7th European Conference on Industrial Engineering and Operations Management
Augsburg, Germany, July 16-18, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/EU07.20240164
Published: July 16, 2024

A Predictive Model for Personalized Healthcare Management for Patients with Chronic Diseases

Meluleki Sibindi, Siphilisiwe Sibanda, Jambaya Luke, Crymore Mugwanda, Majaha Zinyika, Prosper N. Dube, Belinda Ndlovu and Sibusisiwe Dube
Department of Informatics and Analytics
National University of Science and Technology
Bulawayo, Zimbabwe
n0232843d@students.nust.ac.zw, n0235177w@students.nust.ac.zw,
n02312054w@students.nust.ac.zw, n0232843d@students.nust.ac.zw,
n023148311@students.nust.ac.zw, n0233219x@students.nust.ac.zw,
n0238533p@students.nust.ac.zw, belinda.ndlovu@nust.ac.zw , sibusisiwe.dube@nust.ac.zw,

Abstract

Chronic diseases have continued to put immense pressure on healthcare facilities, particularly in the developing world, through inefficient management besides high incidence and prevalence. Modern healthcare systems do not have individualized strategies thus limiting preventative steps and management of chronic diseases. This leads to a very important topic – medication non-adherence. This research intended to improve disease prediction and management by creating an effective model that can predict the possibilities of non-compliance with the medication among chronic disease patients. Using the cross-sectional qualitative data from a developing country and the CRISP-DM method, the study employed Recurrent Neural Networks. The model fared better than other approaches like Naïve Bayes, Decision Trees and Support Vector Machines in terms of predicting chances of missed medication doses among chronic patients. The successful performance of the model validates the usefulness of the proposed method in this critical application domain. The research provides real life recommendations on how communication can be useful in the case of chronic diseases and compliance with medication, which is a concern in health systems around the world. The model offers a better solution in a situation where other chronic cases could have been managed properly and required less facilities and cost.

Keywords
Medication non-adherence, personalized healthcare, Recurrent Neural Network, chronic diseases, predictive model.

1. Introduction
Chronic diseases or non-communicable diseases or, plus diabetes, cardiac diseases, and respiratory illnesses, are among the leading causes of morbidity and mortality worldwide (Maguraushe 2024). In the current world especially the developing countries, the health facilities experience a flood of patients with such diseases (Thandu and Gera 2023). Failure to manage chronic diseases not only reduces the well-being of patients but also the mortality and morbidity rates (Mutunhu et al. 2022). Another challenge that is common in chronic diseases is medication non-compliance, whereby patients do not adhere to the recommended dosage as prescribed. This can lead to worsening of the disease condition, hospitalization, and sometimes death (Mpofu et al. 2024). More importantly in developing countries there are different factors that may affect medication non-adherence among the chronic patients like education, income, and availability of family members. Conventional health management interventions do not incorporate individual plans in handling medication non-compliance thus poor compliance with prescribed treatments (Kim et al. 2019). The growth of technology especially in the area of big data is a solution to enhancing personalized client healthcare for chronic diseases. With the help of big data analytics in healthcare, providers can create models

that will analyse data on patients and find out when they are less likely to stick to the medication schedule. This enables early interventions and individualized care plans to enhance medication compliance and health outcomes (Janssoone et al. 2019). Recurrent Neural Networks (RNN) have been applied to the analysis of health care data and to prognosis of the effectiveness of the compassionate care of patients in the healthcare industry according to studies conducted in the recent past (Kim et al, 2019). RNNs are a kind of feed-forward artificial neural network designed specifically for understanding the recurrence data sequences making it most favorable for analysing the patient's health records, to and from in time. Based on large databases of patient data researchers have been able to design prognostic algorithms with degree of hormonal RNNs for predicting patient outcomes such as medication compliance (Luijten 2019.). Even in condition that the RNNs have several application in healthcare, still there is insufficient literature that gives the entire mode of healthcare managing chronic diseases for the individual. Earlier published papers did not quantitatively synthesize and comprehensively map the use of RNNs more specifically in predicting medication non-adherence and individualised care planning for chronic illness (Li et al. 2024). Also many papers do not look at the factors that can affect medication non-adherence in developing countries to be different from developed countries. The paper intends to address the gap in the current literature by providing an informative review of the current state of personalized healthcare for patients with chronic diseases with an emphasis on identifying factors affecting medication non-adherence among chronic patients in developing countries and identifying effective algorithms for the creation predictive models. For healthcare practitioners, researchers, and policymakers, this SLR aims to describe the present status of the research concerns and opportunities for further research in the future to synthesize a more coherent plan that has not previously been addressed when using social media platforms (Merino-Barbancho et al. 2022)

This study is structured as follows: Section 1 offers the study's context, Section 2 presents a literature review, Section 3 details the methodologies utilized, Section 4 data collection, Section 5 will show the results, Section 6 presents the discussion of those results and then final section will be the conclusion.

1.1 Objectives

The objectives of this research are to:-

i) To collect and analyse qualitative data from a developing country to identify key factors contributing to medication non-adherence.
ii) To develop a predictive model using RNNs that can accurately forecast the likelihood of chronic disease patients skipping a medication dose.
iii) To validate the performance of the predictive model and assess it's potential to enhance chronic disease management and improve patient outcomes.

2. Literature Review

This section will look at AI based Technologies, Iot Technologies, Classification Algorithms and finally RNN Algorithms.

