Engineering and Technology Journal e-ISSN: 2456-3358

Volume 08 Issue 11 November -2023, Page No.- 2954-2961

DOI: 10.47191/etj/v8i11.02, I.F. – 7.136

Applications of Knowledge Discovery and Database (KDD) in Clinical Decision Support System

Nwankwo, U. C[1], Ngene, N.J[2], Onuora, J. N[3]

[1]Lecturer, Department of Computer Science, Caritas University, Amorji Nike Enugu, Enugu State, Nigeria. OCID 0000 0002 9536652

[2]Lecturer, Department of Computer Science, Enugu State University of Science and Technology, Enugu State, Nigeria.

[3]Lectuer, School of General Studies, Caritas University Enugu, Nigeria.

ABSTRACT: Applications of Knowledge Discovery and Database (KDD) in Clinical Decision Support Systems) is a research work that aims to develop a medical system that has the ability to detect and suggest cure for an ailment with minimal effort. Healthcare system has been a major source of worry across the globe in recent time due to emergence of different types of diseases and epidemics. The number of health care personnel in various hospitals falls short of the number required most especially in terms of specialists. The medical system in Nigeria today suffers from lack of fast, accurate, reliable and intelligent software solutions that can help healthcare practitioners make decisions that would solve urgent, and in some cases, complex medical problems in real-time. Also, the cost of processing and analyzing large volumes of data in a medical environment is high most especially in terms of time consumption. So in this research design and implement clinical decision support system using knowledge discovery in database (KDD) was the major focus. The design required a computerized database for storing medical records as well as software for improved intelligence-based medical procedures that were going to be simpler to use, versatile, adaptable, savvy, and effortless at combining and evaluating medical data, allowing medical practitioners at all levels to make realistic, intelligent, and real-time decisions on critical health issues. The system was designed using the Object-Oriented Design and Analysis Methodology (OOADM).

KEYWORD: knowledge discovery in databases, assistance systems, human-computer communication, dynamic decision support systems.

I INTRODUCTION

In medical domain, a vast amount of knowledge is required to solve seemingly simple problems. A physician is required to remember and apply knowledge of a vast array of documented disease presentations, diagnostic parameters, combination of drug therapies and guidelines. However the physician's cognitive abilities are restricted due to factors like multitasking, limited reasoning and memory capacity. Consequently, it is impossible for an unaided physician to make the right decision every time. Ironically, the increasing rate of information generated by medical advances has aggravated the physician's task. Looking at the field of clinical decision support system (CDSS) as a solution to healthcare challenges, the algorithms employed by a CDSS can be characterised by the reasoning paradigm employed by the CDSS, which is primarily of two types. The first type is the knowledge-driven paradigm, which employs an inference engine that applies the formalised clinical knowledge representations stored in the knowledge base to an instance of the patient's data. The knowledge base in this case, can contain rules for treatment or diagnosis, probabilistic associations of symptoms to diseases, drug-drug interactions, or clinical workflows with decision steps. The second type is the data-driven reasoning paradigm which employs methods from machine-learning such as neural networks, support vector machines, statistical methods such as regression models, and pattern recognition methods such as k-nearest neighbors (kNN) to detect patterns in clinical data. Rather than an inference engine, the data-driven approach uses a classifier, and the data contained in the knowledge-based may be thought of as the trained parameters or weights associated with the classifier. Optimal use of CDSSs have the potential to improve healthcare processes and outcomes by ensuring compliance with the most up to date guidelines, reduce clinical errors, and reduce cost without compromising care. A CDSS should be viewed as supportive tool available to the clinician to facilitate their task, and definitely not as her substitute. Hence, this thesis concentrated on the design and implementation of clinical decision support system that is used by physician in the process of taking medical decisions on patients. The system used knowledge discovery and data mining approaches as the optimization technique for the clinical decision support system. The main purpose of the CDSS as developed in this research is to assist clinicians at the point of care This implies that a doctor would work with a CDSS to assist in the diagnosis, analysis, and other aspects of patient data. Previous CDSS ideas proposed using the CDSS to make decisions for the doctor. The physician would enter the information, wait for the clinical decision support system to output the "correct" decision, and then act on that result. The new methodology of using CDSS as implemented in this thesis is to assist the clinician to interact with the CDSS utilizing both the clinician’s knowledge and the CDSS to make a better analysis of the patients data than either human or CDSS could make on their own. Typically, the CDSS would recommend outputs or a group of reports for the clinician to review, and the clinician would select helpful information while rejecting erroneous CDSS suggestions.

