Enhanced Medical Intelligence Process Using Data Visualization

Duru Rosetta., Amanze B.C., Agbasonu V.C and Agbakwuru A.O
Department of Computer Science, Faculty of physical Sciences,
Imo State University, Owerri
amanzebethran@yahoo.com
DOI: 10.56201/ijcsmt.v9.no3.2023.pg277.286

Abstract
This paper focused on developing an expert system for health sector that uses intelligent agent to guide doctors in accurately carrying out disease control procedures. Therefore, knowledge sharing across board for medical practitioners. Also, the ontology-based data integration (OBDI) and virtual data integration (VDI) approaches appear as a promising way to resolve semantic issues in information interoperability in medical record management. the system developed was used to manage a disease registry that consists of the concepts of the domain, the attributes characterizing each disease, the different symptoms, and treatments. Also, a relational database for storing and tracking disease outbreak and control using ontology-based data integration (OBDI) was achieved. A platform for Virtualized Data Access – Connect to different data sources and make them accessible from a common logical data access point was created which integrated an intelligent agent that uses case-based reasoning for determining the disease control procedure to be applied to patient treatment for effective control of the disease. The system achieved integration of various patients' medical records from different hospitals using ontology based and virtual data integration technique that will allow clinic data of one patient collected together to form a combinational resource, and could be accessed by physician if authority is assigned to the physician. The Hybrid technique using both Ontology-based data integration and virtual data integration technique for disease control procedure achieved 95% accuracy in predicting the disease control procedure.

Keywords: *Hospitals, Visualization, Database, Case-Based Reasoning and Patients*

Introduction

Virtualization helps in dealing with data intensive problems. The virtualization process is defined as the mapping of an abstract data set to a virtual space according to three major intertwined steps consisting of data selection for representing the problem space, assumptions definition to define the final virtual space, the mapping between the starting space and the final space through a metaphor (Magali and Michel, 2021). Virtualization features in the management of distributed data with data intensive problems such as;

1. The preservation of data and knowledge in their actual format. The virtual process has no impact on the physical reality. Its putative evolution; that is data and knowledge production can go their own way without any necessary change.
2. The operation-ability implies virtualization is not just abstraction, it allows to recursively transform the physical reality according to the lessons learned in virtual reality. That is it aims to actualize physical reality, such that any change in physical reality has its counterpart in virtual reality.
3. As par heterogeneity especially on the internet, virtualization was a successful solution to the problem. Because protocol heterogeneity of existing computer communication architectures was not able to interoperate although they were built for the same purpose that is for exchanging data between distant computers. Virtualization process allows to integrate existing communication technologies and to preserve the future development.
4. Portability- Virtualization process allows for portability as it allows software to be run on any microprocessor architecture without re-writing it, but just compiling it with ad-hoc compilation.
5. Virtualization can be mathematically represented formally by sets say E and V corresponding respectively to physical reality and virtual reality and the two functions ME and MV.

$$ME : E \rightarrow V, \quad MV : V \rightarrow E$$

ME and MV define the mapping rules between E and V. V is the metaphor that refers to a domain of knowledge. (Magali and Michel, 2021). Virtualization process has been successfully used by information technologies to deal with heterogeneity problems, to increase productivity of information technology tools and spread information technology products to non-information technology skilled users. Data Virtualization layer isolates Business Intelligence tools from the details for accessing the underlying data sources. Through the data virtualization layer a Business Intelligence tool can get access to any data available either in the data warehouse, staging area or in the production systems, enabling for instance single views of customer with information about a customer coming from the customer relation management (CRM) system, historical information about this customer retrieved from the data warehouse, billing information from the billing system, etc.

1. Data Virtualization enables access to real-time data.
2. Data Virtualization enables shorter time to market for delivering reports to users.
3. Data Virtualization enables access to any data type from semi-structured to unstructured sources, internal or external, including recently arising sources such as non-SQL databases.

There is cache in data virtualization layer that can be used to store any data to keep track of historical information when needed. (Francesco et al., 2021).