2.1 AI Based Technologies

A recent study by Mbunge and Batani (2023) reviewed the access to care health infrastructures by deploying recent AI based applications, examining opportunities, trends, and potentials of adopting new AI based models for care institutions. This research also establishes that AI models can be adopted for early identification, assessment, and tracking of prolonged illnesses, the prognosis of diseases and analysis of public health trends for directing health care delivery and enhancing patient care. Nevertheless, challenges such as AI partiality, reduced access to health information, and the absence of supporting frameworks and policies persist, restricting the integration of AI-based models into health systems. The authors underscore the importance of responsible and accountable use of AI and how policies to support the implementation of data determined AI built results can be integrated into health systems. Therefore, this research argues that the application of AI-based technologies in healthcare settings has a huge prospect as elaborated by the following research questions and their findings. Yu et al (2022) developed the Artificial Intelligence Chronic Management System (AICMS) that uses AI, knowledge graphs, large amounts of data, and IoT to provide an effective solution to the analysis and managing of chronic diseases in pediatric patients. The AICMS comprises clinical individuals and regular patients, information and data storage, and decision support tools based on AI services. They scored 74 % for accuracy and 81 % for the F-score. Makroum (2022) reviewed literature on machine learning and smart devices for the management of diabetes and it was revealed that new technologies such as analytics of large data and application of AI to data on diabetes can change the way diabetes and its related difficulties can be prevented and controlled.

2.2 IoT Technologies

In a study by Abidi et al (2023), a new healthcare monitoring system is presented that employs Hadoop MapReduce technology for efficiently analysing data gathered from several wearable sensors worn on the subject's body. The information is then sent to a cloud environment and processed using gadgets that use Internet of Things technologies. The researchers also use an optimization algorithm for feature selection and different deep learning approachs with different classifiers including but not limited to Convolutional Neural Network (CNN), LSTM, and Deep Neural Network (DNN) just to name the few that the researchers used for identifying the physical activities of the elderly persons. This proposed system that use the different classifiers thus shows enhanced performance .It is also 1% better than the other classifiers on the first and second data set. Another study by Ratta et al (2024) presents a Decentralized, Blockchain-based System with IoT Sensors, Machine Learning, and DApps for effective and safe diabetes self-management. The proposed framework is a layered one, with an IoT Sensor Layer responsible for collecting data, a Blockchain Layer that use contracts that are smart that use on the Ethereum for the protected distribution of data and transactions, a Machine Learning Layer responsible for the analysis of the data collected and the diagnosis of diabetes, and a DApps Layer that allows for interaction between patients, doctors, and hospitals. The paper also assesses the effectiveness of nine different machine learning algorithms for the PIMA dataset with logistic regression, KNN, SVM, Decision Tree, Random Forest, AdaBoost, Naive Bayes and stochastic gradient boosting. From the results it is evident that the AdaBoost model had the maximum analytical accuracy of 92%. The first machine learning algorithm used by researchers was the K-Nearest Neighbors, with an accuracy of 64%, while the second was the Decision Tree with an accuracy of 90%.

2.3 Classification Algorithms

The L. Wang et al (2020) also completed a cross-sectional study based in one centre, where they enrolled 446 Chronic Disease patients. The researchers created and tested two machine learning techniques, namely a BPNN and an SVM with the LR technique. The performance metrics were evaluated, and as well as the ROC curve. Similar near the first trial, a classification accuracy test was conducted to compare the performance of the SVM, BPNN, and as well as the models using the LR algorithms. The percentage accuracy obtained for all the models was, 87. Results of the multivariate analysis on the probability of AZA non-adherence demonstrated several factors such as medication concern beliefs, education, anxiety, and depression as key factors that were significant. On the other hand, the analysis revealed that medication necessity beliefs and medication knowledge were some of the factors that posed protection. Carpinteiro (2023) developed models to early detect Diabetes Mellitus (DM) using different classification algorithms and among the selected algorithms SVM, MLP and GBM had highest overall performance in detecting DM early. Omisore et al (2020) developed effective system developed for analysis and modified management of DM. The system consisted of some models specifically a diagnosis model for diagnosis of DM, the diagnosis model diagnosed DM by combining an inference system. The second model is a Diet Recommender, a knowledge based Diet Recommender model for indorsing personalized certain meals based on the analysis. Both models achieved high accuracies in training, validation and prediction, outperforming singular standard models and current ML approaches and techniques used in interrelated studies. Mohana, (2021) developed models to predict chronic diseases at the initial stage using an ML method. Four classification ML algorithms were used which Support SVM, MLD, J48 Decision Trees and K-Nearest Neighbors (KNN) and SVM are obtained the highest accuracy. Mukura, (2023) carried a research on the evaluating the performance of AI in predicting heart disease risk and Random Forest (RF) algorithm was used a primary ML method to develop their predictive models. After model evaluation the RF algorithm yielded the best results with accuracy of about 80% which highlights the effectiveness of RF algorithm in predicting heart disease risk and importance of utilizing such models in healthcare support systems.

