ZAMBIA INFORMATION COMMUNICATION TECHNOLOGY (ICT) JOURNAL

Volume 8 (Issue 1) (2024) Pages 1-6

Artificial Intelligence-Driven Data Science for Enhancing TB Treatment Outcomes and Reducing Mortality Rates

Daniel Katongo
Department of Computer Science
Copperbelt University
Kitwe, Zambia
[1]danielkatongo.dk@gmail.com

Jameson Mbale
Department of Computer Engineering
Copperbelt University
Kitwe, Zambia
[2]Jameson.mbale@gmail.com

Abstract **- Tuberculosis (TB) remains a leading global health concern, ranking among the top causes of death worldwide. It surpasses HIV in mortality, with approximately 1.3 million deaths annually among HIV-negative individuals and 214,000 among those co-infected with HIV** [1]**. Despite progress in reducing TB mortality rates, as evidenced by Zambia's decrease from 759 to 361 deaths per 100,000 population between 2000 and 2017, TB continues to be a significant health challenge in high-burden regions like southern Africa** [1]**. In Zambia's Itezhi-Tezhi district, factors contributing to TB mortality include advanced age, poor treatment adherence, extra-pulmonary TB, and complications related to co-infection with HIV. While systems such as SmartCare and YATHU DR-TB have been developed to manage records for drug-susceptible and drug-resistant TB, there is a critical need for more advanced tools to further enhance treatment outcomes. This paper proposes an AI-driven automated system that utilizes data science techniques to improve TB treatment outcomes and reduce mortality rates. The system employs data mining to track and analyze comprehensive patient records throughout the treatment phases, including intensive and continuation phases. It gathers and evaluates data on patient demographics, drug adherence, treatment progress, and outcomes in comparison with similar cases that have achieved successful treatment. By leveraging AI algorithms to predict treatment outcomes based on historical data, healthcare providers can gain valuable insights into patient progress, enabling more timely and effective interventions. This proactive approach aims to address challenges in TB management, enhance patient monitoring, and ultimately improve overall treatment efficacy. The integration of AI and data science into TB care represents a promising advancement in combating one of the world's most persistent infectious diseases.**

Keywords: Data-Science, Bacteriologically Confirmed, Clinically Diagnosed, Treatment Success, Artificial Intelligence

I. INTRODUCTION

According to the Zambian Ministry of Health 2022 annual statistical health report, tuberculosis (TB) ranks among the top 10 causes of mortalities and is the top cause of mortalities in adults and 8 in children between the ages of 5 to 14. A retrospective study conducted nationwide revealed that tuberculosis mortality remains high at 86 per 100,000 populations, this could be translated to 15,000 TB-related deaths per annum approximately [2]. This shows that tuberculosis plays a major role in the mortalities faced in Zambia. Zambia adopted the World Health Organization's (WHO) strategy to eliminate tuberculosis by 2030 [2]. This Global Strategy comprised of three pillars. The integrated patient-centered care and prevention aimed at providing access to the diagnosis and treatment of tuberculosis globally. The other pillar included bold policies and supportive systems for strengthening the

government leadership, civil society and private sector engagement in providing universal health coverage, social protection and poverty alleviation. The other pillar was the intensified research and innovations which involved the discovery, development and quick uptake of implementation of new tools for the strategy. This meant that even new technologies such as Artificial Intelligence could be applied in combating TB.

Ever since the WHO's end tuberculosis by 2023 strategy was adopted, different plans, policies and implementation tools have been invented and enforced to meet the set target, Tools such as electronic health record systems to handle information of patients on anti-tuberculosis treatment (ATT), e-mobile health technology to enable tracking of patients and sending reminders [3]. For example, the YATHU Drug Resistant Tuberculosis (YATHU DR-TB) system manages records of patients who are drug resistant, recording information such as comorbidities, type of regimen, patient demographics and other relevant information to track the progress of the patients. On the other hand, SmartCare records information of patients who are attended to by the health facility regardless of their tuberculosis status [3]. These tools had some weaknesses which relied on the patients to having access to phones to receive certain information, systems like SmartCare focused mostly on providing quality Antiretroviral (ART) services while not enough on other services like tuberculosis. As technology advances, the involvement of new technologies such as Artificial Intelligence (AI) in combating tuberculosis is steadily increasing.

