

Data-Driven Lifestyle Therapy Model for Hypertension Management: A Comparative Analysis of Algorithms

Abidemi Akanfe

Department of Computer Science and
Engineering, Obafemi Awolowo
University, Ile-Ife, Nigeria
esejoyakhaine@gmail.com

Theresa Omodunbi

Department of Computer Science and
Engineering, Obafemi Awolowo
University, Ile-Ife, Nigeria
tessydunbi@oauife.edu.ng

Oluwadare Ogunlade

Department of Physiological Sciences,
Obafemi Awolowo University, Ile-Ife,
Nigeria
oogunlade@oau.edu.ng

Joy Akhaine

Department of Computer Science and
Engineering, Obafemi Awolowo
University, Ile-Ife, Nigeria
akanfeabidemi@pg-
student.oauife.edu.ng

Rhoda Ikono

Department of Computer Science and
Engineering, Obafemi Awolowo
University, Ile-Ife, Nigeria
rudo@oauife.edu.ng

Abstract— Primary hypertension in young adults in Nigeria is a significant health concern, with many having risk indicators well in advance of the disease's negative consequences. Preventive care is essential for mitigating the early start of the disease, rather than relying on pharmacological therapy. Lifestyle therapy, a preventative intervention, addresses related risk factors to treat, reverse, and avert chronic conditions like hypertension. This research examined risk factors, developed a predictive model for hypertension classification, and provided lifestyle advice based on these characteristics. The model was developed using three machine learning algorithms: extreme gradient (XG) boost, random forest, and support vector machine learning. Data from 303 individuals at the General Out-patient Department of General Hospital, Kubwa, FCT-Abuja was used to train and evaluate the model's performance. The results showed an accuracy of 93.4%, 92.6%, and 91.8% for the random forest, XG boost, and SVM models, respectively. The performance of the developed system showed that 294 participants (93.7%) out of the 303 participant data tested were accurately classified and given appropriate lifestyle therapy recommendations. The model performance indicates that a thorough analysis of risk variables associated with hypertension can yield effective lifestyle therapy suggestions for illness prevention and early management.

Keywords— Hypertension, Preventive Management, Lifestyle Measures Lifestyle Recommendations, Pharmacological Therapy, Risk Factors, Blood Pressure.

I. INTRODUCTION

Critical decisions influence and define daily existence. Accurate and well-defined data is essential for informed decision-making. Timely and accurate decision-making necessitates the right elicitation, presentation, and application of evidence-based data to effectively support decision-makers in making informed choices. A decision support system (DSS) in computing is an information system that aids decision-makers in making precise decisions through efficient data processing [1]. To ensure sufficient support, methodologies such as data mining and artificial intelligence,

particularly in machine learning, are employed. Data mining reveals common characteristics and patterns in data, assisting decision-makers via domain analysis, data object interaction, and historical evaluation [1]. Data mining techniques employed in Decision Support Systems (DSS) can enhance the knowledge of clinicians and hospital administrators, thereby facilitating decision-making by rapidly examining vast quantities of health information [2].

Clinical informatics is an expanding field necessitating decision support, driven by the demand for effective computerized models and evidence-based information systems such as Decision Support Systems (DSS), which facilitate the management of clinical data to enhance medical care [3]. Machine learning algorithms can be applied to create models for accurate decision-making in clinical informatics, which are known as clinical decision support systems (CDSS). The healthcare sector produces thousands of health records each year, supplying ample information to create extremely precise data models. While little knowledge in crucial health investigations may adversely affect patient outcomes, an excessive amount of complex and overwhelming information might yield similar results [4]. The instruments in this field are designed to assist decision-makers in collecting extensive computational data, recommending treatments, detecting patterns, and diagnosing problems, such as adverse pharmacological connections, to enhance data analysis, decision-making, and patient care [5].

A Clinical Decision Support System (CDSS) employs patient datasets, decision-making algorithms, and pattern recognition to furnish evidence-based recommendations to healthcare providers [6]. It evaluates, diagnoses, and prescribes based on input classifications such as medical history, disease interactions, and laboratory findings. For conditions like hypertension, these classifications can forecast risk and propose potential treatments to optimize health outcomes. The objective of implementing a CDSS is to achieve improved and efficient medical management

where necessary. CDSSs also serve as a mechanism to mitigate adverse drug reactions and medical incidents while reducing healthcare costs [4].

The World Health Organization (WHO) defines hypertension as a medical condition marked by persistent elevation of blood pressure. Blood pressure, a measure of heart blood flow and artery resistance, increases with increased blood flow. Hypertension, characterized by abnormally high blood pressure, can cause symptoms like shortness of breath, severe weakness, heart and kidney failure, stroke, and death [7]. High blood pressure can be caused by two classes: primary hypertension, which is naturally developed due to factors like blood plasma volume and hormone activity, and secondary hypertension, which is caused by previous illnesses like diabetes, kidney problems, hyperthyroidism, pregnancy, sleep apnea, and obesity. Hypertension is diagnosed when a patient has 120-139 systolic or 80-89 diastolic blood pressure, which can lead to high blood pressure if not managed properly [8]. In Nigeria, its prevalence has greatly increased over the past decade [9], [10]. Among adults aged 25 and older, 46% of hypertension cases in Africa are reported. An article in the Bulletin of the World Health Organization reported that Nigerians are increasingly confronting elevated blood pressure [11], posing a substantial challenge in a society with a failing healthcare system and an average life expectancy of 55 years.

[12] In Nigeria, the primary approach to managing hypertension is pharmacological therapy. Even at initial and developing stages of the condition, medications or sedatives are commonly prescribed as the primary treatment modality. In certain instances, patients receive minimal guidance, such as recommendations to monitor their diet or abstain from alcohol, often accompanied by inadequate follow-up and prognosis regarding the impact of hypertension risk factors on their health. While pharmacological interventions play a crucial role in effectively lowering elevated blood pressure, they are associated with several adverse effects such as:

i. Socio-economic factors: A study conducted in a tertiary hospital in Lagos showed that the mean monthly cost of antihypertensive prescriptions as at 2016 was ₦6611.47 (US\$44). The lowest cost per month was ₦300.00 (US\$2) while the highest cost per month was ₦27,030.00 (US\$180) [13]. In Nigeria today, these figures have more than doubled due to the ongoing economic upsurge.

ii. Cultural factors: Due to the continual push towards industrial drugs, several people are geared towards using locally made drugs and traditional methods.

iii. Side effects: Because of the concentrated nature of blood pressure drugs, side effects such as cough, diarrhea or constipation, dizziness or lightheadedness, erection problems, feeling nervous, feeling tired, weak, drowsy, or a lack of energy, headache, nausea or vomiting, etc. These symptoms, more often than not are prevalent with moderate dosages (2-3 tablets a week) of hypertension drugs.

