Development of a Knowledge-based Patient Self-assessment and Diagnostic System for Malaria, Typhoid, and Related Diseases using Knowledge Discovery Database Techniques

Samuel Isah Odoh
Faculty of Computing
Federal University, Lafia

Ngbede Barnabas Michael
Faculty of Computing
Federal University, Lafia

Peter Samuel Oche
Department of Comp Sc.
Benue State University

ABSTRACT
Many regions continue to face severe challenges in providing prompt and correct diagnosis of diseases like typhoid, malaria, and associated ailments, especially in underprivileged populations with inadequate medical infrastructure. Conventional diagnostic techniques that depend on laboratory testing and professional advice are frequently unavailable, which causes delays in the start of therapy and unfavorable health consequences. This study suggests creating a knowledge-based patient self-assessment and diagnostic system for typhoid, malaria, and related illnesses as a solution to these problems. The system intends to enable people to evaluate their symptoms, obtain preliminary diagnoses, and obtain prompt recommendations for additional action without the need for direct medical assistance by utilizing database approaches and medical knowledge representation. Key issues addressed include improving healthcare accessibility, ensuring diagnostic accuracy based on patient-reported symptoms and medical guidelines, enhancing user-friendliness and acceptance, implementing robust data security and privacy measures, and integrating seamlessly with existing healthcare infrastructure. Through the development of this knowledge-based system, this study seeks to revolutionize healthcare delivery by providing accessible, accurate, and patient-centric diagnostic solutions for malaria, typhoid, and related diseases. Lastly, knowledge discovery database (KDD) technique is implemented using the Hypertext Pre-Processor (PHP) programming language to build the user application interface for the users of the diagnostic system. By addressing the unique challenges faced by underserved communities, the system has the potential to significantly improve health outcomes and reduce the burden of infectious diseases in these regions.

General Terms
Knowledge-Based Systems, Health Informatics, Diagnostic Systems, Data Mining, Machine Learning, Artificial Intelligence, Medical Decision Support.

Keywords
Malaria, Typhoid Fever, Knowledge Discovery, Database Techniques, Patient Self-assessment, Diagnostic System, Health Technology, Disease Diagnosis, Expert Systems, Clinical Decision Support.

1. INTRODUCTION
Medical diagnosis, (often simply termed diagnosis) refers both to the process of attempting to determine or identifying a possible disease or disorder to the opinion reached by this process. A diagnosis in the sense of diagnostic procedure can be regarded as an attempt at classifying an individual's health condition into separate and distinct categories that allow medical decisions about treatment and prognosis to be made Subsequently, a diagnostic opinion is often described in terms of a disease or other conditions [1].

Clinical decision support systems (CDSS) are interactive computer programs designed to assist healthcare professionals such as physicians, physical therapists, optometrists, healthcare scientists, dentists, pediatrists, nurse practitioners or physical assistants with decision making skills. The clinician interacts with the software utilizing both the clinician's knowledge and the software to make a better analysis of the patient's data than neither humans nor software could make on their own. Typically, the system makes suggestions for the clinician to look through and the he picks useful information and removes erroneous suggestions [2].

Computer-Aided Diagnosis (CAD), are diagnosis made by physicians who uses the output from a computerized analysis of medical data as a second opinion in detecting lesions, assessing disease severity, and making diagnostic decisions, is expected to enhance the diagnostic capabilities of physicians and reduce the time required for accurate diagnosis. With CAD, the final diagnosis is made by the physician [3].

A knowledge is an information that can be used in decision making process. A knowledge base uses knowledge representation formalism to capture the subject matter expert's knowledge and codifying it according to the formalism which is called the knowledge engineering. A knowledge-based system uses a knowledge base containing accumulated experience and a set of rules for applying the knowledge to a particular situation. Knowledge based system/expert system have been developed in the last decade for many different applications by adopting artificial intelligence techniques. The knowledge-based system in medical diagnosis was started in 70's and its usage actually started in 80's. In the last decade it is observed that the use of Knowledge Based System is embedded with other technologies [4].

Despite significant advancements in healthcare, access to timely and accurate diagnosis remains a challenge in many regions, particularly in areas with limited medical infrastructure and resources. Diseases such as malaria, typhoid, and related illnesses continue to impose a substantial burden on public health, often resulting in delayed treatment, complications, and even fatalities. Traditional diagnostic methods, reliant on laboratory tests and expert consultation, are often unavailable or inaccessible in underserved communities. In light of these challenges, there is a pressing need to develop a knowledge-based patient self-assessment and diagnostic system for malaria, typhoid, and related diseases. Such a system would empower individuals to assess their symptoms, receive provisional diagnoses, and access timely recommendations for further action, all without the need for direct medical intervention. Thus, it is on this backdrop that the researchers seek to contribute to the advancement of accessible, accurate, and patient-centric healthcare solutions for malaria, typhoid, and related diseases.

2. LITERATURE REVIEW

[5], observed a prevalent issue with existing medicine query systems—they heavily rely on keyword searches, which often prove ineffective when users are uncertain about specific search terms. Moreover, these systems frequently lack comprehensive explanations of symptoms and ailments, leaving users unsure and dissatisfied. To address these shortcomings, Sung and Chi proposed a novel approach: developing a knowledge base capable of facilitating inference-based searches and integrating data from diverse web sources. Their solution involved constructing an ontology model to delineate relationships between ailments and symptoms, enhancing medicinal product datasets by linking them with the ontology model at a semantic level and creating a data mash-up to amalgamate web resources for users to find relevant references. However, in their approach they experimented on skin ailments as a pilot experiment, the scope of the knowledge base was limited to common skin ailments, relevant symptoms, and OTC medicinal products. Therefore, we propose a tailored solution: a Knowledge-Based Patient Self-assessment and Diagnostic System for Malaria, Typhoid, and Related Diseases Using Database Techniques. This targeted approach aims to empower patients with an effective tool for self-diagnosing and treating Malaria, Typhoid, and Related Diseases.

