ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1172

Predictive Analytics in Health Information Management: The Impact of Electronic Health Record (EHR) Data on Patient Outcomes in Nigerian Tertiary Health Facilities

Julius Ekunke Ajaba[1]; Dallah Plangshak Anthony[2];
Benjamin Bishtu[3]; Samue B. Kaze[4]

[1,4]Newgate University Minna.
[2,3]Jos University Teaching Hospital

Publication Date: 2025/10/04

Abstract:

- ***Background –***

The trend towards the implementation of Electronic Health Records (EHR) into healthcare systems has offered a treasure trove of data which can be used in predictive analytics. Patient care and clinical decision-making can be enhanced by learning the factors that are important in determining patient outcomes and predictive modeling in tertiary health facilities in Nigeria.

- ***Objective –***

The purpose of the study was to determine the primary elements in EHR data, that have a significant impact on patient outcomes and to construct and test predictive models to predict these outcomes, and any such outcome can lead to improved clinical decision-making in Nigerian hospitals.

- ***Methods –***

An analysis of 500 patient records of two tertiary hospitals of Nigeria was carried out in the retrospective manner. The information contained demographics, medical history, diagnosis, laboratory data, and treatment information. Correlation analysis was done to come up with factors that significantly affect patient outcomes. Logistic Regression, Random Forest and Support Vector Machine (SVM) were some of the predictive models that were developed and tested based on accuracy, precision, recall, and F1 score.

- ***Results –***

The most important variables that were regarded as significant predictors of mortality, readmission, and recovery were age, high blood pressure, diabetes, and blood pressure. Random Forest model did better and reported the accuracy and precision of 90 and 89 respectively, recall of 91 and F1 score of 90. It was found that hypertension and diabetes were most strongly correlated with adverse outcomes.

- ***Conclusion –***

The research shows that EHR data may be successfully employed to create predictive models that may be utilized to improve patient outcomes and clinical decision-making. Random Forest model proved to be the most efficient in terms of patient outcome prediction and this hints on its possible further application in the healthcare environment. The study ought to be extended in terms of data volume and incorporating new variables to maximize predictive models to apply in a larger context.

Keywords: *Electronic Health Records, Predictive Analytics, Patient Outcomes, Mortality, Recovery, Readmission, Recovery, Machine Learning, Healthcare Decision-Making, Nigeria.*

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1172

How to Cite: Julius Ekunke Ajaba; Dallah Plangshak Anthony; Benjamin Bishtu; Samue B. Kaze (2025) Predictive Analytics in Health Information Management: The Impact of Electronic Health Record (EHR) Data on Patient Outcomes in Nigerian Tertiary Health Facilities. *International Journal of Innovative Science and Research Technology*, 10(9), 2405-2412. https://doi.org/10.38124/ijisrt/25sep1172

I. INTRODUCTION

➢ *Background to the Study*

The use of data analytics, specifically predictive analytics has become a revolution in the healthcare system today. Due to the development of technology, healthcare systems are currently in a position to access a significant volume of data, with a considerable portion of that information being gathered using Electronic Health Records (EHR) (Shafqat et al., 2020). The predictive analytics role of data analytics in healthcare has been actively developed over the last few years and has demonstrated worthy insights, which can greatly contribute to better patient outcomes, lowering operational expenses, and making better clinical decisions. The use of predictive analytics, a statistical algorithm and machine learning model to analyze past data and predict the future, could transform healthcare delivery (Badawy e al, 2023).

EHR systems are at the core of predictive analytics in medical care. EHRs are a paperless variant of patient charts, which include an abundance of information about their demographics, medical history, treatment plans, lab results and diagnoses, and patient outcomes (Adeniyi et al., 2024). Such records are all-inclusive data that can be used to forecast health trends, patient behavior and possible health risks. With the adoption of EHRs by healthcare systems, predictive analytics on how to utilize the records to enhance healthcare delivery is becoming more important in the healthcare delivery system (Adeniyi et al., 2024)

The rising significance of predictive analytics is due to the ability to predict patient outcomes, early disease warnings, and treatment effectiveness. The combination of predictive analytics will allow healthcare providers to see previously unseen patterns and trends to provide their patients with personalized care and prevent possible health problems proactively (Osei-Frimpong et al., 2018). This healthcare delivery model is not only more likely to enhance patient satisfaction and outcomes but also helps lessen the financial strain on healthcare systems through avoidable hospitalization and interventions (Rahman et al., 2024).

