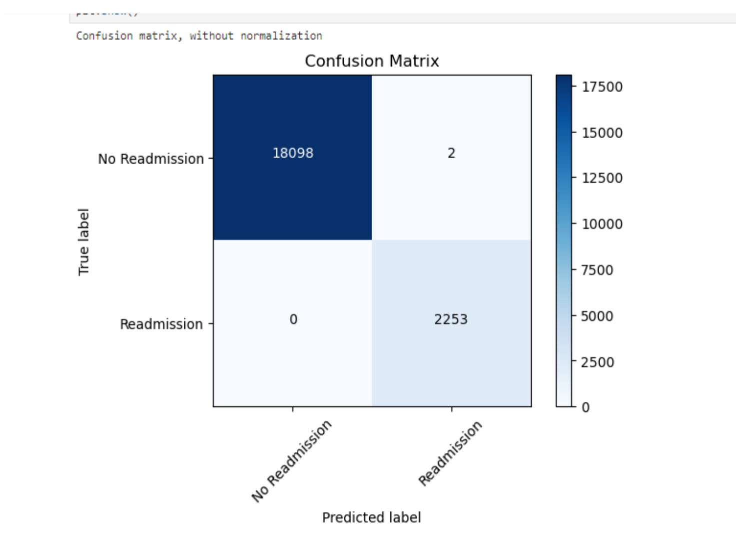

Figure 3. Random Forest Confusion Matrix

5.2 Graphical Results

Random Forest accuracy rises with the test set, suggesting that increasing the size of the training set is unlikely to boost accuracy however, there was no significant difference in the performance between LR and RF. The superiority of random forest over logistic regression appeared to be subtle. Logistic Regression performance was then evaluated on the test set, where it continued to retain an accuracy of 88%. Following the training and testing process, the RF model was embedded into a web service and deployed via Flask. The deployed service acted as the inference point to the model to allow potential users to enter fresh data and get instant responses. The design is straightforward with easy-to-use questions. As shown in Figure 4, the replies are either a numerical input or an input that is rendered accessible via a drop-down menu. The interface not only provides predictions but also alerts users of high-risk patients. This alert serves as an additional measure to assist users in making informed decisions regarding the patient's health. It is worth noting that the interface prioritizes data privacy and confidentiality by ensuring the protection of sensitive information throughout the prediction process. The implementation of this user interface marks a significant milestone in the research, enabling real-world application of the developed model and showcasing its practical utility in identifying and flagging risky patients.

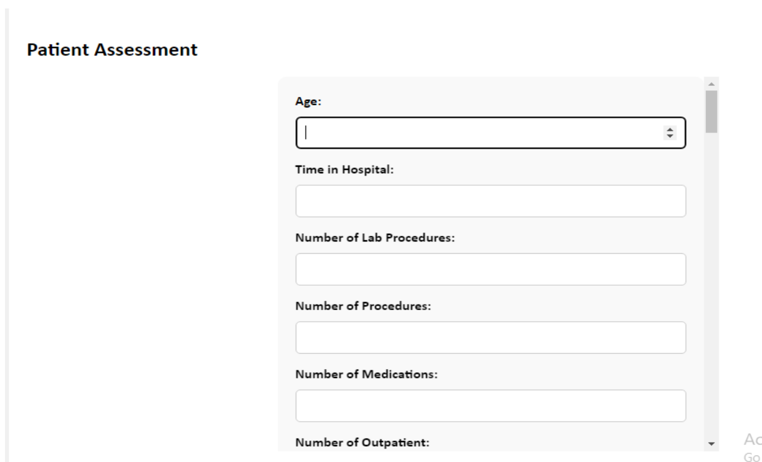

Figure 4. User-Interface

5.3 Proposed Improvements

A healthcare predictive modeling system must be able to explain itself. It is challenging to win over healthcare professionals' trust and integrate predictive models into day-to-day operations in the absence of openness (Yang 2022). Machine learning models must be able to explain why a certain categorization was made in addition to making precise predictions in a clinical context. The user's "trust" in the ML model determines its acceptability, therefore, to gain that trust, it is necessary to justify the reasoning behind the black-box algorithms. In addition, the model may investigate the incorporation of machine learning interpretability techniques to boost openness and encourage healthcare professionals to understand the model's forecasts better. To enhance models and results, the researchers may use a variety of datasets and Deep Learning in subsequent investigations.