In their study,[6], employed a knowledge-based system (KBS), a sophisticated framework designed to integrate and leverage specialized knowledge from diverse sources. This system aids in complex decision-making and problem-solving, proving particularly beneficial in enhancing and streamlining organizational processes and operations. These systems are applied across various sectors, including healthcare diagnostics, financial investment analysis, counselling, and production management.

In her scholarly investigation, [7], undertook an experimental inquiry into the design and construction of a web platform dedicated to diabetes diagnosis, employing the knowledge engineering approach. Through systematic filtration, the knowledge embedded within the system was distilled and subsequently rendered into a structured format conducive to effective interpretation, thereby facilitating the presentation of relevant diagnostic outcomes. Her method involves a comprehensive process, commencing with the acquisition of data through direct interviews with experts, followed by meticulous storage of this data in a database to ensure the accuracy of the conclusions drawn. The application operates through three essential interface stages: member registration, self-diagnosis facilitated by answering relevant questions, and the presentation of conclusive results. Building upon this premise, our research endeavours to operationalize a Knowledge-Based Patient Self-assessment and Diagnostic System tailored to the identification and analysis of Malaria, Typhoid, and Associated Disorders, employing advanced database methodologies.

[8], designed and implemented a user-friendly GUI knowledge-based system aimed at assisting both physicians and patients in the diagnosis and treatment of pneumonia. The acquisition of knowledge for this system involves structured and unstructured interviews conducted with domain experts selected through purposive sampling from Addis Ababa Bethel Referral Hospital. Furthermore, relevant document analysis techniques are employed to capture explicit knowledge. Subsequently, the acquired knowledge is structured using a decision tree to depict the concepts and procedures pertinent to pneumonia diagnosis and treatment, with domain knowledge represented using production rules. Additionally, knowledge about gastritis treatment is extracted from both domain experts and document analysis. We are streamlining our approach by utilizing a similar technique to the Development of a Knowledge-Based Patient Self-assessment and Diagnostic System for Malaria, Typhoid, and Related Diseases Using Database Techniques. Consequently, a prototype of the knowledge-based system is developed, facilitating diagnosis and treatment for gastritis patients.

In an attempt to investigate the opaque functionality of self-diagnosis systems and the prevailing lack of transparency regarding datasets, medical contributions, and symptom assessment algorithms, [9], devised a self-diagnosis system as a proof-of-concept for an open-data platform. Utilizing a symptom-disease knowledge database as their dataset, the system was deployed within a cloud-native environment to facilitate validation by medical professionals and users alike. The concept of establishing a publicly accessible and community-driven medical dataset garnered validation from 93.1% (27 out of 29) of interviewed doctors. Among these, the proof-of-concept received a rating of "correct" from 58.6% (17 specialists) of the surveyed doctors and was deemed "satisfactory" by 69.7% (23 out of 33) of users interviewed. Notably, the application of this proof-of-concept was not explicitly specified to a particular illness or disease domain in their study.

3. METHODOLOGY

The developed system utilizes the Structured System Analysis and Design Methodology (SSADM) in conjunction with the Knowledge Discovery in Databases (KDD) technique. SSADM offers a structured approach to systems analysis and design, essential for developing a reliable and robust health informatics system, while KDD is employed to extract valuable information from large datasets.

The SSADM framework was chosen for its methodical approach, which encompasses several stages. Initially, a feasibility study was conducted to determine the viability of the proposed system. This involved evaluating the technical, operational, and economic feasibility and producing a feasibility report that summarized the findings and recommendations. Following this, a detailed requirements analysis was carried out. This included conducting interviews and focus groups with healthcare professionals and potential users, and developing detailed use cases and user stories to capture both functional and non-functional requirements.

The next phase, system design, focused on creating the system architecture. Sequence, activity and use case diagrams were developed to model information flow within the system. User interface prototypes were designed and iteratively refined based on feedback. Detailed specifications for hardware and software requirements were also prepared, culminating in a thorough system design document. During the implementation phase, the system was built according to the design specifications. This involved setting up the development environment, developing the system modules, conducting unit testing, and integrating the modules to form a complete system.

System testing was then performed to verify and validate the integrated system against the requirements. User acceptance testing (UAT) was conducted with healthcare professionals to ensure usability, and any identified issues were resolved. The results and feedback from these tests were documented comprehensively.

In parallel, the Knowledge Discovery in Databases (KDD) technique was applied to analyze and extract valuable information from the collected data. The KDD process began with data selection, identifying and gathering relevant data from clinical records, patient surveys, and health databases to include diverse cases of malaria, typhoid, and related diseases. The data was then preprocessed to clean and transform it into a consistent format suitable for analysis, followed by normalization and aggregation where necessary. During the data mining stage, clustering algorithms were applied to group similar symptoms and disease cases, classification techniques were used to build predictive models for disease diagnosis, and association rule mining was employed to discover relationships between symptoms and diseases.

Data collection for this study involved both primary and secondary sources. Primary data was gathered through clinical visits, where observations and documentation of symptoms and treatments were made, and interviews with healthcare professionals provided deeper insights into diagnostic processes and treatment protocols. Secondary data collection included an extensive literature review, analyzing research papers, journals, and articles related to malaria, typhoid, and related diseases from online databases such as PubMed and Google Scholar. Additionally, internet research involved searching for relevant case studies, clinical trials, and epidemiological reports from reputable sources like the World Health Organization (WHO) and Centers for Disease Control and Prevention (CDC). By integrating SSADM and KDD methodologies and utilizing comprehensive data collection methods, the study ensured a detailed and structured approach to developing the knowledge-based patient self-assessment and diagnostic system, thereby enhancing its reliability, usability, and effectiveness.

3.1 Analysis of the Proposed System

The primary objective of the developed system is to assist in the diagnosis of patient diseases, particularly in scenarios where a medical expert is unavailable. This system aims to complement the manual diagnostic processes by providing a reliable, automated alternative. Leveraging the records of the patient's past diagnoses and major symptoms specific to each user, the system is designed to maintain a high degree of accuracy. By retaining and analyzing historical patient data, it can identify patterns and correlations that are crucial for accurate diagnostics. This capability is particularly beneficial in regions with limited access to medical professionals or in emergency situations where immediate diagnostic support is required. Thus, the proposed system enhances the efficiency and effectiveness of disease diagnosis, contributing significantly to patient care and health outcomes.