The EHR systems play a critical role in this transformation which is data driven. EHRs enable the enhanced coordination and continuity of care by ensuring patient data is shared and stored across various healthcare providers by means of a centralized and digitalized repository of patient data (Barbieri et al., 2023). When predictive analytics is combined with EHR systems, healthcare providers can build models predictive of patient outcomes, high-risk patients, and point-of-care decision-making. As an illustration, predictive models can detect the probability of a patient to recover after an operation, determine the possibility of readmission after discharge, or even anticipate possible drug interactions before they happen (Amarasingham et al., 2014).

The integrations of EHRs and predictive analytics have considerable benefits to the delivery of healthcare, especially in resource-constrained environments. Predictive analytics can serve as an efficient remedy to enhance patient care and resource management in countries such as Nigeria, where the healthcare system is usually under pressure due to the lack of resources, staff, and infrastructure. Using the EHR data, healthcare providers can better foresee patient needs and prioritize interventions and allocate resources more effectively to minimize wastes and enhance overall healthcare outcomes (Ganesan, 2020). Furthermore, low-resource healthcare settings can use predictive models to detect new factors affecting the health of populations, including the increased prevalence of chronic illnesses, to allow the implementation of the relevant prevention and intervention measures by the public health officials.

Although there are the benefits that are likely to be experienced, there are challenges facing the adoption of predictive analytics in healthcare. The adoption of the use of advanced analytics within health systems presupposes a highly effective technological backbone, such as the availability of reliable data storage and processing facilities, as well as qualified staff trained in data science and machine learning. Moreover, concerns like data privacy and security, particularly when it comes to EHRs, should be handled with a lot of care to achieve patient trust and adherence to the regulatory measures (Bani et al., 2020). Moreover, the data under analysis is vital; incomplete, inaccurate, or inconsistent data may compromise the predictive models, which makes them make incorrect predictions and the results of predicting may be detrimental.

Nevertheless, the fact that predictive analytics could improve healthcare delivery is indisputable. With the help of the huge volumes of the data received in EHRs, healthcare providers are able not only to predict the results of their patients with more accuracy but also to enhance effectiveness and efficiency of healthcare systems overall. Incorporation of predictive analytics in EHR systems is one of the major milestones on the way to more personalized, proactive, and cost-effective healthcare.

➢ *Problem Statement*

Patient outcome improvement is a major challenge of the Nigerian healthcare system where predictive analytics is not used effectively in Electronic Health Records (EHR) systems. The fact that EHRs have the potential to hold all patient data does not imply that predictive models are being used to predict patient outcomes, which also leads to the lack of efficiency in care delivery. The healthcare sector in Nigeria is also vastly understaffed, so there is only 0.35 physicians per 1,000 residents (Onah et al., 2022), and only 35 percent of hospitals use EHRs fully (Adedeji et al., 2018). It has led to elevated readmission rates (15% of patients readmitted 30 days post-discharge, Adebayo et al., 2019) when compared to developed countries (12%), as this sector does not use advanced data analytics. There is a lack of early intervention opportunities due to the failure of healthcare providers to use predictive analytics, which results in low-quality management of high-risk patients and wastes of resources. This paper will set forth to illustrate how predictive analytics on EHR data may enhance patient outcomes, decrease readmissions, and improve the provision of health care in Nigerian tertiary hospitals.

➢ *Aim and Objectives*

- *Aim*

The aim of this study is to explore the impact of Electronic Health Record (EHR) data on patient outcomes in Nigerian tertiary health facilities.

- *Objectives*

✓ To identify key factors in Electronic Health Record (EHR) data that significantly influence patient outcomes in Nigerian tertiary health facilities.
✓ To develop and evaluate predictive models using EHR data to forecast patient outcomes and improve clinical decision-making in Nigerian hospitals.

➢ *Research Questions*

- What are the key factors in Electronic Health Record (EHR) data that significantly influence patient outcomes in Nigerian tertiary health facilities?
- How can predictive models using EHR data be developed and evaluated to forecast patient outcomes and improve clinical decision-making in Nigerian hospitals?

II. LITERATURE REVIEW

Machine learning (ML) algorithms and statistical models are the main predictive analytics in healthcare as they predict patient outcomes based on historical data. Both these strategies are based on the idea of data-driven decision-making, according to which the data are used to predict future tendencies, practices, and results (Sarker, 2021). Logistic regression, decision tree, random forests, support vector machine (SVM), are the common machine learning techniques that are applicable in healthcare and can analyze large and complex data, including the data in Electronic Health Records (EHR) (Dinesh et al., 2014).

Cox proportional hazards models are another type of statistical models commonly used in the healthcare industry to estimate patient outcomes, especially in survival analysis (Gur et al., 2020). Such models can be used to learn the interconnections between variables such as patient demographics, clinical history, and patient outcomes, so it can be predicted that one will encounter a disease or become readmitted again in the future (Pranckevičius & Marcinkevičius, 2017). When combined with the EHR data, these models can enable healthcare providers to make informed and data-driven decisions in predicting risks in their patients.