6. Conclusion

Readmissions to hospitals have a detrimental impact on hospitals' reputations and increase health care expenses. Therefore, it is quite interesting to forecast hospital readmissions among diabetics. In this research, machine learning was proposed as a useful method for anticipating hospital readmissions in patients with diabetes. When Random Forest is used and assessed against real-world data, it performs better than other machine learning algorithms. The quality of treatment given to patients and health service providers is greatly impacted by hospital readmission. The secret to success is preparation and training, and to guarantee accurate forecasts, our research adopted a thorough, rigorous approach. To generate functional web services, Random Forest, the algorithm with the highest performance, is used. Healthcare professionals can enter information and receive outcomes. This study aimed to contribute to the existing literature on hospital readmissions related to diabetes. Using hospital data on diabetes presents prospects for predictive modelling. An interface expedites the delivery of information by delivering the data to healthcare practitioners so that informed choices can be made.

References

Alajmani, S., and Elazhary, H., Hospital Readmission Prediction using Machine Learning Techniques A Comparative Study, 2019. Available: www.ijacsa.thesai.org.

Deng, L., and Laub, D., Predictive Modeling of Hospital Readmission for Diabetic Patients Using Machine Learning, 2022.

Farajollahi, B., Mehmannavaz, M., Mehrjoo, H., Moghbeli, F., and Sayadi, M.J., Predicting hospital readmission of diabetic patients using machine learning, *Frontiers in Health Informatics*, vol. 10, 2021, https://doi.org/10.30699/fhi.v10i1.266.

Goudjerkan, J., and Jayabalan, M., Predicting 30-day hospital readmission for diabetes patients using multilayer perceptron, *International Journal of Advanced Computer Science and Applications*, 2019. Available: http://researchonline.ljmu.ac.uk/www.ijacsa.thesai.org.

Hammoudeh, A., Al-Naymat, G., Ghannam, I., and Obied, N., Predicting hospital readmission among diabetics using deep learning, in *Procedia Computer Science, Elsevier B.V.*, pp. 484–489, 2018. Available at: https://doi.org/10.1016/j.procs.2018.10.138

Huang, Y., Talwar, A., Chatterjee, S., and Aparasu, R. R., Application of machine learning in predicting hospital. Readmissions: a scoping review of the literature, *BMC Medical Research Methodology*, vol. 21, no. 1, December 2021, Available at: https://doi.org/10.1186/s12874-021-01284-z.

Ismail. M., Praveen, D., and Srikanth, K.S., A Machine Learning Technique for Reducing Hospital Readmissions for Diabetic Diseases, *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 2, pp. 4878–4884, Dec. 2019, Available at: https://doi.org/10.35940/ijitee.B6475.129219.

Kumar V., and Garg, M. L., Predictive Analytics: A Review of Trends and Techniques, 2018,

Lu H, and Uddin S., Explainable Stacking-Based Model for Predicting Hospital Readmission for Diabetic Patients, Information (Switzerland), vol. 13, no. 9, September 2022, Available at: https://doi.org/10.3390/info13090436.

Mureyi, D., Katena, N and Monera-Penduka, T., Perceptions of diabetes patients and their caregivers regarding access to medicine in a severely constrained health system: A qualitative study in Harare, Zimbabwe, *PLOS Global Public Health*, vol. 2, no. 3, March 2022, Available at: https://doi.org/10.1371/journal.pgph.0000255.

Mutunhu, B., Chipangura, B., and Twinomurinzi, H., A Systematized Literature Review: Internet of Things (IoT) in the Remote Monitoring of Diabetes, *Proceedings of Seventh International Congress on Information and Communication Technology*, pp. 649–660, 2022, Available at: https://doi.org/10.1007/978-981-19-1610-6_57.

Neto et al, C., Different Scenarios for the Prediction of Hospital Readmission of Diabetic Patients, *Journal of Medical Systems*, vol. 45, no. 1, January 2021, Available at: https://doi.org/10.1007/s10916-020-01686-4

Robbins, T. D., Keung, L. C., Arvanitis, T. N., and Robbins, T., Risk Factors for Readmission of Inpatients with Diabetes: A Systematic Review, *Journal of Diabetes and its Complications*, vol. 33, no. 5. Elsevier Inc., May 01, 2019. Available: https://doi.org/10.1016/j.jdiacomp.2019.01.004

Rodriguez-Gutierrez, R., Herrin, J., Lipska, K. J., Montori, V. M., Shah, N. D., and McCoy, R. G., Racial and Ethnic Differences in 30-Day Hospital Readmissions among US Adults with Diabetes, *JAMA Network Open*, vol. 2, no. 10, October 2019, Available at: https://doi: 10.1001/jamanetworkopen.2019.13249.

