Research Article

Journal of AI & Machine Learning

Advancing Hypertension Risk Prediction in African Healthcare: A Machine Learning Approach with Web-Based Visualization and Interpretability

Damilare Timileyin, Daramola[1*], Folasade Mojisola, Dahunsi[2] and Ponmile Idaresit, Ogunjemite[3]

[1]*Department of Computer Engineering, Federal University of Technology Akure, Nigeria.*

[2]*Department of Computer Engineering, Federal University of Technology Akure, Nigeria.*

[3]*Founder and Executive Director, Cardinal Healthcare Foundation (Pioneer Medical Initiative), Lagos, Nigeria.*

***Corresponding author**

Damilare Timileyin, Daramola,
Department of Computer Engineering,
Federal University of Technology Akure,
Nigeria.

Submitted : 1 Apr 2025 ; Published : 15 Jul 2025

Citation: Damilare., Daramola *et.al.,* (2025). Advancing Hypertension Risk Prediction in African Healthcare: A Machine Learning Approach with Web-Based Visualization and Interpretability. *J AI & Mach Lear., 1*(2):1-8.

Abstract
Hypertension is a critical public health issue, especially in African countries where early detection and management are hampered by resource constraints. This study introduces a robust, machine learning-powered web platform designed for the prediction and management of hypertension risk in African populations. The system leverages a Decision Tree model, achieving a 97% accuracy rate, and is embedded in a user-friendly web interface for real-time risk assessment and patient management. The platform supports personalized dashboards, secure data handling, and seamless integration with electronic health records (EHRs). Data were gathered from outreach programs across Nigeria, encompassing key parameters such as age, sex, systolic and diastolic blood pressure, and body mass index. Comparative analysis with Logistic Regression and Support Vector Machine models highlights the superior performance and interpretability of the Decision Tree approach. The platform is engineered for scalability and adaptability to diverse healthcare environments. Future enhancements will focus on expanding the dataset, incorporating additional health indicators, and exploring advanced machine learning techniques for improved predictive power. This research demonstrates a significant step forward in digital health innovation for Africa, offering a scalable, interpretable, and practical solution for hypertension risk management.

Keywords: Africa, Electronic Health Records (EHR), Hypertension, Machine Learning, Medical Informatics

Introduction

HYPERTENSION, commonly known as high blood pressure, is a widespread health concern globally, contributing significantly to cardiovascular diseases and other complications. According to the World Health Organization (WHO, 19 September 2023), hypertension is a leading cause of premature death worldwide, and its prevalence continues to rise. Early detection and intervention are crucial for effective management and prevention of complications associated with (Bosu et al., 2019). Providing healthcare professionals in Africa with a reliable tool for hypertension prediction, empowers them to make informed decisions and recommendations. This research seeks to bridge the gap between technological innovation and practical application in healthcare settings, fostering a collaborative approach between artificial intelligence experts and medical professionals.

Hypertension is an important public health challenge all over the world (Juliet et al., 2007), it is estimated that over 20% of all adults across the world have hypertension (National Research Council [NRC], 2011). It is the leading driver of cardiovascular disease deaths in Africa and its prevalence is highest in older populations where three out of five older adults in rural and urban Africa have raised blood pressure yet, they receive little attention in many African Countries (Bosu et al., 2019).

Due to its lack of severe symptoms and potentially fatal effects, hypertension has been known as a silent killer (Meher et al., 2023). The following factors can lead to its prevalence: Excessive Salt Intake, Obesity, Diet and Physical Fitness, Sex, Family History, Illiteracy, Socioeconomic Status, Alcohol Consumption.

Healthcare State of Africa

This solution will take into account unique limitations in the Africa society and provide a tool that can be widely accessible and used, even in resource-constrained settings. In Africa, there is a need for scalable and technology-driven solutions that can facilitate the early identification of individuals at risk of hypertension and its complications hence; the engineering challenge is to develop effective and culturally sensitive tools to tackle this health issue (Nematollahi et al., 2023).