➢ *EHR Data in Health Information Management*

Health Information Management (HIM) is based on EHR systems, which are a virtual representation of the medical chart of the patient. EHRs also contain all the essential details about patients, including medical history, demographics, diagnoses, lab results, medications, and treatment plans, which allows healthcare professionals to obtain real-time and correct data (Yadav et al., 2018). The systems are critical in enhancing patient care and operational efficiency by mitigating the possibility of human error, clinical workflow, and enhancing the coordination of healthcare providers (Carayon et al., 2018).

Although the main use of EHR systems is to gather and handle data, they can be augmented with the help of predictive analytics. Predictive models can be applied to EHR data to enable healthcare professionals to predict patient outcomes, control risks, and make clinical decisions in real-time (Shafqat et al., 2020). These benefits prevent the full harnessing of EHR data in predictive analytics; in developing nations such as Nigeria, adoption of EHR is still in its infancy (Oni et al., 2018).

➢ *Predictive Analytics in Healthcare*

Predictive analytics accumulated with EHR data have demonstrated a potential in enhancing patient care and healthcare delivery. There has been an increase in the use of predictive models to forecast such outcomes as hospital readmissions, deterioration of patients, and disease progression. As an example, a study by Parikh et al. (2016) and its results showed how random forest algorithms can be used to predict patient deterioration with high accuracy rates. In the same way, Luo et al. (2015) used the MSO to EHR data to predict the occurrence of hospital readmissions demonstrating the potential of predictive analytics to decrease readmission rates and streamline the allocation of hospital resources. Successful applications of predictive analytics have also been made to patient stratification which is an identification of the high-risk patients who need more intensive care or early intervention. Jordan. (2025) emphasized a study in which predictive models assist in identifying patients at risk of chronic conditions to enable the healthcare provider to intervene early, manage the patients proactively, and save on costs incurred in long-term healthcare.

Nevertheless, in low- and middle-income countries (LMICs) such as Nigeria, the use of predictive analytics in health care is still low. Although the advantages of the integration of predictive analytics with EHR data in Nigerian hospitals are evident in high-resource environments, the lack of proper infrastructure, trained staff, and data quality are limiting to the application of this tool (Tilahun et al., 2022).

➢ *Patient Outcomes*

Healthcare delivery focuses on patient outcomes which are affected by many factors including demographics, clinical

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1172

history and treatment interventions. It has been revealed that age, gender, and underlying conditions are the main factors in determining patient outcomes, especially when it comes to chronic diseases like diabetes, hypertension, or cardiovascular conditions (Mohammed et al., 2016). An example of this is that older patients who have numerous comorbidities are more prone to negative outcomes in terms of recovery and mortality (Ahluwalia et al., 2012). In the same vein, the clinical history such as prior surgeries or illnesses is important to establish the response of patients to the treatments (Risse & Warner, 1992).

Predictive analytics has the potential to combine all these factors, have a more insightful picture of the state of a patient and have better outcomes predictions. Using EHR data as the input, predictive models have the potential to determine the high-risk patients, personalize their treatment regimens, and issue early alerts regarding conditions that might deteriorate (Rahman et al., 2023). This individualized approach to health care improves patient care and utilization of resources through avoiding unnecessary hospital stays and complications.

➢ *Research Gaps*

Although predictive analytics have potential, it is not fully used in the healthcare environment in Nigeria. The implementation of EHR systems in Nigeria is slow, and therefore only 35% of hospitals adopt EHRs completely (Oni et al., 2018). Hospitals that use EHRs have no integration with predictive analytics, and due to the absence of it, it is difficult to predict patient outcomes in a manner of forecasting rather than merely accepting information.

III. METHODOLOGY

➢ *Research Design*

The research design adopted in this study is ex-post facto research design to determine the impact of EHR data on patient outcomes. The design is suitable because it does not manipulate variables by analyzing available data. It concentrates on the interaction between the demographics of the patients, their clinical history, and treatment data (independent variables) and such outcomes as recovery, mortality, and readmission (dependent variables).

➢ *Data Collection*

In this research, the secondary data sources were Electronic Health Records (EHR) in the chosen Nigerian tertiary health facilities. The data covered demographics of patients, clinical history, diagnosis, lab results, medications, treatment plans, and patient outcomes in terms of recovery, mortality and readmission rates. Data retrieval was done in a retrospective manner and was taken within the past 5 to 10 years so as to have a comprehensive sample size in predictive modeling.

