

DEVELOPMENT OF AN ENHANCED ARCHITECTURE OF CLASSIFICATION OF PATIENT MEDICAL RECORDS

¹ILOFULUNWA C .G

ilochinonye@yahoo.com

²MGBEAFULIKE I.J

ike.mgbeafuike@gmail.com

³OKIDE S.O

samuelokide@yahoo.com

^{1,2} Computer Science Department, Chukwuemeka Odumegwu University, Uli.

³Computer Science Department, Nnamdi Azikiwe University, Awka, Anambra State

ABSTRACT

The tedious nature of searching or locating for a doctors previous prescriptions or patient review in medical records for a particular illness and the process of searching and reviewing previous patients medical prescriptions to a large extent has been very difficult. The inquirers of such records are being confronted with the problems of information overload, extracting relevant patients previous medical records prescriptions, loss of stored records, slow information retrievals process and analysis for previous prescription, which the conventional method of storing patients medical record do not seem to be able to tackle. This project proposed a data mining model that can classify patients medical records based on the similarity of previous illness and prescription using clustering technique. The results obtained shows that medical records prescriptions can be classified based on the similarity of previous patient diagnosis, illness and prescriptions which helps medical personnel to easily take decision on the accurate use of drugs to treat their patients and manage records in their achieves effectively.

Keywords: Medical records, Classification, data warehouse

1.0 INTRODUCTION

Medical records management is the part of records management that relate to the operation of a healthcare practice. It is the field of management that is responsible for all records throughout their lifecycle from creation, receipt, maintenance, and use to disposal. Medical records management can involve many kinds of records about a practice and patients, including but not limited to a patient's history, clinical findings, diagnostics test results, patient's progress, medication and even email. Additionally, it involves making sure all of these items are accessible, safe and secure.

Medical records can help both the physician and the patients. For the patients, the records can outline one's history and treatment plan in an easily accessible way. For the physician, it can provide support about the correctness of that treatment plan. A good medical records management system can be the difference between life and death if someone needs to check a test, a medication, or make quick medical decision. Data management is integral to every clinical trial, for many years, computer technologies including the internet and the World Wide Web have been used to support the management of data in clinical trials. Advances in data collection software, wireless connectivity, and encrypted data transfer allow for web-based data entry systems and storage of data in a centralized location. These structures permit researchers immediate access to data, reduce participant burden and human entry error and maintain the integrity of clinical trial data. Generally, computer has played an

important in global economy in various ways. The computer is used in many different fields such as in education, for learning, in entertainment industries, for management, for problem solving and in business.

The advent of internet has made it possible for online medical records to come up with a way in which the problem of computerizing medical records effectively and sharing it can be solved and computer and internet has also made it possible for millions to access proper medical diagnosis and administer correct treatment procedure. Today, several algorithm, models and technologies are put in place to ensure accurate and precise diagnosis on patient which in turns has reduce the rate of mortality in the country and world at large. One of such technologies is data mining (clustering). In medicine and radiology data mining technology utilizes the available data in radiology information system (RIS) and hospital information system HIS). It provides meaningful information adding value to diagnosis, plans further patient management, save time and reduces costs for the healthcare industry (Howell, 2012). The application of information mining techniques to the medical domain are useful in extracting medical knowledge for diagnosis, decision-making, screening, monitoring, therapy support and patient management records. Data mining techniques have been widely used in diagnostic and health care applications. Online medical diagnosis software automates the whole process of patients' diagnosis with accurate and precise result using a well detail and proven algorithm to achieve it desire output, one of such algorithm is data mining (clustering), using the software will remove all such complexities as it helps you to retrieve your information accurately. Implementation of the web-based medical records management system will have a dramatic impact on the overall quality of health delivery in developing countries.

2.0 Related Literature

The related works are employed in research to review related contents or research done by other scholars in other to identify changing issues with the concerned domain of the research. Below, we discuss some of the related works reviewed and to assist to achieve the aim of this work. Computer-Based Clinical Information System, Hunt, (2004) stated that the key advantage of shifting to computer based Patients record is the opportunity to strengthen the link between the hospital records and management information system so that resources used and quality of care can be analyzed using hospital database which increase physicians efficiency, and reduce costs, as well as promote standardization of care.

