Journal of Applied Artificial Intelligence
https://jaai.sabapub.com
ISSN: 2709-5908

2025 Volume 6, Issue 2: 28–43
DOI: https://doi.org/10.48185/jaai.v6i2.1849

A Bi-Model Machine Learning Driven Application for Diagnosing the Dominant Illness among Typical Nigerian University Students

Dauda Olorunkemi Isiaka [1*], Malik Adeiza Rufai [2], Abubakar Aliyu [3]

[1,2,3]Department of Computer Science, Faculty of Science, Federal University Lokoja, Kogi State, 058, Nigeria.

Received: 14.10.2025 • Accepted: 26.12.2025 • Published: 30.12.2025 • Final Version: 30.12.2025

Abstract: The implementation and deployment of machine learning models for the diagnosis of dominant illnesses among students require significant investment in technology and infrastructure, which is among the barriers for healthcare organizations with limited resources. In order to increase its adoption, this research suggests the creation of a Bi-Model Machine Learning Driven Application that will enable university students to get diagnosed with common ailments. The plan is to apply a high-level model using a hybrid methodology that combines the development of Machine Learning Models with Agile Software Development. In order to do this specifically, Python was used to implement exploratory data analysis, classification, and regression models, as they have proven to be highly effective in both diagnosing the primary illness and predicting the length of hospital stay. The bi-model were built with four different algorithms each, so as to adopt the ones with best performance for the deployment. The model built with Gradient Boosting Classifier has 100% accuracy, 100% precision, 100% recall as compared to other three algorithms through three repeated training of the model. On the prediction of admission duration task, Gradient Boost Regression works best, and this is because it has the least Root Mean Square Error of 0.57 and Mean Absolute Error to be 0.423 among other compared three algorithms, as measured. This was achieved through the use of fresh localized dataset from the Federal University Lokoja Health Center, which was pre-processed, and stored in the file manager/internal storage for visualization and modelling. Furthermore, the completed models was deployed to a web application using flask and Mysql Lite Database. In the end, the application reduced human error in diagnosis and care management of the student population while they are pursuing their education by enabling evidence-based awareness, educated public health policy, and individualized treatment.

Keywords: Bi-model, application, dominant illness, machine learning, deployment

1. Introduction

An increasing percentage of people worldwide are afflicted with common illnesses these days, and caring for them comes at a steep financial cost. According to [1], over half of the population in a given nation, state, neighborhood, organization, or school is afflicted with a common illness such as malaria, diabetes, or hypoglycemia.

In educational settings, students are among the most susceptible patient groups because of the rapid deterioration of their health that can occur from living alone for the first time, eating irregularly, not knowing about a common sickness, and other causes [2]. Numerous such ailments lower patients' quality of life, causing stress to their family members and caregivers, and divert students from their studies.

[1*] Corresponding Author: dauda.isiaka@fulokoja.edu.ng

Instead of explicitly deploying the models with a localized dataset to demonstrate their practicability, prior studies have been on developing machine learning models that have an impact on the safety and quality of patient care. These models can then be integrated into user-friendly applications that will be affordable for the majority of small health centers, which lack the funding that tertiary health institutions do. Moreover, documentation can be done electronically, on paper, or with both in order to gather data on the patient's vital signs, diagnosis, course of treatment, and progress pertinent to a patient's healthcare needs (Adeyemo and Olaogun, 2013). These enormous amounts of data would save costs and enhance the quality of healthcare services. Additionally, it supports community health management, disease surveillance, and clinical decision support [4].

Precision medicine, chronic disease or prevalent illness management, medical imaging, population health, and other areas of care have all benefited from the usage of machine learning applications in the medical field. This might expedite the diagnosis and prediction of sickness and increase the efficiency of medical delivery by lowering administrative costs. As a result, diagnosing using information gathered from patient interactions is time-consuming and ineffective since it frequently depends on the doctor's subjective judgment, which is not error-free. Therefore, it is possible to create and implement a machine learning-driven application that will allow for the accurate diagnosis of patients within a community.

As a result that, machine learning had improved the quality of patient diagnosis, there are still a number of obstacles to overcome, including poor data quality, a lack of trained AI teams, the high cost of putting machine learning solutions into practice, and the use of biased data [5]. Furthermore, it has never been demonstrated that employing randomized controlled trials and firsthand work experience to reduce the main consequences of frequent illness incidents is particularly beneficial in this context [6]. Likewise, data analytics might be used effectively to establish an application for managing the medical care of those with dominant illnesses.

This study expands on earlier works [7], [8] that offered background reports focusing on Exploratory Data Analysis and model building. Based on the use of more advanced tool to integrate the bi-model built from the complex localized data, a strong conclusion on the significance of creating a strategy for the creation of a real world bi-model machine learning driven application for the diagnosis of dominant disease among students of the university would be achieved.

Regarding [7], the following enhancements were made specifically:

i. The models built will be integrated into a frontend of the Web App, and a thorough and seamless operation of the bi-model will also be controlled by one prediction button from the interface. Compared to the framework in [7], which has the models to be used separately, is not user friendly and technically not right for user to do prediction from the engine, as they will experience difficulty using the model irrespective their level of computer savvy, and also [8] whose emphasis was exploratory data analysis for insights and features' engineering. A more user friendly interface for the suggested application's running will be made possible by the one button to control all models within the system.

ii. To boost the output quality of a suggested application, the two tasks—classification and an admission's runtime prediction module—have been combined as bi-model control by one prediction button on the interface.

The purpose of this study is to devise a system for showing ailment that are often experienced by undergraduate students through the integration of a Bi-Model into a Web App for Diagnosing the Dominant Illness.

This research makes a contribution by suggesting an efficient strategy for deploying a diagnosis application driven by Bi-model machine learning technique via;

i. curating a localized medical dataset from a Typical University Health Centre.

ii. building of four machine learning algorithms each for the two tasks (Bi-Model) predictions based on the data curated.

iii. evaluation of the four machine learning algorithms each for the two tasks built; and

iv. deploying the best performed Bi-Model into a Web App for Diagnosis predictions.

2. Related Work

[9] note that the model's accuracy has to be increased in their research titled "Theoretical model of health management utilizing data-driven decision-making: the future of precision medicine and health." They examined the potential applications of EHR data to offer data-driven recommendations to physicians and patients to support improved decision-making. It was done using formal concept analysis (FCA), albeit it was not made clear how accurate the results were. In order to improve patient outcomes, the model's ability to influence wise healthcare decisions should be illustrated with actual data from EHR databases. The proposed framework made use of data pre-treatment methods as cleaning, transformation, and the Naive Bayes classifier. To verify the accuracy of the proposed framework, confusion matrix (CM) and receiver operating characteristic (ROC) tests were employed.