3.2 System Design

The system is made up of several forms with their various countless procedures and modules, one user control and a user document. Each of which has name relating to what it does and significantly, they are easily filled.

Most of them contain command buttons that will enable the user either to submit or to reset the form for new entry. Each of which was designed to its level of importance and what it can do. A patient cannot register as a doctor because he/she is not given the right in the proposed system. The holder of the key to the master file is the administrative body holding the affairs of which they as well are given their numbers by the mouth diagnostic system administrator to enable them access the database without fear.

All records manipulation must only be done by the admin of the system. Patients and doctors only have limited access to update some of their records and if they wish to change some records they need to mail the admin to help them make those changes. And there is also a platform where customer can actually message the staff or admin in case of any complaint.

3.3 High-level Model of the Proposed System

A Sequence diagram is a graphical visualization of sequences of messages between objects i.e. sequence of method invocation of objects which results in accomplishing some tasks. The emphasis in a sequence diagram is on the sequence of messages. A Sequence diagram is a structured representation of behavior as a series of sequential steps over time. It is used to depict work flow, message passing and how elements in general cooperate over time to achieve a result. The sequence diagram for this system is shown in figure 1 below.

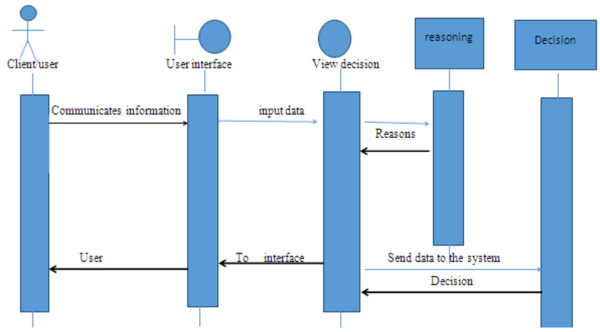

Figure 1: Sequence diagram of the knowledge base setting

3.4 Database Design

Apart from the knowledge base that is a major characteristic of any expert system for storing rules and facts about the object, the proposed system makes use of a conventional database table. This table serve for the of storing information about patients whose health status has been diagnosed and the test result thereof, this can be useful in further researches, for example one may decide to query out from the database the number of patients with mouth diseases to assess the rate of the pandemic, or to use the previous results of patients as Case base reasoning for further diagnosis or treatment.

Table 1. Patient Database

Variable name	Field length	Field type
Patient's names	50	Varchar
Sex	3	Varchar
Username	20	Varchar
Password	20	Varchar
Date of birth	10	Date
Symptoms	250	Varchar

Table 2. Symptoms

Variable name	**Field length**	**Field type**
Symptom_id	4	integer
Type	50	Varchar
Description	250	Varchar

Table 3. Admin Information's

Variable name	**Field length**	**Field type**
Admin's names	50	Varchar
Admin's address	50	Varchar
Sex	3	Varchar
Phone number	15	Varchar
Username	20	Varchar
Password	20	Varchar

3.5 Use Case Diagram
Figure 2 below is a use case diagram for all the activities to be carried out by the user in the proposed system.

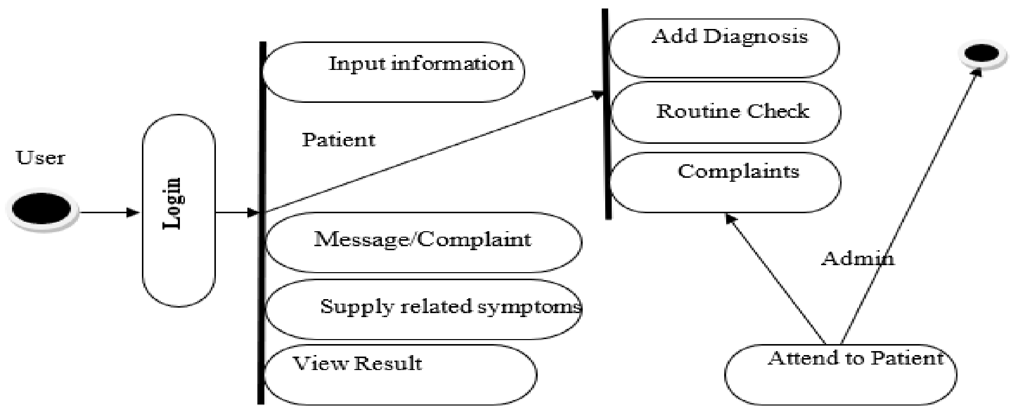

Figure 2: Use case diagram

3.6 Activity Diagram
Figure 3 below is a use activity diagram for all the activities to be carried out by the user in the proposed system.