Reduce cost: Users delay of acquaintance with the electronic Information technology. Many healthcare providers have found that electronic health records help improve medical practice. Efficiencies and cost saving. A national survey of doctors who are ready for meaningful use offers important evidence 79% of providers reports that with an EHR, their practice functions more efficiently 82% report that sending prescriptions electronically (e-prescribing) save time.

Modernizing Health-care through Electronic Medical Record, Miller (2000) says that health care has been an issue of growing importance for national government. Many health care plans have been developed in order to Control the cost, quality and the availability of health care for all citizens.

Cost: failure of electronic patient system may cost negative effect on medical record. A question from every healthcare provider is: where will the money come from to pay for this?

Quality: Effective computerized systems and procedures ensure proper utilization of restricted resources towards quality health care.

Availability: Bed space availability is a difficult task due to the involvement of paper work.

Evaluation of the Quality of Information Retrieval of Clinics Finding from a Computerized Patient Database using a Semantic Technology Brown (2000) opined that the application of electronic clinical information system (ECIS) has generated useful insight into the quality of data accuracy and health care provision. This is partly one of the adapted style and approach to data entry presented by the resent structure and also subsequent follow up of Patients clarification on the policy and frequent treatment.

Information System Development to Enhance the Work of Medical Practitioners in Hospital and Clinics, Tumba (2006) developed a database for drugs that improved the effectiveness and efficiency to eliminate Drugs wastage through over stocking or expiring of drugs. They intend to develop not only a database for drugs but computerized medical records that would encompass other related records of Patients within the health care institution.

Medication waste is a worldwide problem that is not only associated with financial impact. It also negatively impacts safety, health and environment. Better understanding of the problem can enable the healthcare system and society to intervene. Thus automated system will track and manage inventories of pharmaceuticals and other medicals.

Nevertheless the decision support systems will give health care provider real-time advice on diagnosis and treatment options based on continuously updated information. As hospitals implement the technologies described, they will be able to improve the quality of patient care by streamlining processes, reducing duplication and minimizing the risk of medical errors such as administration of the wrong drug or dosage or even the performance of a wrong surgery.

Medical Documentation of a Patient History and Care Laubbel (2012) defined medical documentation of a Patient's Medical History and Care as "medical records" used both as the physical folder of Patients and for the body of information which comprises of total of each Patient's health history. This is a special concern in this work and it forms part of the implementation. It is of great important to reflect faster on the historical information of patients in order to give a good report of diagnosis and for proper treatment of the patients.

2.1 Concept of Classification

Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, software can be developed to classify the data items into groups. For example, we can apply classification in the application that given all records of employees who left the company, predict who will probably leave the company in a future period. In this case, we divide the records of employees into two groups that named "leave" and "stay", and then we can ask our data mining software to classify the employees into separate groups. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data, for example, a classification model could be used to identify loan applicants as low, medium, or high credit risks.

A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time. In addition to the historical credit rating, the data might track employment history, home ownership or rental, years of residence, number and type of investments, and so on. Credit rating would be the target, the other attributes would be the predictors, and the data for each customer would constitute a case.

The simplest type of classification problem is binary classification, in binary classification the target attribute has only two possible values: high credit rating or low credit rating. Multi-class targets have more than two values: low, medium, high, or unknown credit rating. In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships these relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown.

Classification models are tested by comparing the predicted values to known target values in a set of test data. The historical data for a classification project is typically divided into two data sets: one for building the model; the other for testing the model. Classification has many applications in customer segmentation, business modeling, marketing, credit analysis, and biomedical and drug response modeling.

In technical term, classification in data mining defines as assigning an object to a certain class based on its similarity to previous examples of other objects. The classification process comes under the predictive method. With classification, new samples of data are classified into known classes, for classification the data required must be already labeled one. Examples of classification are:

- i. A marketing manager of a company needs to analyze the customer with available profile that who will buy a new computer.
- ii. A bank officer wants to predict that which loan applicants are risky or which are safe.

<https://www.quora.com/what-is-classification-in-data-mining/29/11/18>

2.2 Organization of Data Warehousing

The term "Data Warehouse" was first coined by Bill Inmon in 1990. According to (Inmon, 2008), a data warehouse is a subject oriented, integrated, time-variant, and non-volatile collection of data. This data helps analysts to take informed decisions in an organization. A data warehouse is a database, which is kept separate from the organization's operational database. There is no frequent updating done in a data warehouse and it possesses consolidated historical data, which helps the organization to analyze its business. A data warehouse helps executives to organize, understand, and use their data to take strategic decisions. Data warehouse systems help in the integration of diversity of application systems. Finally, a data warehouse system helps in consolidated historical data analysis.