2.4 RNN Algorithms

The literature reviews that deep learning and ensemble learning are being used to develop analytical models for tailored healthcare aimed at the management of chronicle diseases. Y.Gu et al (2021) developed models for predicting medication adherence among patients living with DM, centering on a perplexing class of patients who were-administering their own medication at their own home through injections. The literature reviewed among the algorithms used for developing A Predictive Models for Personalized Healthcare Management for Patients with Chronic Diseases Long Short-Term Memory (LSTM) models demonstrated high accuracy in predicting patient failure to adhere to their medication compared to other algorithms. The models had been trained and as well as weighed using a validation method called the 5-fold method to assess their performance in predicting patients adherence to self-administered injection , RNN also yielded better performance metrics in terms of specificity compared to Multi-Layer Perception(MLP).W. Hsu (2022) carried a study that focused in the development of predictive model for personalized care management of Cardiovascular disease management, the algorithms that were Simple Neural Network (SNN), RNN, Ridge Classifier (RC) and LR and the LSTM. The LSTM algorithm was also best performing model for predicting medication adherence in the study. In order to build the model, the dataset was composed from routine healthcare data sources from patients with cardiovascular disease. Cho et al (2020) developed an individual glucose prediction model to support medical staffs to check the glucose level of patients and regulate the quantity of insulin dosages. Multiple machine learning algorithms were used to determine the one with best performance. The GRU was established to be satisfactory at forecasting the glucose level of a patient. The study by Ismukhamedova et al (2024) includes a comprehensive comparison and enhancement of various modelling approaches like Decision Trees such as Random Forest, Adaboost and multiple Deep Learning algorithms. This study utilized a subset of an already existing database, which has a large dataset of health data, and found that the GMB was the most successful for diabetes diagnosis and that the RNN had the highest performance of all the deep learning techniques.

3. Methods

This study used the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework to design the medication non-adherence predictive model for chronic disease patients. One of the strengths of CRISP-DM is that it is systematic and covers all critical aspects of data mining projects (Studer et al. 2021).

.

I. Business Understanding

The initial stage or the first phase focused on the problem domain analysis and the objectives of the project. This phase involved a review of literature on chronic disease management, medication noncompliance, and issues affecting health systems in developing countries.

II. Data Understanding

During the data understanding phase, the research team gathered qualitative data from a developing country, exploring the reasons for medication non-adherence among chronic disease patients. The data was obtained from a dataset that was downloaded from the internet.

III. Data Preparation

The data preparation phase included cleansing, transforming, and aggregating the gathered data to fit the modelling stage. This involved dealing with the missing values, transforming categorical data into numerical, and constructing features from the raw data. The prepared dataset was then divided into training and validation datasets to make sure that the model is not overfitting.

IV. Modelling

During the modelling stage, the research team designed an RNN model to estimate the number of chronic disease patients missing a dose of medication. This is because the RNN architecture enable it to learn the temporal and sequential characteristics of the patient's medication usage behaviour. The prepared dataset was used for training and fine-tuning the model with several hyper parameters evaluated to achieve the best model performance.

V. Evaluation

The evaluation phase was aimed at analysing the results of the model's work, such as the accuracy. The model was further tested and finalized on the validation dataset to check the general performance of the model.

4. Data Collection

To understand the dataset used for the research, we conducted the exploratory data analysis process. The dataset was cleaned to eliminate duplicates, replaced missing values and cleaning the outliers. The parameters values were encoded to numerical labels for easy model building. To visualize the parameters correlation a heatmap was used.

We used dataset downloaded online from https://data.mendeley.com/datasets/3t39fctrzv/1, the dataset has 609 rows and 52 columns. Then, feature engineering methods were used to find and include the most significant features/columns that clearly contribute to the desired result. The following Table 1 will show the attributes.

Table 1. Parameter Description

Parameter	Description
gender _mapping	Patient's sex(male, female)
Adherent _level	Patient's recommended treatment plan (LOW_ADHERENT, MEDIUM_ADHERENT, NON_ADHERENT, HIGH ADHERENT)
Marital status	Patient's state of marriage or partnership (married, Single, Divorced, Widow/Widower)
Occupation	Patient's type of work engaged(Professional, Trader, Farmer, Student, Retirees, Artisan)
Health	Patient's presence or absence of specific medical condition
Caregiver	Patient's individual who provides assistance to patients with chronic disease(Spouse, Parent ,Children, Others)
Perception	Patient's eye sight (BAD PERCEPTION,FAIR PERCEPTION,GOOD PERCEPTION,UNDESIRABLE)
How many tablets do you take in a day	This is the number of tablets taken within one day
Educational Attainment	The type of Education that the patient was able to get.
AGE	This is the age of the patient
Behaviour Level	How forgetful is the patient

5. Results

5.1 Descriptive Statistics

This dataset was created from patients suffering from chronic diseases in the southeast region of Nigeria. The dataset includes two gender categories - "female" and "male", with a count of 567 for each. There are five distinct age groups in the dataset; "Above 50" is the most common, with 247 entries. There are four distinct categories for marital status in the dataset, with "married" being the most common with 414 individuals. The dataset contains five distinct kinds of educational attainment, the most common of which is "WASC/SSCE" (213 persons), which is a form of educational qualification. There are 23 distinct types of health conditions in the dataset; "Hypertension" is the most common, affecting 241 people. The dataset has four distinct non-adherent level classifications. Out of all of them, "MEDIUM ADHERENT" is the most prevalent, comprising 202 individuals as show in Figure 1 and Figure 2.