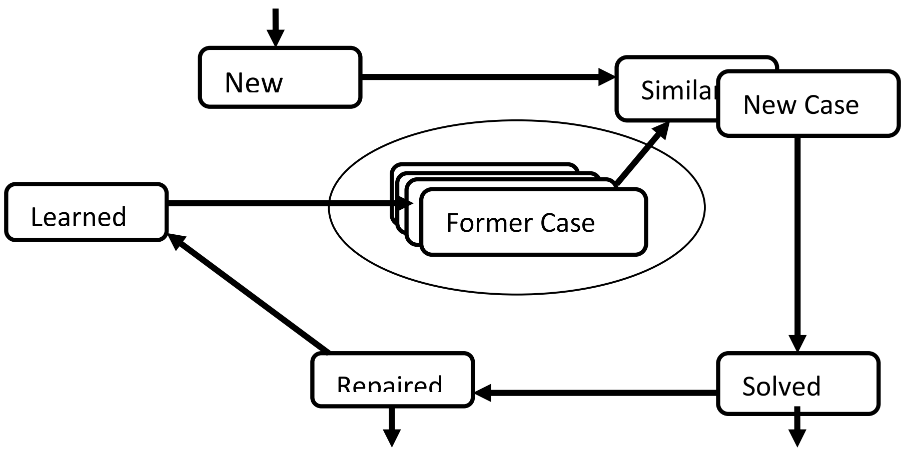
II. LITERATURE SURVEY

[1] In his research work titled “A framework for chronic kidney disease diagnosis based on case based reasoning” presented a case-based reasoning (CBR) technique suitable for problems that depend on experiences. The first step in building CBR system is preparing a comprehensive case base from patients’ electronic health records (EHRs). EHR data need quality improvement steps, such as normalization, feature selection, feature weighting, and outlier detection. The depiction of the resultant case base using codified ideas and terminology is critical in the medical industry. The manuscript proposes a methodology for diagnosing the chronic kidney disease based on an ontology reasoning mechanism. In the paper, they first prepare the chronic kidney dataset of 400 real cases with 25 features by utilizing a set of data mining algorithms. Next, they construct an ontology structure to represent this case base in the W3C web ontology language (OWL) ontology format and populate this ontology with the individual cases. The fuzzy rough set algorithm used achieved the highest accuracy for selecting the most suitable feature set. The resulting OWL ontology is based on disease ontology (DO) semantics, which is the most common and standardized ontology in the medical field. This paper concentrated on the preprocessing step of kidney dataset and representing the resulting data in the form of ontology. The followed steps were numerical dataset normalization, feature reduction and selection, treating missing data values, and features weights assignment. In each of these preprocessing steps, they adopted many algorithms. In addition, they recorded the results of classification accuracy of these techniques. These steps were done on a dataset of kidney patients. The resulted high-quality data can be considered as a case base knowledge, which can improve the retrieving process in CBR systems. We then focused on representation the produced dataset in ontology structure. In their conclusions, they recommend that future works will concentrate on applying the remaining steps in case base preparation like handling fuzziness in the case base. Also, they recommended the implementation of CBR system for diagnosing patients with kidney disease. In addition, they seek to enlarge the size of kidney case base for improving the retrieval results of the CBR system in future works. According to [2], digitally integrated healthcare settings may be attained if doctors are encouraged to employ smart technologies for knowledge production and exchange in clinical decision-making systems (CDSS). While CDSSs are moving towards intelligent settings, they lack support for physicians' abstraction of technology-oriented information. As a result, in order for physicians to produce shareable and interoperable information for CDSS processes, abstractions within the manner of some user-friendly and flexible writing environment are necessary. The proposed system provides a user-friendly authoring environment to create Arden Syntax MLM (Medical Logic Module) as shareable knowledge rules for intelligent decision-making by CDSS. They stated that present systems are unfriendly to physicians or lack interoperability and information sharing. They present the Intelligence Knowledge Creating Tool (I-KCT) in this study, an expert authoring environment that addresses the aforementioned shortcomings. Share ability is achieved by utilizing Arden Syntax to create a repository of information from MLMs. Standardized data models and terminology improve interoperability. However, creating shareable and interoperable information without abstraction using Arden Syntax increases complexity, which makes it hard for physicians to use the publishing environment. Therefore, physician friendliness is provided by abstraction at the application layer to reduce complexity. This abstraction is regulated by mappings created between legacy system concepts, which are modeled as domain clinical model (DCM) and decision support standards such as virtual medical record (vMR) and Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT). They represent these mappings with a semantic reconciliation model (SRM). The objective of the study is the creation of shareable and interoperable knowledge using a user- friendly and flexible I-KAT. As a result, researchers evaluated the system based on its completion or satisfaction criteria, which were examined utilizing system- and user-centric assessment techniques. They contrasted the execution of clinical information modeling system specifications with the system that was proposed and with existing systems for system-centric evaluation. According to the results, their system supported 82.05% of the criteria completely, 7.69% partially, and 10.25% not at all. 35.89% of criteria were completely provided in current systems; 28.20% were somewhat accepted; and 35.89% were not fully supported. 'Ease of use' was the assessment criterion for user-centric evaluation. The suggested approach outperformed the present system in terms of MLM generation time by 15 times [2]. The proposed system uses vMR schema classes as a standard data model with standard SNOMED CT terminologies to enhance shareability. To increase user friendliness, the system

