International Journal of Engineering Research and Technology. ISSN 0974-3154, Volume 12, Number 6 (2019), pp. 745-752

Data Privacy Preserving Model for Health Information System

Ebiesuwa Seun
[1] *Computer Science Department, Babcock University, Nigeria.*

Adekunle Y.A
[2] *Computer Science Department, Babcock University, Nigeria.*

Omotosho, O.J
[3] *Computer Science Department, Babcock University, Nigeria.*

Adebayo, A.O
[4] *Computer Science Department, Babcock University, Nigeria.*

Tayo Omolara
[5] *Computer Science Department, Babcock University, Nigeria.*

Abstract

Information systems are employed by organizations for the collection, filtering, and processing of data, and creation and distribution of information. In healthcare delivery, patients are required to share information with certain categories of health personnel to facilitate correct diagnosis and to determine appropriate treatment. There have been cases of unauthorized access to and misuse of patient information by health personnel. Some of these personnel eventually cause great harm to the patient by divulging sensitive information. The existing Data Privacy Preservation (DPP) models are designed for Clinical Decision Support Systems with inadequate information available for DPP in Health Information Systems (HIS) in Nigeria. This research, therefore focused on the development of a model for DPP in HIS to address this inadequacy.

A model for DPP in HIS was developed using the iterative design technique. The model developed comprises a local database that contains the health information of patients, the Random Forest Decision Tree (RFDT) algorithm, an attribute blocking module that employs the RFDT algorithm, an attribute unblocking module which also uses the RFDT algorithm and a module for the computation of time elapsed in unblocking attributes. Mandatory Role-based Access Control was used to restrict the access health professionals have to patient data; each category of health worker can only view the attribute(s) needed for them to provide the service required to fulfill their role. An application based on the RFDT algorithm, was developed to instantiate the model following the Waterfall Software Development Life Cycle. Netbeans Integrated Development Environment, MySQL server, Java Development Kit 8, Scenebuilder 2.0, and Navicat 8 query editor constitute the programming environment. The application was evaluated against the machine learning approach to DPP that employed the classification technique, by comparing its efficiency with the Waikato Environment for Knowledge Analysis (WEKA) version 3.8 software in ensuring DPP using the RFDT algorithm.

The model developed in this study provides a generic framework for DPP in HIS that reveals the necessary components. This model provides a template that could be adapted for use in studies on DPP in HIS. The application provides the health personnel with Graphical User Interfaces that depict the professional's access to the patient database while restricting access to attributes not allowed for such category of health workers. The use of the RFDT algorithm in WEKA for DPP gave an efficiency of 73.77% while the approach that employed the application gave an efficiency of 78.32%.

The model presented in this study would help preserve sensitive patient data from being accessed by health workers who are not authorized to do so. The study showed that the application is more efficient than the WEKA software in ensuring DPP using the RFDT algorithm. The DPP model proposed in this study could also be employed in other domains outside the health sector to curb the challenges resulting from weak DPP.

I. INTRODUCTION

Health Information Systems (HIS) provide the bedrock for health-related decision-making and has four key functions: data generation, data compilation, data analysis and synthesis, and data communication and use. The HIS gathers data from the health sector and other relevant sectors, analyzes the data and ensures their overall relevance, quality, and timeliness, and converts data into information for health-related decision-making. In addition to being essential for monitoring and evaluation, the information system also provides early warning capability, supports patient and health facility management, facilitates planning, supports and stimulates research, permits health situation and trends analysis, supports global reporting, and underpins communication of health challenges to diverse users (WHO, 2009).

To improve the quality of medical care around the globe, efforts are being made to increase the practice of evidence-based medicine through the use of an HIS called Clinical Decision Support Systems (CDSSs). CDSS provides clinicians, patients, or caregivers with clinical knowledge and patient-specific information to help them reach decisions that enhance patient care (Osheroff, Teich & Middleton, 2011). The patient's information is matched to a clinical knowledge base, and patient-specific appraisals are then communicated effectively at appropriate times during patient care. Some CDSS include forms and templates for entering and documenting patient information, and alerts, reminders, and order sets for providing suggestions and other supports. The use of CDSS comes with many potential benefits. Importantly, CDSS can increase adherence to evidence-based medical knowledge and can reduce unnecessary variation in clinical practice. CDSS can also assist with information management to support the physicians' decision-making abilities, reduce their mental workload, and improve clinical workflows (Karsh et al., 2010). When well designed and implemented, CDSS have prospects that can improve health care quality, and also increase efficiency and reduce health care costs (Berner, 2010).