Primary hypertension lacks a definitive remedy; nevertheless, efficient antihypertensive drugs can assist in identifying risk factors related to blood pressure. However, insights from a patient's medical history can facilitate the implementation of management strategies for early hypertension. Clinical decision support systems employing data mining and machine learning algorithms can facilitate the management and prevention of specific diseases. By

examining individual risk factors, distinct categories of hypertension can be identified, facilitating the use of management treatments to ease the disease in the initial phases.

Analyzing risk factors affecting early hypertension in young adults, clinical information systems can be developed to provide effective lifestyle recommendations for preventive management through the use of classification model like extreme gradient (XG) boost machine learning algorithm [12]. The research was further worked on in this paper, by using two additional machine learning algorithms which are random forest model and support vector machine model to create predictive models for the classification of hypertension based on its associated risk.

This article is structured as follows; Section 1.0 provides an introduction and background on hypertension and how a lifestyle predictive model can reduce its adverse effect; Section 2.0 reviews existing works on usage of machine learning in the prediction and classification of hypertension risk.; Section 3.0 discusses the methods employed in carrying out the research work; Section 4.0 presents the results obtained and their discussions; Finally, Section 5 concludes the research work.

II. LITERATURE REVIEW

CDSS can be classified into Knowledge-based and Non-knowledge-based systems. Knowledge-based systems create rules using existing literature, practice, or patient evidence. Non-knowledge-based systems use data references but use AI, machine learning, or statistical pattern recognition for decision-making [14]. These systems are trained on the relationship between clinical symptoms and associated diseases, rather than programming to follow expert medical knowledge. Predictive modelling (or predictive analytics) in healthcare uses technology and statistics to search and find/identify predicting outcomes for individual patients based on past stored data. Predictive modelling is used to formulate regression models, classification models and patient similarity models to support clinical decision support activities [15].

A study found that integrating CDSS for hypertension recognition in medical centers led to significant improvements in disease identification [16]. In the research, within a 6-month period, healthcare workers identified hypertension in 54.9% of medical wards with CDSS adoption, compared to 21.3% in centers without CDSS adoption, demonstrating the benefits of CDSS adoption. Other studies also show the value of CDSS in promoting, preventing, assisting treatment, and improving patient care, with an overall satisfaction (among healthcare workers) with the CDSS's [17],[18],[19].

A plethora of investigations have been executed in the fields of health informatics and machine learning in the prediction and classification of hypertension risk. Some of the research carried out was centered solely around the prevention of primary hypertension, while others focused on predicting the hypertension risk of research participants. A study on the prevalence of high blood pressure among young adults was carried out in Ile-Ife, Nigeria [20]. For the research, seven hundred and two (702) students participated (412 males and 290 females) ranging from age 18 to 40 years of age. The results showed that 50.4% of the participants had normal

blood pressure, 45.2% had prehypertension, 3.7% had stage 1 hypertension, and 0.7% had stage 2 hypertension. The study concluded that regular blood pressure screening for young adults in the society will play an important role when attempting to discover and manage the disease early.

Machine Learning was used to predict hypertension from a Clinical Dataset [21]. An artificial neural network was adopted in the formulation of the model to classify the hypertensive and control patients. The results of the neural network model showed an 82% accuracy. Another predictive model was developed for the classification of hypertension risk using the Decision Trees Algorithm [22]. The method adopted was a decision trees machine learning algorithm which was used to classify the risk of hypertension as low, moderate or high risk given the identified risk factors. After testing the model and gathering results, an evaluation of the performance of the model gave 100% accuracy using the ID3 algorithm and 86.36% accuracy for C4.5 algorithms.

Also, a hypertension prediction system was also developed using a naïve bayes classifier [23]. The model, which was tested on live data, achieved an accuracy of 83.67%. However, only 2% of the dataset was for patients under 36, indicating its effectiveness for younger populations. The classifier could only diagnose hypertension or not have hypertension without proposed therapy.

Again, a hypertension risk monitoring system was developed using an adaptive neuro-fuzzy inference system (ANIFS) model [24]. The model had 100% accuracy during training and 82.4% during testing. Most patients had diabetes mellitus, which triggers secondary hypertension. Preventive management should avoid secondary illnesses or inducing factors.

Furthermore, [25] developed four (4) machine learning models for the automatic classification of hypertension types based on personal features. Four machine learning algorithms were adopted to create four models: C4.5 decision tree classifier, random forest, linear discriminant analysis (LDA), and linear support vector machine (LSVM). The results showed that the c4.5 decision tree classifier and random forest classifier both had a 99.5% accuracy, with the LDA at 96.3% accuracy, and 92.7% accuracy for the LSVM.

Previous research focused on predicting and monitoring hypertension risk times over the systematic prevention of the

disease, captured a relatively small amount of risk factors used in the classification of hypertension risk. The study addresses the issues highlighted above by identifying and capturing fifteen risk factors to help identify areas of improvement and mitigate the early onset of hypertension. The study also focused on predicting and monitoring hypertension risk in young adults aged 18-39 to enable early detection and management of risk factors.

III. METHODOLOGY

The study developed a computational model for hypertension classification and management using a random forest-supervised machine learning classifier. Knowledge was gathered from the Federal Ministry of Health, the Nigerian Hypertension Society, and medical experts. Participants were chosen based on inclusion criteria and data was collected through closed-ended questionnaires and interviews. The dataset collected was cleaned, pre-processed, and used for model formulation. The study developed a lifestyle therapy model for early onset high blood pressure management using three machine learning algorithms: random forest classification, extreme gradient boost algorithm, and support vector machine. The three (3) models were then compared and contrasted and the model with the best performance was selected as the resulting lifestyle therapy model. The Jupyter Notebook software was used in the model formulation and evaluation for the three models formulated. The model data was divided into train and test data, and the results were used to classify hypertension levels and propose lifestyle therapy recommendations for each level.