This paper proposes the utilization of AI in predicting tuberculosis treatment outcomes such as lost to follow, died and treatment success or failed in order to provide timely intervention during the course of treatment. This would help reduce the high number of mortalities caused by tuberculosis in Zambia.

A. Problem Statement

Among the Sub-Saharan countries highly burdened by tuberculosis, Zambia is among the top 30 [2]. In Zambia tuberculosis ranks among the top 10 causes of mortalities. A lot of studies focus on the application of Artificial Intelligence in the diagnosing of tuberculosis and developing systems like the Computer Aided Detection for tuberculosis (CAD4TB) [2]. When it comes to improving tuberculosis treatment outcomes, such as treatment success, treatment after lost to follow, reducing treatment failure and lowering the mortality rate, diagnosing tuberculosis is one of the major tasks. The other crucial task is monitoring the patients on anti-tuberculosis treatment (ATT). Hence, being able to predict the treatment outcome of the patient on anti-tuberculosis treatment with the help of Artificial Intelligence and Data Science based on the patient's clinical records would really improve the tuberculosis treatment outcomes and reduce tuberculosis -related mortalities.

B. Research Objectives

The following research objectives will be addressed in this study.

1. To determine the best suited AI Model for predicting TB treatment outcomes
2. To analyze the feature importance of other comorbidities in TB mortalities
3. To evaluate the impact of AI in enhancing TB treatment outcomes

The following research questions will help us effectively address the objectives of this study ultimately providing insights to improve tuberculosis management and patient care.

1. How can we determine the most effective AI model for predicting TB treatment outcomes?
2. What comorbidities are commonly associated with Tuberculosis Mortalities?
3. What effects does AI have on the enhancement of TB treatment outcome?

C. Significance of Research

This research will ensure the health sector provides the optimal service to their tuberculosis patients based on their needs. It will provide a different approach on how to tackle the challenging issue of tuberculosis related mortalities by applying AI, one of the emerging technologies. With this research, resources in the health sector could be focused on how to prevent the undesirable outcome based on the model prediction, enabling clinicians to respond timely to changing needs of the patients. This would also move the health sector a step closer towards achieving the end TB target in the adopted strategy.

II. RELATED WORKS

[4]used Artificial Intelligence to predict the adverse effects of anti-tuberculosis drugs on patients under anti-tuberculosis treatment despite the study conducted in 2004 attributed the anti-tuberculosis drugs contributing to 86% of the treatment success. The adverse effects of the anti-tuberculosis drugs targeted were hepatitis, respiratory failure, and mortality. Algorithms such as XGBoost, Random Forest, Multilayer Perceptron (MLP), light GBM, logistic regression, and Support Vector Machine (SVM) helped predict the adverse effects early in the patients during treatment [4]. In most studies, the AI algorithms have not predicted acute respiratory failure and hepatitis. Since anti-tuberculosis drugs such as isoniazid and rifampicin can cause a patient to have acute respiratory failure and hepatitis, the model built was used to predict these and help out in narrowing down which patients should be stopped and put on other drugs as well as empower the clinicians with the timely change of the treatment of the patient. Machine Learning was applied in the TB therapy in Colombia to help support TB diagnosis in patients at their health facilities. The automated Machine Learning was tested to figure out which models were the best while the Tree-based pipeline Optimization Tool (TPOT) was used to determine the best detectors [5]. It was discovered that the Multilayer Perceptron (MLP) had best specificity, Logistic Regression (LR) model had best accuracy. The essence of the study was to prove the effectiveness of the use of Machine Learning in Tuberculosis diagnosis support.

In Korea, a retrospective study was carried out on the TB Patients to help improve the treatment outcomes via the analysis of the data of patients with pulmonary Tuberculosis and culture conversion [6]. The study used 6 referral centers to collect data. Among the data collected, Chest X-rays were collected and analyzed together with the other comorbidities and patient demographics. In this study, Treatment success was considered when the patient completed treatment, for the bacteriologically confirmed, their culture had to convert twice, both at month two and month five [6]. [7] investigated the utilization of Artificial Intelligence to predict therapeutic efficacy in pulmonary Tuberculosis (PTB). The study revealed that the use of sociodemographic, image data such as chest x-rays, clinical and genomic data held encouraging outcomes when employed in Artificial Intelligence systems that used deep learning neural networks. A retrospective analysis on data from 2017 to 2018 was conducted to investigate the performance of Computer-Aided Detection Digital Chest X-ray in reading active tuberculosis among people with a history of previously treated tuberculosis in Zambia [8]. It was found that the accuracy for CAD4TB among people who were previously treated for TB substantially decreased when a fixed abnormality threshold was applied.