[7], suggested an improved data analytics framework for the predictive diagnosis of frequent ailments affecting college students be created, so that users can interact with the system through a user-friendly user interface. Further recommended that research should also be focused on integrating established models into a web application. And also that, it is a less explored field of study that is deserving of future investigation. They only employed the CRISP-DM (Cross Industry Standard Process of Data Mining) technique. According to the findings, they developed a machine learning model for the diagnosis and prediction of common ailments, but they only used it for two jobs and didn't implement it for practicality or user-friendliness. In order for a disease to be widespread or recurrent among a population, measures must be taken to significantly lower its incidence or hospitalization rate, as well as any associated distractions [8]. The model's involvement in the identification of this common ailment through an intuitive interface will enable evidence based awareness on the prevalent illness, personalized treatment, and reduce human-prone errors.

The framework with emphasis on the ensemble Gradient Boosting classifier and regression had 100% accuracy and mean absolute error of 0.18, respectively, when compared to other examined frameworks that employed survey datasets, standardized or online repositories' dataset. Because it can handle large and small data sets without affecting the model's performance, it is also stable. The improved outcomes made possible by the localized dataset show how useful it is to use local data sources when creating models for the diagnosis and prognosis of common illnesses that affect people in any given location.

A machine learning method was created by [10] to differentiate between normal and infected blood cells. A variety of ML algorithms are available to ML designers for use in creating new concepts. The type of data used to train the model, the quantity of features, and—most importantly—the total amount of data available are some of the factors that affect the decision of which machine learning approach to utilize [11]. Many types of data, including as biological and environmental data, have been used in supervised learning to predict a particular sickness outcome. Nonetheless, social determinants of health, which have gotten less attention, may be significant indicators of public health problems including malaria and childhood anemia [12]. For instance, commonly used diagnostic techniques like the Rapid Diagnosis Test have limitations [13]. When taking into account the effectiveness of the most widely used machine learning models for the task of Malaria occurrence prediction based on sign and symptom information, machine learning algorithms—particularly Naive Bayesian—look promising [13]. According to [14], there are several barriers to data capture in biomedical applications, such as hardware limitations, small clinical spaces, patients who don't follow study protocols, and difficulties in collecting large amounts of biomedical data. These obstacles compel ML designers to work with imperfect data and devise fixes. In order to power the application, the teenage dataset must also be collected in a way that accurately depicts the educational setting and associated components.

3. Materials and Methods

This contains the following sub-sections: Description of existing system, Architectural Design of the proposed application, High Level Model of the proposed Application, and implementation method that was used to realize the Bi-Model Machine Learning Driven Application. Below is a full description of all the sub-sections;

3.1. Description of Existing System

The process of prevalent illness diagnosis is to identify if an individual from the population sample (undergraduate students) is suffered with the identified prevalent illness. The class of diagnosis could be either you are suffered by the prevalent illness disease or other illness. The existing system was designed such that the patient shows up quickly or visits the clinic; the vital signs would be taken, after which the doctor diagnosis starts and recommends, if a test is needed to conclude the treatment, and if not, symptomatic treatment approach is often used to treat the patients.

3.2. Architectural Design of the Bi-Model Machine Learning Driven Application

Knowing that the objective is to create a Bi-Model Machine Learning Driven Application, with the capacity to categorize patients as being afflicted by the prevalent sickness or not, as well as anticipate the length of time that patients with such illness would need to be admitted, we advocated the use of electronic inputs from a user interface and storage utilizing CSV files and data analytics and prediction strategy.

The user interface (Read Data) module feeds in the input that is made of comma separated variable (CSV) file into the file manager, then immediately, the process of data understanding kicks up leading to the data preparation stage where dataset is divided into the training and testing data. The training data is used to develop the models and they are validated by the testing data, and thereafter used by the Random Forest, Gradient Boosting and Support Vector Classifiers and Regression algorithms for the classification and prediction for the final output of the models. The evaluation of the respective algorithms for accuracy would be done, so as to show that Gradient Boosting Classifier and Gradient Boosting Regressor are suitable for the data analytic model for patient care management in a typical university health centre. Furthermore, the model is to be deployed into Web app using flask and MySQL Lite database. The input entries are processed to produce output of both models on result page, and which all entries are stored in the database for easy access, view and retrieval.

Figure 1 illustrates the architectural design of the Bi-Model machine learning driven application and lists the activities that take place there.

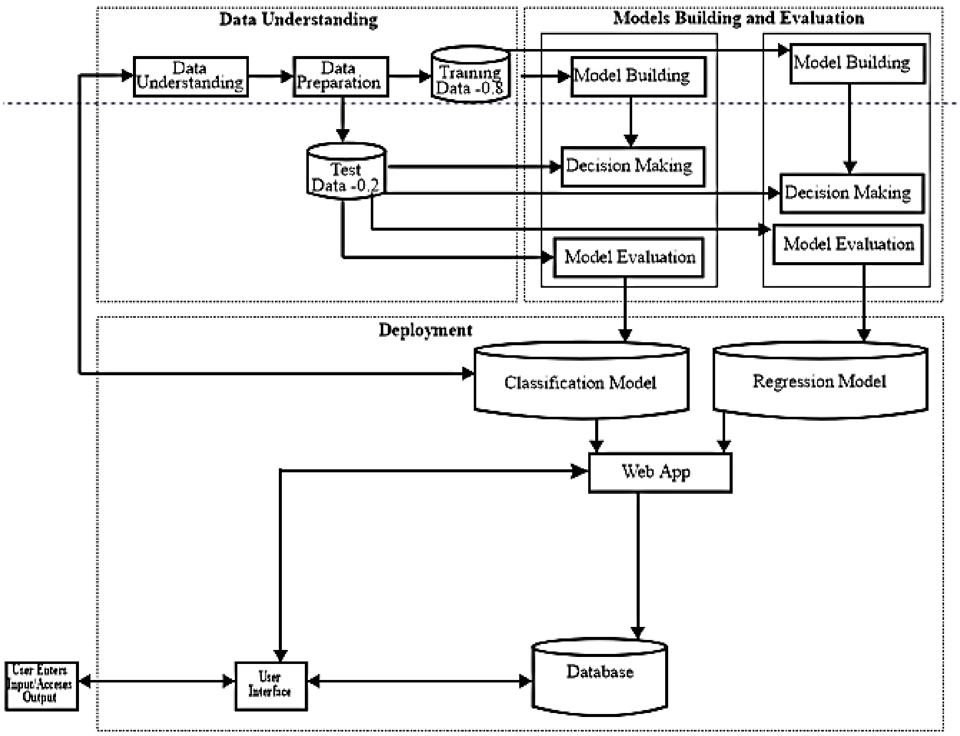

Figure 1. The architectural design of the Bi-Model machine learning driven application being proposed.

The general model and its constituent parts are shown in Figure 2 below.