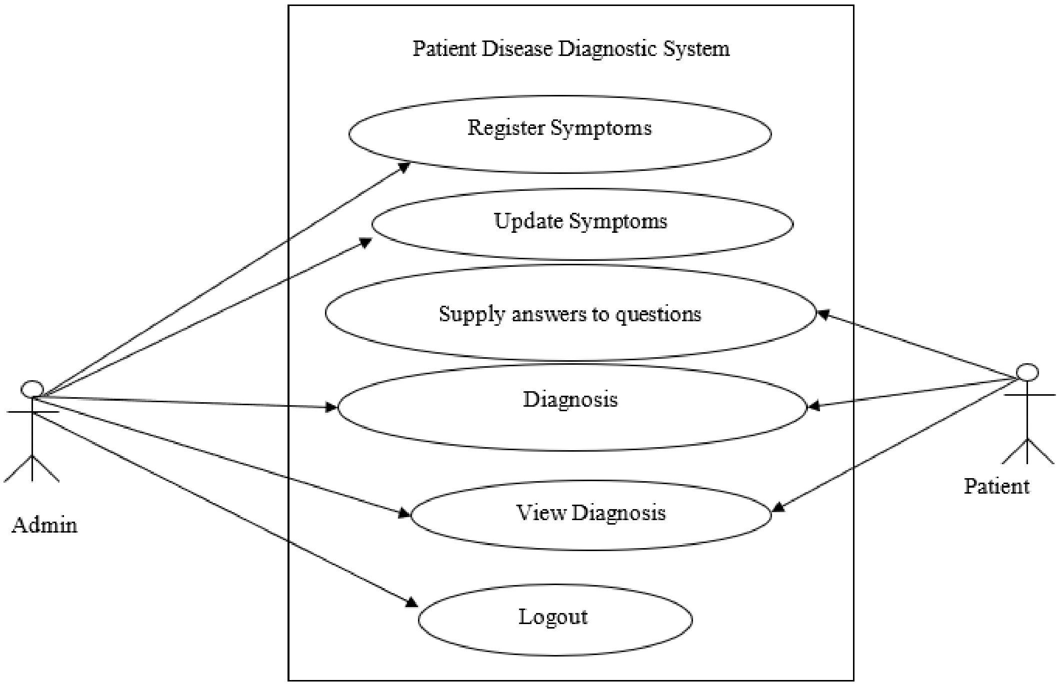

Figure 3: Activity diagram

4. RESULTS

The summarized result of the system is shown on figure 4-13.

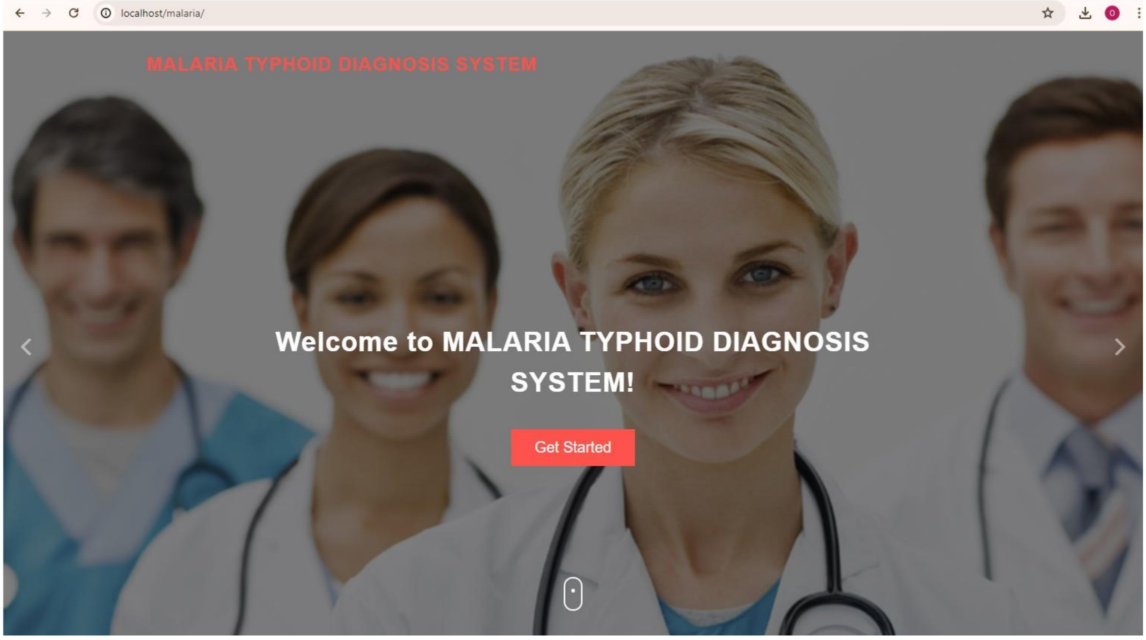

Figure 4: System Homepage

Figure 4 above represents the system homepage which is the first page users of the diagnostic system will see before proceeding to carry out their designated task or activities.

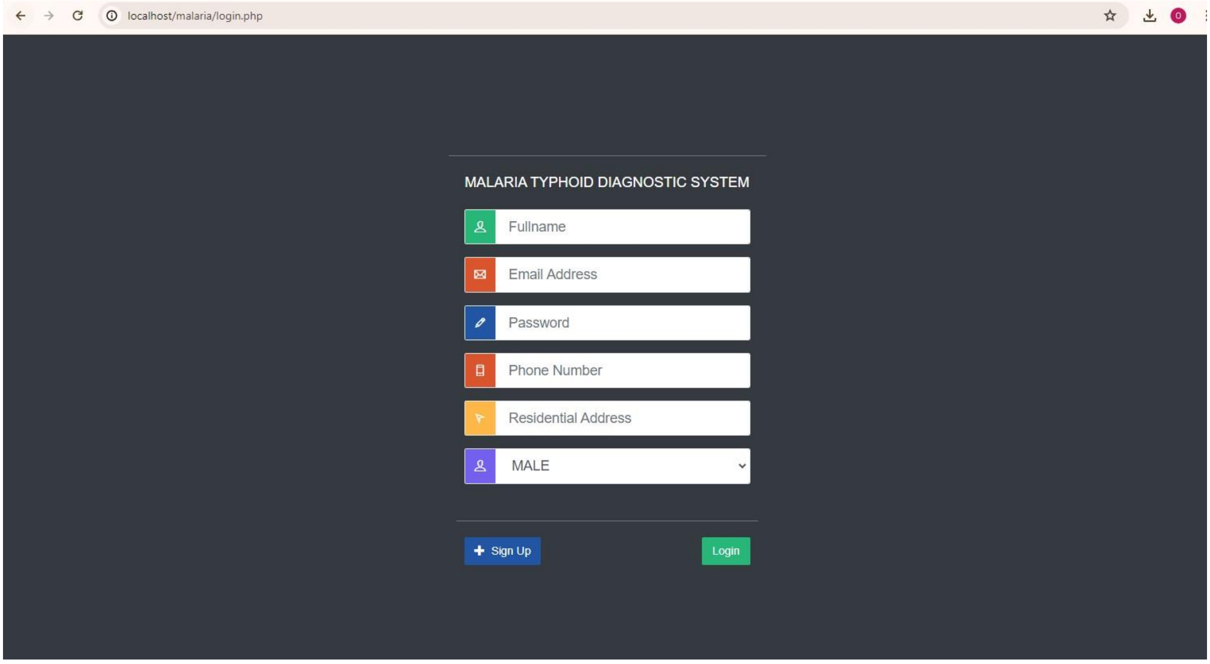

Figure 5: User Registration

Users of the Diagnostic system have to create an account on the system by supplying relevant information as shown on figure 5.