Despite the promise of CDSS, there are several barriers that can hinder their development and implementation. Medical knowledge base is essentially incomplete in part because of insufficient clinical evidence (Englander & Carraccio, 2014). Moreover, methodologies are still being designed to convert the knowledge base into computable codes, and interventions for conveying the knowledge to clinicians in a way they can easily use in practice are nascent. Low clinician demand for CDDS is another encumbrance to its broader adoption. Clinicians' lack of motivation to use CDSS appears to be related to its usability issues, its lack of integration into the clinical workflow, concerns about autonomy, and the legal and ethical implications of adhering to or overriding recommendations made by the CDSS (Berner, 2010). In addition, in many cases, acceptance and use of CDSS are hinged upon the adoption of electronic medical records (EMRs), because EMRs can include Clinical Decision Support applications as part of Computerized Provider Order Entry (CPOE) and electronic prescribing systems. There have been cases of unauthorized access to and misuse of patient information by health personnel (citations required). Some of these personnel eventually cause great harm to the patient by divulging sensitive information. The existing Data Privacy Preservation (DPP) models are designed for Clinical Decision Support Systems with inadequate information available for DPP in Health Information Systems (HIS) in Nigeria. This study, therefore, proposes a DPP model for HIS. In order to guarantee the secrecy of sensitive patient data domiciled in a HIS, the study involved the development of an application named Schizoapp which was used to instantiate the proposed DPP model and effected data privacy by blocking attributes on a patient database based on the Mandatory Role-Based Access Control (MAC) model that assigns access rights to health professionals based on their role in the hospital. The study also compared the use of the application (Schizoapp) developed in this study for data privacy preservation with the machine learning approach to data privacy preservation which employed the Random Forest Decision Tree algorithm embedded in the WEKA software.

II. RELATED WORK

II.I Machine Learning Approach to DPP

This study addressed the issue of privacy preserving data mining by considering a scenario in which two parties owning confidential databases wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information. The study employed the popular ID3 decision tree algorithm. The work was motivated by the need to both protect privileged information and enable the use of these databases for research or other purposes. Since the scenario used is a secure multi-party computation, the study posited that it can be solved using known generic protocols (Lindell, Pinkas, Smart & Yanai, 2015). However, data mining algorithms are typically complex and, furthermore, the input usually consists of massive data sets. The generic protocols in such a case are of no practical use and therefore more efficient protocols are required. Thus, the claim made in this study as regards the possibility of using generic protocols for the kind of scenario presented is doubtful. An algebraic technique-based scheme to privacy preserving data classification problem in a CDSS was introduced. The algebraic technique is a privacy intrusion technique that is capable of reconstructing the private data in a relative accurate manner. Compared to the randomization approach, the proposed new scheme can build classifiers with better accuracy but disclose less private information. The study also claimed that the proposed scheme is immune to privacy intrusion attacks (Sun, Wang, Shen, & Zhang, 2015). The study however failed to provide any substantial proof or empirical evidence to buttress its claim about the immunity of the proposed scheme to privacy intrusion attacks.