The societal and economic impact of hypertension-related complications is substantial mostly in Africa, and by developing accurate predictive models, this research aims to contribute to the broader efforts of Africa public health initiatives. Timely identification of individuals at risk enables targeted interventions, potentially reducing the incidence of severe complications and alleviating the strain on Africa healthcare resources (Bosu et al., 2019).

African countries experience challenges in healthcare infrastructure which cause a significant and growing burden of hypertension, with various factors such as shortage of medical professionals (average of 2 professionals to 1000 people - (WHO, 19 September 2023; Adeloye et al., 2017), limited access to medical facilities, and disparities in healthcare distribution all contributing to its prevalence. Rapid urbanization, changes in lifestyle, and limited access to healthcare resources in some regions have resulted in an increased incidence of hypertension.

Abbreviations and Acronyms

SBP : Systolic Blood Pressure,
DBP : Diastolic Blood Pressure,
BMI : Body Mass Index,
EHR : Electronic Health Records,
WHO : World Health Organisation.

Materials and Methods

Overview of the System

The Hypertension Risk Prediction System is designed to predict the risk level of hypertension in patients based on specific clinical parameters, such as age, sex, systolic blood pressure (SBP), diastolic blood pressure (DBP), Weight, Height and Body Mass Index (BMI). The system utilizes machine learning algorithms to classify patients into three categories: good health, mild risk of hypertension, and high risk of hypertension. This prediction model is integrated into a web application to make it accessible to healthcare providers and patients for real-time risk assessment.

System Architecture

The system architecture consists of several key components, including data collection, preprocessing, model training, prediction, and user interaction through a web interface. The architecture can be broken down as shown below.