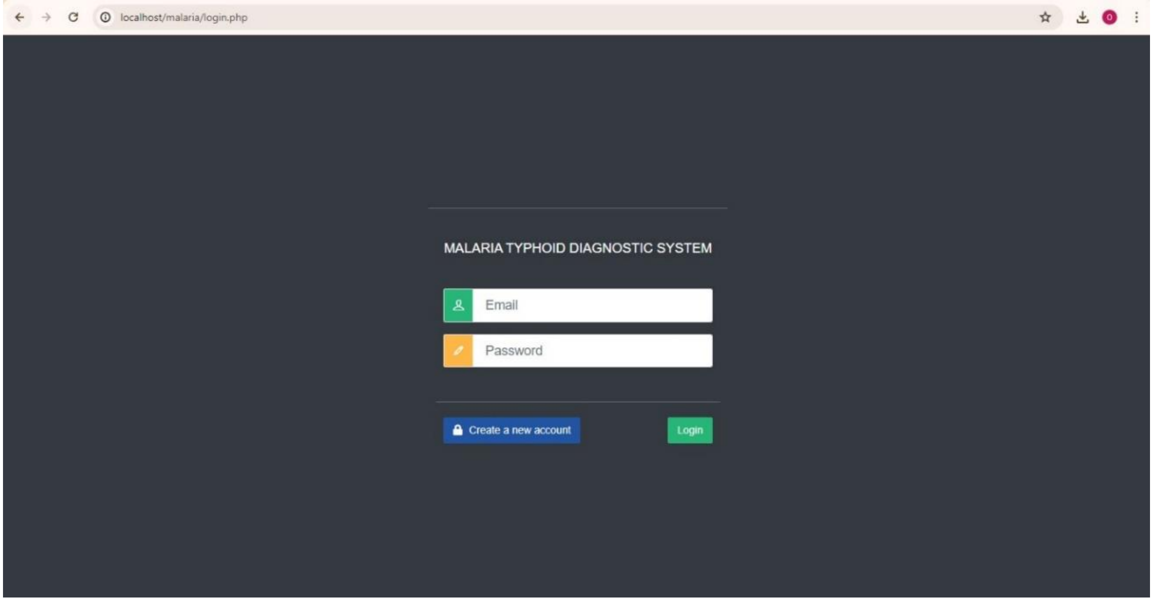

Figure 6: User Login

Figure 6 represents the user login page. Users of the system are required to supply their email address and password as submitted during creation of account on the system in order to have access to the user dashboard.

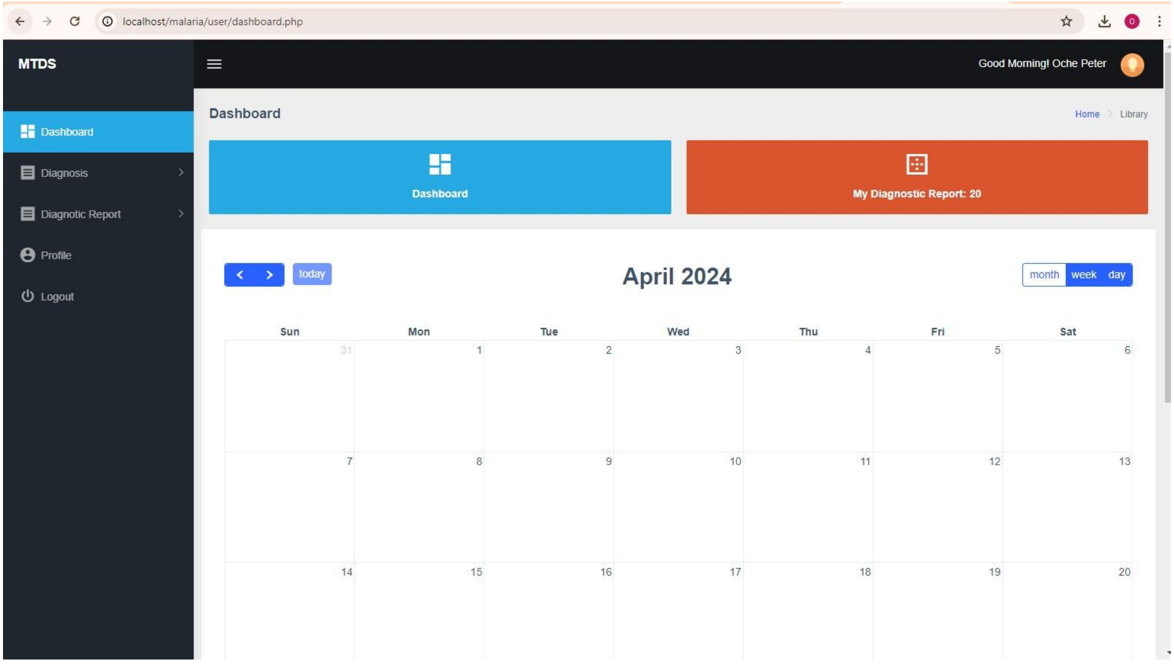

Figure 7: User Dashboard

Figure 7 represents user dashboard, this is where after successfully submitting the login credential the user is redirected to in order to perform all the designated functions such as diagnosis, viewing of diagnostic report, updating profiles etc.