Review of Related Works

Taqdir et al. (2020) Proposed that technologically integrated healthcare environments can be realized if physicians are encouraged to use smart systems for the creation and sharing of knowledge used in clinical decision support systems (CDSS). While CDSSs are heading toward smart environments, they lack support for abstraction of technology-oriented knowledge from physicians. Therefore, abstraction in the form of a user-friendly and flexible authoring environment is required in order for physicians to create shareable and interoperable knowledge for CDSS workflows. 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 pointed out that existing systems are not physician friendly and lack interoperability and share-ability of knowledge. In this paper, they proposed Intelligent-Knowledge Authoring Tool (I-KAT), a knowledge authoring environment that overcomes the above-mentioned limitations. Shareability is achieved by creating a knowledge base from MLMs using Arden Syntax. Interoperability is enhanced using standard data models and terminologies. However, creation of shareable and interoperable knowledge using Arden Syntax without abstraction increases complexity, which ultimately makes it difficult for physicians to use the authoring 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. Therefore, they evaluated the system using completeness and user satisfaction criteria, which was assessed through the system- and user-centric evaluation processes. For system-centric evaluation, they compared the implementation of clinical information modelling system requirements in the proposed system and in existing systems. The results suggested that 82.05% of the requirements were fully supported, 7.69% were partially supported, and 10.25% were not supported by their system. In the existing systems, 35.89% of requirements were fully supported, 28.20% were partially supported, and 35.89% were not supported. For user-centric evaluation, the assessment criterion was ‘ease of use’. The proposed system showed 15 times better results with respect to MLM creation time than the existing (Taqdir, 2020). 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. Additionally, they recommend that an endeavor to integrate ongoing research on maintenance and validation of MLMs into the current system is required. Richter and Weber (2020) 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 (Richter, 2020). 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 (Gu, 2020). 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 (Gu, 2020). 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. Among the studies conducted, which focused on the missing data problem, was by Jagannathan and Petrovic (2021). The authors of this study illustrated that case base, which contains missing data, is a big problem. It affects negatively on the CBR system's performance. So, these missing values must be treated using imputation approaches like mean/mode method, KNN imputation, etc. However, the missing data values of some attributes have been handled while others are not treated. The performance of CBR applications is enhanced when applying preparation algorithms on the case base, such as feature selection. Every CDSS system, which is based on knowledge, needs a step of preprocessing to produce a high-quality dataset. For obtaining superior results during the retrieval process, cleaning, and normalizing data steps are done on retrieval algorithms like KNN algorithm (Jagannathan, 2021). A clinical expert system has been proposed in (Asma, 2020) which adopts decision tree approach for prediction of presence and absence of diabetes among the instances. The author used PIMA Indians Diabetes Data Set, which collects the information of patients with and without developing diabetes. Their research went through two phases. The first phase is data pre-processing including attribute identification and selection, handling missing values, and numerical discretization. The second phase is a diabetes prediction model construction using the decision tree method and they used Weka software throughout all the phases of their study. Nassim (2022) investigated and evaluated a model framework, for diagnostic decisions, based on a cognitive process and a Semantic Web approach. Fuzzy cognitive maps (FCM) are a cognitive process applying the main features of fuzzy logic and neural processors to situations involving imprecision and uncertain descriptions, in a similar way to intuitive human reasoning. The research explored the use of this method for modeling clinical practice guidelines, using Semantic Web tools to implement these guidelines and for the formalization process. Twenty-five clinical and 13 diagnosis concepts were identified, to represent the problem of urinary tract infection diagnosis. With the proposed methodology, each cause-effect relationship between an observation and diagnosis is described by one or more fuzzy rules, thus producing the rule-based FCMs. From each fuzzy rule, an inference is generated, depicting the degree of confidence the research has in the influence concerned. Moreover, based on the available fuzzy rules, some of the relationships (weights) may change in value or degree of confidence before the final diagnosis is reached, taking into account cases and observable states that contribute to a different diagnosis. This is an important issue in FCM construction and diagnosis, as specific initial states of observables can be assessed for patients. The result of the research work is summarized in the development of a platform able to interact with heterogeneous data and formalize knowledge from the model. The results presented here follow the steps in the methodology described above and the inference mechanism for FCM (Eulersharp/plugin). 92% (65/70 patients: 37 males and 28 females) diagnosis proposed by the system was in fully agreement with the guidelines.