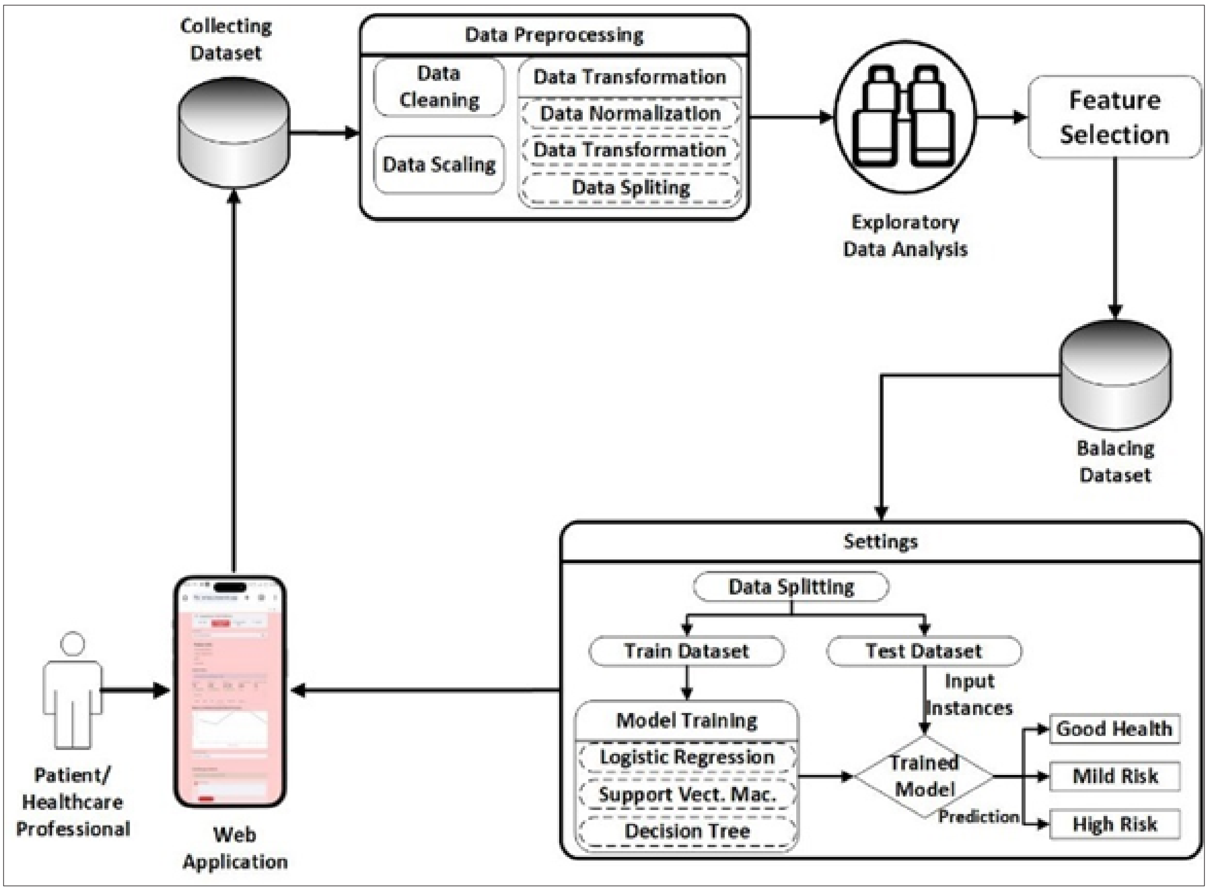

Figure 1: System Architecture Diagram

The system architecture for predicting hypertension using machine learning is integrated with Streamlit - an open source framework that turns data scripts into web applications - it is designed to be comprehensive, scalable, and user-friendly. By integrating machine learning techniques into an intuitive web interface, the system aims to improve early detection, enhance preventive measures, and optimize the available healthcare resources in the management of hypertensive and its related complications - stroke, heart attack, heart failure, kidney disease and loss of life among others.