A. Data Set

The study involved 303 participants from General Out-Patient Department (GOPD) of General Hospital Kubwa, Abuja, who provided consent for data collection. They were measured for height, weight, and blood pressure, with elevated blood pressure levels acknowledged. Data was electronically stored in Microsoft Excel spreadsheets, with each participant designated an identification number as shown in the sample data set in Figure 1.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	ID	PHYSICAL ACTIVITIES	DAILY EXERCISE	ALCOHOL	TOBACCO	FRUITS	VEGETABLES	SUGAR	SALT	BMI	SLEEP	CLASS A	CLASS B	CLASS C	CLASS D	RECC
2	1	2	1	1	1	2	2	1	1	3	2	0	0	1	0	3
3	2	1	1	1	1	1	1	2	1	3	2	1	0	1	1	13
4	3	3	1	1	1	2	1	1	2	2	1	1	0	1	1	13
5	4	1	2	1	1	3	2	3	1	1	3	1	0	0	1	7
6	5	1	1	1	1	1	1	2	2	3	1	1	0	1	1	13
7	6	3	1	1	1	1	1	3	3	3	1	0	0	1	1	10
8	7	1	1	1	1	3	2	1	1	2	2	1	0	0	0	1
9	8	1	1	1	1	1	1	3	3	3	2	1	0	1	1	13
10	9	3	2	2	1	2	2	2	3	3	2	0	1	1	1	14
11	10	1	1	3	1	1	1	1	1	2	2	1	1	1	0	11
12	11	1	1	2	2	3	2	3	3	3	1	1	1	0	1	12
13	12	1	1	1	1	2	1	1	1	3	2	1	0	1	0	6
14	13	2	2	2	2	3	3	1	1	3	1	0	1	0	0	2
15	14	3	1	1	1	1	3	3	2	2	1	1	0	1	1	13
16	15	1	1	1	1	1	1	2	1	2	2	1	0	1	1	13
17	16	1	2	1	1	1	2	1	2	3	1	1	0	1	1	13
18	17	1	1	2	1	2	2	2	2	1	2	1	1	1	1	15
19	18	1	1	1	1	1	1	1	2	3	2	1	0	1	1	13
20	19	1	1	1	1	1	1	3	1	3	3	0	0	1	1	10
21	20	2	2	1	1	1	1	2	2	3	3	0	0	1	1	10
22	21	2	1	2	3	2	3	2	3	2	1	1	1	0	1	13
23	22	1	1	1	1	3	2	2	2	3	1	1	0	0	1	7
24	23	3	1	1	1	2	1	1	1	2	1	1	0	1	0	6
25	24	1	1	1	1	1	2	1	1	3	2	1	0	1	0	6
26	25	3	2	1	1	3	2	2	2	3	1	0	0	0	1	4
27	26	2	2	1	1	2	2	3	2	3	1	0	0	1	1	10
28	27	1	1	1	1	3	2	1	1	3	1	1	0	0	0	1
29	28	1	1	2	1	1	1	1	3	2	1	1	1	1	1	15
30	29	2	1	1	1	1	1	3	3	1	2	1	0	1	1	13
31	30	1	2	2	1	3	2	2	1	3	1	1	1	0	1	13

Fig. 1. Sample Dataset [12]

Three (3) main factors served as the inclusion criteria for participants from whom data was collected for the course of this research. They are as follows:

1. Age: The study involved participants aged 18-39, who are typically young adults due to their low prevalence of high blood pressure, lower stress levels, lower alcohol consumption, and faster fat burning, making them more likely to effectively manage hypertension risk.
2. Prevalence of hypertension/ high blood pressure: The study aimed to evaluate the effectiveness of lifestyle therapy in treating primary hypertension, focusing on participants without or without high blood pressure.
3. Prevalence of secondary hypertension inducing diseases: The research excluded participants with hypertension-inducing diseases, as kidney, adrenal, thyroid, and obstructive sleep apnea-related conditions trigger secondary hypertension.

B. Identification and Description of Variables

From the review of previous literature as well as the knowledge elicited both from medical experts in OAUTHC and the Federal Ministry of Health, certain risk factors that possess the potential of triggering the early onset of hypertension were identified. The identified variables include includes age range, marital status, BMI, high blood pressure prevalence in family members, secondary illness, daily exercise, physical activity, alcohol, tobacco, fruits and vegetables, sugar and salt consumption, sleep, and blood pressure level. These variables were collected from participants and summarized in Table 1. A 3-point Likert scale as described in Table 2 was employed to transform and classify the variables. Data transformation of acquired variables was documented for efficient calculation and manipulation during the data mining process of model development. Each of the identified variables was classified as rarely, sometimes, and often.

C. Target Class Mapping

Each data point in the dataset has its input variables assigned to a corresponding target class, which is the resultant lifestyle therapy class. The resultant class indicates the suggested lifestyle modifications for each dataset according to the detected risk factors. Four principal classes and eleven hybrid classes were defined, as illustrated in Table 3. Class A included daily exercise, daily physical activity, daily sleep, and body mass index as areas of improvement. Class B included alcohol consumption and tobacco consumption as areas of improvement. Class C included fruit consumption

and vegetable consumption as areas of improvement. Class D included sugar consumption and salt consumption as areas of improvement. Other classes were a hybrid of the primary classes for participants who had multiple areas of improvement.

These classes were defined and mapped to the input variables based on the hypertension management guidelines defined by medical experts and the Federal Ministry of Health. This means that for each participant group of data, there was a medical recommendation class for the management of hypertension.

D. Extreme Gradient Boost Model Formulation

Extreme Gradient (XG) boost model, an ensemble technique that uses sequential learning was used to develop a predictive model for hypertension management [12]. The model uses classification trees, with each tree mapping an input data point to a leaf with a continuous score. The training process is iterative, with new trees added to forecast residuals or errors. Gradient boosting uses a gradient descent method to minimize loss. 70% of the dataset was used for model training, while 30% was used for data testing. The model was built by building a base model (m_0) using the average of input variables and a second model (m_1) fitted along the residuals found in the base model. The sequential model formulation process continues until the minimum error rate is reached, known as loss optimization. The final classification is determined by calculating the sum of residuals and taking their average. Figure 2 depicts the framework of the lifestyle therapy model developed with XG Boost.

The model formulation is a sequential recursive learning process where each new tree attempts to get a predicted value closest to the actual target value.

The pseudocode for the model formulation using XG boost is given below:

Given the input variables as “n” number of features

1. Create a single-leaf tree
2. Compute the average of the target variable as a prediction and calculate the residuals using the desired loss function for the first tree.
3. Calculate the residuals for the next and subsequent trees from the prediction made by the previous tree.
4. Calculate the similarity score, “S”
5. Use “S” to select the appropriate node
6. Use “S” to calculate Information gain IG
7. Repeat steps 1 – 6 for “t” number of times to create “t” trees
Predict the residual values using the “t” trees constructed.
8. Calculate a new set of residuals
Repeat steps 1 – 9 to get the final prediction.