provides a high-level abstraction of vMR schema classes in the form of a DCM. The use of different models and terminologies to represent clinical knowledge requires mapping among them. They recommended that future plan to extend the system to support complex Arden Syntax artifacts such as loops and aggregate functions are needed. To begin, they can be used at the level of each patient to communicate risk and aid in clinical decision-making by placing patients into various therapeutic option categories or to determine whether additional testing is needed to arrive at a suitable decision [11]. Additionally, they recommend that an endeavor to integrate ongoing research on maintenance and validation of MLMs into the current system is required. The research done by [3] illustrated that to have high-quality content of case base, they should be prepared using effective and dependable sources. There must be a regular or a uniform distribution of cases of the problems [3]. When cases are not appropriately distributed or fairly among problems, this results in the existence of any problems without solutions while others may have redundant and useless cases. The CBR systems' accuracy will be enhanced when the medical dataset of EHR is pre-processed. The step of data pre-processing is about CBR as well as applying the many techniques of artificial intelligence, such as genetic algorithm, k-nearest neighbor (KNN), Bayesian network, and fuzzy approach [4]. The steps of data preprocessing steps involve handling missing data, feature selection and weighting, data integration, discretization of data, normalization, and outliers' detection and removal [4]. These data preprocessing steps are applied on EHR for converting database structure of EHR to case base structure and transforming EHR generic data to specified case base.

Based Reasoning Systems

Case-based Reasoning (CBR) has become a successful technique for knowledge-based systems in many domains, while in medical domains some more problems arise to use this method. Case-based Reasoning means to use previous experience in form of cases to understand and solve new problems. A case-based reasoner recalls previous situations that are comparable to the present problem and seeks to adapt their answers to the present case (Fig. 2). Shows the Case-based Reasoning cycle developed by [5]. The underlying idea is the assumption that similar problems have similar solutions. Though this assumption is not always true, it holds for many practical domains. CBR consists of two main tasks: The first is the retrieval which is the search for or the calculation of most similar cases. If the case base is small, a sequential computation is possible; otherwise, non-sequential index or classification techniques (such as ID3 or the closest neighbor match) should be used. Much research has been conducted in recent years for this endeavor, and it has become similarly easy to discover powerful CBR retrieval algorithms sufficient for practically any type of application challenge. The second task, the adaptation (reuse and revision) means a modification of solutions of former similar cases to fit for a current one. For comparable instances, straightforward transfers are frequently sufficient, although adaptation can be challenging. There are no generic procedures or algorithms, and specialists' knowledge is a combination of textbook information and experience {{6}, with typical and exceptional situations included in their reasoning. In medical knowledge based systems there are two sorts of knowledge, objective knowledge, which can be found in textbooks, and subjective knowledge, which is limited in space and time and changes frequently.