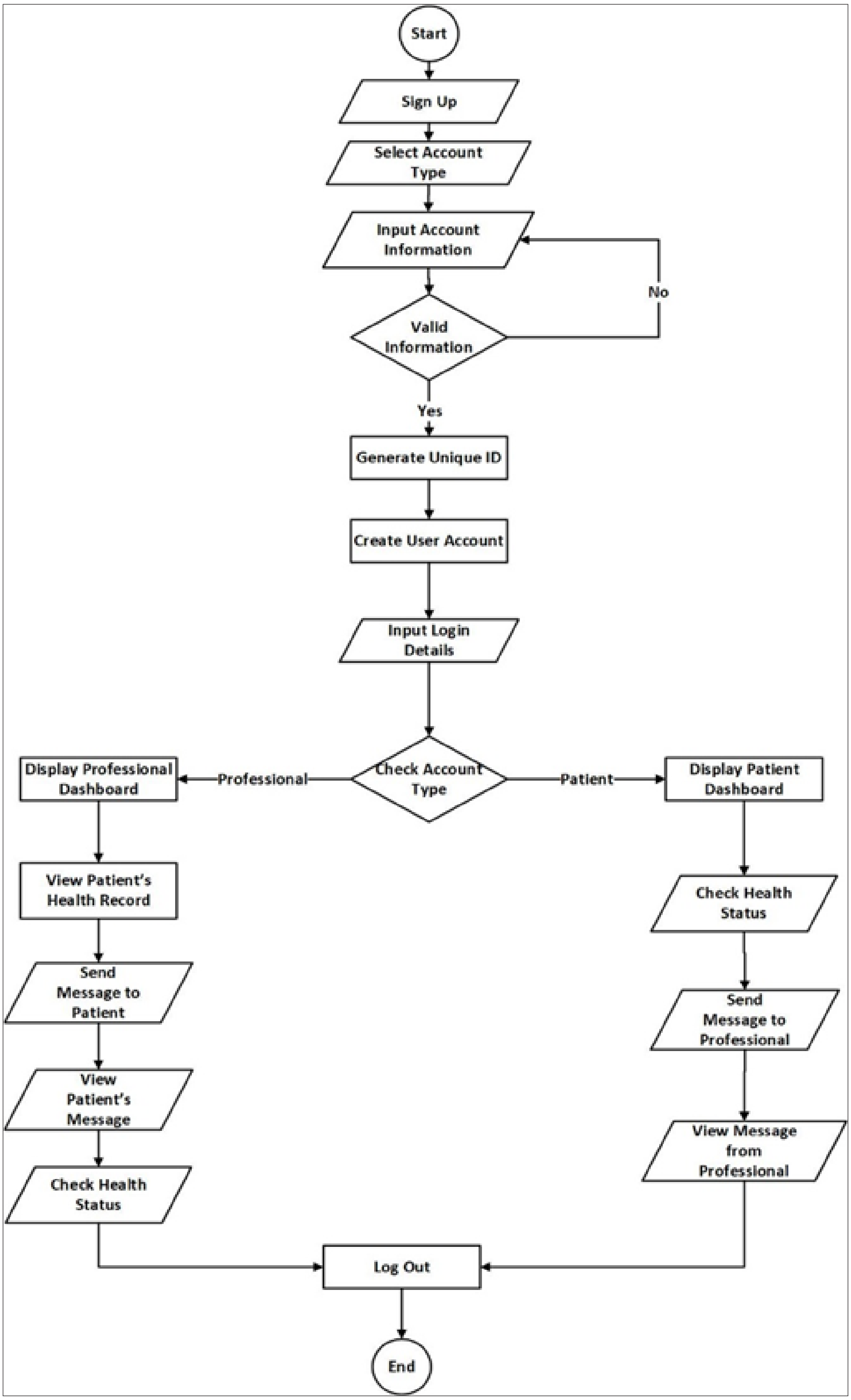

Figure 2: Web Application Flow Chart

The primary language for data processing, model training, and web application development due to its extensive libraries and support for machine learning is Python. Other frameworks used include Streamlit, scikit-learn, Pandas, NumPy, Matplotlib & Seaborn Libraries. MongoDB database was used to stores patient data and model predictions.

Method of Data Collection

The dataset used in this study comprises health records collected from various outreaches within a space of four years (2021 - 2024) across several locations in Nigeria, these outreaches focused on providing basic health checks and screenings for general wellbeing among adults in several communities. The data was utilized to provide an accurate, accessible, and scalable solution for early detection and management of hypertension in Africa.

The dataset consists of 800 records with 350 males and 450 females, and each record represents an individual patient's health profile, all personally identifiable information was removed to ensure patient privacy and comply with ethical standards.

Model Training

Three different machine learning algorithms were trained and evaluated on the preprocessed dataset to determine the most suitable model for deployment. The algorithms tested include Logistic Regression, Support Vector Machine (SVM), and Decision Tree.

1. Logistic Regression: The Logistic Regression model achieved an accuracy of 66%. Although the model was straightforward and provided insights into the relationship between features and outcomes, the accuracy was not sufficient for deployment.
2. Support Vector Machine (SVM): The SVM model achieved an accuracy of 72%. This was an improvement over Logistic Regression, indicating that the decision boundary for hypertension risk prediction might be non-linear. However, the accuracy still fell short of the desired threshold for reliable predictions in a clinical setting.
3. Decision Tree: The Decision Tree algorithm achieved a significantly higher accuracy of 97%, making it the best-performing model among those tested. The model's ability to accurately classify patients into the three risk categories (good health, mild risk, high risk) demonstrated its suitability for deployment in the hypertension risk prediction system.

Model Deployment

Based on the performance metrics, the Decision Tree model was selected for deployment into a Web Application Interface developed using Streamlit, this interface allows users to input data, view predictions, and interact with their health records. This application is hosted on a web server, ensuring accessibility and scalability.

This selection process ensures that the deployed model is both accurate and practical for predicting hypertension risk, providing valuable insights that can aid in the early detection and management of the condition. Model Inputs: The model requires a set of health-related parameters to accurately predict the risk of hypertension. Age, Sex, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Weight, Height, Body Mass Index (BMI), Model Outputs: A categorical output classifying the individual into risk categories based on the risk score which are Good Health, Mild Risk and High Risk. This simplifies the interpretation of the risk score for healthcare professionals and patients.

The system integrates various components, including data collection, data preprocessing, machine learning model development, and deployment, along with user interfaces for healthcare professionals and patients. This comprehensive approach ensures that the system is both technically robust and practically applicable in resource-constrained settings.