TABLE 1. IDENTIFIED VARIABLES

ATTRIBUTES	DATA TYPE	UNIT(S) OF MEASURE
Age range	Numeric	Years
Marital status	Nominal	Yes, no, others
Weight	Numeric	Kilogram (kg)
Height	Numeric	Inches (in)
Family history	Nominal	Yes, no
Secondary illness	Nominal	Yes, no
Daily exercise	Numeric	Minutes (m)
Daily Physical activity	Numeric	Hours (h)
Alcohol consumption	Nominal/ Numeric/ Numeric	Yes, no/ Bottles/ Weeks
Tobacco consumption	Nominal, Numeric	Yes, no/ Weeks
Salt consumption	Numeric	Days
Sugar consumption	Numeric	Days
Fruit consumption	Numeric	Days
Vegetable consumption	Numeric	Days
Daily sleep	Numeric	Hours
Blood pressure level	Numeric	Millimetre of Mercury (mmHg)

TABLE 2. DEVELOPED REFERENCE STANDARD FOR RISK FACTORS [12]

VARIABLE (RISK FACTOR)	CLASSIFICATION	ASSIGNED VALUE
Daily physical activity	Rarely (< 3 hours)	1
	Sometimes (3 - 6 hours)	2
	Often (>= 6 hours)	3
Daily exercise	Rarely (< 30 minutes)	1
	Sometimes (30 – 59 minutes)	2
	Often (>= 60 minutes)	3
Alcohol consumption	Rarely (< 1 bottle weekly)	1
	Sometimes (1 bottle weekly)	2
	Often (>=2 bottles weekly)	3
Tobacco consumption	Rarely (No consumption)	1
	Sometimes (Occasional consumption)	2
	Often (Daily or weekly consumption)	3
Salt consumption	Rarely (<1 teaspoon daily)	1
	Sometimes (1 – 2 teaspoons daily)	2
	Often (>2 teaspoons daily)	3
Sugar consumption	Rarely (<1 cube daily)	1
	Sometimes (1 – 2 cubes daily)	2
	Often (>2 cubes daily)	3
Vegetable consumption	Rarely (Occasional consumption)	1
	Sometimes (Weekly consumption)	2
	Often (Daily consumption)	3
Fruit consumption	Rarely (Occasional consumption)	1
	Sometimes (Weekly consumption)	2
	Often (Daily consumption)	3
Daily sleep	Rarely (< 6 hours)	1
	Sometimes (6 - 8 hours)	2
	Often (> 8 hours)	3
BMI	Rarely (Underweight/ Normal weight)	1
	Sometimes (Overweight)	2
	Often (Obese)	3

TABLE 3. RECOMMENDED LIFESTYLE THERAPY CLASS GROUPING [12]

NUMERICAL REPRESENTATION	CLASS NAME	RECOMMENDATIONS	THERAPY
1	A	1. Daily exercise 2. Daily physical activity 3. Daily sleep 4. Body mass index	1. At least 30 minutes of daily exercise 2. At least 60 minutes of daily physical activity 3. At least 8 hours of daily sleep 4. $18.5 > \text{BMI} < 24.9$
2	B	1. Alcohol consumption 2. Tobacco consumption	1. < 1 bottle of alcohol weekly 2. No tobacco consumption
3	C	1. Fruits consumption 2. Vegetables consumption	1. Daily fruit consumption 2. Daily vegetable consumption
4	D	1. Sugar consumption 2. Salt consumption	1. < 1 teaspoon salt daily 2. < 1 tablespoon sugar daily
5	AB	Class A and B	Class A and B
6	AC	Class A and C	Class A and C
7	AD	Class A and D	Class A and D
8	BD	Class B and D	Class B and D
9	BC	Class B and C	Class B and C
10	CD	Class C and D	Class C and D
11	ABC	Class A, B and C	Class A, B and C
12	ABD	Class A, B and D	Class A, B and D
13	ACD	Class A, C and D	Class A, C and D
14	BCD	Class B, C and D	Class B, C and D
15	ABCD	Class A, B, C and D	Class A, B, C and D

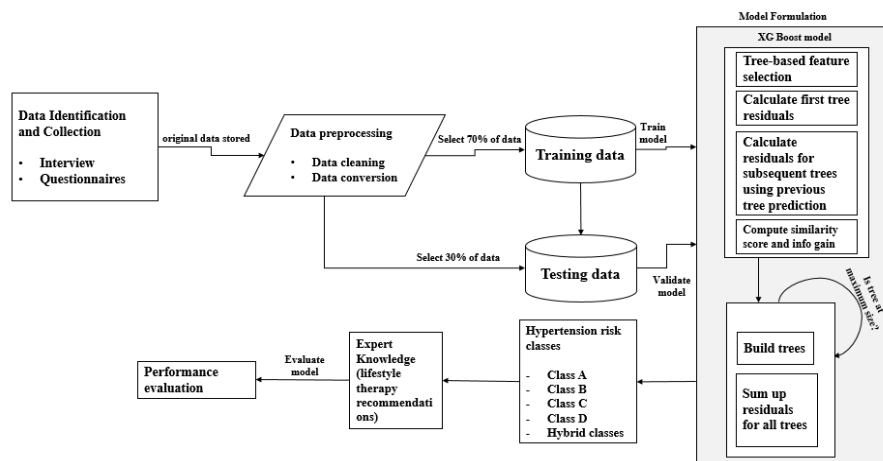


Fig. 2. Framework of Lifestyle Therapy Model developed with XG Boost [12]

E. Random Forest Classifier Model Formulation

The random forest supervised machine learning algorithm is an ensemble method that creates a forest of trees using randomly selected data samples. 70% of the dataset was used for training, while 30% was used for testing. The algorithm uses a large number of uncorrelated trees to perform better than individual trees. The underlying principle behind the random forest classifier is that a large number of relatively

uncorrelated trees operating as a group (in a forest) will perform more effectively than any of the individual trees that make up the forest. This means that in formulating the model, multiple decision trees are generated to classify each risk factor fed into the model as an input variable. For the input variable salt consumption, a decision tree would be generated to answer the question: Does higher salt intake lead to

hypertension class A, or any other hypertension class? This process is repeated for all input variables and hypertension classes. Afterwards, the model gets prediction from each tree and selects the best solution by means of voting. Each separate decision tree is created using the splitting criteria known as gini index, after which each tree votes. The most popular class is chosen as the final result. Figure 3 depicts the framework of the lifestyle therapy model developed with random forest classifier.