Confirmed Solution Proposed Solution

Fig. 2: Cycle of Case-Based Reasoning [5]

The challenge of upgrading subjective knowledge can be partially overcome by introducing new, up-to-date situations progressively [6]. Both types of knowledge may be clearly distinguished: Subjective information is included in situations, but objective textbook knowledge may be expressed as rules or functions. As a result, the following are the reasons in favour of case-oriented methods:

1. Reasoning with cases corresponds with the decision making process of physicians.
2. Incorporating new cases necessitates the automated updating of portions of the changing knowledge.
3. Objective and subjective information can be distinguished.
4. Integration with hospital communication systems is simple because cases are frequently stored.

CBR has mostly been used in medicine for both diagnostic and therapeutic purposes. Related methods have been used in further fields, case-oriented methods for tutoring and retrieval methods to search for similar images. One of the earliest medical expert systems that uses CBR techniques is CASEY [7]. It deals with heart failure diagnosis. The system involves three steps: identifying similarities, comparing current scenarios, and addressing significant discrepancies in diagnosis. Casey employs a rule-based domain theory if no analogous case can be discovered or when all modification attempts fail. The most intriguing part of CASEY is its audacious effort to tackle the adaptation challenge via generic adaptability operators. However, as many features have to be considered in the heart failure domain and as consequently many differences between cases can occur, not all differences between former similar and current cases can be handled by the developed general adaptation operators. The FLORENCE system [8] deals with health care planning in a broader sense, for nursing, which is a less specialized field. It fulfils all three basic planning tasks: diagnosis, prognosis, prescription Individual patient features may be erased by the dissemination of health record data in medical networks, exposing solely generic, utilitarian conditions for computer-supported reasoning applications. Prognosis seeks to answer the question: "How may the health status of this patient change in the future? " Here a Case based approach is used. The present patient is contrasted to a similar earlier patient whose health status progression is known. Similar patients are sought first for their general health condition and then for their particular health markers. As the further development of a patient depends not only on his situation (current health status, basic and present diseases), but additionally on further treatments, several individual research studies for different treatments are generated. Prescription employs extensive knowledge and case-based strategies to improve the health of patients. It integrates rule-based and case-based approaches using procedural comprehension packets with diagnostic memory organisation packets. Memory stores schemas, specific instances of assessment, and scenes representing pattern manifestations in specific circumstances. This memory organization and retrieval allows a reasoner to find the most specific problem-solving procedures available. To use Case-based Reasoning a few problems have to be solved: A representation form for cases has to be determined, an appropriate retrieval algorithm has to be selected and an infinite growth of the case base has to be avoided e.g. by clustering cases into prototypes and removing redundant cases or by restricting the case base to a fixed number of cases and updating the case base during an expert consultation session, but the main problem of Case-based Reasoning is the adaptation task. Little research has been undertaken on this topic and only formal adaptation models [9], but no general methods have been developed so far. The adaptation still depends on domain and application characteristics. Sometimes no adaptation is necessary, because e.g. the field and the cases are as unspecialized as in FLORENCE, sometimes the adaptation is a simple solution transfer or only a little bit more, sometimes just a few constraints have to be checked, but sometimes many differences between current and former similar cases have to be considered (e.g. CASEY). The latter situation is not only a problem for medical applications. However, in medicine it increases, because cases often consist of an extremely large number of features. In nonmedical CBR applications, the adaptation is usually solved by a set of specific adaptation rules, which usually have to be acquired during expert consultation sessions. As these rule sets have to consider all possible important differences between current and former similar cases, for medical applications it is mostly impossible to generate such sets. So, some adaptation solutions have been developed that are not limited to, but are rather typical for medical domains