Data Warehouses are an important asset for organizations to maintain efficiency, profitability and competitive advantages. Organizations collect data through many sources such as Online, Call Centre, Sales Leads, and Inventory Management. The data collected have degrees of value and business relevance. As data is collected, it is passed through a 'conveyor belt', call the Data Life Cycle Management. Let us examine an organization's data life cycle management's policy will dictate the data warehousing design and methodology.

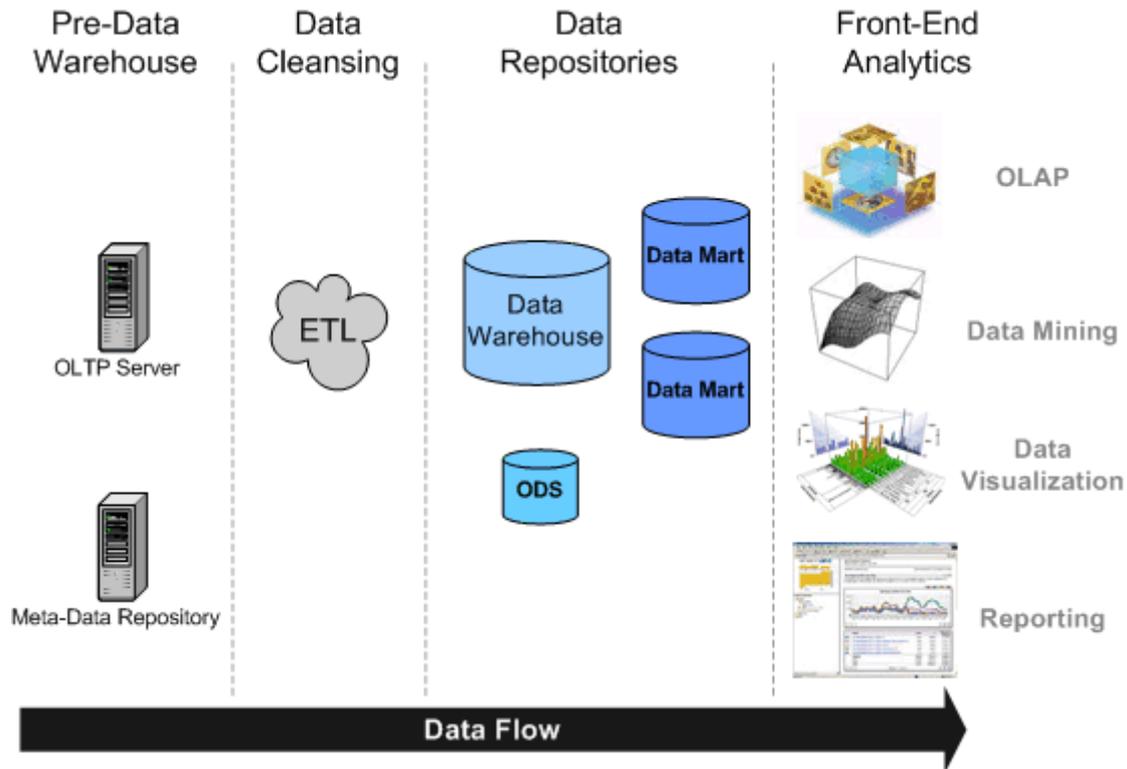


Figure 2: Overview of Data Warehousing Infrastructure

1. **Pre-Data Warehouse:** The pre-Data Warehouse zone provides the data for data warehousing. Data Warehouse designers determine which data contains hospital information value for insertion.

- a) OLTP databases are where operational data are stored.
- b) Metadata ensures the sanctity and accuracy of data entering into the data lifecycle process. Meta-data ensures that data has the right format and relevancy.

2. **Data Cleansing:** Before data enters the data warehouse, the extraction, transformation and cleaning (ETL) process ensures that the data passes the data quality threshold. ETLs are also responsible for running scheduled tasks that extract data from OLTPs.