A. Limitations

In Columbia, the study conducted only emphasized on the use of machine learning on the diagnosing of TB, rather than providing an AI model that focuses on predicting TB treatment outcomes. This indicated that machine learning can be used in TB diagnosis support. In Korea, the retrospective study conducted, focused on diagnosing TB using Chest X-Rays in addition to other patient information such as comorbidities to only improve PTB treatment outcomes rather than those with extrapulmonary TB (EPTB). [4] focused on predicting adverse effects of drugs on patients with tuberculosis, its main aim was to identify the right drugs for a TB patients rather than predict their likely TB treatment outcome, making it a great tool when trying out new drugs on TB patients.

Looking at Zambia, there has not been any use of AI in the prediction of TB treatment outcomes to help prevent the undesirable TB treatment outcomes. Hence, this research aims to incorporate AI in the prediction of TB treatment outcomes in Zambia to compliment the systems like CAD4TB that use AI to diagnose TB and analyze how the prediction of TB treatment outcomes using medical data such as patient demographics, sputum culture results, treatment duration, type of regimen with the application of AI could reduce TB related mortalities in Zambia. The next section highlights the methodology of the research.

III. METHODOLOGY

We used Python and libraries such as Scikit Learn, pandas, and TensorFlow to create a model. The algorithms used in developing the model were logistic regression, support vector machine, random forest, decision tree, and the neural network. We evaluated the models on four performance metrics, precision,

accuracy, f1-score, and recall. The dataset acquired was imbalanced, this led to the use of techniques such as Synthetic Minority Over-sampling Technique (SMOTE) to balance the dataset.

The key features in the dataset were the weight, age, gender, Human Immunodeficiency Virus (HIV) status, type of regimen, sputum culture results and the tuberculosis history which indicated if one had tuberculosis before. The categorical features in the dataset were assigned numeric values using the one hot encoding technique. Sputum culture had some missing values; to deal with this, we used the mode imputation technique to fill in the missing values of the categorical data. SMOTE was used to up-sample the minority class where samples in a class were more than 2 without overfitting the model.

To assess the performance of the models, the K-fold cross validation technique was applied. The dataset was split into 80% (654) training and 20% (163) testing, exposing the models to more data to enable them predict treatment outcomes efficiently.

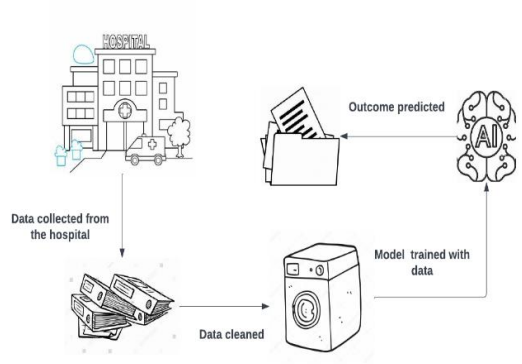

Fig .1. shows the stages leading to TB treatment outcome prediction.

A. Data Collection

With the clearance from the Copperbelt University, data was collected at the Livingstone Central Hospital from various sources, including the DR-TB treatment register and the YATHU DR-TB system. The YATHU DR-TB system contained patient information from different provinces and districts. This helped in acquiring a larger set of patient information due to the low number of DR TB notified cases in Zambia. The dataset had 1817 records in total. The study focused on patients who were actively receiving treatment during the period of 2017 to 2023. To ensure the accuracy and relevance of the data, patients who had been transferred out, transferred in, were excluded. This led to the reduction of records in the dataset to 817. This exclusion criterion was essential to maintain a consistent and reliable dataset, which is crucial for drawing meaningful conclusions. The distribution of outcomes in the dataset acquired is as shown in Fig 2.