Results

Result Overview

The study evaluated three machine learning algorithms—Logistic Regression, Support Vector Machine, and Decision Tree—for predicting hypertension risk. The performance metrics are summarized below:

Accuracy of Machine Learning Models

- Logistic Regression: Achieved an accuracy of 66%, serving as a baseline model. Its moderate performance suggests that while it captures some linear relationships, it may not adequately model complex interactions within the data.
- Support Vector Machine: With an accuracy of 72%, SVM demonstrated improved performance over Logistic Regression. This improvement indicates its effectiveness in handling non-linear relationships through kernel functions.
- Decision Tree: Significantly outperformed the other models with a 97% accuracy rate. The high accuracy underscores the model's ability to capture intricate patterns and interactions between features, making it highly effective for hypertension risk prediction.

Confusion Matrix

The confusion matrix for each model presents a breakdown of predicted versus actual classifications, allowing for an assessment of model performance in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This breakdown is critical for understanding how well each model identifies individuals at risk of hypertension.

- **Logistic Regression:** The confusion matrix indicates moderate performance, with some misclassifications evident. The model may have higher false negatives, suggesting it sometimes fails to identify individuals who are actually at risk.

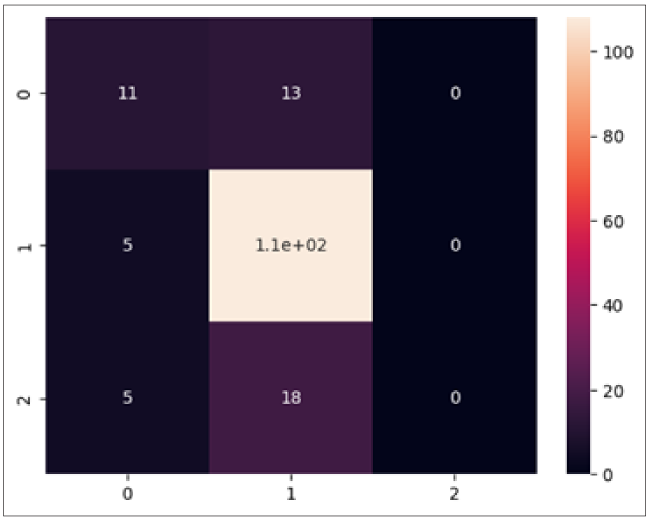

Figure 3: Heatmap Showing Logistic Regression Model Performance

Support Vector Machine: The SVM model shows improved classification accuracy compared to Logistic Regression. However, it may still exhibit some false positives, indicating that it incorrectly labels some non-risk individuals as at risk.

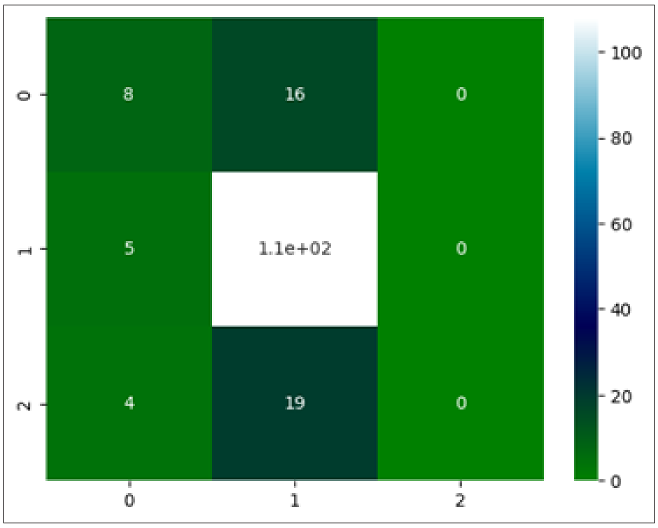

Figure 4: Heatmap Showing Support Vector Machine Model Performance

Decision Tree: This model demonstrates the highest accuracy in the confusion matrix, with a significantly lower rate of both false positives and false negatives. This suggests that the Decision Tree is particularly effective in distinguishing between individuals who are at risk and those who are not.