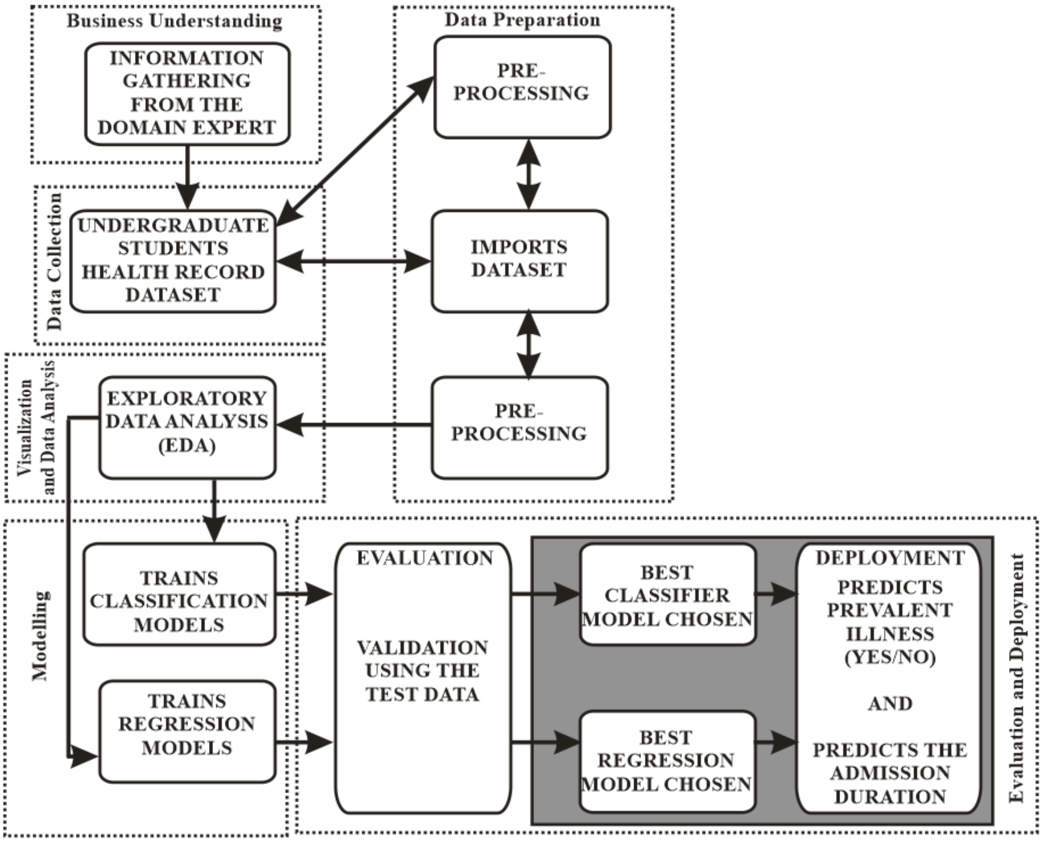

Figure 2. The High-Level Model of the Bi-Model machine learning driven application being proposed.

The diagram above depicts the generic model with all the components for the realization of it.

i. Business Understanding: This phase of the model is where we would visit the domain experts to seek their views to determine the prevalent illness among students, and seek for knowledge on their mode of operations in tackling it. Other information that can assist the model would also be gathered.

ii. Data Collection: After determining the prevalent illness from the data gathered from the health workers, a health record dataset of students with the complete features would be collected and a better understanding on what the various features are, and their relevance.

iii. Data Preparation: The dataset collected is pre-processed, imported, transformed, structured and make ready for further process in the next phase.

iv. Data Analytics and Visualization: An exploratory data analysis would be employed to do feature engineering so as to gain more insight visually into the dataset through plotting of graphs, tables and matrix.

v. Modelling: The models would be built with the appropriate supervised machine learning algorithms for prediction and classification of the prevalent illness (task 1) and prediction on the duration of admission (task 2).

vi. Evaluation and Deployment: The model developed would be evaluated and validated at this point, such that the outcomes of the metrics determine whether to do some adjustment for a further improvement or not. If the desirable performance is achieved, the model would be deployed into a Web App using flask and MySQL Lite Database. The user would login, input the data of the patients through a user interface, and then submit for the system to perform the two tasks. The entries by user would then be converted into numeric values or equivalents assigned to the responses, store them into the database in a comma separated values (CSV) format in the form of table that can be accessed, viewed and retrieved.

3.3. Methodology

A Hybrid Methodology was adopted, because the system has the data model end and the application software end.

For the data model end, a Cross Industry Standard Process of Data Mining (CRISP-DM) methodology would be followed, because of the data-driven nature of the end.

Meanwhile, the agile methodology was used during the software development, because it focuses on streamlining the system development life cycle (SDLC). Also, much of the modelling and documentation overhead are eliminated. Agile methodology is also useful in the development of systems with short time schedule and also system with unclear user requirements.

There are several popular approaches to agile methodology, but for the implementation of the application software end of this proposed system, the Extreme Programming (XP) approach was used.

3.4. An explanation of Dataset

The data for the network training and testing, which are composed of the characteristics to train the model and the target variable to categorize the output of the classification network together with their values and measurements, are contained in the table, which has 1048 sets of observations. To develop models for classification and regression, respectively, which result in the values of the dependent variable, the characteristics may be categorical, nominal, or ordinal in type.

Table 1 below displays the sample of dataset gathered from the university:

Table 1. A sample of the localized dataset from the Health Service Directorate, Federal University Lokoja – Nigeria.

S/N	FILE NO	SEX	AGE	TYPE OF ATTENDANCE	COMPLAINTS	DIAGNOSIS	OUTCOME
1	150	F	23	NEW	CATARRH, COUGH, BODY PAIN, FEVER	MALARIA	TREATED
2	958	F	30	FOLLOW UP	FEVER, JOINT PAIN, HEADACHE	MALARIA	TREATED
3	962	M	23	NEW	FEVER, JOINT PAIN, HEADACHE	MALARIA	TREATED
4	957	F	20	NEW	JOINT PAIN, DYSENTRY STOOL, ABDOMINAL PAIN	DYSCENTRY	TREATED
5	565	F	18	NEW	PV ITCHING	VULVO-VAGINA	TREATED
6	961	F	18	NEW	FEVER	MALARIA	TREATED
7	391	F	20	NEW	V/S DOBLU	LUMBARYO	TREATED
8	658	M	24	NEW	COLD, CATARRH	FLU	TREATED
9	859	M	21	NEW	COLD, FEVER	MALARIA	TREATED

Based on this dataset, female students record the highest rates of hospitalization and Malaria was the outcome of most diagnosis for that week.

4. Result

4.1. Data Transformation and Cleaning

In order to make sure there are no null data, the data gathered from the Health Service Directorate was cleaned by checking for missing values, contaminants, and data imputation. An interval scale "age group" feature was created from the "age" column, and also, the "complaints" column was split into all the possible complaints as features towards data exploration and model building.