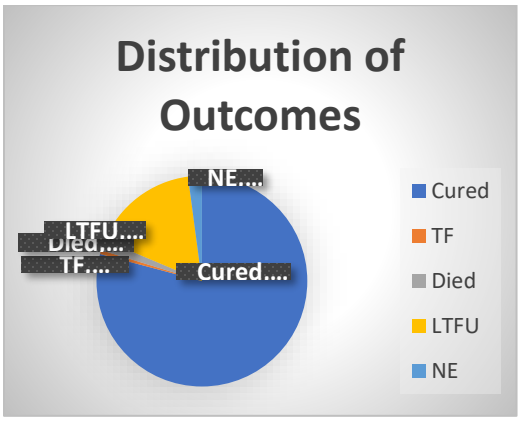

Fig .2. distribution of TB treatment outcomes in the dataset acquired.

IV. RESULTS

After assessing the models, it was seen that the outcome predicted mostly accurate was the cured as compared to the other outcomes. The mean accuracy of the models across all folds was high except for the decision tree model, which was quite low. The error bars in Fig. 3 for all models were small, suggesting that the accuracy results were consistent across the folds in the model.

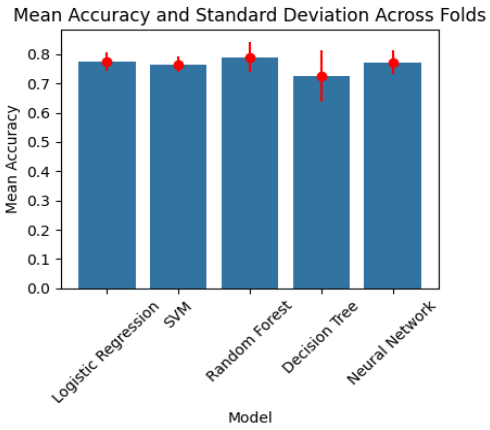

Fig. 3. Shows performance of each model based on mean accuracy and standard deviation across all folds

From the predictions, we used a kernel density plot displaying all the machine learning's model distribution of accuracies across all folds in the k-fold cross validation technique. The accuracies are represented in the x-axis ranging from 0.60 to 0.90 while the density which indicates the frequency of occurrences of accuracy across the models. Fig. 4 shows the kernel plot.

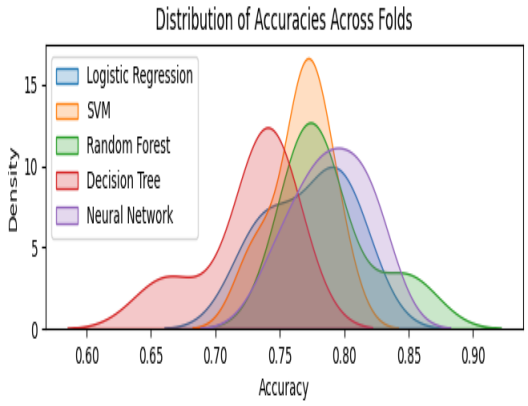

Fig. 4. distribution of accuracies of prediction of TB treatment outcomes for each model across folds

The SVM has a sharper peak around 0.78 indicating, that its accuracy was more consistent across the folds. Random Forest model has peak near 0.75, but with a longer tail indicating it occasionally achieves higher accuracies. The Neural Network and Logistic Regression model had a smoother distribution with peaks around 0.75-0.78. Only the decision tree shows a wider. The performance metrics, accuracy, precision, recall and f1 score seem to have a great performance in predicting the cured outcome. Fig 5 shows the metrics performance in predicting cured outcomes.

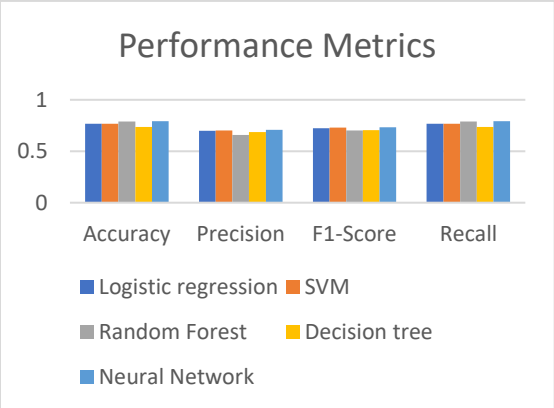

Fig 5: shows the performance metrics in predicting TB treatment outcomes of each model

Table I summarizes the information about the average performance metrics values across all folds in the models.