Fig 1 shows the Architecture of a Cloud-Based eHealth Model for Privacy Preserving Data Integration

II.II Data Privacy Preservation Models

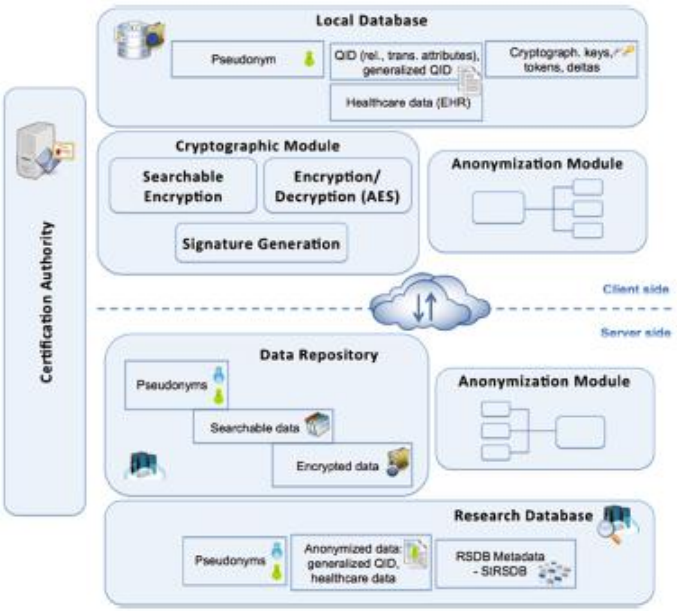

Figure 1 Architecture of a Cloud-Based eHealth Model for Privacy Preserving Data Integration
(Source: Dubovitskaya et al., 2015)

The drawbacks of this model are that:

The Cloud-Based eHealth Model employs only encryption to preserve data privacy. However, encryption alone is not a very effective way to preserve the privacy of data because once the private key of a user of the system is obtained by an unauthorized person, the privacy of data belonging to the legitimate user is in jeopardy.

The Cloud-Based eHealth Model does not include any module that evaluates the Cryptographic Module to give a measure of how good the encryption approach is in preserving data privacy.

Fig 2 depicts the Architecture of a Data Privacy Preservation Model for a Clinical Decision Support System

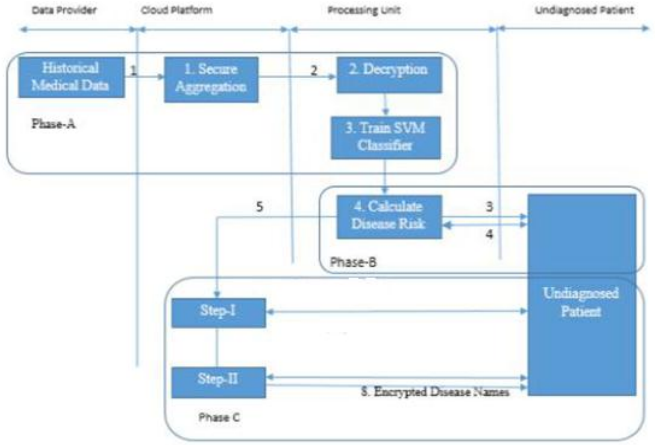

Figure 2 Architecture of a Data Privacy Preservation Model for a Clinical Decision Support System
(Source: Deshmukh, Tijare & Sawalkar, 2016)

The flaws of this model are that:

It attempts to preserve the privacy of patients' data by ensuring that both the symptoms sent by the undiagnosed patient to the Cloud platform and the diagnosis result sent back to the patient are encrypted. However, a patient's result can be accessed by an imposter if he/she obtains the private key of the patient. The model does not also include any module that evaluates the homomorphic encryption technique for its effectiveness in preserving data privacy.

Fig 3 displays the Architecture of a Privacy Preserving Data Classification Model.

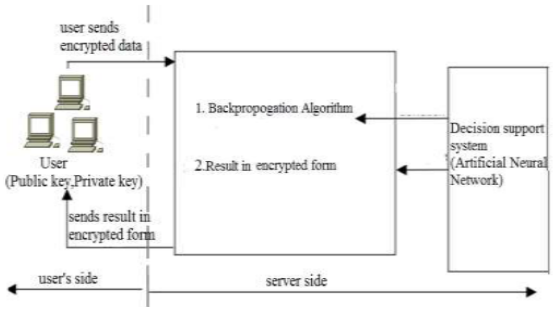

Figure 3 Architecture of a Privacy Preserving Data Classification Model (Source: Desale & Javheri, 2016)

The drawbacks of this model are that:

It attempts to preserve the privacy of patients' data by employing the pailler homomorphic encryption technique. However, the decision sent to a user can be accessed by an unauthorized individual if he/she obtains the private key of the user and can thus breach the privacy of the user's data. The model does not also include any module that evaluates the paillier cryptosystem to give a measure of how good the pailler homomorphic encryption approach is in preserving data privacy.