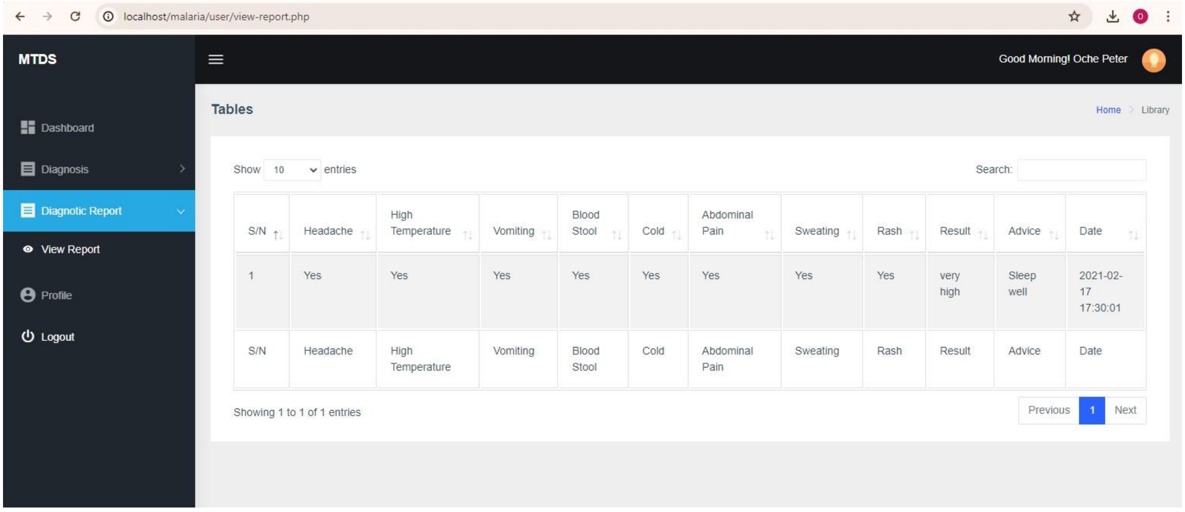

Figure 8: Diagnostic Answers

This table represents all the answers supplied by the patient during diagnosis. This is where an inferential deduction is made for diagnosis using the knowledge discovery database (KDD) technique.

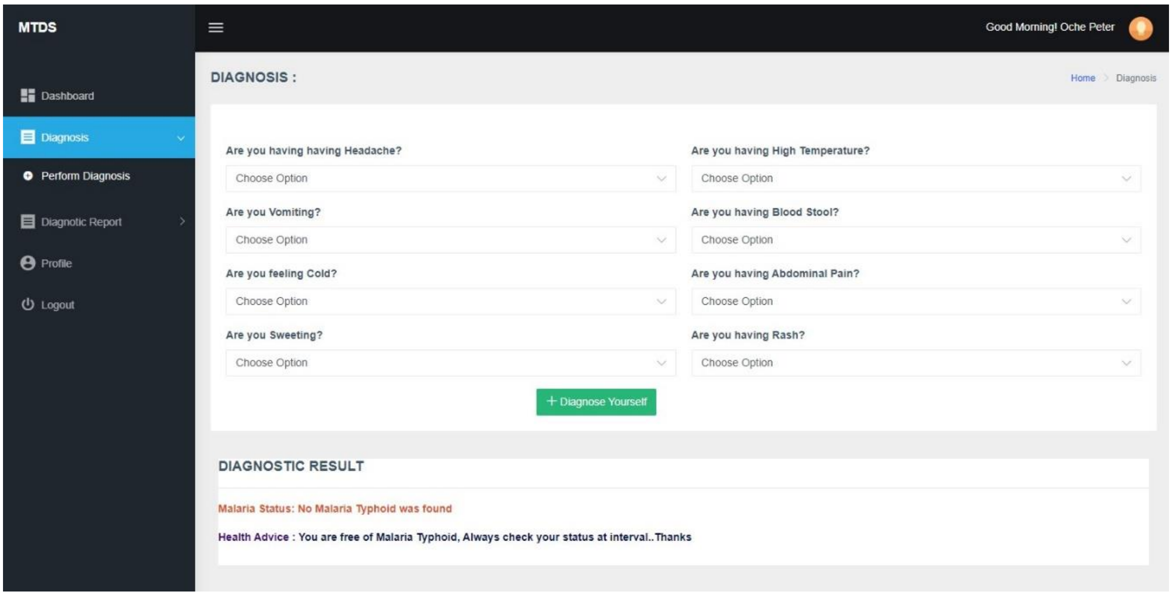

Figure 9: Diagnosis symptoms

Figure 9 represent the diagnostic tool (symptoms) which are list are arrays of relevant questions where patients select from based on their symptoms.

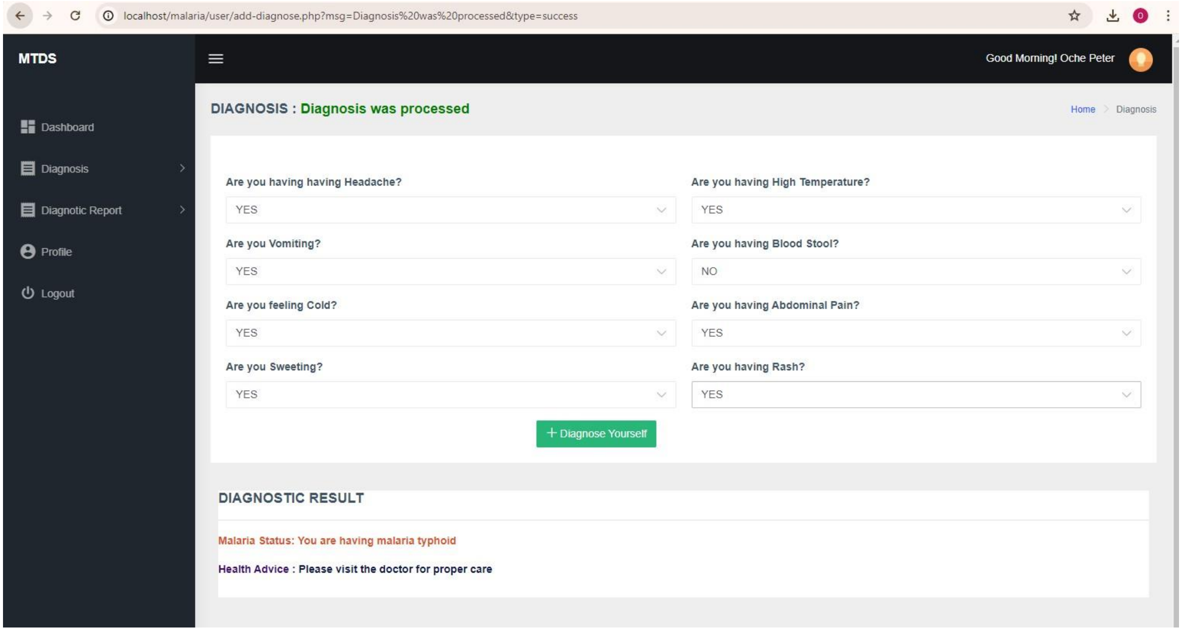

Figure 10: Processed Diagnosis

Figure 10 represent the Processed Diagnosis which is the inference drawn after a user provide relevant answers based on the questions and the related symptoms they feel.