	GENDER	AGE	Marital Status	Educational Attainment	Occupation	Who is your care giver	Health Condition	How many tablets do you take in a day	PERCEPTION LEVEL	BEHAVIOUR LEVEL	NON ADHERENT LEVEL
count	567	567	567	567	567	567	567	567	567	567	567
unique	2	5	4	5	7	5	23	4	4	4	4
top	female	Above 50	married	WASC/SSCE	Professional	Others	Hypertension	Above three	FAIR PERCEPTION	MODERATELY FORGETFUL	MEDIUM ADHERENT
freq	310	247	414	213	164	127	241	195	163	173	202

Figure 1. Distribution of dataset

5.2 Feature Importance Analysis

Figure 2 shows a bar chart showing the feature importance for a machine learning model is displayed in this image. The value or influence of each input feature in the model's prediction or decision-making process is measured by its feature significance. The Health Condition feature, with a value of about 0.14, is the most important. This suggests that the model's predictions are significantly influenced by the individuals' health status. From "Who is your care giver" to "BEHAVIOUR LEVEL" features suggests that while not as influential as Health Condition feature still affect the machine learning model. However the last two, "GENDER" and "Marital Status" are the lowest and suggest that both will have no effect on the machine learning model when it is trained.

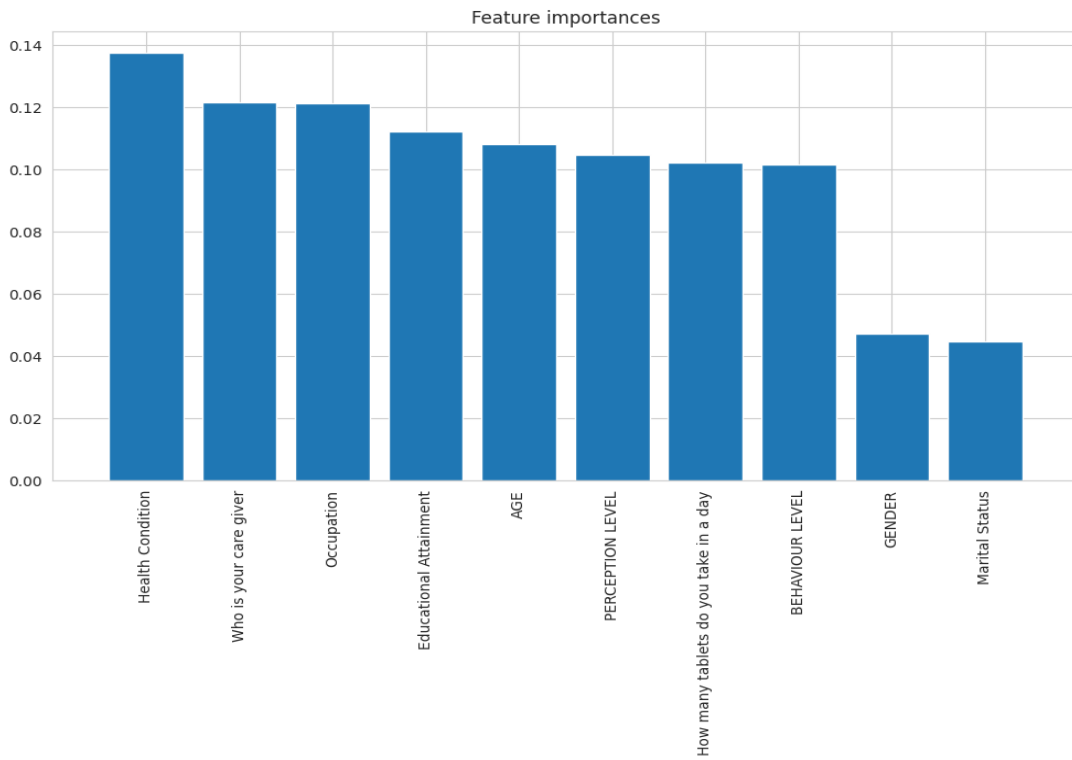

Figure 2. Features in the dataset

5.3 Model Building

In order to make the training of the data easier we label encoded the data so that the LSTM was to be trained. Figure 3 shows how the data was encoded.

```
# Filter rows where gender is either 'male' or 'female'
df_filtered2 = df_filtered[df_filtered['GENDER'].isin(['male', 'female'])]
df_filtered2 = df_filtered[df_filtered['Marital Status'].isin(['married', 'single', 'Divorced', 'widow/widower'])]
df_filtered2 = df_filtered[df_filtered['NON ADHERENT LEVEL'].isin(['LOW ADHERENT', 'MEDIUM ADHERENT', 'HIGH NON-ADHERENT', 'HIGH ADHERENT'])]
df_filtered2 = df_filtered[df_filtered['Occupation'].isin(['Professional', 'Trader', 'Farmer', 'Student', 'Retirees', 'Artisan'])]
df_filtered2 = df_filtered[df_filtered['How many tablets do you take in a day'].isin(['one', 'two', 'three', 'Above three'])]
df_filtered2 = df_filtered[df_filtered['Health Condition'].isin(['Diabetes', 'Mental', 'HIV', 'Hypertension'])]
df_filtered2 = df_filtered[df_filtered['Who is your care giver'].isin(['Spouse', 'Parent', 'Children', 'Relatives', 'Others'])]
df_filtered2 = df_filtered[df_filtered['PERCEPTION LEVEL'].isin(['BAD PERCEPTION', 'FAIR PERCEPTION', 'GOOD PERCEPTION', 'UNDESIRABLE'])]


gender_mapping = {'male': 1, 'female': 0}
marital_status = {'married': 0, 'single': 1, 'Divorced': 2, 'widow/widower': 3}
adherant_level={'LOW ADHERENT': 1,'MEDIUM ADHERENT': 2, 'HIGH NON-ADHERENT': 3, 'HIGH ADHERENT':4}
occupation = {'Professional': 0, 'Trader':1, 'Farmer': 2, 'Student': 3, 'Retirees':4, 'Artisan': 5}
tablets = {'one': 1, 'two': 2, 'three': 3, 'Above three': 4}
health = {'Diabetes': 1, 'Mental': 2, 'HIV':3, 'Hypertension':4}
caregiver = {'Spouse':1, 'Parent': 2, 'Children':3, 'Relatives':4, 'Others': 5, '':6}
perception = {'BAD PERCEPTION':1, 'FAIR PERCEPTION':2, 'GOOD PERCEPTION':3, 'UNDESIRABLE':4}
occupation = {'Professional': 0, 'Trader':1, 'Farmer': 2, 'Student': 3, 'Retirees':4, 'Artisan': 5}



# Replace gender values with numerical labels
df_filtered2['GENDER'] = df_filtered2['GENDER'].replace(gender_mapping)
df_filtered2['Marital Status'] = df_filtered2['Marital Status'].replace(marital_status)
df_filtered2['NON ADHERENT LEVEL'] = df_filtered2['NON ADHERENT LEVEL'].replace(adherant_level)
df_filtered2['Occupation'] = df_filtered2['Occupation'].replace(occupation)
df_filtered2['How many tablets do you take in a day'] = df_filtered2['How many tablets do you take in a day'].replace(tablets)
df_filtered2['Health Condition'] = df_filtered2['Health Condition'].replace(health)
df_filtered2['Who is your care giver'] = df_filtered2['Who is your care giver'].replace(caregiver)
df_filtered2['PERCEPTION LEVEL'] = df_filtered2['PERCEPTION LEVEL'].replace(perception)
```