4.2. Exploration and Preparation of Data

In this stage, an interpretation of the data was used to comprehend the data in order to describe what the dataset contains by tabulating all relevant parameters and also to depict the dataset's behaviors using both univariate and multivariate analysis approach. Furthermore, Pearson correlation coefficient was used to find out whether there are correlations amongst the features of the dataset.

Exploratory Data Analysis (EDA) was used to visualize and gain more insights from the data collected.

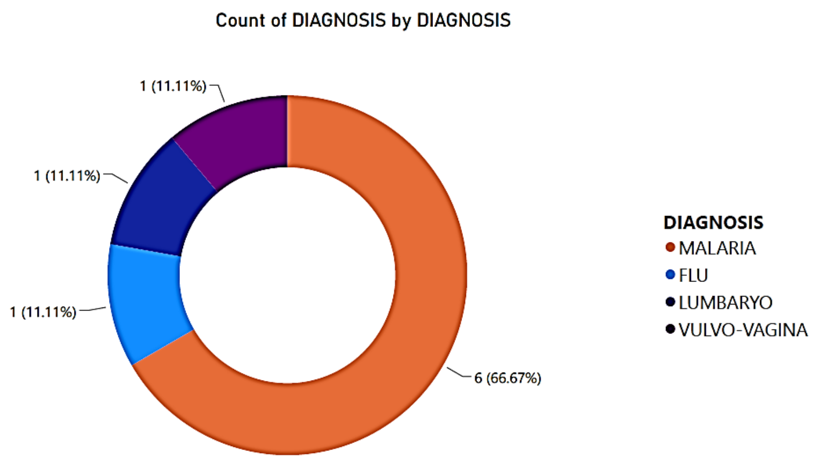

Figure 3: Doughnut Chart to show the Frequencies of Diagnosis Results

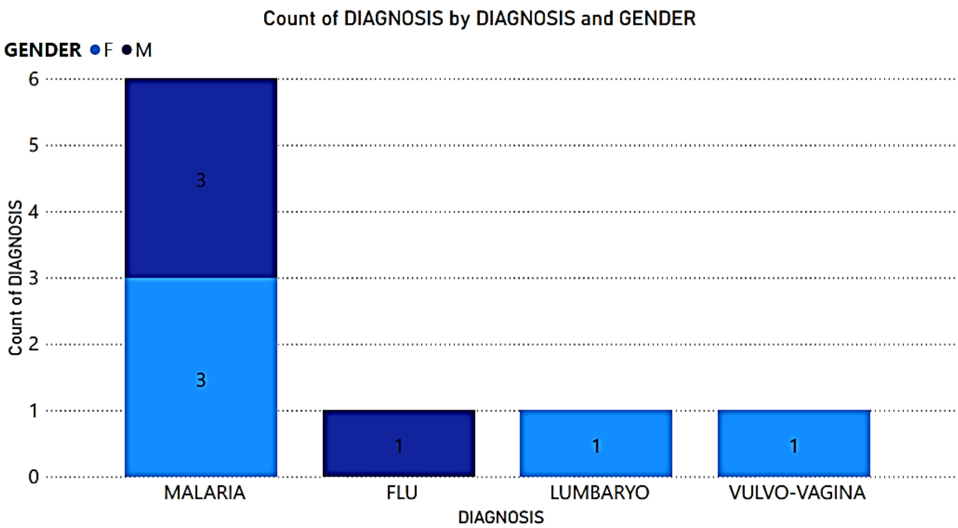

Figure 4: Bar Graph to show the Relationship between Gender and the Diagnosis Results

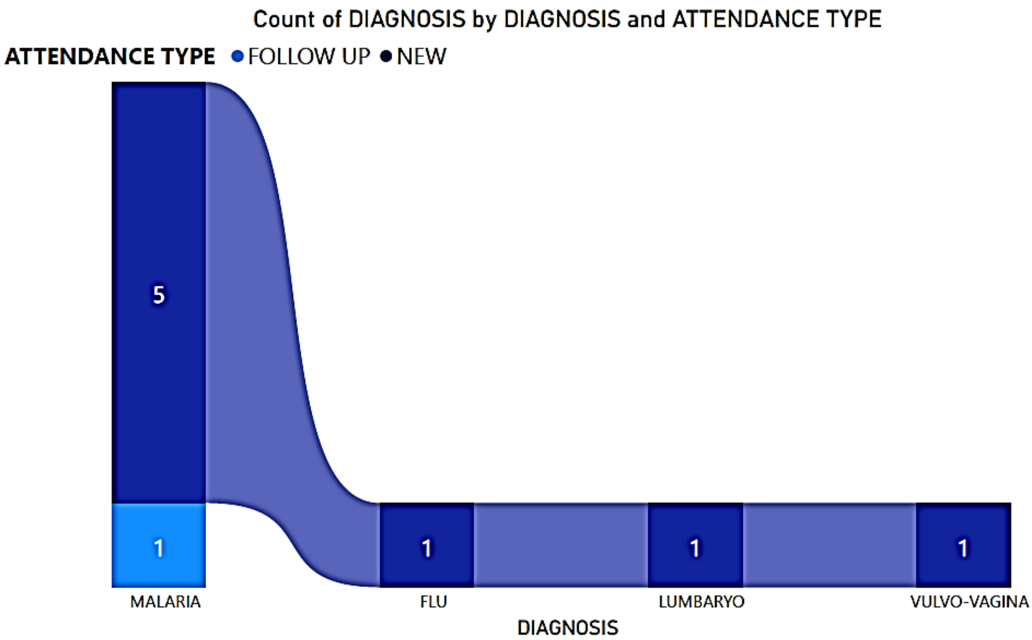

Figure 5: Bar Graph to show the Relationship between Visit Type and the Diagnosis Results

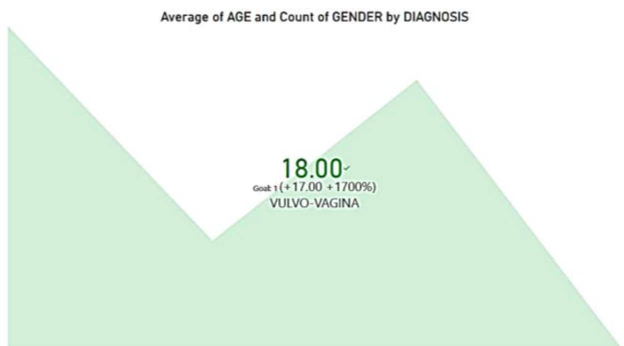

Figure 6: Trend Graph to show the average age of students, Gender and the Diagnosis Results