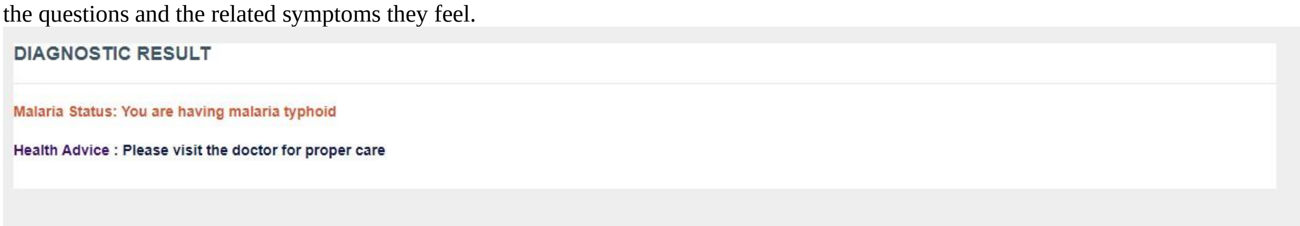

Figure 11: Diagnostic Result

Figure 11 represent the diagnostic result, which is the output of the inference drawn after a user provide relevant answers based on the questions and the related symptoms they feel.

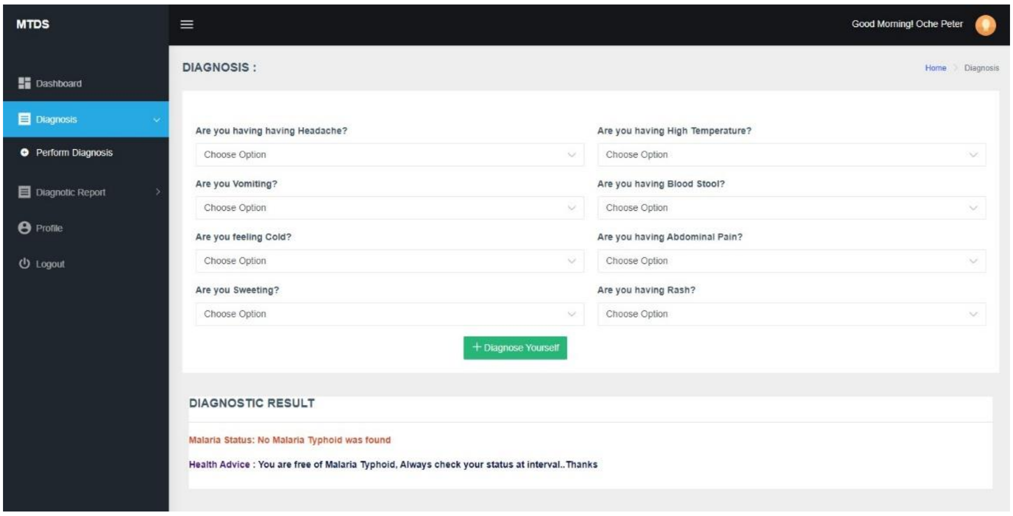

Figure 12: Processed Diagnosis

Figure 12 represent another Processed Diagnosis which is the inference drawn after a user provide relevant answers based on the questions and the related symptoms they feel.

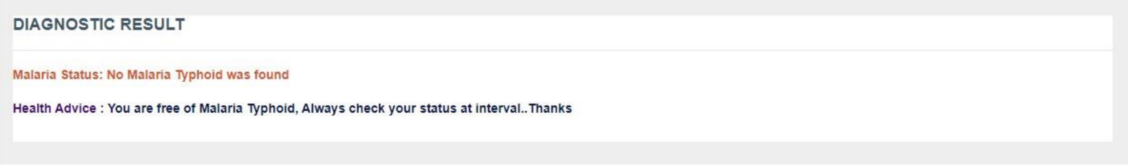

Figure 13: Diagnostic Result

Figure 13 represent the diagnostic result, which is the output of the inference drawn after a user provide relevant answers based on the questions and the related symptoms they feel.

4.1. Discussion

The testing of the application was duly carried out and the test result obtained where users are registered. The developed system demonstrates its functionalities through various user interfaces and diagnostic processes. The key components of the system include the System Homepage, User Registration, User Login, User Dashboard, Diagnosis Symptoms, Diagnostic Answers, Processed Diagnosis, and Diagnostic Result. Each of these components plays a critical role in the user experience and overall effectiveness of the system.

The System Homepage serves as the entry point for users, providing an intuitive and user-friendly interface. It offers easy navigation to various features, including user registration and login. The design ensures that users can quickly access the necessary functions without confusion, making the system accessible to a wide range of users.

The User Registration process is designed to be straightforward and secure. New users are required to provide essential information to create an account. This step ensures that the system can maintain individual patient records, which are crucial for accurate and personalized diagnostics. The registration process also sets the foundation for secure data management, protecting user information from unauthorized access.

The User Login functionality allows registered users to securely access their accounts. This feature ensures that patient data is protected and only accessible by authorized individuals. The login process also facilitates the retrieval of past diagnostic records, contributing to the system's ability to provide accurate and consistent diagnostic results. By maintaining a history of user interactions and diagnoses, the system can offer more informed and reliable assessments.

Once logged in, users are directed to their personalized User Dashboard. This dashboard provides an overview of their account, including past diagnoses and any pending actions. It serves as a central hub where users can manage their health information and initiate new diagnostic sessions. The dashboard's design prioritizes ease of use, ensuring that users can effortlessly navigate through their health data and access the system's diagnostic tools.