Figure 3. Label Encoding

The LSTM model was built with three dense model and 1 LSTM model as shown in Figure 4. This is due because the data is categorical and the researchers needed to extract relevant features in the dataset.

```
# Define the model (rest of your model definition)
model = Sequential()
model.add(LSTM(64, input_shape=(X_train.shape[1], 1)))
model.add(Dense(32, activation='relu'))
model.add(Dense(64, activation='relu'))
#model.add(Dense(32, activation='relu'))
model.add(Dense(1, activation='sigmoid'))

# Compile the model
model.compile(optimizer=Adam(learning_rate=0.001), loss='binary_crossentropy', metrics=['accuracy'])

# Train the model
model.fit(X_train, y_train, epochs=63, batch_size=32, validation_data=(X_test, y_test))
```

Figure 4. Model Building

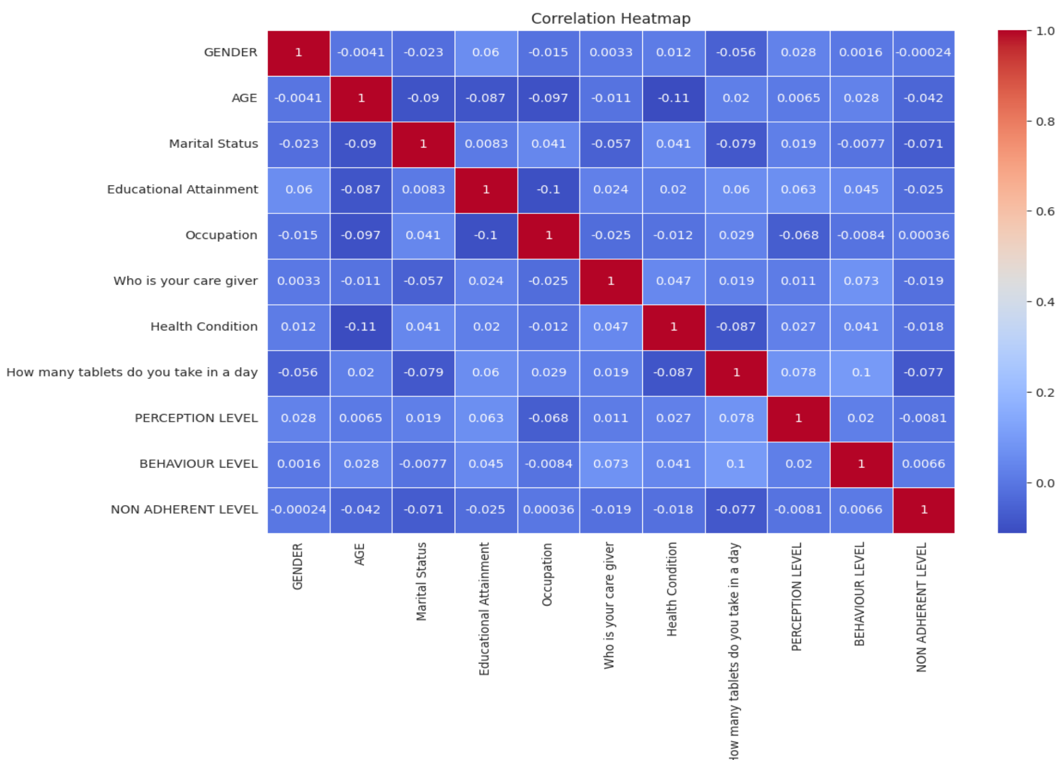

Figure 5. Dataset Heat map

Figure 5 image is a correlation heat map, which is a visual representation of the correlation between different variables in a dataset. The heat map shows the correlation coefficients between various factors related to patient's health and medication non-adherence. The rows represent the different factors being analysed, such as gender, age, marital status, and educational attainment, occupation, who is the care giver, health condition, and number of tablets taken per day, perception level, behaviour level, and non-adherent level. The columns also represent the same set of factors. The colour of each cell in the heat map represents the strength and direction of the correlation between the corresponding factors. The scale ranges from -1 to 1, with red indicating a strong negative correlation, blue indicating a strong positive correlation, and white indicating no correlation. Gender has a very weak correlation with most other factors, with correlation coefficients close to 0 Educational attainment has a weak positive correlation with occupation, perception level, and behaviour level, suggesting that higher educational attainment is associated with better occupational outcomes and improved perception and behaviour. The behaviour deals with whether a patient is forgetful or not while perception deals with eyesight. Occupation has a weak negative correlation with health condition, perception level, and behaviour level, implying that certain occupations may be associated with poorer health and lower perception and behaviour levels. The number of tablets taken per day has a weak positive correlation with health condition and a weak negative correlation with non-adherent level, suggesting that patients who take more tablets may have poorer health and are less likely to be non-adherent. Non-adherent level has a moderate negative correlation with perception level and behaviour level, implying that patients who are more non-adherent tend to have lower perception and behaviour levels.

Table 2. Comparison of Algorithms

PERFORMANCE MATRICS	ALGORITHMS			
	RNN(LSTM)	Naive Bayes Classification	Decision Tree	and Support Vector Machine (SVM)
Accuracy	0.61	0.36	0.35	0.42
Precision	0.37	0.37	0.36	0.42
Recall	0.61	0.36	0.35	0.42
F1 Score	0.46	0.35	0.34	0.40

Three other models were built and trained in order to compare them to the LSTM that was built. The reason why the other three were chosen was because these three are the best when it comes to categorical data. Table 2 above shows multiple different models used to compare with our LSTM model we used. It shows that the RNN outperforms the other classification algorithms when using the Nigerian data. The RNN has an accuracy of 0.61 which did better than the Naives Bayes Classification, Decision Tree, and the Support Vector Machine.

6. Discussions