Analysis of the System

The system flowchart as shown in figure 1 entails the hybrid of the ontology-based and virtual data integration process. The process flow of the system would be boosted with the software agents as indicated in the data integration layer of the system design. The data integration layer is where the ontology-based and virtual data integration process takes place and it is the aspect of the business intelligence process that the research is enhancing.

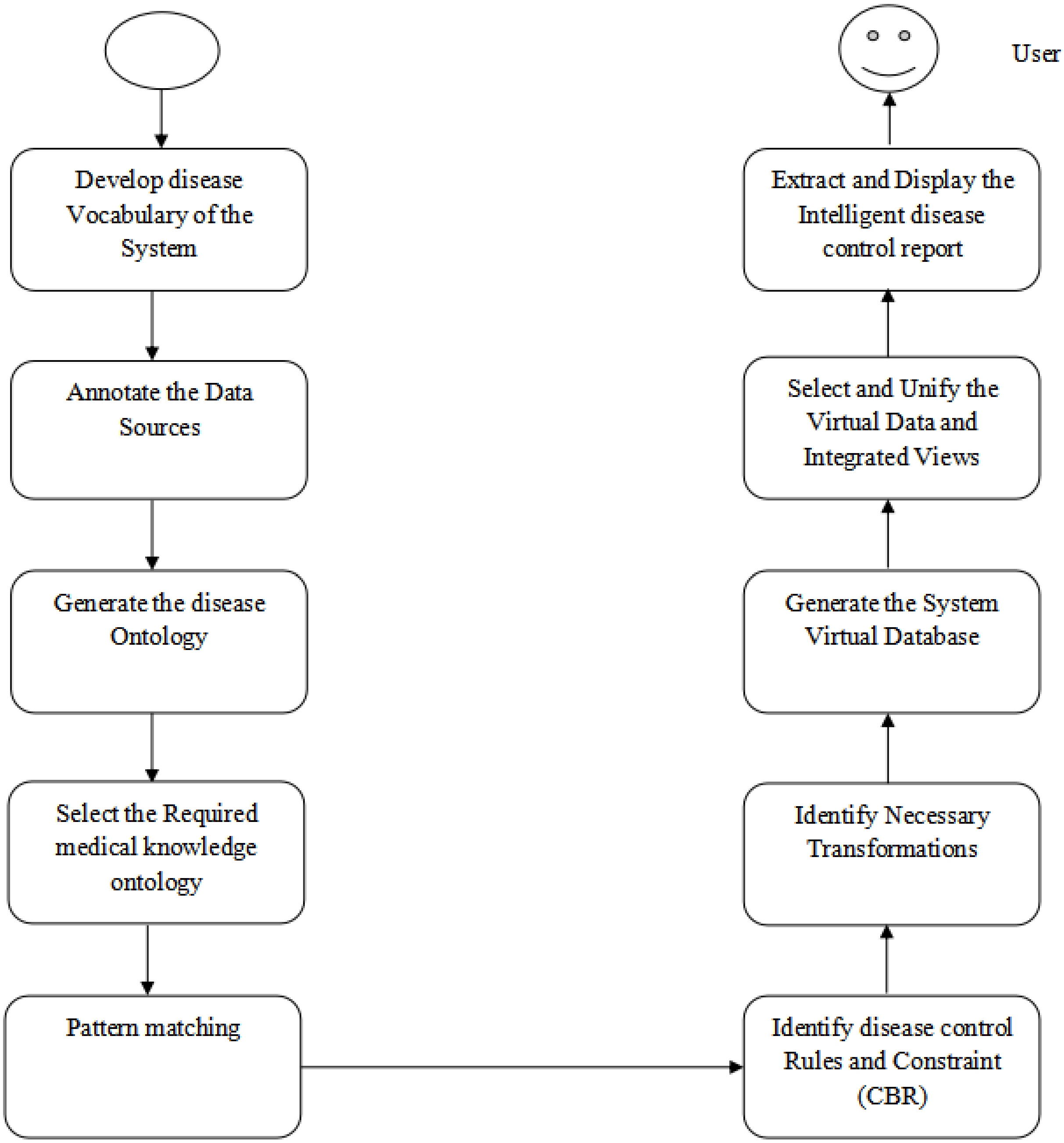

Figure 1: System Flowchart

Justification of the System

The key concept of the proposed medical intelligence hybrid model adopting ontology-based (OBDI) and virtual data integration (VDI) techniques would have the ability to ensure abstraction of data that comes from multiple sources in varying schemas, syntactic accuracy and to have a seamless transition from data into information then into action.

The advantages of the proposed system include the following;

1. The system ensures seamless transition from a practical workspace into a virtual business-oriented analysis world expected by business users.
2. The system is based on hybrid architecture and also relies on elements such as system vocabulary and local ontology per each heterogeneous data source.
3. The system reduces syntax errors, structural and semantic heterogeneity and redundancy which leads to increased availability and degree of completeness.
4. With the layers of ontology-based and virtual data integration both in the system seamless transition and hiding of technical jargons from users is feasible.
5. With both approach to data integration in the system, there would be reduce cost of processing, maintenance and risk in the project as well as increased availability will be feasible.
6. The system would ensure real-time processing, analyzing and accessing of data.
7. The system would be intelligent, reliable, adaptive, flexible and agile.
8. The ontology aspect would see to the reduction in latencies in the decision process which would allow users to take and make faster or fastest business decisions accessing current business data in its proper level of abstraction, thereby enhancing the ability of organization to adapt it as new necessities or in business changes.

These justify the implementation of the new system to help health sector manage and control diseases and patients record.

High-Level Model of the System

The high-level model of the proposed system as shown in figure 2 applied ontology-based data integration technique as it relates to seamless transition, solving inconsistencies in semantics and accuracy issues in syntactic and that of virtual data integration technique as it relates to hiding of technical jargons from users and unifying integrated and reconciling views of data residing at different sources as well as at different location for users.