The pseudocode for the random forest classification algorithm is given as:

Given the input variables “n” number of features

- i. Select “f” features randomly from the total number of features “n”
- ii. From “f” features selected, calculate the node “d” using the best split point
- iii. Divide the node into daughter nodes using the best possible split
- iv. Iterate through steps (i) to (iii) until “l” number of nodes has been reached. This forms one (1) tree in the forest
- v. Build the forest by iterating through steps (i) to (iv) for “t” number of times to create “t” trees
- vi. Use the rules of each randomly created tree to predict the outcome of the tree
- vii. Store each predicted outcome (of the tree) as the target class
- viii. Compute results for every predicted target
- ix. Consider the high result predicted target as the final prediction from the random forest classifier.

F. Support Vector Machine Model Formulation

The Support Vector Machine Model (SVM) model is a linear SVM (LSVM) that classifies input variables into two classes along a graph. It does not require feature selection and uses a hyperplane to divide the data points. The Decision boundary is a line that divides the two classes, and the optimum hyperplane is one with a large distance between the hyperplanes of both classes. 70% of the preprocessed data was used to train the model, while 30% was used for testing the model's performance. The first step in model formulation is particle fitness evaluation, where multiple hyperplanes are used to divide the input data into two classes. The distance between each hyperplane is measured to determine the "best fit," which is the hyperplane that best separates the data equally along the graph. The optimal velocity and position of the hyperplane are adjusted to better fit the model, and the data value of the hyperplane best fits the model is used as the final classification value. Figure 4 depicts the framework of the lifestyle therapy model developed with SVM.

The pseudocode for the model formulation using SVM is given below:

Given the input variables as “n” number of positions

1. Initialize the particle with random position, r and velocity, v
2. Evaluate the fitness function (F) for each particle's position (r)
3. If $F(r) > F(p_{best})$, set $q_{best} = r$
4. Set best of p_{best} as g_{best}
5. Update velocity (v) and position (r) using p_{best} and g_{best}
6. Repeat steps 1 – 5 until the condition to terminate is satisfied
7. Output the optimal parameters

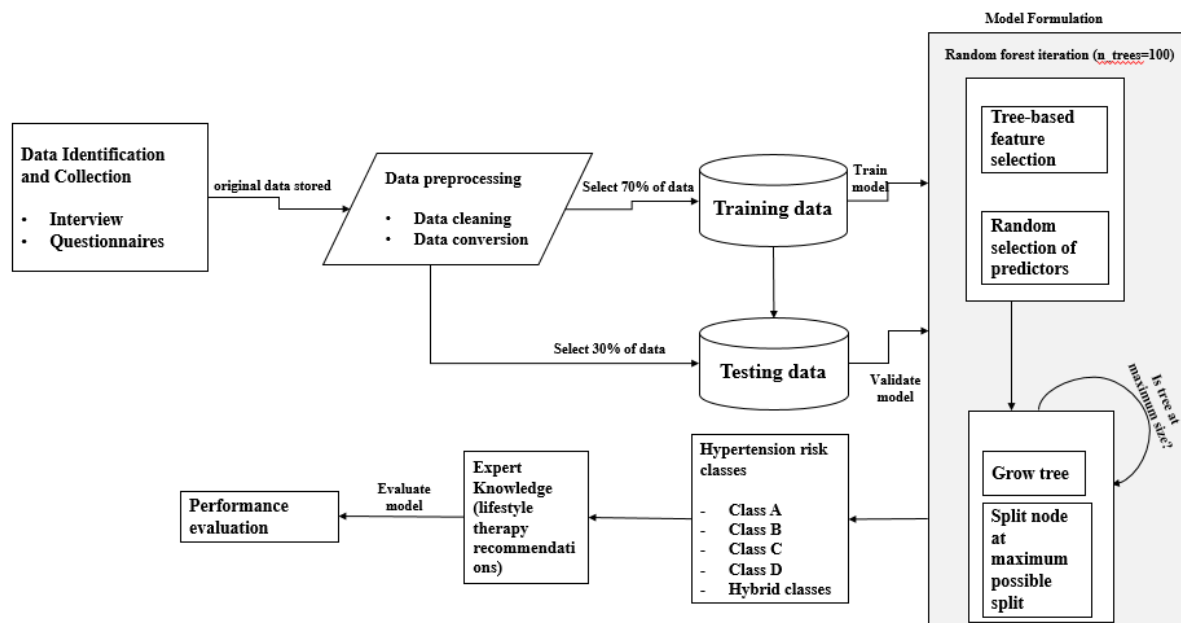


Fig. 3. Framework of Lifestyle Therapy Model developed with Random forest

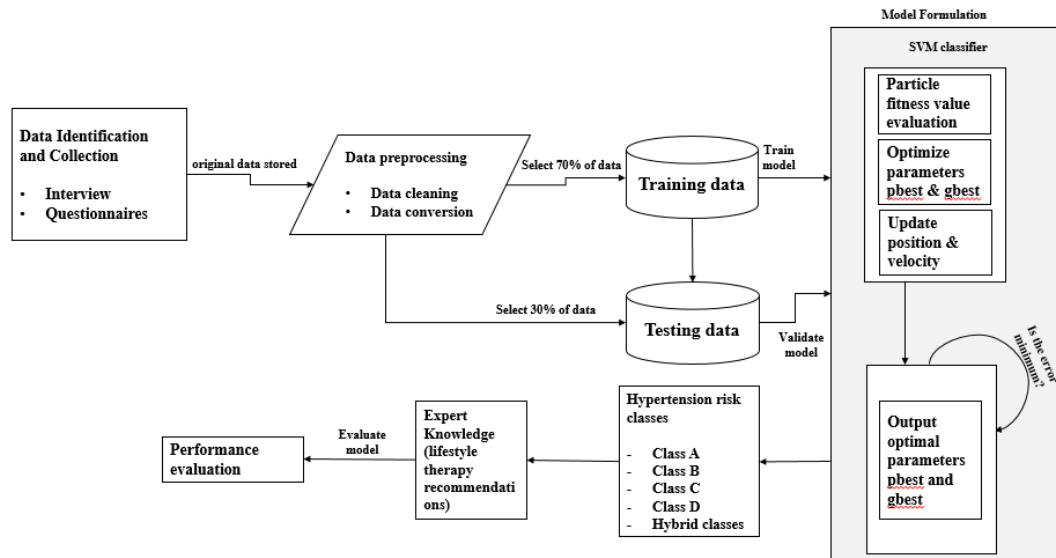


Fig. 4. Framework of Lifestyle Therapy Model developed with SVM

IV. RESULTS AND DISCUSSION

Using the risk factors recorded from the participants, each were given a resulting class of lifestyle therapy recommendations. The summary of the lifestyle therapy recommendations given to the participants is represented in Table 4. The results showed that fifteen (15) participants were given Class A recommendation, six (6) participants were given Class B recommendation, fifteen (15) participants were given Class C recommendation, Twelve (12) participants were given Class D recommendation, six (6) participants were given Class AB recommendation, fifteen (15) participants were given Class AC recommendation, fifteen (15) participants were given

Class AD recommendation, six (6) participants were given Class BD recommendation, three (3) participants had Class BC recommendation, sixty-three (63) participants were given Class CD recommendation, six (6) participants were given Class ABC recommendation, nine (9) participants were given Class ABD recommendation, eighty-seven (87) participants were given Class ACD recommendation, twenty-four (24) participants were given Class BCD recommendation, and eighteen (18) participants were given class ABCD recommendations. Three (3) participants were given no recommendations, and as such were considered “medically fit”.