After Comparison of different models, the study highlighted the superiority of the LSTM model against other models.This is further supported by the significant positive relationship between age and number of chronic conditions showing that as the number of chronic diseases increase so too does the risk for non-adherence due to multiple morbidities, increasing medication burden, and effects of aging. Further, the positive but weak relationship between medication compliance and socioeconomic status indicates that patients with low incomes or low education levels may potentially have more difficulties in obtaining and adhering to their chronic diseases' medications. Interestingly, the study found out that patient- provider discussions bear negative relationship with non-adherence thus supporting the notion that communication plays vital role in enhancing medication adherence.

The results of the investigation of the predictive model confirm the efficiency of the proposed LSTM architecture that has a significantly higher accuracy of the chances of the missed medication doses among chronic patients compared to the rest of the widely used techniques. Hence, this LSTM model becomes the most effective way to tackle the medication non-adherence prediction indicting the significance of this model in better chronic diseases management in the healthcare organization.

6.1 Numerical Results

The results of this study show that the proposed approach based on LSTM achieved the better results as compared to such standard approaches as Naive Bayes, Decision Trees, and Support Vector Machines in terms of predicting the chronic disease patients' tendency to miss the medication doses. The features selection model obtained an accuracy of 61% while Naive Bayes was at 36%, Decision Tree at 35% and Support Vector Machine at 42%. The findings presented in this paper show that the proposed LSTM model is a good solution for determining whether chronic disease patients are likely to miss their doses as compared to the three algorithms. The results of the feature importance analysis are informative in identifying the influential factors for non-adherence to medication among patients with chronic diseases in developing countries like shown in Fig. shows "Health condition", "Occupation" and as well as "educational attainment" affect how medication adherence among patients. The discovery of the features emerged as significant determinants of non-adherence is consistent with previous studies. These ideas may therefore help design effective strategies to manage non-adherence specifically, and overall chronic illnesses.

6.2 Graphical Results

The ROC curve shown in Figure 6 is a graphical representation of the performance of a binary classifier, where the true positive rate (TPR) is plotted against the false positive rate (FPR) at various threshold settings. The x-axis represents the False Positive Rate (FPR), which is the proportion of negative instances that are incorrectly classified as positive. The y-axis represents the True Positive Rate (TPR), which is the proportion of positive instances that are correctly classified as positive. Fig. shows that AUC for the LSTM is 0.79 and is the highest AUC meaning it performs better than all the other models meaning that researchers, W. Hsu (2022) was right that compared to the other classification algorithms the LSTM performs better at predicting medication non adherence.

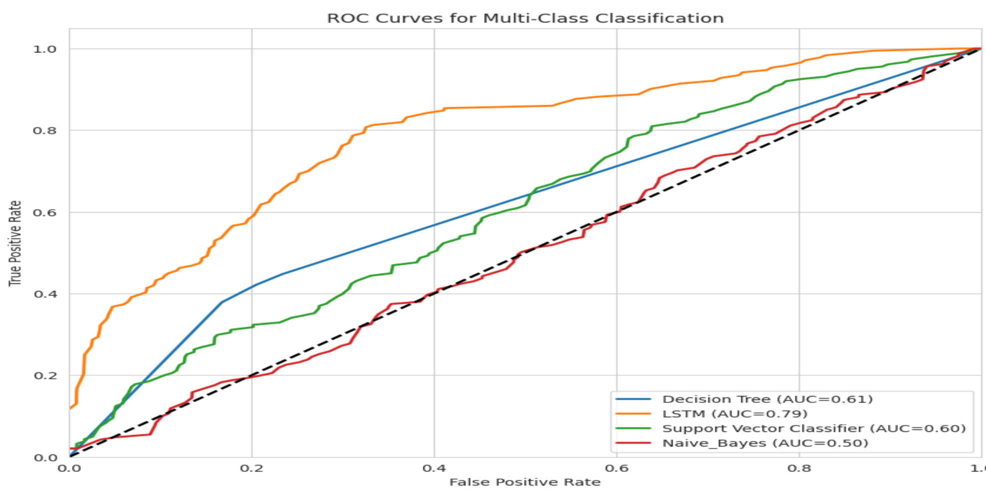

Figure 6. Dataset Heat map

5.3 Limitations

The present study has several limitations that should be considered when interpreting the results. Firstly, the dataset used was from Mendeley Data as such information from local hospitals are either not easily accessible, it may take time to be cleared from respective authorities before getting permission to access such information , it maybe there is no proper data fragmentation or it may be also be issue of privacy and security concerns of patients information. Second, because this study only included data from one developing nation, it may be challenging to generalise the findings to other contexts or healthcare settings.