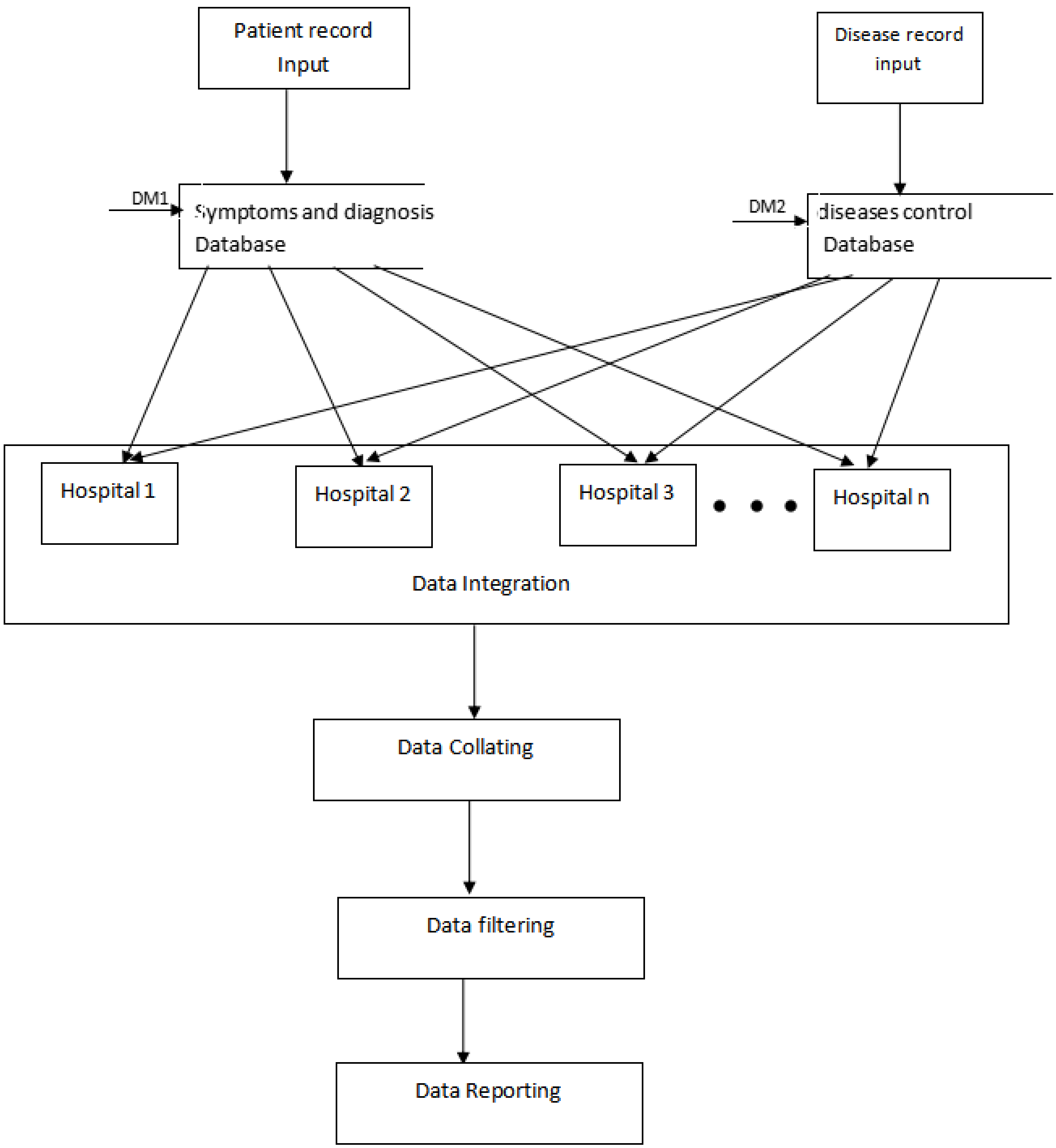

Figure 2: High Level Model of the Proposed System

Data Visualization

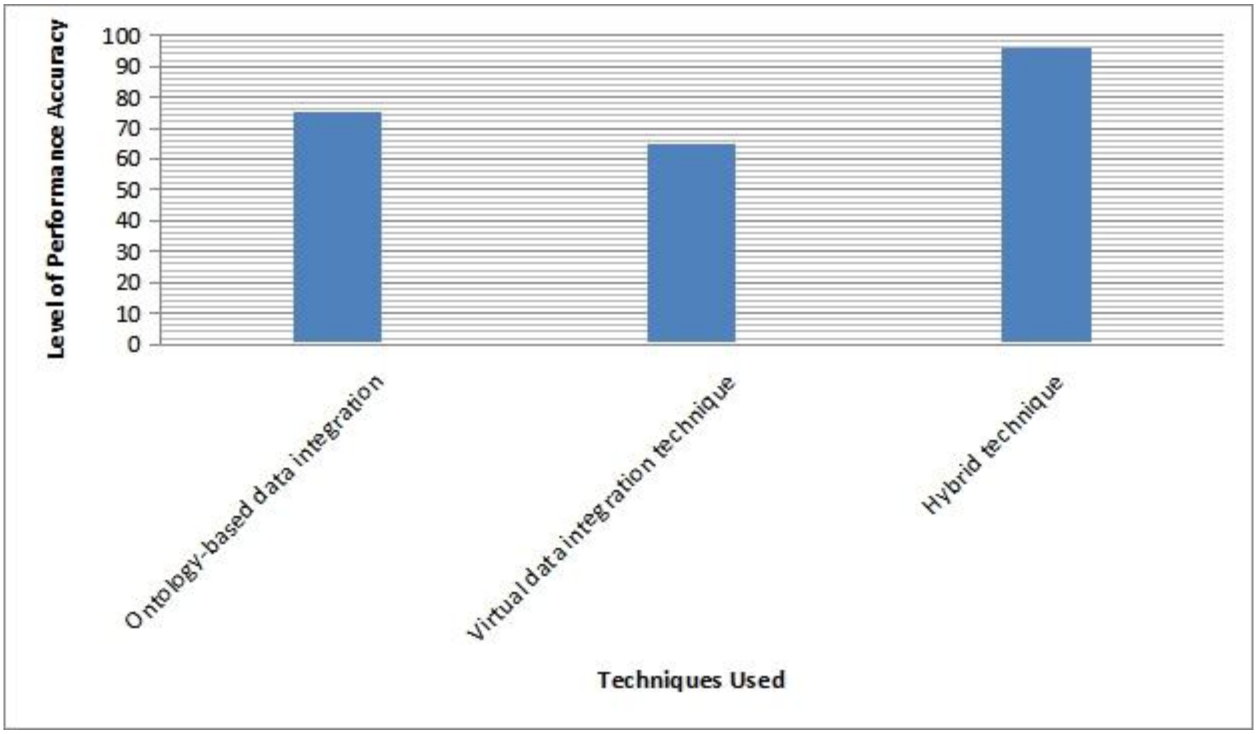

Figure 3: Comparison of level of prediction accuracy using various techniques

From figure 3, we can see that the Ontology-based data integration technique for disease control procedure has 75% accuracy in predicting the disease control procedure; Virtual data integration technique for disease control procedure has 65% accuracy in predicting the disease control procedure; while Hybrid technique using both Ontology-based data integration and virtual data integration technique for disease control procedure has 95% accuracy in predicting the disease control procedure.

Conclusion

In conclusion, virtualization of those federated sources at endpoint application means that end users are able to access data live from multiple databases and systems, aligned just the way they need it to make decisions or perform their specific duties. The idea of virtualization is that data never actually moves anywhere. This shows that the Hybrid technique outperforms the existing techniques with (95 – 75) 20%, i.e. there is 20% improvement from the existing technique.

Reference

Magali, R. & Michel, S., (2021). Virtualization in System Biology: Meta Model & Modelling Language for Semantic Data Integration" retrieved on May 29, 2021

Francesco, D. T., Ezio, L. & Filippo, T., (2021). Academic Data Warehouse Design Using Hybrid Methodology" Computer Science & Information System 12(1):135-160 DOI:10.2298/c815140325087D www.csisn3p135-160.pdf/ retrieved.