Table I: **average model performance across all folds using the K- fold validation technique**.

Model	Accuracy	Precision	F1-Score	Recall
Logistic Regression	0.7669	0.7000	0.7249	0.7669
SVM	0.7669	0.7027	0.7289	0.7669
Decision Tree	0.7359	0.6872	0.7051	0.7359
Random Forest	0. 7899	0.7078	0.7315	0.7913
Neural Network	0.7913	0.7078	0.7315	0.7913

V. DISCUSSION

The models evaluated had a bias in predicting the cured treatment outcome. Even after applying data balancing techniques such as SMOTE, applying the mode imputation to missing categorical data and mean imputation to the missing numerical data the prediction of the other outcomes was very low. The key features selected in the prediction of the treatment outcomes, age, weight and height played a significant role. These three features attributed to whether the patient was likely to complete treatment due to his or her health status. When it came to comorbidities, we only had access to the HIV status, while the other comorbidities were not available due to some data entry issues. We discovered that in the dataset, this feature did not play an important role probably because there was not enough information in a dataset to ensure that the HIV status could play a major role in TB related mortalities with the help of another key feature like drugs taken. Implementing this model would be a challenge to due data entry issues, leading to delayed predictions resulting in untimely response by the clinicians.

A. Limitations

The data collected had missing information in some key features selected, which led to the exclusion of relevant features such as type of drugs which would

have had a great impact on the prediction model. The dataset acquired did not have enough records with a variety features. Records from SmartCare were not fully populated, this became difficult to collect accurate data due to missing records from the pulled reports. We could not access patient files because each patient returned with their file to their respective home. Hence, the data from SmartCare had to be excluded. From the dataset collected, the major comorbidity that was focused on was HIV, hence limiting the study to how HIV impacts TB mortalities.

VI. CONCLUSIONS

In conclusion, AI has had a great impact in the combating of TB mortalities. From the systems built that incorporate the use of AI in diagnosing TB. We can see that number of notified cases have increased and timely initiation is being done to ensure that we do not lose more lives to TB. In the future, the use of advanced models such as the XGBoost, Multilayer Perceptron (MLP), as well as the use of a large dataset that has a variety of features can be done to implement a model that is very accurate and responds quickly to information gathered without being biased to the most appearing class in the features.

VII. REFERENCES

[1] Ministry of Health, "Annual health statisitical report," Zambia, 2022.

[2] Ministry of Health, "statistical anuual health report," 2021.

[3] World Health Organization, "Global Tuberculosis Report," 2022.

[4] K.M. Liao, "Using an Artificial Intelligence approach to predict the adverse effects and prognosis of tuberculosis," PubMed Central, p. 15, 2023.

[5] A. Orjuela, A. Juantico, C. Awad, E. Vegrgara and A. Palencia, "Machine learning in the loop for tuberculosis diagnosis support," Frontiers in Public Health, p. 10, 2022.

[6] Hyung Jun Kim, "Artifcial intelligence based radiographic extent analysis to predict tuberculosis treatment outcomes: a multicenter cohort study," Scientific Reports- Nature, p. 8, 2024.

[7] F. Zhang, F. Zhang, L. Li, Y. Pang, "Clinical utilization of artificial intelligence in predicting therapeutic efficacy in pulmonary tuberculosis," Journal of Infection and Public Health, p. 10, 2024.

[8] M. Kaguje et al, "The Performance of Computer-Aided Detection Digital Chest X-ray Reading Technologies for Triage of Active Tuberculosis Among Persons With a History of Previous Tuberculosis," Clinical Infectious Diseases, 2023.

[9] J. Bwembya et al, "Mortality among persons with tuberculosis in Zambian hospitals: A retrospective cohort study," PLOS Glob Public Health., p. 17, 17 June 2024.

[10] J. Bwembya et al, "Mortality among persons receiving tuberculosis treatment in Itezhi-Tezhi district of Zambia: A retrospective cohort study," PLOS Glob Public Health., p. 13, 22 february 2023.

[11] K. K. Dasu, "Predictive analysis of tuberculosis treatment outcomes using machine learnig : A Karnakata TB data study at a scale," 13 March 2024.