The data was gathered by cooperating with hospitals that had applied EHR systems. A request to access anonymized data of patients through ethical approval and consent was made to the concerned institutions, and the ethical approval guaranteed that the data privacy rules were adhered to, and patient confidentiality was preserved. Data cleaning and preprocessing were done to manage missing data, inconsistencies, and outliers in order to come up with correct analysis.

➢ *Sample Selection*

A two-stage sampling technique was used to select hospitals for data collection.

- *Stage 1: Stratification*

Hospitals were stratified by Nigeria's six geopolitical zones: North, South, East, West, Central, and South West, ensuring regional diversity.

- *Stage 2: Purposive Sampling*

Two hospitals were selected based on:

✓ Full EHR implementation.
✓ Availability of comprehensive data from the past 5 years.
✓ Willingness to provide anonymized data.

The selected hospitals were:

✓ Federal Medical Centre (FMC), Lokoja (North Central).
✓ Obafemi Awolowo University Teaching Hospitals Complex (OAUTHC) (South West).

- *Stage 3: Data Extraction*

The information was taken out of the EHR systems of the chosen hospitals. Each hospital was searched to obtain 500 patient records, and this made 1 000 cases of patient records to be analyzed. These records contained numerous demographic, clinical and outcome-related data years ago up to the last 5 years; thus, sufficient data to predictive model.

This paper has chosen tertiary health institutions in Nigeria that adopted Electronic Health Records (EHR) systems and had up to date patient data in the last 5 to 10 years. The hospitals were selected according to the ability to submit adequate amounts of EHR data and the desire to participate in the research.

- *Inclusion Criteria:*

✓ Tertiary health facilities with fully implemented EHR systems.
✓ Hospitals with a minimum of 5 years of comprehensive, historical EHR data.

- *Exclusion Criteria:*

✓ Hospitals without fully functioning EHR systems or with incomplete data.
✓ Facilities unable to provide retrospective patient data for analysis due to insufficient records.

➢ *Variables*

The independent variables are EHR data variables (patient demographics (age, gender), clinical history (past conditions, treatment), diagnosis codes, lab outcomes,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1172

treatment information (medications, procedures). Dependent variables are centered around patient outcomes, mortality rates, recovery rates, readmission rates and patient complications (infections or adverse reactions). Some of the control variables encompass such variables as socio-economic status, co-morbidities, and hospital characteristics.

➢ *Data Preprocessing*

Preprocessing of data entailed the treatment of missing values by either imputing them or deleting them, treatment of outliers using z-scores and the conversion of categorical variables into numerical variables using one-hot encoding. Numical data were normalized or standardized.

➢ *Predictive Modeling*

The predictive models were done using logistic regression, random forests, and support vectors machine (SVM). Model performance was measured by accuracy, precision, recall and F1 score with the aim of determining the accuracy of prediction, positive prediction and balance.

➢ *Ethical Considerations*

The personal identifiers were eliminated via data anonymization, which is mandatory by NDPR and privacy standards. Each of the hospitals was required to obtain institutional review board (IRB) approval to guarantee ethical use of data and to keep the data confidential.

IV. RESULTS

➢ *Demographic and Clinical Characteristics*

Table 1 Demographic and Clinical Characteristics of the Sample

Variable	Options	Frequency	Percentages
Age			
	Less than 20 years	49	9.8
	20-25	73	14.6
	26-30	128	25.6
	31-35	101	20.2
	36-40	74	14.8
	Above 40	75	15
Gender			
	Male	199	39.8
	Female	301	60.2
Common Diagnosis			
	Hypertension	251	50.2
	Diabetes Mellitus	149	29.8
	Asthma	51	10.2
	Chronic Obstructive Pulmonary Disease	24	4.8
	Heart Failure	35	7
	Acute Respiratory Infections	19	3.8
	Stroke	16	3.2
	Pneumonia	30	6
	Peptic Ulcer Disease	10	2
	Chronic Kidney Disease	38	7.6

The demographic and clinical sample characteristics demonstrate a wide age group with the highest age category (26-30 years old) constituting 25.6 per cent, the second one (31-35 years old) constituting 20.2 per cent and the rest of the age groups being evenly distributed. The sample is mostly female (60.2%), and males constitute 39.8 percent. The prevalence of chronic conditions is high with Hypertension being most prevalent (50.2% of the sample) with Diabetes Mellitus (29.8) coming in second. Other prevalent diagnoses are Asthma (10.2%), chronic kidney disease (7.6%), and Heart Failure (7%), and such conditions as Stroke, Pneumonia, and Peptic Ulcer Disease are not common. Such results indicate that the sample is highly impacted by chronic diseases, especially Hypertension and Diabetes which have proven to have a significant effect on patient outcomes and thus, are significant predictive variables in this study.