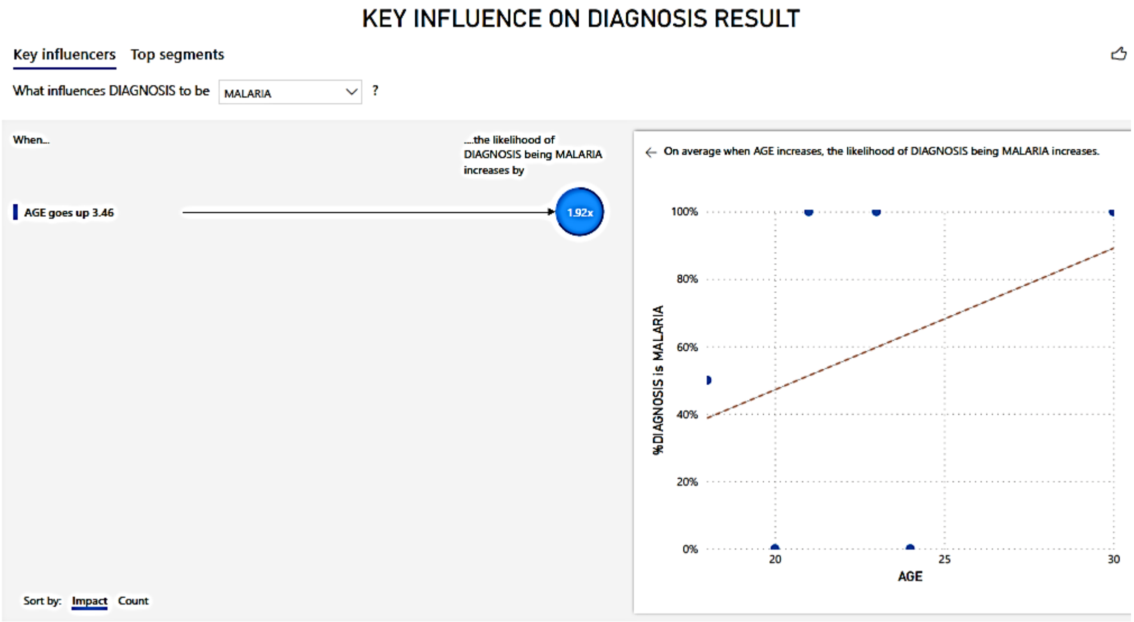

Figure 7: Graph to show that Age of the Students has Key Influence on the Diagnosis Results

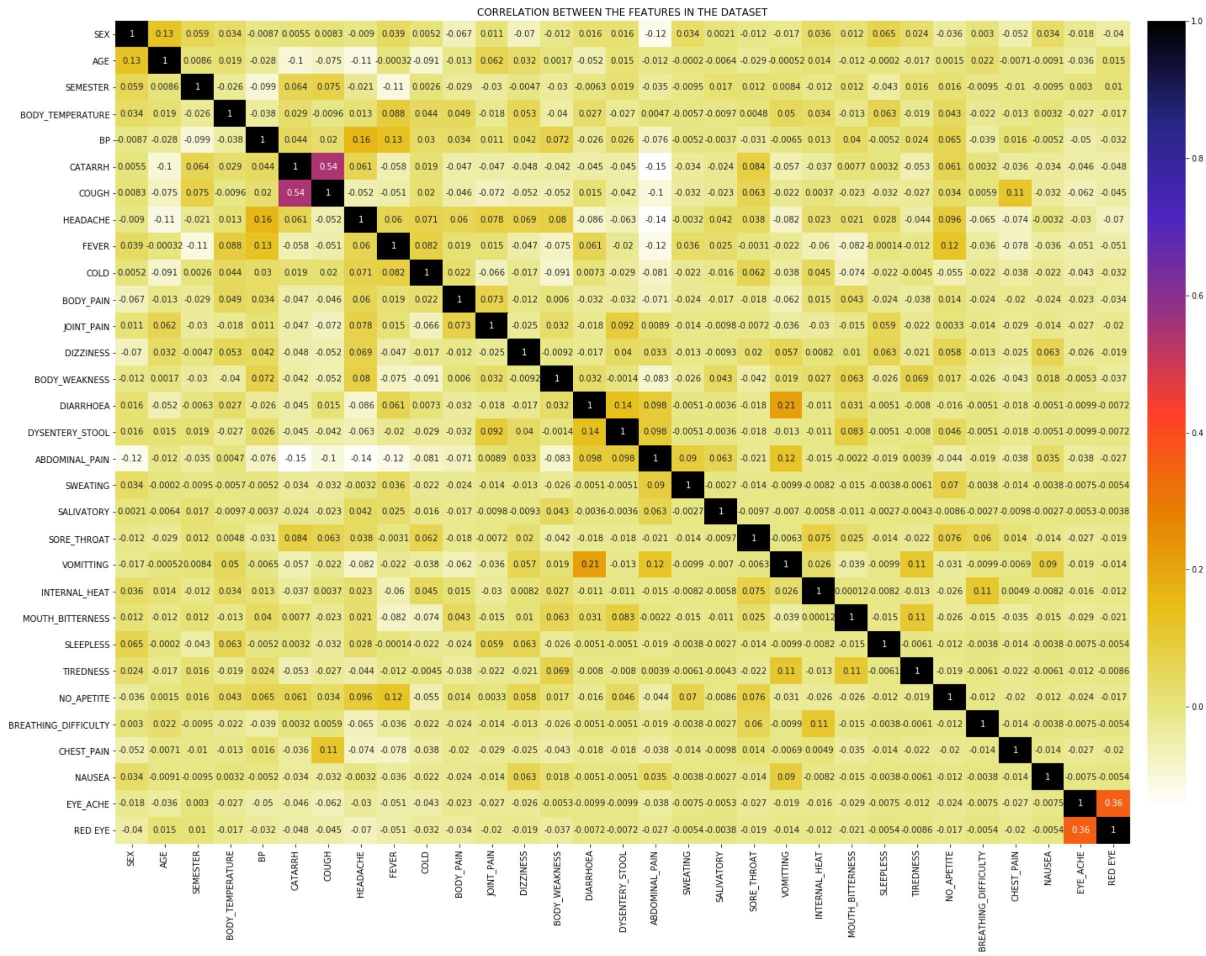

Figure 8: Feature Engineering through Correlation Matrix.

4.3. Model Training and Implementation

The models were trained such that, the models of the data for task 1 used the Random Forest Classifier, K-Neighbor Classifier, Gradient Boosting Classifier and Support Vector Classifier from the "modelselection" library of the "SKlearn" package installed into the Python.

Similarly, the regression task model was trained with the data using the Random Forest Regression, Linear Regression, Gradient Boosting Regression and Support Vector Regression from the "modelselection" library of the "SKlearn" package installed into the Python.

Table 2. The output of the Bi-model prediction of the two tasks as one output.

	FILE_NO	MALARIA	DURATION_OF_ADMISSION
0	250	YES	2.127195
1	258	YES	2.167702
2	2662	YES	2.158686
3	257	NO	0.969604
4	2165	NO	2.079713

4.4. Step 7: Model Evaluation

Confusion Matrix was used to determine the accuracy, error rate, precision, recall and F1 score of the four (4) classifiers algorithms used to build the task 1 model, as shown below in the table 2.