TABLE 4. LIFESTYLE THERAPY RECOMMENDATIONS

CLASS LABEL	RECOMMENDATION	NO. OF. PARTICIPANTS	PERCENTAGE (%)
A	Increased physical activity Increased daily exercise More sleep Reduced body mass index (BMI)	15	4.95
B	Reduced alcohol consumption Reduced/ no tobacco consumption	6	1.98
C	Increased fruit consumption Increased vegetable consumption	15	4.95
D	Reduced sugar consumption Reduced salt consumption	12	3.96
AB	Class A and B	6	1.98
AC	Class A and C	15	4.95
AD	Class A and D	15	4.95
BD	Class B and D	6	1.98
BC	Class B and C	3	0.99
CD	Class C and D	63	20.79
ABC	Class A, B, and C	6	1.98
ABD	Class A, B, and D	9	2.97
ACD	Class A, C, and D	87	28.71
BCD	Class B, C, and D	24	7.92
ABCD	Class A, B, C, and D	18	5.94
NONE	No recommendation	3	0.01
TOTAL		303	100

A. Extreme Gradient Boost Model Result

The XG boost model was developed using a percentage split technique to evaluate its performance. The model was divided into two sets for training and testing, with 70% of the data used for training and 30% for testing. The confusion matrix as shown in Figure 5 summarizes the classification results, with true positives indicating that the classified value matches the actual value, true negatives indicating that the classified value matches the actual value, and false positives indicating that the predicted value is falsely predicted.

The model achieved a precision of 100% in classifying participants with various recommendations, with a precision of 50% in class ABD, 92% in class ACD, 73% in class BCD, and 0 in class BC. The model had a weighted accuracy of 92.6%, weighted precision of 92%, weighted recall of 93%, and weighted f1-score of 92%. Table 5 shows the results for the individual class recommendations. The model had a mean absolute percentage error of 97.71%, a mean absolute error of 0.22 degrees, a mean squared error of 0.861, and a root mean squared error of 0.928.

B. Random Forest Model Result

The accuracy, recall, support, and precision of the predictive classification model were evaluated using a confusion matrix. The confusion matrix as shown in Figure 6 identifies true positive (TP), true negative (TN), false positive (FP), and false negative (FN) cases of the model's classifications. The results were used to calculate the accuracy, precision, support, and recall of the model. Error rates were calculated, including root mean error, root mean squared error, and mean absolute error.

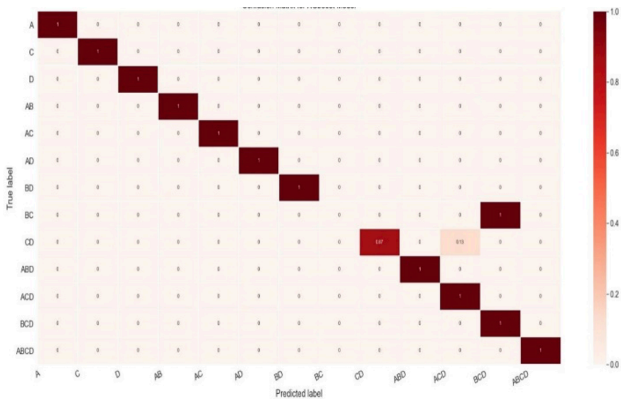


Fig.5. Confusion Matrix for Lifestyle Therapy Model using Extreme Gradient (XG) Boost Algorithm [12]

TABLE 5. CLASSIFICATION RESULTS OF THE XG BOOST MODEL [12]

CLASS LABEL	CLASS NAME	PRECISION	RECALL	F1-SCORE
1	A	100%	100%	100%
2	B	100%	100%	100%
3	C	100%	100%	100%
4	D	100%	100%	100%
5	AB	100%	100%	100%
6	AC	100%	100%	100%
7	AD	100%	100%	100%
8	BD	100%	100%	100%
9	BC	0	0	0

10	CD	100%	88%	94%
11	ABC	100%	100%	100%
12	ABD	50%	100%	67%
13	ACD	92%	92%	92%
14	BCD	73%	100%	84%
15	ABCD	100%	100%	100%

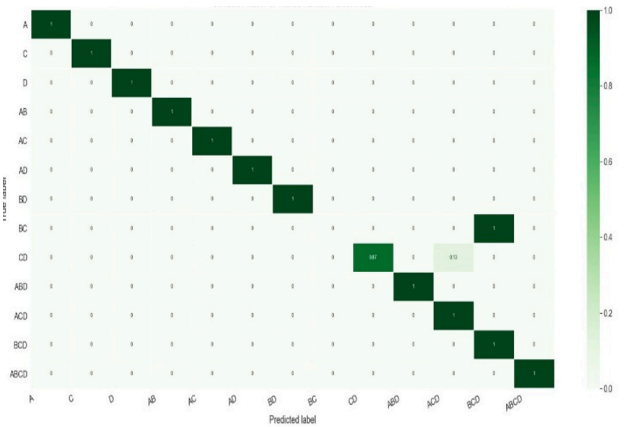


Fig.6. Confusion Matrix for Lifestyle Therapy Model using Random Forest Classifier

Table 6 showcases the performance of the developed model in testing the data. The model did not include class B or ABC in the classification process, as none of the participants had the resulting class B or ABC. The model performed well, with an accuracy of 93.4%, f1-score of 92%, and recall of 93%. The model had a precision of 100% in classifying participants with class A, C, D, AB, AC, AD, BD, CD, ABD, and ABCD recommendations, 89% in class ACD, 70% in class BCD, and 0 in class BC recommendations. The model had a mean absolute error of 0.26 degrees, mean squared error of 1.12, and mean root squared error of 1.06.