➢ *Key Factors in Electronic Health Record (EHR) Data that Significantly Influence Patient outcomes in Nigerian Tertiary Health Facilities.*

Table 2 Correlation Between Key Variables and Patient Outcomes

Variable	Mortality	Readmission	Recovery
Age	0.45	0.40	-0.55
Gender (Male = 1)	-0.20	-0.10	0.15
Hypertension (Yes = 1)	0.60	0.55	-0.30
Diabetes Mellitus (Yes = 1)	0.50	0.45	-0.40
Blood Pressure (High)	0.55	0.50	-0.35

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1172

As indicated in the analysis, age is positively correlated with mortality (0.45) and readmission (0.40) and negatively correlated with recovery (-0.55), which means that older patients are at risk of poor results. Gender is weakly negatively correlated to mortality (-0.20) and readmission (-0.10) and positively correlated with recovery (0.15), implying that it has a small impact on outcome. Hypertension has the strength of a positive correlation to mortality (0.60) and readmission (0.55) and a negative correlation to recovery (-0.30) hence is one of the principal risk factors. Diabetes Mellitus exhibits its effects on adverse outcomes as it has positive correlations with mortality (0.50) and readmission (0.45), and negative correlation with recovery (-0.40). On the same note, the correlation between high blood pressure and mortality (0.55) and readmission (0.50) is positive, whereas recovery (-0.35) is negative, which underlines the contribution of high blood pressure to poor patient outcomes.