The Diagnosis Symptoms component allows users to input their symptoms. This step is crucial as it forms the basis of the diagnostic process. The system provides a comprehensive list of symptoms related to malaria, typhoid, and related diseases, ensuring that users can accurately describe their condition. By capturing detailed symptom information, the system enhances its diagnostic precision and effectiveness.

After inputting symptoms, users proceed to the Diagnostic Answers section. Here, the system may ask additional questions to gather more detailed information. This interactive process ensures that the system has all necessary data to make an accurate diagnosis. By engaging users in a dialogue, the system can clarify ambiguous symptoms and gather context-specific details that are vital for accurate disease identification.

The Processed Diagnosis component is where the system analyzes the input data. Using the Knowledge Discovery in Databases (KDD) technique, the system processes the symptoms and any additional information provided. This step involves applying data mining algorithms to identify patterns and correlations that are indicative of specific diseases. The use of advanced data analysis techniques ensures that the system can discern subtle patterns in the data, leading to more accurate diagnostic outcomes.

Finally, the Diagnostic Result is presented to the user. The system provides a detailed report of the diagnosis, including the identified disease and recommended next steps. This result is based on the comprehensive analysis conducted during the processed diagnosis stage and leverages the historical data stored in the system to enhance accuracy. By delivering clear and actionable diagnostic reports, the system empowers users to take informed steps towards managing their health. Hence, the results demonstrate that the system effectively supports users in diagnosing diseases such as malaria and typhoid. Each component works together to provide a seamless and accurate diagnostic process. The system's ability to maintain and utilize patient history, combined with its interactive and user-friendly interface, ensures that it can serve as a valuable tool in scenarios where medical experts are not readily available. Future enhancements could include expanding the range of diseases covered and further refining the diagnostic algorithms to improve accuracy and user experience.

5. CONCLUSION

The implementation of the knowledge-based patient self-assessment and diagnosis approach for typhoid, malaria, and related illnesses is a noteworthy development in the accessibility and provision of healthcare. This system has effectively solved major issues in diagnosis and treatment initiation, especially in underprivileged populations with limited access to standard healthcare facilities, by integrating database approaches with medical knowledge representation. The system's capacity to reduce obstacles relating to geographic distance, healthcare infrastructure, and medical expertise has shown how effective it is at enhancing accessibility to healthcare. The technology has enabled prompt access to healthcare suggestions, resulting in early treatment initiation and better health outcomes by enabling individuals to remotely assess their symptoms and receive provisional diagnosis. Furthermore, thorough testing and validation procedures have validated the system's diagnosis accuracy based on patient-reported symptoms and medical criteria. Through precise identification and prioritization of possible diagnoses, the system has made it possible to use healthcare resources more efficiently and lessen the strain on overworked healthcare workers and institutions. Clear instructions, culturally appropriate methods, and an intuitive interface design have all improved the system's acceptability and user-friendliness. All users now have equal access to healthcare information and services since the system can accommodate users with different degrees of health literacy and technology skills. Lastly, the adoption of strong data security and privacy protocols has protected patient privacy and adherence to pertinent healthcare laws. Because the system places a high priority on protecting personal health information, users have become more confident and trusting of it, which has helped it become widely accepted. Therefore, the effective deployment of the knowledge-based patient self-assessment and diagnostic system represents a noteworthy advancement in the delivery of easily available, precise, and patient-focused healthcare treatments for typhoid, malaria, and associated illnesses. The system has the ability to change the way healthcare is delivered, enhance patient outcomes, and lessen the prevalence of infectious diseases in underserved areas by addressing the particular difficulties these populations confront. As further enhancements and refinements are made, this system holds promise for continued advancements in healthcare accessibility and delivery worldwide. However, it is recommended that future studies conduct a more extensive evaluation by considering various datasets and scenarios. This could involve testing the system with more diverse datasets, such as those from different geographical locations, patient demographics, and varying disease prevalence rates.

6. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

7. REFERENCES

[1] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nat. Commun.*, vol. 11, no. 1, pp. 1–9, 2020, doi: 10.1038/s41467-020-17419-7.

[2] F. Hak, T. Guimaraes, and M. Santos, "Towards effective clinical decision support systems: A systematic review," *PLoS One*, vol. 17, no. 8 August, pp. 1–19, 2022, doi: 10.1371/journal.pone.0272846.

[3] E. Triantaphyllou, "A Systematic Survey of Computer-Aided Diagnosis in Medicine : Past and Present To cite this paper or access its final version please use the following information : Yanase J , and Triantaphyllou E , 2019 , A Systematic Survey of Computer-Aided Diagnosis i," no. August, pp. 1–46, 2019.

[4] M. Biglarbegian, "Formalization and Knowledge Representation in Advanced Engineering Informatics," *J. Comput. Nat. Sci.*, vol. 2, no. 1, pp. 8–14, 2022, doi: 10.53759/181x/jcns202202002.

[5] H. Y. Sung and Y. L. Chi, "A knowledge-based system to find over-the-counter medicines for self-medication," *J. Biomed. Inform.*, vol. 108, no. 200, p. 103504, 2020, doi: 10.1016/j.jbi.2020.103504.

[6] S. A. Khan, I. Naim, S. Kusi-Sarpong, H. Gupta, and A. R. Idrisi, "A knowledge-based experts' system for evaluation of digital supply chain readiness," *Knowledge-Based Syst.*, vol. 228, 2021, doi: 10.1016/j.knosys.2021.107262.

[7] S. Rohajawati, "Optimizing Prediabetes Diagnosis Through Knowledge-Based Systems," *Asian J. Eng. Soc. Heal.*, vol. 3, no. 2, pp. 330–342, 2024, doi: 10.46799/ajesh.v3i2.244.

[8] Y. Yohannes, "DEVELOPING KNOWLEDGE BASED SYSTEM FOR DIAGNOSIS AND TREATMENT OF GASTRITIS DISEASE," 2023.

[9] V. A. Stefanescu, "Autonomous Self-Diagnosis System," 2020.