TABLE 6. CLASSIFICATION RESULTS OF THE RANDON FOREST MODEL

CLASS LABEL	CLASS NAME	PRECISION	RECALL	F1-SCORE
1	A	100%	100%	100%
3	C	100%	100%	100%
4	D	100%	100%	100%
5	AB	100%	100%	100%
6	AC	100%	100%	100%
7	AD	100%	100%	100%
8	BD	100%	100%	100%
9	BC	0	0	0
10	CD	100%	87%	93%
12	ABD	100%	100%	100%
13	ACD	89%	100%	94%
14	BCD	70%	100%	82%
15	ABCD	100%	100%	100%

C. Support Vector Machine Model Result

The confusion matrix outlines the classification results for the model as shown in Figure 7. A true positive (TP) shows that the classified value matches the actual value, the actual value was positive and the model classified a positive value. A true negative (TN) shows that the classified value matches the actual value, that is, the actual value was negative and the model classified a negative value. A false positive (FP) shows that the predicted value

was falsely predicted (i.e. the actual value was negative but the model classified a positive value. A false negative (FN) shows that the classified value was falsely predicted, that is, the actual value was positive but the model classified a negative value. Table 7 shows the results for the individual class recommendations.

The results showed that the model had a precision of 100% in classifying participants with class B, D, AB, AC, AD, BD, CD, ACD, BCD, and ABCD recommendations, a precision of 71% in classifying participants with class A, a precision of 50% in classifying participants with class C recommendation, a precision of 50% in classifying participants with class ABD recommendation, and a precision of 0 in classifying participants with class BC recommendation. The SVM model gave a weighted accuracy of 91.8%, weighted precision of 93%, weighted recall of 92%, and weighted f1-score of 91%. A mean absolute percentage error of 95.95% was recorded. A mean absolute error of 0.32 degrees was recorded. A mean squared error of 1.648 was recorded. A root mean squared error of 1.284 was recorded.

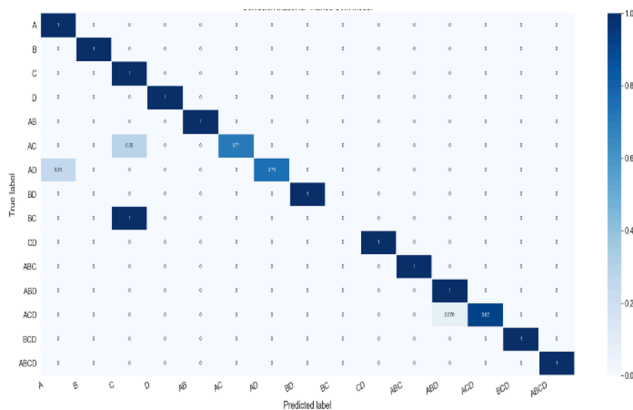


Fig.7. Confusion Matrix for Lifestyle Therapy Model using Support Vector Machine (SVM) Algorithm

TABLE 7. CLASSIFICATION RESULTS OF THE SUPPORT VECTOR MACHINE (SVM) ALGORITHM

CLASS LABEL	CLASS NAME	PRECISION	RECALL	F1-SCORE
1	A	71%	100%	83%
2	B	100%	100%	100%
3	C	50%	100%	67%
4	D	100%	100%	100%
5	AB	100%	100%	100%
6	AC	100%	71%	83%
7	AD	100%	75%	86%
8	BD	100%	100%	100%
9	BC	0	0	0
10	CD	100%	100%	100%
11	ABC	100%	100%	100%
12	ABD	50%	100%	67%
13	ACD	100%	92%	96%
14	BCD	100%	100%	100%
15	ABCD	100%	100%	100%

D. Comparison of Models Result

Table 8 shows the results of the performance evaluation carried out on the three (3) lifestyle therapy models developed. The performance of the models were measured

based on eight (8) criteria: accuracy, mean absolute percentage error, mean absolute error, mean squared error, mean root squared error, precision, recall, and f1-score. The results showed an accuracy of 93.4%, 92.6%, and 91.8% for the random forest, XG boost, and SVM models respectively. The results showed a mean absolute percentage error (MAPE) of 97.18%, 97.71%, and 95.95% for the random forest, XG boost, and SVM models respectively. The results showed a mean absolute error (MAE) of 0.26, 0.22, and 0.32 for the random forest, XG boost, and SVM models respectively. The results showed a mean squared error (MSE) of 1.12, 0.86, and 1.65 for the random forest, XG boost, and SVM models respectively. The results showed a root mean squared error (RMSE) of 1.06, 0.93, and 1.28 for the random forest, XG boost, and SVM models respectively. The results showed a precision of 91%, 92%, and 93% for the random forest, XG boost, and SVM models respectively. The results showed a recall of 93%, 93%, and 92% for the random forest, XG boost, and SVM models respectively. The results showed an F1-score of 92%, 92%, and 91% for the random forest, XG boost, and SVM models respectively.

From the comparison of performance metrics used in evaluating the performance of the formulated models, the lifestyle therapy model formulated with the XG boost algorithm significantly outperformed the models formulated with the random forest classifier and SVM algorithm. This is also understandable as the technique used in XG boost (boosting), was developed as an improvement to the technique used in random forest (bagging). Boosting enables learners in the model formulation “learn” consecutively and adaptively to improve the model’s classifications, while bagging involves homogeneous weak learners in model formulation “learning” from each other independently in parallel, and then, combining them for deriving the model average-which is given as the classification.

TABLE 8. MODEL PERFORMANCE FOR LIFESTYLE THERAPY MODELS DEVELOPED

PERFORMANCE METRIC	RF MODEL	XG BOOST MODEL	SVM MODEL
Accuracy	93.4%	92.6%	91.8%
Mean Absolute Percentage Error	97.18%	97.71%	95.95%
Mean Absolute Error	0.26	0.22	0.32
Mean Squared Error	1.12	0.86	1.65
Root Mean Squared Error	1.06	0.93	1.28
Precision	91%	92%	93%
Recall	93%	93%	92%
F1-Score	92%	92%	91%

V. CONCLUSION

This research presented a computational procedure for the predictive management of primary hypertension in Nigeria, a disease that causes sudden heightened blood pressure in young adults. The study used live data from 303 young adults without primary hypertension to identify significant health risk factors associated with early onset.

These risk factors are then processed and used to formulate three lifestyle therapy models using three machine learning algorithms: random forest classification, extreme gradient (XG) boost algorithm, and support vector machine algorithm. The XG boost algorithm outperforms the random forest classifier, while the SVM algorithm has the least performance. The research concludes that by studying and analyzing risk factors affecting early onset of hypertension among young adults, clinical information systems can be developed to provide effective lifestyle recommendations for preventive management.