Medical Records in Computer Supported Reasoning

Medical records typically involve: first, clinical information obtained through the case history and physical examination; second, data obtained from different diagnostic processes; third, facts associated with various therapeutic interventions; and, finally, administrative and financial data such as insurance, medical treatment costs, hospitalization costs, and so on. [10]. Medical records are used in a variety of ways and they serve a multiplicity of purposes. The first of these functions, is that of the use of the records in the treatment process. On the basis of information included in the record, the physician reaches a diagnosis and charters the course of the intervention, and, on the basis of the record, a patient is enabled to make informed decisions, concerning the proposed treatment. Another important point is that physicians, nurses, and other types of health care specialists must be proficient in information derived from clinical practice alongside formal education and information generated from books and other publications. Information registered in patient's records must be precise and comprehensive, since data concerning clinical trials of novel therapeutic schemata and statistical data involved in epidemiological research, derive from these records. Knowledge bases, used in various decision supporting systems, are composed by the selective employment of components of the records, and consequently, various research programs and their methods, are bound to the structure, the availability and the handling of the medical records. Most of these data can be found primarily, only in medical records. The exchange of medical record data throughout medical networks has the potential to erase specific patient traits, leaving only generic and depersonalized situations via computer-supported reasoning applications. The optimization of the decision making processes in Medicine, requires continuous training. The spread of healthcare record data in medical networks has the potential to obliterate specific patient features, leaving only abstract and impersonal circumstances in computer-supported reasoning applications.

On the other hand, hypertext and multimedia courseware, are gaining importance in Medical Education and they can be used in order to offer clinical-practice oriented training. Web-based and other emerging technological alternatives promise to reach various groups, offering them continuous education services. These groups may comprise also of those, who are already engaged in professional work, such as physicians, nurses, engineers, physicists, technicians etc. Case based reasoning systems can play an additional role, in Medical Education, that of contributing to the acquisition and dissipation of clinical expertise since, first, they familiarize the trainee with a rich empirical content, often not available in individual clinics, second, by relating this content to the theoretical aspects of the specific cases, and third, by revealing the " diagnostic feeling" involved in diagnosis and treatment. This supportive role, as well as the critical role that reasoning based on case systems should play in the organization, methodology, and content of healthcare education, highlight the need for more study in both theoretical and practical elements of such systems

III. Methodology

Object-oriented analysis and design methodology (OOADM) which is adopted in this research work is a set of standards for system analysis and application design. It employs a formal, systematic approach to information system analysis and design. Object-oriented design (OOD) develops analytical models into implementation requirements. The main difference between object-oriented evaluation and other methods of evaluation is that the object-oriented technique organizes requirements around objects, which include actions (procedures) or states (data) modelled on objects in the real world in which the system interacts. Other or classic analytical approaches analyze the two aspects—procedures and data—independently. Data, for example, may be represented using ER diagrams, whereas actions can be expressed through flow diagrams or structure charts. The following are the main responsibilities in object-oriented analysis (OOA):

• Locate the items;
• Organize the objects and explain how they interact;
• Define the objects' behavior; • Define the objects' internals.

Case scenarios and object models are common models in OOA.

IV. SYSTEM DESIGN

Use cases are situations that specify standard area tasks that the system must do. Object models explain the primary objects' names, class relationships (e.g., circles have a kind of subclass of shape), actions, and properties.