Shang, Y et al., The 30-day hospital readmission risk in diabetic patients: predictive modelling with machine learning classifiers, *BMC Medical Informatics and Decision Making*, vol. 21, July 2021, Available at: https://doi.org/10.1186/s12911-021-01423-y

Sharma, A., Agrawal, P., Goyal, S., and Madaan, V., Prediction on diabetes patient's hospital readmission rates, in *ACM International Conference Proceeding Series, Association for Computing Machinery*, June 2019, Available at: https://doi.org/10.1145/3339311.3339349.

Sheila, M., Magboo, A., Coronel, A., Sheila, M.A., and Coronel, A.D, 30-DAY HOSPITAL READMISSION PREDICTION MODEL FOR DIABETIC PATIENTS WITHIN THE 30-70 AGE GROUP. Available: https://www.researchgate.net/publication/335928725

Soh, J. G. S., Wong, W.P., Mukhopadhyay, A., Quek, S.C., and Tai, B. C., Predictors of 30-day unplanned hospital

Readmission among adult patients with diabetes mellitus: A systematic review with meta-analysis, *BMJ Open Diabetes Research and Care*, vol. 8, no. 1. BMJ Publishing Group, August 11, 2020, Available at: https://doi.org/10.1136/bmjdrc-2020-001227.

Tavakolian, A., Rezaee, A., Hajati F., and Uddin, S., Hospital Readmission and Length of Stay Prediction Using an Optimized Hybrid Deep Model, 2023, Available at: https://doi.org/10.20944/preprints202307.0320.v1

Yang, C.C., Explainable Artificial Intelligence for Predictive Modeling in Healthcare, *Journal of Healthcare Informatics Research*, vol. 6, no. 2, pp. 228–239, June 2022, Available at: https://doi.org/10.1007/s41666-022-00114-1.

Zaky, E. H., Soliman, M. M., Elkholy, A. K., and Ghali, N. I., Enhanced Predictive Modelling for 30-Day Readmission Diabetes Patients Based on Data Normalization Analysis, *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 4, pp. 204–216, August 2021, Available at: https://doi.org/10.22266/ijies2021.0831.19.

Biographies

Luyanda Mpofu is a National University of Science and Technology student pursuing an informatics degree. She has gained valuable experience as an IT Technician and Statistician at Simbisa Brands. She also has a CCNA qualification, and she is working towards other credentials right now.

Belinda Mutunhu Ndlovu is a Ph.D. in Information Systems student at UNISA. She holds an MSc in Information Systems and a BSc in Computer Science. She is a seasoned software developer and academic. She has published several papers in the fields of Data Analytics, Health Informatics, ICT4D, and 4IR.

Sibusisiwe Dube is an experienced lecturer of Information Systems and Computer Science courses. She holds a PhD in Information Systems, an MSc in Computer Science, and a BSc in Information Systems. She has been lecturing since 2004. She is also an active researcher and supervisor of Postgraduate dissertations and undergraduate student projects.

Fungai Jacqueline Kiwa holds a Doctor of Philosophy Degree in Cultural Heritage and Information Technology from Chinhoyi University of Technology, complemented by a Post Graduate Diploma in Higher Education. Additionally, she holds a Master of Science degree in Information Systems, a Bachelor of Technology (Honors) degree in Computing and Information Technology, and an Advanced Diploma in Computing and Information Technology. With a wealth of academic achievements, Dr. Kiwa has authored 17 publications, encompassing articles, conferences, thesis, and dissertations. Currently, she is a candidate for a Master of Mechatronics and AI at the University of Zimbabwe. Her expertise extends to Artificial Intelligence, creative IoT framework designing, and intensive programming skills, particularly in Java, Python, and C++.

Martin Muduva is currently pursuing his PhD at the Chinhoyi University of Technology, Martin is also concurrently engaged in Master's programs in Innovation and Entrepreneurship at the Bindura University of Science Education and Mechatronics and Artificial Intelligence at the University of Zimbabwe. He holds a Master's degree in Leadership and Corporate Governance from Bindura University of Science Education, along with expertise in Big Data Analytics and Information Security. Martin's commitment to education and professional development is evident through his Certificate in Higher and Tertiary Education. His multidisciplinary background and dedication make him a valuable asset in technology, innovation, and corporate governance.

Kudakwashe Maguraushe is an experienced lecturer in multiple computing-related modules. He holds a PhD in Information Systems, an MSc in Information Systems and a BSc (Hons) in Computer Science. He has supervised many students at both undergraduate and postgraduate levels. He has research interests in information privacy and security, healthcare systems, emerging technologies (artificial intelligence, machine learning and social media) and digital transformation.