Below are the mathematical assessment metrics formulae used for each of the classifiers mentioned above:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}\ X\ 100 \quad (1)$$

$$\text{Error Rate} = \frac{FP+FN}{TP+FP+TN+FN}\ X\ 100 \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP}\ X\ 100 \quad (3)$$

Where;

TP = True Positive

FP = False Positive

FN = False Negative

TN = True Negative

In tabular form, the respective parameters measured and their values are shown below:

Table 2. Summary of the evaluation results on the four (4) classifier algorithms

Algorithm	Accuracy	Error Rate	Precision
Gradient Boosting	1.0	0.0	1.0
Random Forest	1.0	0.0	1.0
K-Neighbors	0.924	0.076	0.75
Support Vector Machine	0.733	0.267	0

The individual algorithms for the classifiers are listed in the table above. These algorithms' accuracy, margin of error, precision, recall, and F1-score were measured. These numbers were calculated using the produced values from each confusion matrix to complete the formulas for accuracy, margin of error, precision, recall, and F1-score. They have a range of 0 to 1 for their actual or acceptable value, which has been converted to a percent (multiples of 100) like this: 0 to 100.

Mathematically, the evaluation metrics formulas applied on all the above respective regressions' algorithms are given below:

The evaluation metrics formulae used for each of the corresponding Regression methods are provided below mathematically:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}| \tag{4}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2 \tag{5}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2} \tag{6}$$

Where,
$\hat{y}$ − *predicted value of y*

Table 4. Summary of the evaluation results on the four (4) regression algorithms

Algorithm	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE/ MAPE)
Gradient Boosting	0.305	0.180/18%
Random Forest	0.319	0.181/18.1%
Linear Regression	21100830507.337	1461474950
Support Vector Machine	0.385	0.239/23.9%

The corresponding methods for the regression model are listed in the table above, and their Mean Absolute Error/ Mean Absolute Percentage Error and Root Mean Square Error were calculated. These values were calculated using the derived values for the relevant metrics used to assess regression models. The rating with the lowest value is frequently preferable to those with greater values.

4.5. **Step 8: Deployment of the Bi-Model**

This phase was done as another round of process for software development, because the models developed needs to be deployed into a web app. So, the development process started with planning, analysis, and then the design phase where the physical and logical design of the system was produced.

The system is made up of the following modules; login, diagnosis and result modules.

i. Login Module: Users will have to login before the diagnosis and result module can be accessed.

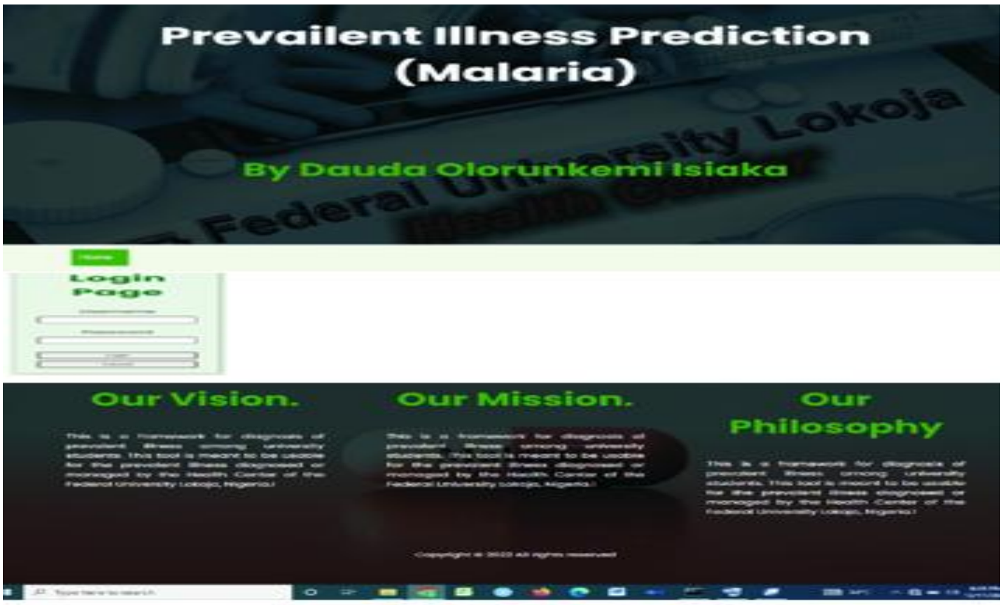

Figure 13: The Login Page

ii. Diagnosis Module: Users will have to enter all the required entries for the diagnosis of patients and submit, so as to receive the result or outcome of the diagnosis.