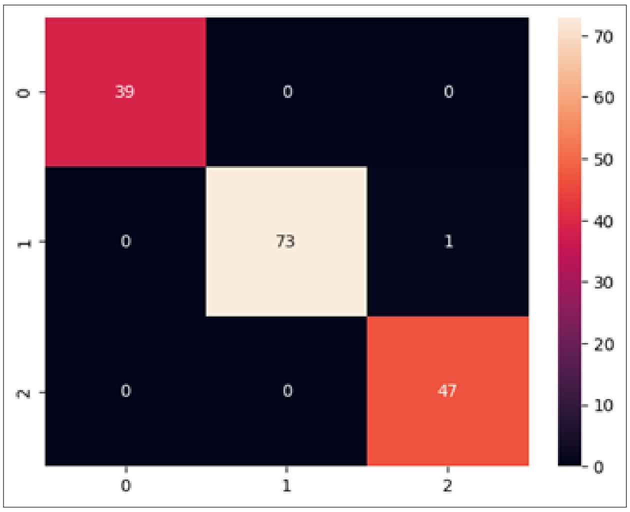

Figure 5: Heatmap Showing Decision Tree Model Performance

Classification Report

The classification report provides additional metrics such as precision, recall, F1-score, and support for each class (risk vs. no risk). These metrics offer a more nuanced view of each model's performance.

Logistic Regression

- **Precision:** Indicates the proportion of positive identifications that were actually correct. The precision might be lower, reflecting its tendency to misclassify some cases.
- **Recall:** Suggests that the model struggles to identify all actual positive cases, leading to a lower recall score.
- **F1-score:** A balance between precision and recall; likely indicates moderate performance overall.

Table 1: Logistic Regression Model Classification Report

	Precision	Recall	F1-score	Support
0	0.52	0.46	0.49	24
1	0.78	0.96	0.86	113
2	0.00	0.00	0.00	23
accuracy			0.74	160
macro avg	0.43	0.47	0.45	160
weighted avg	0.63	0.74	0.68	160

Support Vector Machine

- **Precision:** Generally higher than that of Logistic Regression, showing better accuracy in identifying true positives.
- **Recall:** Improved recall compared to Logistic Regression, indicating better sensitivity in detecting at-risk individuals.
- **F1-score:** Higher than Logistic Regression, suggesting a more balanced performance between precision and recall.

Table 2: Support Vector Machine Model Classification Report

	Precision	Recall	F1-score	Support
0	0.47	0.33	0.39	24
1	0.76	0.96	0.84	113
2	0.00	0.00	0.00	23
accuracy			0.72	160
macro avg	0.41	0.43	0.41	160
weighted avg	0.60	0.72	0.65	160

Decision Tree

- **Precision:** The highest among the three models, indicating that when it predicts a positive case, it is more likely to be correct.
- **Recall:** Also high, meaning it successfully identifies a large proportion of actual positive cases.
- **F1-score:** The best score among all models, highlighting its overall effectiveness in predicting hypertension risk.

Table 3: Decision Tree Model Classification Report

	Precision	Recall	F1-score	Support
0	1.00	1.00	1.00	39
1	1.00	0.99	0.99	74
2	0.98	1.00	0.99	47
accuracy			0.99	160
macro avg	0.99	1.00	0.99	160
weighted avg	0.99	0.99	0.99	160

The Decision Tree model outperforms both Logistic Regression and Support Vector Machine in predicting hypertension risk based on the results from the confusion matrix and classification report. Its high accuracy rates suggest it is a reliable tool for early detection of hypertension risk within the studied population. Conversely, while Logistic Regression and SVM show potential, they exhibit limitations in sensitivity and specificity that could impact their practical application in healthcare settings. These findings underscore the importance of selecting appropriate machine learning models tailored to specific health prediction tasks.

Comparative Analysis

- **Performance Comparison:** The Decision Tree model's superior accuracy compared to Logistic Regression and SVM highlights its robustness in handling the dataset's characteristics. The substantial gap in performance suggests that the Decision Tree effectively manages feature interactions and non-linearities that other models may overlook.
- **Model Interpretability:** Beyond accuracy, the interpretability of the Decision Tree is a critical factor in healthcare applications. Healthcare professionals can trace the decision-making process, understand the influence of each feature, and validate the model's predictions against clinical knowledge.