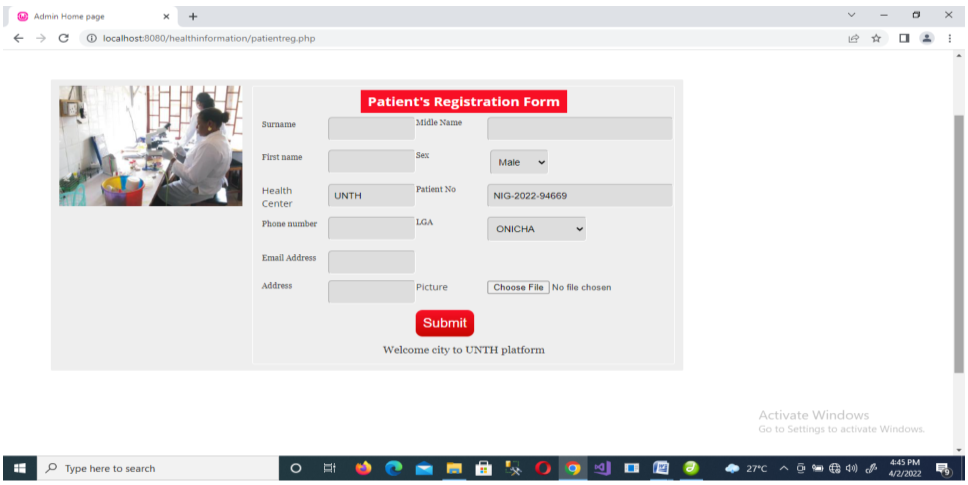


Fig.4.2: Patient Registration Form

Figure 4.2 shows the form used to record patient's data in the hospital. It captures the details of the patient and the hospital where the registration took place.

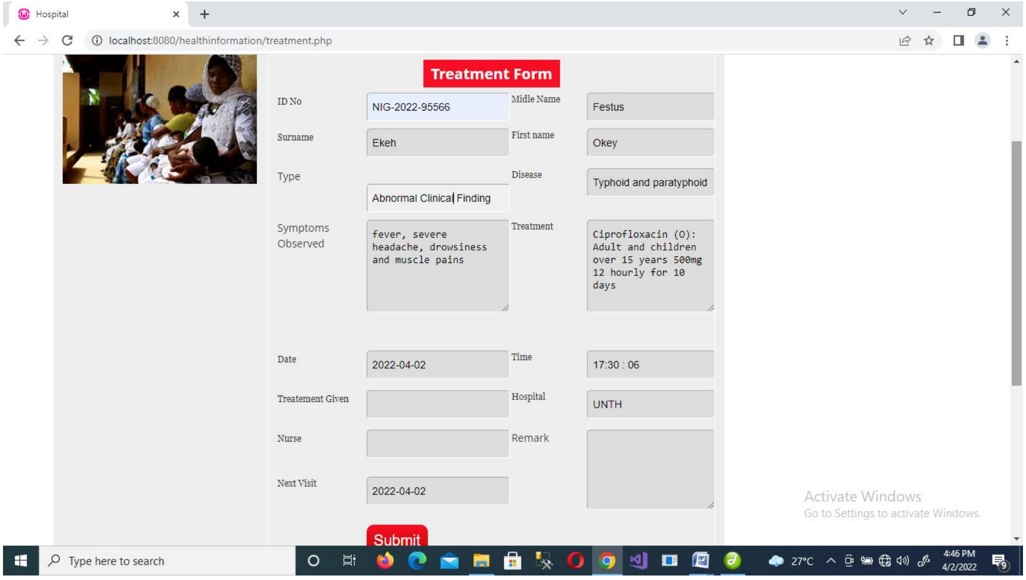


Fig.4.3: Patient Treatment Form

The patient's treatment form as shown in figure 4.3 is used to record each drug of treatment given to a patient and the time and date it was administered on the patient.