Third, since the patients provided their own information, recollection and social desirability bias may have an impact on the data. Thankfully, the study's possible shortcomings are noted. Fourthly, the model may have overfitted due to the limited dataset (609 cases), which is why the accuracy score was low even though the model outperformed other classification techniques.

In conclusion the RNN is good at predicting when the patient is not adhering to their medication. However they are ways to improve its accuracy so that it can be higher.

5.4 Recommendations

Zimbabwe should consider enforcing creation and maintenance of National Health Data Repositories, also establish systems that use such Predictive models for personalized healthcare management for patients with chronic diseases. Future research could focus on seamlessly integrating predictive models with Electronic Health Records (EHR) to facilitate real-time data extraction and prediction. Additionally foster self quantification practices amongst people with non-communicable diseases such as diabetes as the increase of health care personlisation (Mutunhu,Chipangura, Singh 2024). Create and put into action customised adherence-promotion plans, such as better patient-provider contact, medication management support, and personalised patient education, depending on the insights gleaned from the prediction model and assess how well these strategies work to improve patients' health outcomes and medication adherence in chronic illness patients.

7. Conclusion

The goal of this research was to create a reliable research model to predict times when chronic disease patients should not be expected to fully adhere to their prescribed medications schedule. The observations made and implications outlined in this study provide valuable understanding of the determinants of medication noncompliance in chronic diseases. Consequently, the study recommends that it would be beneficial for the healthcare systems to implement specific prevention campaigns and information dissemination plans, especially for individuals with multiple chronic diseases or population groups with complex or difficult social conditions. Such interventions should use patient-specific adherence assessment models that consider factors such as age, comorbidities, socioeconomic status and the interaction between the patient and the healthcare provider to facilitate the adherence to recommended evidence-based strategies.

References

Abidi, M. H., Umer, U., Mian, S. H., and Al-Ahmari, A., Big Data-Based Smart Health Monitoring System: Using Deep Ensemble Learning. *IEEE Access*, *11*, 114880–114903. 2023. https://doi.org/10.1109/ACCESS.2023.3325323

Ismukhamedova, A., Uvaliyeva, I., and Belginova, S., Integrating machine learning in electronic health passport based on WHO study and healthcare resources. *Informatics in Medicine Unlocked*, *44*, 101428. 2024. https://doi.org/10.13026/C2XW26

Mbunge, E., and Batani, J., Application of deep learning and machine learning models to improve healthcare in sub-Saharan Africa: Emerging opportunities, trends and implications. *Telematics and Informatics Reports*, *11*. 2023. https://doi.org/10.1016/j.teler.2023.100097

Ratta, P., Abdullah, and Sharma, S., A blockchain-machine learning ecosystem for IoT-Based remote health monitoring of diabetic patients. *Healthcare Analytics*, *5*. 2024. https://doi.org/10.1016/j.health.2024.100338

Wang, L., Fan, R., Zhang, C., Hong, L., Zhang, T., Chen, Y., Liu, K., Wang, Z., and Zhong, J., Applying machine learning models to predict medication nonadherence in crohn's disease maintenance therapy. *Patient Preference and Adherence*, *14*, 917–926, 2020. https://doi.org/10.2147/PPA.S253732

Yu, G., Tabatabaei, M., Mezei, J., Zhong, Q., Chen, S., Li, Z., Li, J., Shu, L. Q., and Shu, Q., Improving chronic disease management for children with knowledge graphs and artificial intelligence. *Expert Systems with Applications*, *201*.2022. https://doi.org/10.1016/j.eswa.2022.117026

Y. Gu *et al.*, "Predicting medication adherence using ensemble learning and deep learning models with large scale healthcare data," *Sci. Rep.*, pp. 1–13, 2021, doi: 10.1038/s41598-021-98387-w.

W. Hsu, J. R. Warren, and P. J. Riddle, "Medication adherence prediction through temporal modelling in cardiovascular disease management," *BMC Med. Inform. Decis. Mak.*, vol. 9, pp. 1–21, 2022, doi: 10.1186/s12911-022-02052-9.

Kim, C., Son, Y., and Youm, S., Chronic disease prediction using character-recurrent neural network in the presence of missing information. *Applied Sciences (Switzerland)*, *9*(10). 2019. https://doi.org/10.3390/app9102170

Thandu, A. L., and Gera, P., A Comprehensive Review of Healthcare Prediction using Data Science with Deep Learning. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Issue 12). 2023. www.ijacsa.thesai.org

Janssoone T, Bic C, Kanoun D, Rinder P, and Hornus P., *AI for Patient Support: Predictive Model of Medication Non-Adherence*. 2019. https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/sniiram/

Luijten, S. J. A., *Improving therapy adherence using smart persuasive interventions generated by an RNN-RL framework*. 2019.