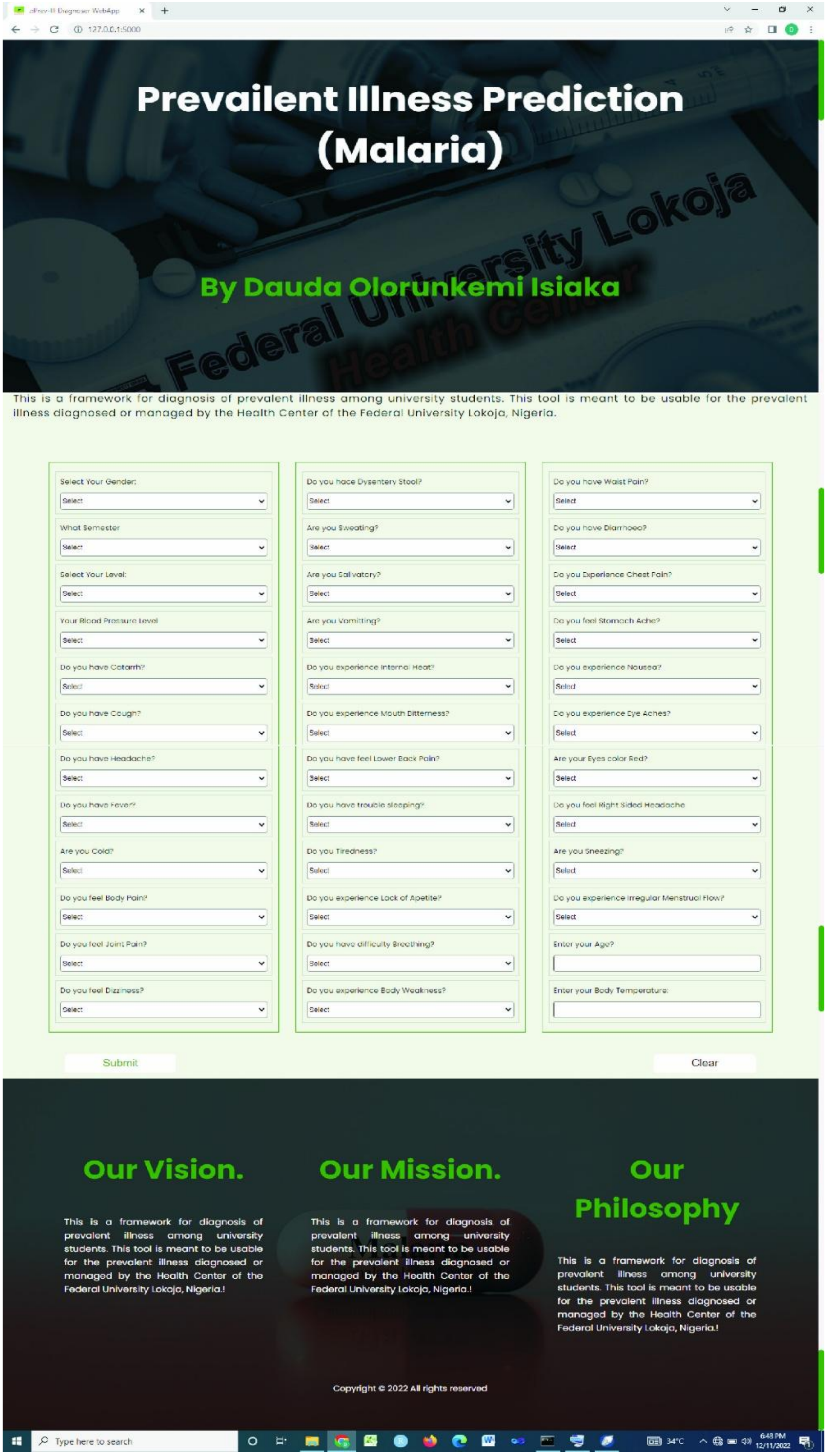

Figure 14: The Diagnosis Page

iii. Result Module: User will view the result of all successfully submitted diagnosis requests here.

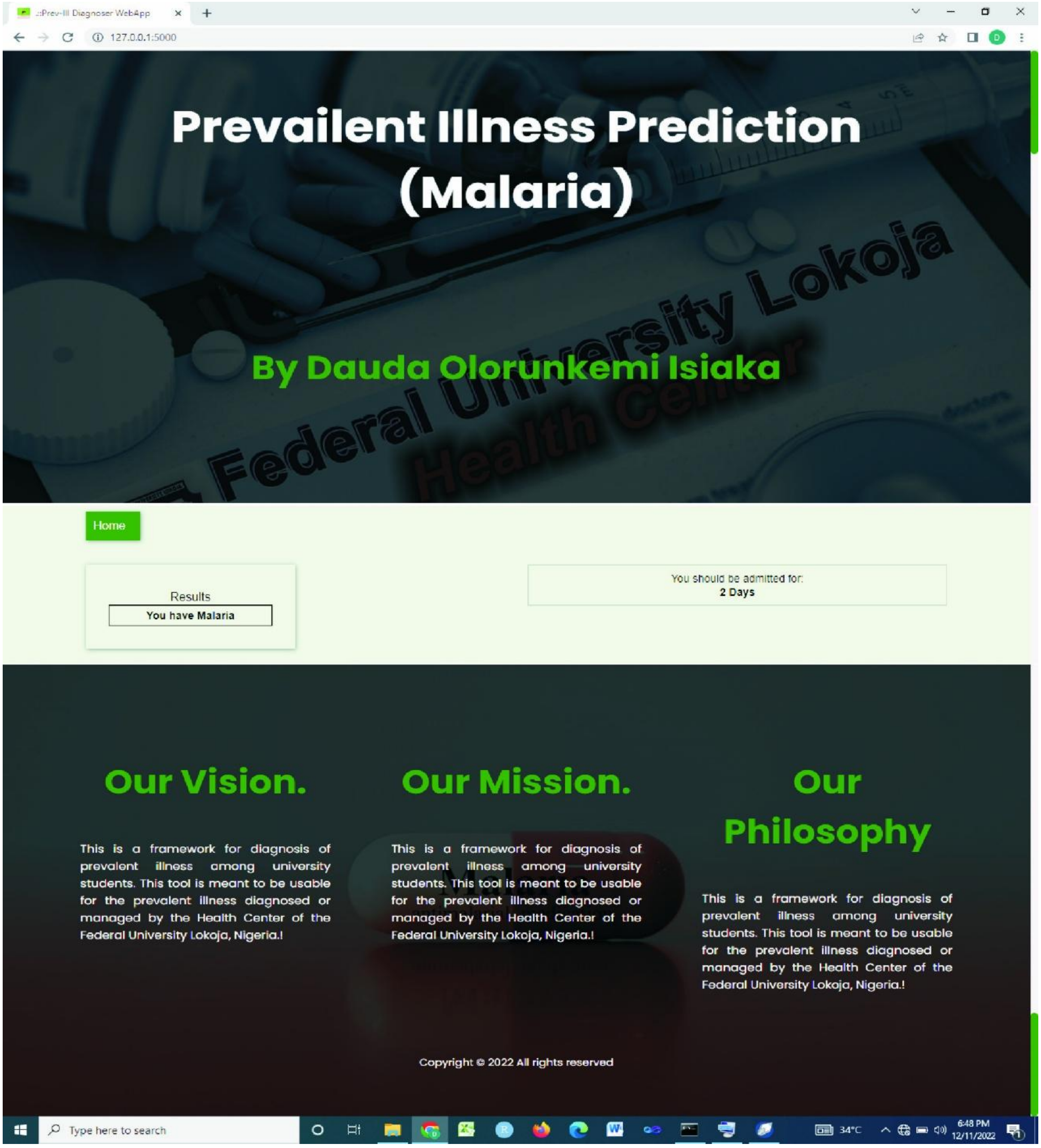

Figure 15: The Result Page

5. Discussion

Figure 3 illustrates the distribution of various illnesses diagnosed among students. Malaria was identified as the most dominant illness, significantly occurred more than other ailments. This dominance indicates the endemic nature of malaria within the student population, necessitating urgent preventive interventions. The prevalence of a single illness suggests a concentrated focus on targeted health policies, such as widespread mosquito control programs and education on anti-malarial practices. The implication of this high incidence of malaria among students leads to repeated hospital visits, affecting their academic performance and overall well-being. Institutional support through regular fumigation, distribution of treated nets, and environmental cleanliness is essential to reduce its occurrence.

In Figure 4, the bar graph demonstrates nearly equal diagnosis rates for male and female students. This balance signifies that both genders are equally vulnerable to health challenges. However, specific interventions should consider gender-sensitive issues, such as differential living conditions or stress levels that may indirectly affect health. As a matter of fact, this implies that gender-inclusive health initiatives are necessary to address vulnerabilities comprehensively. Such programs should equally promote awareness about preventive measures and encourage regular health checks for both male and female students.