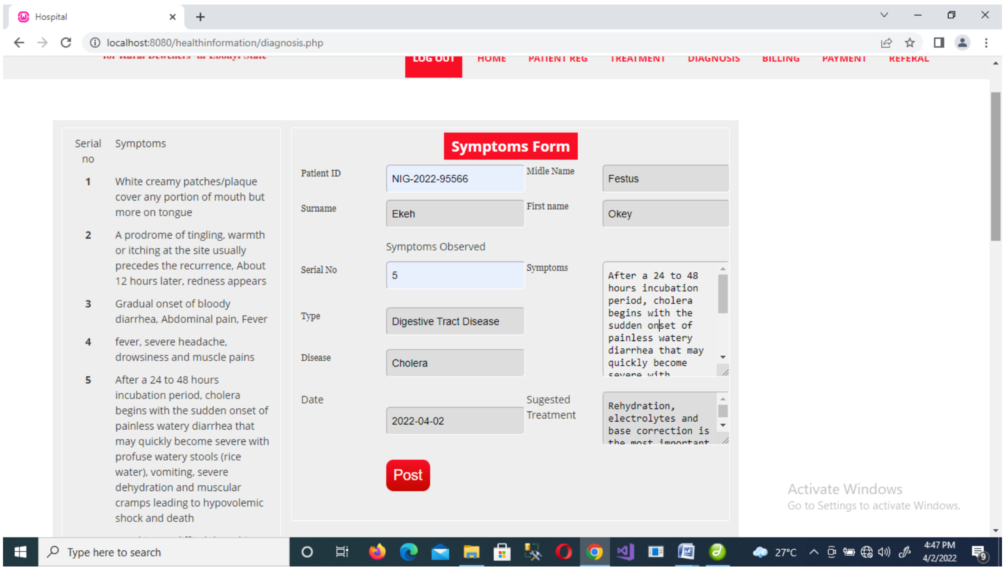


Fig.4.4: Patient Diagnosis Using KDD Form

Figure 4.4 is used to capture the patient symptoms and use KDD to find the type of disease, the suggested treatment as recommended and other patient's details.

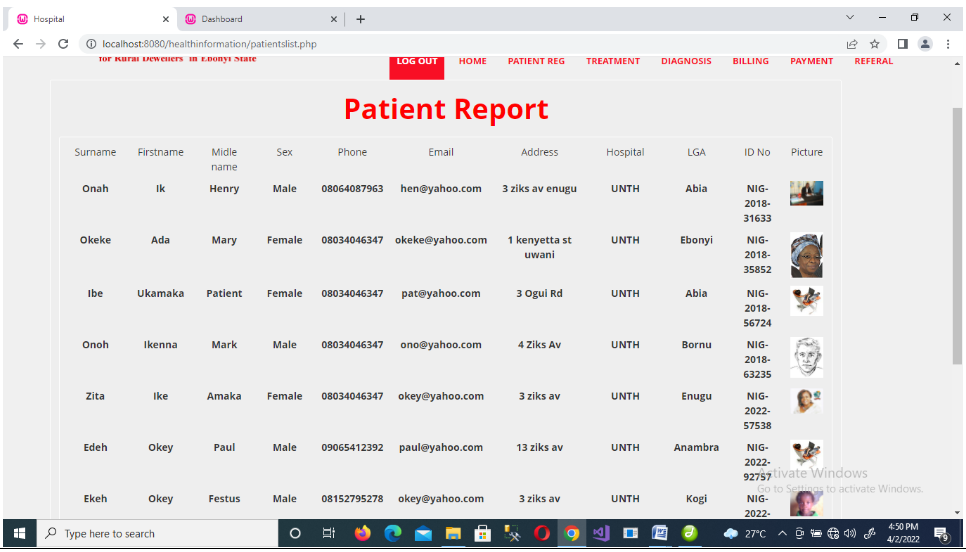


Fig.4.5: List of Patients

Figure 4.5 is the patient's registry showing the records of patients in the hospital