Merino-Barbancho, B., Cipric, A., Arroyo, P., Rujas, M., Martín Gómez Del Moral Herranz, R., Barev, T., Ciccone, N., and Fico, G., *Methods and computational techniques for predicting adherence to treatment: a scoping review*. 2022. https://doi.org/10.1101/2024.06.10.24308540

Li, Y., Xiong, Y., Fan, W., Wang, K., Yu, Q., Si, L., van der Smagt, P., Tang, J., and Chen, N., *Sequential Model for Predicting Patient Adherence in Subcutaneous Immunotherapy for Allergic Rhinitis*. 2024. http://arxiv.org/abs/2401.11447

Kim, D. Y., Choi, D. S., Kim, J., Chun, S. W., Gil, H. W., Cho, N. J., Kang, A. R., and Woo, J., Developing an individual glucose prediction model using recurrent neural network. *Sensors (Switzerland)*, *20*(22), 1–15.2020. https://doi.org/10.3390/s20226460

Makroum, M. A., Adda, M., Bouzouane, A., and Ibrahim, H., Machine Learning and Smart Devices for Diabetes Management: Systematic Review. In *Sensors* (Vol. 22, Issue 5). 2022. MDPI. https://doi.org/10.3390/s22051843

Rao, M. N., Madana Mohana, R., Talasila, V., and Sureshkumar, M., PREDICTION OF CHRONIC DISEASES AT AN EARLY PHASE USING MACHINE LEARNING APPROACH. *Turkish Journal of Physiotherapy and Rehabilitation*, *32*(3). 2021. www.turkjphysiotherrehabil.org

Maguraushe, K., Ndayizigamiye, P.,Towards a Smart Healthcare System for Non-Communicable Diseases (NCDs) Management: A Bibliometric Analysis. In: Masinde, M., Möbs, S., Bagula, A. (eds) Emerging Technologies for Developing Countries. AFRICATEK 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 520. 2024. Springer, Cham. https://doi.org/10.1007/978-3-031-63999-9_7

Mpofu, L., Ndlovu, B., Dube, S., Muduva, M., Jacqueline, F., & Maguraushe, K., *Predictive Model for Hospital Readmission of Diabetic Patients*. 2024. https://doi.org/10.46254/AF05.20240252

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.

Mutunhu, B., Chipangura, B., & Singh, S., Towards a quantified-self technology conceptual framework for monitoring diabetes. *South African Journal for Science and Technology*, *43*(1), 69–84. 2024. https://doi.org/https://doi.org/10.36303/SATNT.2024.43.1.970E

N. W. C. Mukura and B. Ndlovu, "Performance Evaluation of Artificial Intelligence in Decision Support System for Heart Disease Risk Prediction .," no. Who 2018, pp. 83–93, 2023.

Carpinteiro, C., Lopes, J., Abelha, A., and Santos, M. F., A Comparative Study of Classification Algorithms for Early Detection of Diabetes. *Procedia Computer Science*, *220*, 868–873, 2023. https://doi.org/10.1016/j.procs.2023.03.117

Omisore, O. M., Ojokoh, B. A., Babalola, A. E., Igbe, T., Folajimi, Y., Nie, Z., and Wang, L., An affective learning-based system for diagnosis and personalized management of diabetes mellitus. *Future Generation Computer Systems*, *117*, 273–290. 2021. https://doi.org/10.1016/j.future.2020.10.035 .

Studer, S., Bui, T.B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S. and Müller, K.R., Towards CRISP-ML (Q): A Machine Learning Process Model with Quality Assurance Methodology', Machine Learning and Knowledge Extraction, 3(2), pp. 392–413, 2021. Available at: https://doi.org/10.3390/make3020020.

Biographies

Belinda Mutunhu Ndlovu is a Ph.D. in Information Systems student at UNISA. She holds an MSc in Information Systems, a BSc in Computer Science, and a Postgraduate Diploma in Education. Additionally, she works as a lecturer and project coordinator for postgraduate students at the National.

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

Majaha Zinyika is currently a Big Data Science graduate student in the Department of Informatics and Analytics at the National University of Science and Technology. He holds a BSc in Informatics. His research interests are in the fields of Big Data and data analytics.

Sibanda Siphiliwe is a Health Information Officer in the ministry of Health and Childcare. She holds BSc in

Informatics, HND in Information Technology. Currently she pursuing her education in Master of Science degree in Big Data Science.

Jambaya Luke is a financial advisor for Zimnat Life assurance, Educational consultant for Chilstar Pvt. He holds a BSc degree in Statistics and Operations Research. Currently pursuing education in MSc in Big Data Science.

Meluleki Sibindi is a qualified Statistician, with BSc in Statistics and currently pursuing MSc in Big Data. Alongside his academic journey, Meluleki serves as a Strategic Information and Evaluation Officer at an International NGO dedicated to HIV prevention, care and treatment.

Mugwanda Crymore is a qualified Computer Science Facilitator since 2015 under the Ministry of Primary and Secondary with a BSc Honors in Computer Science, National Diploma in IT, Diploma in Education specializing in IT (Secondary) and currently is pursuing his MSc in Big Data.

Prosper N. Dube has a BSc Honors in Operations research and statistics, post graduate degree in Monitoring and Evaluation current working towards a master's degree in Big Data Science. He has worked as a Monitoring and Evaluation Intern at Hope for a child in Christ and a Sales Data Analyst at National Foods Ltd. Currently he works as an Administrative assistant at International Medical Corps.