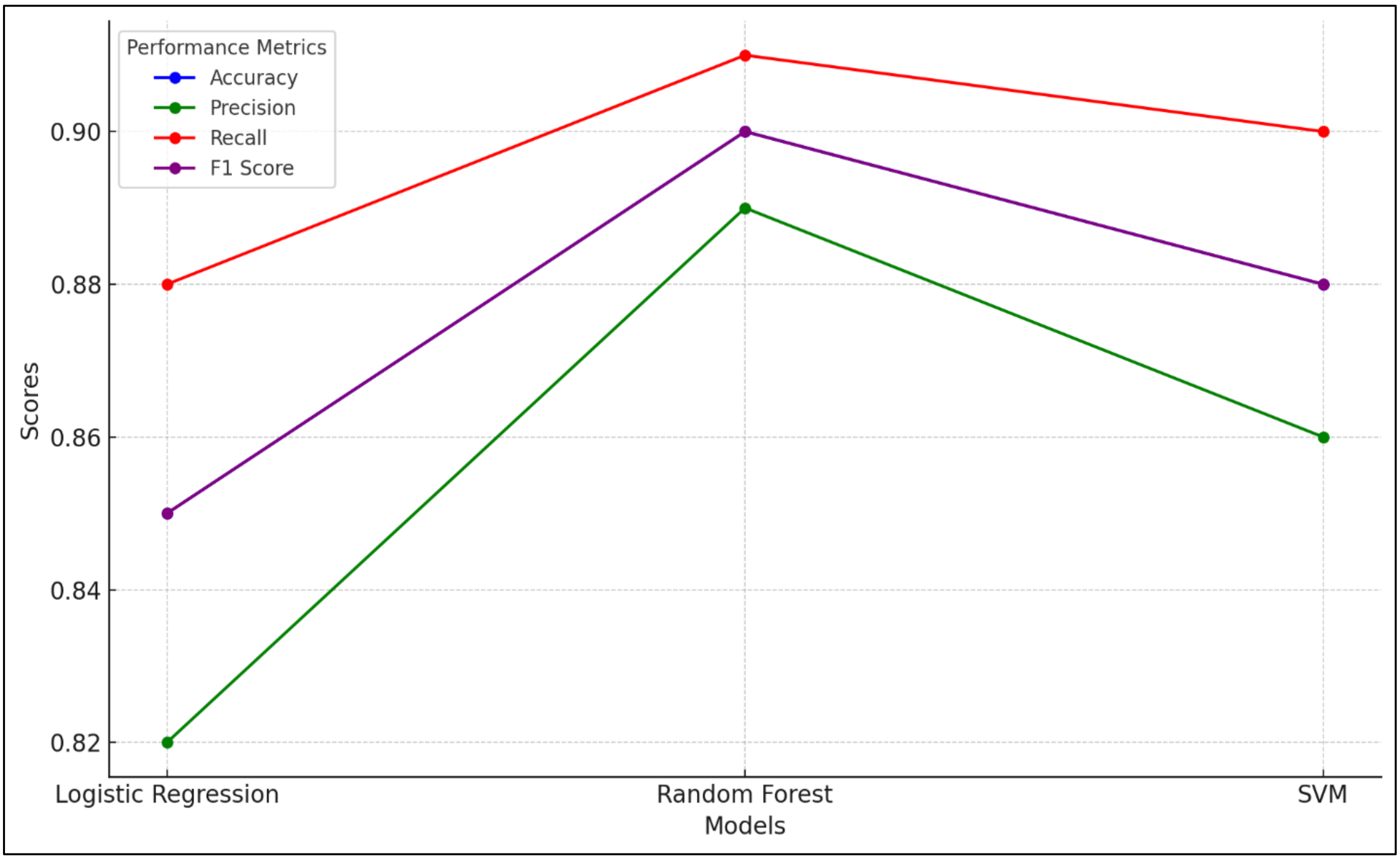

Fig 1 Model Performance Comparison Chart

The chart compares the behaviours of three predictive models, namely Logistic Regression, random forest and Support Vector Machine (SVM) in four significant evaluation metrics including Accuracy, Precision, Recall and F1 Score.

- Precision: The Rand Forest model is the most accurate with a 90 rate which means that it was able to predict patient outcome more frequently compared with the other models. The SVM model has the highest accuracy of 88, and Logistic Regression has had the lowest accuracy of 85, indicating that it has the low overall success of prediction.

- Precision: Random Forest once again tops the charts with an 89 percent precision, which is the least number of false positives. The precision of SVM was 86 and the Logistic Regression was 82, which means that it was worse in false positives than the other models.

- Recall: Random Forest and SVM performed the same in terms of recall (91 percent) which implies that they were able to correctly identify most of the cases of positivity (e.g., patients at risk of dying or readmission). The recall of Logistic regression was slightly lower at 88% with the model missing some positive cases as opposed to the other models.

- F1 Score: The highest scores were achieved by the random forest and SVM models at 90, which means that there was a balance between the precision and recall. The worse in this measure was the Logistic Regression with the F1 Score of 85% that indicates the not the best balance of precision and recall.

The advantage of the other models is that the other models are not as accurate as the Random Forest in all measures making the random forest much more effective than the other models when predicting patient outcomes. SVM is also highly effective, especially on the aspect of recall, whereas, the lowest performance is on the aspect of precision and recall in the case of Logistic Regression.

Volume 10, Issue 9, September – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25sep1172