REFERENCES

- [1] M. Bhuma, "Health informatics: Prospective and Current research areas," Student Blogs From Karolinska Institutet, 2017. Available from <https://studentblogski.wordpress.com/2016/05/30/health-informatics-prospective-and-current-research-areas/> (accessed on April 12, 2019).
- [2] C. Castaneda, K. Nalley, C. Mannion, P. Bhattacharyya, P. Blake, A. Pecora, and K. S. Suh, "Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine," *Journal of Clinical Bioinformatics*, vol. 5, pp. 1–16, 2015.
- [3] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, and P. C. Tang, "Clinical decision support systems for the practice of evidence-based medicine," *Journal of the American Medical Informatics Association*, vol. 8, no. 6, pp. 527–534, 2001.
- [4] M. Moore, and K. Loper, "An Introduction to Clinical Decision Support Systems," University of Miami Scholarly Repository, 2011. Available from <https://pdfs.semanticscholar.org/88f2/10826ccb7b36925e7e50321274e41e4982e3.pdf> (accessed on April 12, 2019).
- [5] J. Bresnick, "Understanding the basics of clinical decision support systems," *Health IT Analytics*, 2017. Available: <https://healthitanalytics.com/features/understanding-the-basics-of-clinical-decision-support-systems>. Accessed: Apr. 12, 2019.
- [6] R. Renugadevi, P. R. Kumar, G. Kalaiarasi, A. A. E. Raj, S. Settu, and R. Ruthravarshini, "Enhancing healthcare decision support systems with advanced analytics and machine learning techniques," in *Cybersecurity and Data Management Innovations for Revolutionizing Healthcare*, IGI Global, pp. 51–80, 2024.
- [7] A. Felman, "Everything you need to know about hypertension," *Medical News Today*, 2019. Available: <https://www.medicalnewstoday.com/articles/150109>. Accessed: Jul. 30, 2021.
- [8] M. MacGill, "Everything you need to know about hypertension," *Medical News Today Newsletter*, 2018. Available: <https://www.medicalnewstoday.com/articles/150109.php>. [Accessed: Jan. 20, 2019].
- [9] J. T. Akinlua, R. Meakin, A. M. Umar, and N. Freemantle, "Current prevalence pattern of hypertension in Nigeria: A systematic review," *PloS One*, vol. 10, no. 10, pp. e0140021, 2015.
- [10] N. U. Okubadejo, O. B. Ozoh, O. O. Ojo, A. O. Akinkugbe, I. A. Odeniyi, O. Adegoke, B. T. Bello, and O. P. Agabi, "Prevalence of hypertension and blood pressure profile amongst urban-dwelling adults in Nigeria: A comparative analysis based on recent guideline recommendations," *Clinical Hypertension*, vol. 25, no. 7, 2019.
- [11] M. Bello, "Nigerians wake up to high blood pressure," *Bulletin of the World Health Organization*, vol. 91, no. 4, pp. 242–243, 2013.
- [12] T. O. Omodunbi, J. E. Akhaine, O. Ogunlade, A. M. Akanfe, and R. N. Ikono, "Data-driven lifestyle therapy model for hypertension management," in *Proc. 17th Int. Multi-conference on ICT Application. (AICTTRA)*, October, 2024.
- [13] O. Q. Bakare, M. R. Akinyinka, O. Goodman, Y. A. Kuyinu, O. K. Wright, A. Adeniran, O. O. Odusanya, and A. Osibogun, "Antihypertensive use, prescription patterns, and cost of medications in a Teaching Hospital in Lagos, Nigeria," *Niger. J. Clin. Pract.*, vol. 19, no. 5, pp. 668–672, 2016.
- [14] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, "An overview of clinical decision support systems: Benefits, risks, and strategies for success," *NPJ Digital Medicine*, vol. 3, no. 17, 2020.
- [15] R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," *International Journal of Medical Informatics*, vol. 77, no. 2, pp. 81–97, 2008.
- [16] E. O. Kharbanda, S. E. Asche, A. R. Sinaiko, H. L. Ekstrom, J. D. Nordin, N. E. Sherwood, and P. O'Connor, "Clinical decision support for recognition and management of hypertension: A randomized trial," *Pediatrics*, vol. 141, no. 2, 2018.
- [17] J. Asiedu, "Clinical decision support systems, prospects and challenges to healthcare practitioners and patients outcomes in Africa (Low and Medium Income countries (LMIC)),," 2019. Available: https://www.researchgate.net/publication/337339582_Clinical_decision_support_systems_prospects_and_challenges_to_healthcare_practitioners_and_patients_outcomes_in_Africa_Low_and_Medium_Income_countries_LMIC. [Accessed: May 30, 2021].
- [18] T. K. Groenhof, F. W. Asselbergs, R. H. Groenwold, D. E. Grobbee, F. L. J., and M. L. B., "The effect of computerized decision support systems on cardiovascular risk factors: A systematic review and meta-analysis," *BMC Medical Informatics and Decision Making*, vol. 19, p. 108, 2019.
- [19] S. Vuppala and C. Turer, "Clinical decision support for the diagnosis and management of adult and pediatric hypertension," *Current Hypertension Reports*, vol. 22, no. 9, p. 67, 2020.
- [20] M. A. Asafa and O. Ogunlade, "Prevalence of high blood pressure among young adults in Ile-Ife, Nigeria," *International Journal of Analytical, Pharmaceutical, and Biomedical Sciences*, vol. 4, no. 2, pp. 69, 2015.
- [21] D. LaFreniere, F. Zulkernine, D. Barber, and K. Martin, "Using machine learning to predict hypertension from a clinical dataset," in *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1–7, 2016.
- [22] P. A. Idowu, "Predictive model for the classification of hypertension risk using decision trees algorithm," *American Journal of Mathematical and Computer Modelling*, vol. 2, no. 2, pp. 48–59, 2017.
- [23] B. O. Afeni, T. I. Aruleba, and I. A. Oloyede, "Hypertension prediction system using Naive Bayes classifier," *Journal of Advances in Mathematics and Computer Science*, vol. 24, no. 2, pp. 1–11, 2017.
- [24] N. C. Egejuru, O. Ogunlade, and P. A. Idowu, "Development of a mobile-based hypertension risk monitoring system," *International Journal of Information Engineering and Electronic Business*, vol. 2019, no. 4, pp. 11–23, 2019.
- [25] M. Nour and K. Polat, "Automatic classification of hypertension types based on personal features by machine learning algorithms," *Mathematical Problems in Engineering*, vol. 2020, pp. 2742781, 2020.