Discussion

Discussion of Results

This research successfully developed a robust and culturally sensitive predictive model for hypertension, utilizing machine learning techniques. The model addresses the unique challenges faced by African societies in the prevention, early detection, and management of hypertension. The research achieved its objectives by developing a predictive model capable of identifying individuals at risk of developing hypertension in diverse African populations, evaluating the model's performance, and integrating it into a user-friendly web application interface for easy access.

The web application effectively enhances the management and monitoring of hypertension risk for both patients and healthcare professionals, with the following key results:

For Patients

- **Personalized Health Dashboard:** The platform offers a customized dashboard that allows patients to track their blood pressure, BMI, and other vital health statistics, empowering them with up-to-date information on their health status.
- **Direct Communication:** Patients benefit from direct, secure communication channels with their healthcare professionals, ensuring timely advice and support whenever needed.
- **Data Security:** The platform prioritizes the confidentiality of patient data, with robust security measures ensuring that only authorized professionals can access sensitive information.
- **Real-time Health Monitoring:** Patients can input daily health data and receive immediate feedback on their hypertension risk, helping them stay proactive in managing their health.

For Healthcare Professionals

- **Comprehensive Patient Management:** Healthcare professionals can efficiently manage multiple patients, accessing detailed health records and tracking patient progress over time.
- **Data-Driven Insights:** The platform provides analytical tools that allow professionals to interpret patient data effectively, identify risk trends, and make informed decisions on patient care.
- **Efficient Patient Assignment:** The system simplifies patient assignment, making it easy for professionals to be matched with patients.
- **Seamless Integration:** The platform integrates smoothly with existing electronic health record (EHR) systems, allowing professionals to easily access and update patient information, ensuring continuity of care.

The high accuracy achieved by the Decision Tree model indicates its potential as a reliable tool for early hypertension risk detection. Early identification allows for timely interventions, which can significantly reduce the risk of complications associated with hypertension.

The model's ability to accurately classify patients into good health, mild risk, and high risk categories aligns with clinical practices. This alignment ensures that the model's predictions are actionable and can be seamlessly integrated into healthcare workflows.

Limitations

The research is limited by inaccuracies or missing data on key variables, potential biases from outreach-collected data, and differences in healthcare access between rural and urban populations. These factors affect the model's accuracy and generalizability.

- **Dataset Size:** While the Decision Tree performed exceptionally well on the current dataset, the relatively small sample size (800 records) may limit the model's generalizability. Future studies with larger datasets are necessary to validate the model's performance.
- **Overfitting Risk:** The high accuracy may also be indicative of overfitting, where the model performs well on training data but may not generalize to new, unseen data. Techniques such as cross-validation and pruning can help mitigate this risk.

Additionally, acceptance of machine learning-based healthcare solutions may vary due to cultural influences.

Contribution to Knowledge

This research contributes to the existing body of knowledge in several significant ways:

- **Machine Learning in Hypertension Prediction:** The study showcases the potential of machine learning algorithms, particularly Decision Trees, in accurately predicting hypertension risk based on routine clinical parameters. This adds to the growing evidence supporting the use of AI in healthcare.
- **Contextual Application:** By focusing on data collected from health outreaches in Africa, this research highlights the importance of tailoring machine learning models to the specific demographic and health profiles found in underrepresented populations, thus addressing a gap in the global application of AI in healthcare.
- **Feature Relevance:** The study underscores the importance of features such as SBP, DBP, and BMI in predicting hypertension risk, reinforcing their relevance in clinical assessments and risk stratification models.