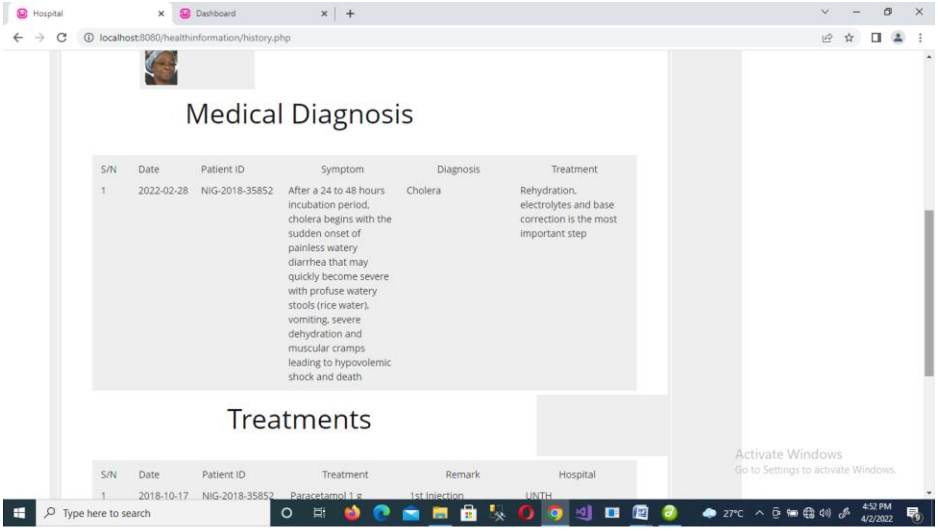


Fig.4.6: Patient's Medical History

Figure 4.6 shows the patient's medical history

V. SUMMARY

This study offered a clinical assistance system. The research investigated the obstacles encountered in the medical sector, particularly in identifying and treating patients. According to the research, problems with medical information processing are largely due to late information retrieval and the wrong medical diagnosis. So, utilizing KDD, this study created a framework for patient diagnosis. Upon extracting the symptoms of the patient, the new system used the KDD algorithm to advice treatment. As a result, the system will be adaptable, intelligent, agile, and autonomous in its integration and analysis of medical data.

RECOMMENDATION

This research's application can be used by hospitals, clinics, and health facilities

CONCLUSIONS

This research's implementation will aid physicians in identifying and administering treatment to patients. The key benefit of this program is that doctors will have a complete history of their patients' health state, and patients will be able to carry their data with them wherever they go. Through the patient knowledge discovery database, the suggested system would also assist medical doctors in speeding up the evaluation and treatment of patients.

REFERENCES

1. Seham, A., Mohammed, E., Shaker, E. and Abdel, N. H. (2018) A framework for chronic kidney disease diagnosis based on case based reasoning International Journal of Advanced Computer Research, Vol 8(35) ISSN (Print): 2249-7277 ISSN (Online): 2277-7970 http://dx.doi.org/10.19101/IJACR.2018.834003
2. Taqdir, A., Maqbool, H., Wajahat. A., Muhammad, A., Jamil, H., Rahman, A., Waseem, H. , Arif, J., Byeong, H., Sungyoung, L. (2017) Multi-model- based interactive authoring environment for creating shareable medical knowledge, *Department of Computer Science and Engineering, Kyung Hee University, Seocheon-dong, Giheung-gu, Yongin-si 446-701, Gyeonggi-do, Republic of Korea*
3. [3] Richter, M.M., Weber, R.O. (2016) Case-based reasoning. Springer-Verlag Berlin. A Hand, D., Heikki, M., Smyth, P. (2011). Principles of Data Mining. Cambridge: The MIT Press
4. Gu, D.X., Liang, C.Y., Li, X.G., Yang, S.L., Zhang, P. (2010) Intelligent technique for knowledge reuse of dental medical records based on case-based reasoning. Journal of Medical Systems; 34(2):213-22.
5. Aamodt, A. (2014) Case-based reasoning: Foundation issues. *AICOM* 7 39-59
6. Gierl, L. (2012) Classification in expert system in medicine. ISBN: 0-7734-4054-2, Mellen Univ. Press, Lewiston.
7. Koton, P. (1988) Reasoning about evidence in causal explanations. In: J. Kolodner (ed.): First Workshop on CBR. ISBN: 0-934613-93-1, Morgan Kaufmann, San Mateo pp. 260-270.
8. [Bradburn, C. and Zeleznikow, J. (2013) The application of case-based reasoning to the tasks of

health care planning. In: S. Wess *et al*.: Proc of European Workshop on CBR, Springer, Berlin, pp. 365-378

9. Bergmann, R. and Wilke, W. (2018).Towards a new formal model of transformational adaptation in case-based reasoning. In: L. Gierl and M. Lenz (eds.): Proc of 6th German Workshop on CBR, University of Rostock, pp.43-52
10. Papagounos, B. S. (1999). The Multifarious Function of Medical Records: Ethical Issues, *Method Inform Med* 38, 317–20.
11. Ezeaku-Ezeme, C. N., Cornelius, N. U., & Akpan, A. M. (2021). A medical predictive system for comparative analysis of fetal parameters. *GSC Advanced Engineering and Technology*, *1*(2), 021-030.