Figure 5 shows that follow-up visits for re-hospitalization far outnumber initial visits. The disparity highlights challenges in complete recovery or adherence to prescribed treatments during the first visit. This trend is particularly significant for illnesses like malaria, where incomplete treatment can lead to recurrence. As such, re-hospitalization will strain healthcare resources and disrupts academic activities. Addressing this issue requires a dual approach: improving treatment protocols using user-friendly diagnosing apps and educating students on the importance of adhering to medical advice. Regular follow-up systems and comprehensive post-treatment care could reduce re-hospitalization rates.

Figure 6 show the trend in which age was identified as a critical factor influencing diagnosis results. Students aged 15–20 are disproportionately affected, aligning with their transition to independent living and associated stressors. Male and female students within this age range show comparable vulnerabilities, reinforcing the need for age-targeted interventions. In addition, Figure 7 further emphasized the relationship between age and health outcomes. Younger students face higher risks of illnesses due to lifestyle adjustments, academic pressures, and environmental factors. The trend diminishes as age increases, likely reflecting improved adaptation and maturity. So, younger students require support systems to navigate the challenges of independence. Orientation programs focused on stress management, self-care, and healthy lifestyles can significantly improve health outcomes in this demographic can mitigate these risks to prevent long-term health issues.

The model built was from GBC and GBR algorithms which allow the creation of an ensemble machine learning model using basic models and achieve excellent results that are comparable to those of models that consume a lot of resources, such neural networks.

Finally, the study emphasizes the utility of machine learning in diagnosing common illnesses among university students. By utilizing a localized dataset, the research integrates two tasks—classification of the dominant illness and prediction of admission duration—into a single application. This Bi-Model approach combines Gradient Boosting techniques for optimal performance in both classification and regression tasks. The deployment of the application into a web-based interface ensures ease of use, enabling real-time and accurate diagnosis, and streamlining healthcare processes in university health centers.

5.1 Final Thoughts and Future Work

This research offered a Bi-Model approach that has combined the Gradient Boosting techniques for optimal performance in both classification and regression tasks, which was deployed into a web-based interface. It is an extension of the study by (Isiaka, et al., 2022), but was made realistic by the deployment to provide a stronger conclusion.

Also, the application has been enhanced with the combination of two tasks; classification and prediction module controlled by a single button, resulting in an easy to use application to boost it acceptability.

The following are the paper's main contributions:

i. Deployment of the models into a web application using Flask, enhancing accessibility and user interaction.

ii. Introduction of a single prediction button interface, simplifying the diagnostic process and reducing the need for extensive technical expertise.

Further research must concentrate on including a Blockchain as a means of decentralized storage and security. Real-time health monitoring through wearable devices that can provide continuous data input to enable dynamic diagnostics through the integration of Internet of Things. So, also, more effective method of using cutting-edge machine learning algorithms to categorize the types of malaria that are dominant among the students of a college or university, because, if the prediction is correct, there would be no need to further identify the type of malaria for more personalized treatment.

Acknowledgment

We appreciate management of Federal University Lokoja for providing the dataset used in this work. We also appreciate the unrelenting efforts of the Software Laboratory Unit of the Department of Computer Science, Federal University Lokoja for providing the enabling platforms and tools for the development, analysis and testing of the application.

References

[1] Liu N. & Kauffman R. J., (2020). Enhancing Healthcare Professional and Caregiving Staff Informedness with Data Analytics for Chronic Disease Management, *Information and amp; Management.* doi: https://doi.org/10.1016/j.im.2020.103315.

[2] Williams, E., Gartner, D. & Harper, P., (2021). A survey of OR/MS models on care planning for frail and elderly patients. *Operations Research for Health Care*, 31, 100325, ISSN 2211-6923, https://doi.org/10.1016/j.orhc.2021.100325.

[3] Adeyemo, F.O. & Olaogun, A. A. (2013). Factors affecting the use of Nursing Process in Health Institutions in Ogbomoso Town, Oyo State. *International Journal of Medicine and Pharmaceutical Sciences,* 89-96.

[4] Ravikumaran, P., Vimala Devi, K., Kartheeban, K., & Narayanan Prasanth, N. (2020). Health Data Analytics: Framework & Review on Tool & Technology. *Materials Today: Proceedings*. doi:10.1016/j.matpr.2020.10.131

[5] Smiti, A. (2020). When machine learning meets medical world: Current status and future challenges. *Computer Science Review*, 37, 100280. doi:10.1016/j.cosrev.2020.100280

[6] Pilz, G. F., Weber, F., Mueller, W. G., & Schaefer, J. R. (2021). Statistical Methods to Support Difficult Diagnoses. *Diagnostics (Basel, Switzerland), 11*(7), 1300. https://doi.org/10.3390/diagnostics11071300

[7] Isiaka, D. O., Agbogun, J. B. ., & Kolajo, T. (2022). A Framework for Predictive - Diagnosis of Prevalent Illness among University Students. *Journal of Applied Artificial Intelligence, 3*(2). https://sabapub.com/index.php/jaai/article/view/667

[8] Isiaka, D. O., Agbogun, J. B. ., & Kolajo, T. (2023). Data Analytics Framework for the Diagnosis of Prevalent Illness among University Students. *FUW Trends in Science & Technology Journal, 8*(1), 107 – 114. https://www.ftstjournal.com

[9] Kriegova, E., Kudelka, M., Radvansky, M. & Gallo, J. (2021). A Theoretical Model of Health Management using Data-Driven Decision-Making: The Future of Precision Medicine and Health. *J Transl Med, 19*, 68. https://doi.org/10.1186/s12967-021-02714-8

[10] Molina, A., Alferez, S., Boldu, L., Acevedo, A., Rodellar, J. & Merino, A. (2020). Sequential Classification System for Recognition of Malaria Infection using Peripheral Blood Cell Images, *J Clin Pathol, 73*(10), 665-670, https://doi.org/10.1136/jclinpath-2019-206419.

[11] Emmert-Streib, F. & Dehmer, M. (2019). Evaluation of Regression Models: Model Assessment, Model Selection and Generalization Error. *Mach. Learn. Knowl. Extr., 1*, 521–551.

[12] Sow, B., Mukhtar, H., Ahmad, H. F. & Suguri, H. (2020). Assessing the Relative Importance of Social Determinants of Health in Malaria and Anaemia Classification based on Machine Learning Techniques. *Informatics for Health & Social Care, 45*(3), 229–241

[13] Mbaye, O., Ba, M. L. & Sy, A. (2021). On the efficiency of machine learning models in malaria prediction. *Studies in health technology and informatics,* 281, 437–441. https://doi.org/10.3233/SHTI210196

[14] Mujahid, O., Contreras, I. & Vehi, J. (2021). Machine Learning Techniques for Hypoglycemia Prediction: Trends and Challenges. *Sensors MDPI AG, 21*(2), 546. http://dx.doi.org/10.3390/s21020546