3. **Data Repositories:** The Data Warehouse repository is the database that stores active data of hospital value for an organization. The Data Warehouse modelling design is optimized for data analysis. There are variants of Data Warehouses such as Data Marts and ODS. Data Marts are not physically any different from Data Warehouses. Data Marts can be thought of as smaller Data Warehouses built on a departmental rather than on a company-wide level. Data Warehouses collects data and is the repository for historical data. Hence it is not always efficient for providing up-to-date analysis. This is where ODS, Operational Data Stores, come in. ODS are used to hold recent data before migration to the Data Warehouse.

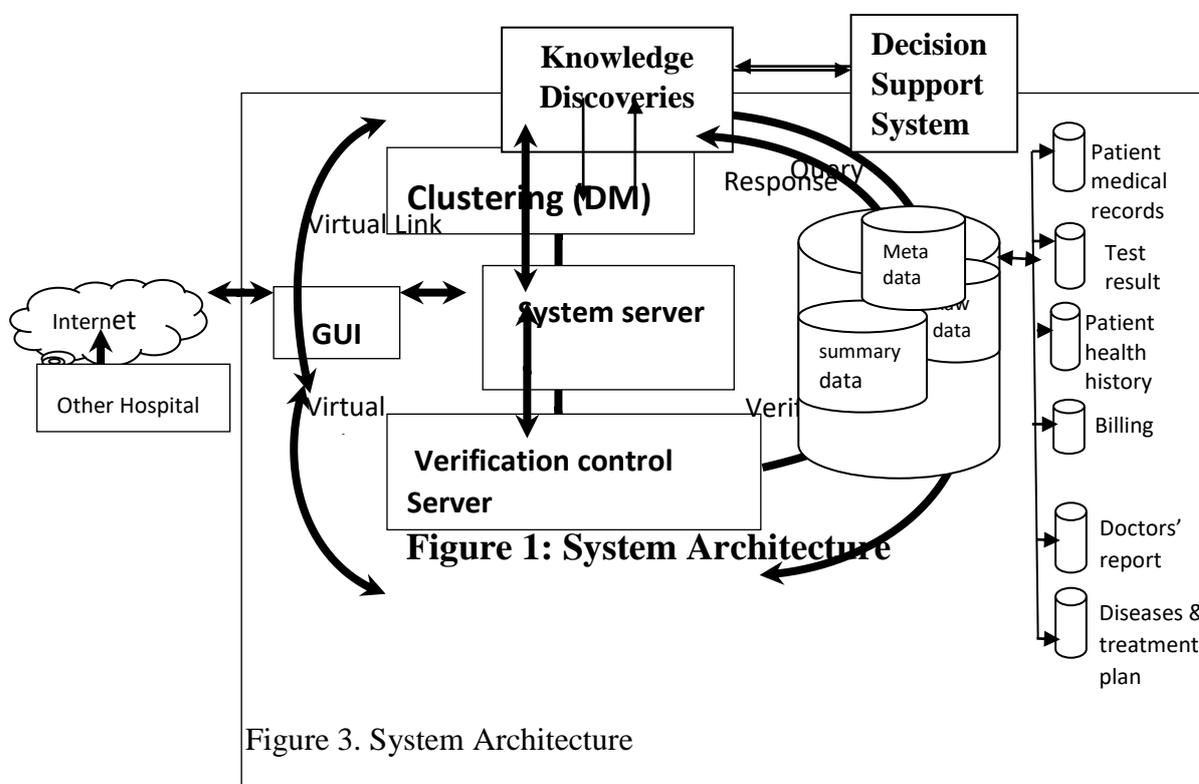
4. **Front-End Analysis:** The last and most critical portion of the Data Warehouse overview are the front-end applications that business users will use to interact with data stored in the repositories.

3.0 Methodology

The methodology used is Object-Oriented Analysis and Design Methodology (OOADM). The work of designing OOADM was started by Grady Booch, James Rumbaugh and Ivar Jacobson in (1994). In Object-Oriented approach we organize requirements around objects, which integrate both behaviors (processes) and states (data) modeled after real world objects that the system interacts with unlike other methodologies where processes and data are considered separately. The aim of OOADM is to find the objects, organize the objects and describe how the objects interact. In the use of object oriented methodology we employ case diagram, the Unified Modeling Language (UML), the system sequence diagram and the activity diagram.