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

Richter, M.M., Weber, R.O. (2020). Case-based reasoning. Springer-Verlag Berlin

Rick, V. D. L., (2020). Data Virtualization for Business Intelligence Systems", www.r20.nl Retrieved from www.3-s2.0-B978...000010.pdf/ on Dec.7, 2020

Jagannathan, R., Petrovic, S. (2021). Dealing with missing values in a clinical case- based reasoning system. In international conference on computer science and information technology (pp. 120-4). IEEE.

Kapil, W., Mrudula , G. and Snehlata, D. (2020) . Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network", IEEE International Conference on Computer and Communication Technology (ICCCT),Vol.978,pp.741–745

Nassim, D., Elpiniki, I. P., Jos, D., Hans, C. and Marie-Christine, J., (2022) · Clinical Decision Support System based on Fuzzy Cognitive Maps, INSERM UMR_**S** 872, Eq 20, Medicine Faculty, Pierre and Marie Curie University, Paris 6, France

Ouwens, M. , Wollersheim, H. , Hermens, R. , Hulscher, M. , Grol, R. (2020). Integrated care programmes for chronically ill patients: a review of systematic reviews. Int J Qual Health-Care, 17 (2) (2020), pp. 141-146

Rosse, C. , Mejino, J.L.V. (2019). A reference ontology for biomedical informatics: the Foundational Model of Anatomy J Biomed Inform, 36 (2003), pp. 478-500 Article Download PDF View Record in Scopus Google Scholar

Asma, A., A. (2020). Decision Tree Discovery for Diagnosis of Type II Diabetes.", IEEE International Conference on Innovations in Information Technology (IIT), pp.303–307