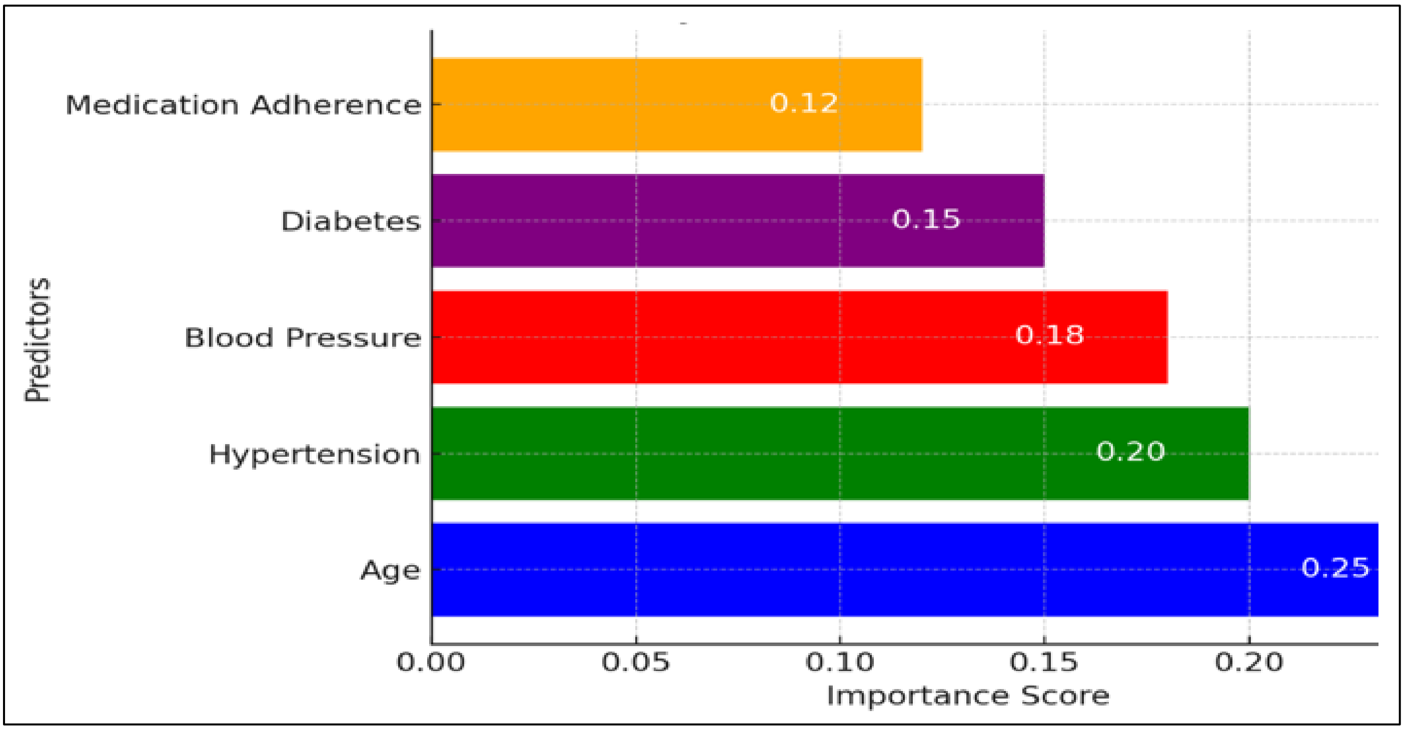

Fig 2 Bar Chart for Key Predictors

Figure 2 shows the relative significance of different factors in patient outcomes prediction. The most prominent predictor is age (importance score 0.25), which means that old age is crucial in determining the outcomes of patients, especially mortality and recovery, and older patients are at a higher risk. In its turn, hypertension comes next with the importance score of 0.20, outlining the significant role that the latter has in patient outcomes, particularly in the context of cardiovascular morbidities such as heart failure and stroke. Blood pressure has the third position with an importance score of 0.18, which once again supports the significance of blood pressure in predicting patient outcomes especially heart health and disease progression. The importance of diabetes has a score of 0.15 implying that it is a determinant factor especially to patients with co-morbidities which can make recovery complex. Finally, medication adherence scores the lowest concerning its importance of 0.12, meaning that although it is an important factor, its influence is lower than the effects of such demographic and clinical variables as age and hypertension. These results emphasize the importance of the chronic management of conditions, including hypertension and diabetes, in enhancing their care and outcomes.

V. DISCUSSION

The results of this paper are consistent with available literature because it identified age, hypertension, diabetes, and high blood pressure as important predictors of mortality, readmission, and recovery. The age and mortality (0.45) and readmissions (0.40) positive correlations are aligned with the prior studies according to which older patients are at a greater risk of poor outcomes (Graham et al., 2020). In the same way, the fact that hypertension is closely related to mortality (0.60) and readmission (0.55) is confirmed by research by Mills et al. (2016), who discovered that hypertension is associated with the risk of getting cardiovascular complications and low recovery rates. The results of diabetes and blood pressure are consistent with Wu et al. (2019) and Kjeldsen et al. (2020), who concluded that these factors lead to unfavorable outcomes, such as slow recovery and death.

The model with the best performance in terms of accuracy was the Random Forest model with an accuracy of 90, and then SVM. This aligns with the results of Boulesteix et al. (2017), who found that the simple models lost to the ensemble models in prediction in healthcare. The performance of Logistic Regression was not as strong, indicating that it is not an effective method to accomplish the task of addressing complex relationships in healthcare data (Dastidar et al., 2016).

VI. CONCLUSION

This research study has illustrated the importance of age, hypertension, diabetes and high blood pressure in determining the patient outcomes in the Nigerian tertiary health facilities. Random Forest model turned out to be the most successful with regards to the process of predicting the mortality, readmission, and recovery, which means that it will be able to make a substantial contribution to clinical decision-making. With the help of EHR data and predictive analytics, medical workers will be able to identify high-risk patients at an earlier stage and take more active measures. This paper has shown the usefulness of predictive models to the enhancement of patient care, particularly on patients with chronic conditions. In order to expand on these findings, the next study must comprise a larger sample and will have to factor in more variables that will help to better adjust the model to its nature and guarantee the further expansion of its application in the healthcare context.

REFERENCES

[1]. Adedeji, P., Irinoye, O., Ikono, R., & Komolafe, A. (2018). Factors influencing the use of electronic health records among nurses in a teaching hospital in Nigeria. *Journal of health informatics in developing countries*, *12*(2).

[2]. Adeniyi, A. O., Arowoogun, J. O., Chidi, R., Okolo, C. A., & Babawarun, O. (2024). The impact of electronic health records on patient care and outcomes: A comprehensive review. *World Journal of Advanced Research and Reviews*, *21*(2), 1446-1455.