Recommendations

Based on the findings of this study, several recommendations can be made:

- **For Healthcare Practitioners:** Incorporating machine learning models like Decision Trees into clinical practice can enhance the accuracy of hypertension risk predictions. Practitioners should consider integrating these models into electronic health records (EHR) systems for real-time risk assessment.
- **For Policymakers:** Investment in AI-driven healthcare technologies should be prioritized, especially in regions with limited access to specialized medical care. Policymakers should support initiatives that leverage AI to improve preventive healthcare in rural and underserved communities.
- **For Future Research:** Future studies should focus on expanding the dataset to include a broader range of demographic and clinical variables. Additionally, exploring ensemble methods or deep learning approaches could further enhance prediction accuracy and robustness.

Future Work

Several avenues for future research have been identified based on the findings and limitations of the current study:

- **Data Expansion and Diversity:** Future work should involve collecting larger and more diverse datasets within the Africa population, including additional clinical parameters (e.g., cholesterol levels, family history of hypertension) to improve model generalization across different populations.
- **Model Optimization:** Further research could explore the optimization of the Decision Tree model, or the development of ensemble models such as Random Forests or Gradient Boosting Machines, which might offer even better performance.
- **Real-time Implementation:** Implementing the model in a real-world clinical setting as part of a web application using Streamlit, and evaluating its performance in practice, could provide valuable insights into its practical utility and impact on patient outcomes.
- **Longitudinal Studies:** Future studies could also focus on longitudinal data to predict not just the current risk of hypertension, but also the progression of the condition over time, thereby enabling early intervention strategies.

By addressing these areas, future research can build on the foundation laid by this study, leading to more effective and widespread use of machine learning in predicting and managing hypertension in diverse populations.

Implications

- **Healthcare Practice:** The deployment of the Decision Tree model can aid healthcare providers in identifying high-risk individuals, enabling targeted interventions and resource allocation. This proactive approach can improve patient outcomes and reduce the burden of hypertension-related complications.
- **Policy Making:** Accurate prediction models can inform public health policies by identifying at-risk populations, guiding preventive measures, and allocating resources effectively to areas with higher hypertension prevalence.
- **Technology Integration:** Integrating the Decision Tree model into user-friendly applications, such as web-based platforms or mobile apps, can facilitate widespread access to hypertension risk assessments, empowering individuals to monitor and manage their health proactively.

References

1. WHO, 19 September 2023. [Online]. Available: https://www.who.int/news/item/19-09-2023-first-who-report-details-devastating-impact-of-hypertension-and-ways-to-stop-it. [Accessed 14 December 2023].
2. Bosu W.K., Reilly S.T., Aheto J.M.K. & Zucchelli E., (2019). “Hypertension in older adults in Africa: A systematic review and meta-analysis,” PLoS ONE, vol. 14, no. 4, 2019. DOI: 10.1186/s12872-019-1147-7
3. Nematollahi, M.A., Jahangiri, S. & Asadollahi, A., (2023).”Body composition predicts hypertension using machine learning methods: a cohort study,” Sci Rep 13, 6885, 2023. DOI: 10.1038/s41598-023-34127-6
4. Adeloye, D., David, R. A., Olaogun, A. A., Auta, A., Adesokan, A., Gadanya, M., Opele, J. K., Owagbemi, O., & Iseolorunkanmi, A., (2017). “Health workforce and governance: the crisis in Nigeria,” Human resources for health, vol. 15, no. 1, p. 32, 2017. DOI: 10.1186/s12960-017-0205-4
5. Juliet Addo, Liam Smeeth, & David A. Leon, (2007).”Hypertension In Sub-Saharan Africa: A Systematic Review,” Hypertension, vol. 5, pp. 1012-1018, December 2007. DOI:10.1161/HYPERTENSIONAHA.107.093336
6. National Research Council, (2011).”Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease,” National Academies Press, 2011. DOI: 10.17226/13284.
7. Meher, M., Pradhan, S., & Pradhan, S. R., (2023).”Risk Factors Associated With Hypertension in Young Adults: A Systematic Review,” Cureus, vol. 4, p. 15, 2023. DOI: 10.7759/cureus.37467
8. Montagna, S., Pengo, M. & Ferretti, S., (2023). “Machine Learning in Hypertension Detection: A Study on World Hypertension Day Data,” J Med Syst 47, 1, vol. 1, 2023. DOI: 10.1007/s10916-022-01900-5.