3.1 System Architecture

The existing systems have a lot of limitations which necessitated the designing of the new system. The new system will have a well-defined user-interface. Thus our architecture is illustrated Figure 3.



The module of the system is explained as follows:

2. **Data warehouse:** this module contains the data being processed, used, to be used and needed by the system
3. **Verification Control Server:** This module is in charge of providing fit-back on the last command of an operation previously performed in the system
4. **Knowledge Discoveries (Knowledge representation):** It is the final phase in which the discovered knowledge is visually presented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results.

5. **Decision Support System:** A knowledge discovery assignment must begin with clear objectives in mind for decision support system.
6. **Metadata:** ensures the sanctity and accuracy of data entering into the data lifecycle process, meta-data ensures that data has the right format and relevancy.
7. **Raw data:** consist of a single list
8. **Summary data:** is data summarized all observation in a category (eg) sum of observation, number of occurrence.
9. **Internet:** is the backbone of web, it let you transfer information in different Hospitals in seconds.
10. **Clustering:** Clustering will classify patient medical records from large records and analyzed them based on their grouping.

4.0 Discussion of Result

The health sectors have number of time been face with issues in handling patients medical information and analysis for previous prescription and retrieval of patients medical records using particular query are not easy to a greater extent, extracting relevant patients medical records that will help doctors arrive at effective decisions as in case of emergency. This would have been easily resolved, if patient medical information were stored in centralized location, synchronized with a remote information system and makes records accessible like classifying patient records for easy access for medical practitioner to enable them take urgent decision that required to save life. In this research we proposed a classification algorithm to cluster patient medical records and classify them for easy extraction in st Joseph hospital. We implemented the proposed system using MySQL development environment; we analyzed patient medical prescription with classification algorithm. The developed system will facilitate an understanding of how clustering technique will be used to develop classification of patients' medical prescription. As the user run the program, the data are extracted from the MYSQLQ database and assigned this data items to the clusters. The data that was stored in MYSQL database were extracted using clustering, the result shows the most effective drugs to use to cure a particular illness, the system extracted the drugs that were used for a particular ill health, the drugs that failed and the drugs that proves effective, more so the proposed system clustered patients' medical prescription based on their similarity and classify them for easy decision making.

5.0 CONCLUSION

Due to the convenience of the Web along with its richness and diversity of information sources, more and more people are using it for serious learning. It is important that effective and efficient systems be built to discover and to organize knowledge to help medical practitioner easily take decision that can enhance the recovery of patient on treatment. We have proposed and implemented a novel architecture of an enhanced web-based medical information management system, which help medical practitioners to discover the best possible treatment for ill health based on previous treatment and prescriptions. This enhanced web-based medical information system help medical doctors organize previous medical prescriptions on a particular ill health into similar drugs components to reduce information overload and improve access to the most effective drug in treating illness. The application of clustering as a data mining technique in solving this problem, help to classify patient records and separate them for easy extraction. We have also built a prototype system that implements

the proposed techniques, given a patient ill health, the system automatically discovers salient concepts on drugs and previous prescriptions of the treatment for easy decision taking by medical personnel and improve the qualitative and standardization of healthcare provision.

6.0 REFERENCE

- Brown P.J. (2000). "Evaluation of the quality of information retrieval of clinics finding from a computerized patient Database using a semantic technological"
- Howell, W.L. (2012) "Data Mining and Analytics in Radiology," *PACS and Informatics, RIS*.
- Hunt, Dereck, Hayners L, R.B. Hanna S.C, Smith and kristia (1999).Effects of computer based clinic. Decision system physician.<http://www.va.gov./133h>.
- Inmon, W H - DW 2.0 ARCHITECTURE FOR THE NEXT GENERATION OF DATA WAREHOUSING, Morgan – Kaufman, 2008.
- Laubbel A.C. (2012). "Define medical health care or medical documentation of a patient history and care" An international journal of advanced computer technology, 2 (3), March-2013
- Miller R.J. (2000). "Modernizing Health care through Electronic Medical Record" information system; <http://www.clinictools.org>
- Tumba I.A. (2006) "information system development to enhance the work of medical practitioners in Hospital and clinics".
<https://www.quora.com/what-is-classification-in-data-mining/29/11/18>