[3]. Ahluwalia, S. C., Gross, C. P., Chaudhry, S. I., Ning, Y. M., Leo-Summers, L., Van Ness, P. H., & Fried, T. R. (2012). Impact of comorbidity on mortality among older persons with advanced heart failure. *Journal of general internal medicine*, *27*(5), 513-519.

[4]. Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., & Xie, B. (2014). Implementing electronic health care predictive analytics: considerations and challenges. *Health affairs*, *33*(7), 1148-1154.

[5]. Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: a survey. *Journal of Electrical Systems and Information Technology*, *10*(1), 40.

[6]. Bani Issa, W., Al Akour, I., Ibrahim, A., Almarzouqi, A., Abbas, S., Hisham, F., & Griffiths, J. (2020). Privacy, confidentiality, security and patient safety concerns about electronic health records. *International nursing review*, *67*(2), 218-230.

[7]. Barbieri, C., Neri, L., Stuard, S., Mari, F., & Martín-Guerrero, J. D. (2023). From electronic health records to clinical management systems: how the digital transformation can support healthcare services. *Clinical Kidney Journal*, *16*(11), 1878-1884.

[8]. Dinesh, P., Vickram, A. S., & Kalyanasundaram, P. (2024, May). Medical image prediction for diagnosis of breast cancer disease comparing the machine learning algorithms: SVM, KNN, logistic regression, random forest and decision tree to measure accuracy. In *AIP Conference Proceedings* (Vol. 2853, No. 1, p. 020140). AIP Publishing LLC.

[9]. Ganesan, T. (2020). Deep learning and predictive analytics for personalized healthcare: unlocking ehr insights for patient-centric decision support and resource optimization. *International Journal of HRM and Organizational Behavior*, *8*(3), 127-142.

[10]. Jordan Nelson, A. J. (2025). Predictive Analytics in Risk Stratification and Early Diagnosis.

[11]. Mohammed, K., Nolan, M. B., Rajjo, T., Shah, N. D., Prokop, L. J., Varkey, P., & Murad, M. H. (2016). Creating a patient-centered health care delivery system: a systematic review of health care quality from the patient perspective. *American journal of medical quality*, *31*(1), 12-21.

[12]. Onah, C. K., Azuogu, B. N., Ochie, C. N., Akpa, C. O., Okeke, K. C., Okpunwa, A. O., ... & Ugwu, G. O. (2022). Physician emigration from Nigeria and the associated factors: the implications to safeguarding the Nigeria health system. *Human Resources for Health*, *20*(1), 85.

[13]. Parikh, R. B., Kakad, M., & Bates, D. W. (2016). Integrating predictive analytics into high-value care: the dawn of precision delivery. *Jama*, *315*(7), 651-652.

[14]. Pranckevičius, T., & Marcinkevičius, V. (2017). Comparison of naive bayes, random forest, decision tree, support vector machines, and logistic regression classifiers for text reviews classification. *Baltic Journal of Modern Computing*, *5*(2), 221.

[15]. Rahman, A., Karmakar, M., & Debnath, P. (2023). Predictive analytics for healthcare: Improving patient outcomes in the US through Machine Learning. *Revista de Inteligencia Artificial en Medicina*, *14*(1), 595-624.

[16]. Rahman, M. H., Uddinb, M. K. S., Hossanc, K. M. R., & Hossaind, M. D. (2024). The role of predictive analytics in early disease detection: a data-driven approach to preventive healthcare. *Journal of the Learning Sciences*, *32*(2), 2024.

[17]. Risse, G. B., & Warner, J. H. (1992). Reconstructing clinical activities: patient records in medical history. *Social History of Medicine*, *5*(2), 183-205.

[18]. Sarker, I. H. (2021). Data science and analytics: an overview from data-driven smart computing, decision-making and applications perspective. *SN Computer Science*, *2*(5), 377.

[19]. Shafqat, S., Kishwer, S., Rasool, R. U., Qadir, J., Amjad, T., & Ahmad, H. F. (2020). Big data analytics enhanced healthcare systems: a review. *The Journal of Supercomputing*, *76*(3), 1754-1799.

[20]. Tilahun, H., Abate, B., Belay, H., Gebeyehu, A., Ahmed, M., Simanesew, A., ... & Wondarad, Y. (2022). Drivers and barriers to improved data quality and data-use practices: an interpretative qualitative study in Addis Ababa, Ethiopia. *Global Health: Science and Practice*, *10*(Supplement 1).

[21]. Yadav, P., Steinbach, M., Kumar, V., & Simon, G. (2018). Mining electronic health records (EHRs) A survey. *ACM Computing Surveys (CSUR)*, *50*(6), 1-40.