III. METHODOLOGY

III.I Study Data Set

The dataset for this study was collected from two Psychiatric hospitals in Nigeria. They are Federal Neuropsychiatric Hospital, Yaba in Lagos State and Neuropsychiatric Hospital, Aro, Abeokuta in Ogun State. Two hundred and sixty three (263) anonymous records of persons that have visited the hospital earlier on account of showing symptoms suggestive of schizophrenia were gotten from Federal Neuropsychiatric Hospital, Yaba while Two hundred and forty eight (248) records of persons that have visited the hospital earlier on account of showing symptoms suggestive of schizophrenia were obtained from Neuropsychiatric Hospital, Aro, Abeokuta. Thus, giving a total of five hundred and eleven (511) records for the study. The background information used in this study was extracted from the hospital records of persons who have at one time or the other visited the hospital to interact with a psychiatrist. The information extracted from the hospital records include the following variables: Age, Gender, State of Origin, Marital Status, Genotype, Bloodgroup, Display of Alogia, Show of Apathy, Third Person auditory Hallucination, Delusions of Control and Thought Echo, Insertion or Withdrawal.

III.I.I Research Methods

III.I.I.I Method of Proposed Model for HIS Data Privacy Preservation

Having studied and having identified the flaws of the three models discussed earlier, a model for implementing data privacy preserving in a CDSS was developed through augmentation and a fluid iterative cycle of awareness (recognition and articulation of the problem), suggestion (leap from curiosity to offering a very tentative idea for solving the problem), development (tentative idea is developed), evaluation (assessment of the model for its worth and deductions from expectations), and conclusion (Vaishnavi & Kuechler, 2004). The model consists of a local database that contains the health information of patients, the Random Forest Decision Tree (RFDT) algorithm, an attribute blocking module that employs the RFDT algorithm, an attribute unblocking module which also uses the RFDT algorithm and a module for the computation of time elapsed in unblocking attributes. A Mandatory Role-based Access Control was used to restrict the access health professionals have to patient data; each category

of health worker can only view the attribute(s) needed for them to provide the service required to fulfil their role.

III.I.I.II Method to Design and Develop a Prototype Application for Data Privacy Preservation

An application based on the RFDT algorithm, was developed to instantiate the model following the Waterfall Software Development Life Cycle. Netbeans Integrated Development Environment, MySQL server, Java Development Kit 8, Scenebuilder 2.0, and Navicat 8 query editor constitute the programming environment.

The dataset for this study was collected from two Psychiatric hospitals in Nigeria. They are Federal Neuropsychiatric Hospital, Yaba in Lagos State and Neuropsychiatric Hospital, Aro, Abeokuta in Ogun State. Two hundred and sixty three anonymous records of persons that have visited the hospital earlier on account of showing symptoms suggestive of schizophrenia were gotten from Federal Neuropsychiatric Hospital, Yaba while two hundred and forty eight records of persons that have visited the hospital earlier on account of showing symptoms suggestive of schizophrenia were obtained from Neuropsychiatric Hospital, Aro, Abeokuta. Thus, giving a total of five hundred and eleven records for the study. From the dataset for the study, three attributes out of the eleven indicate the likelihood of a patient being schizophrenic depending however on how many of these three attributes are exhibited by the patient in question. The three attributes are i) third person auditory hallucination ii) thought echo, insertion or withdrawal iii) Delusions of control.

Table 1 shows the various categories of healthcare professionals and their access levels to the system database.

Table 1 Healthcare Professionals and their access levels to the system database

Healthcare Professional	Access Level
Doctors	Level 1
Psychologist	Level 2
Nurse	Level 3
Social worker	Level 4

Table 2 displays the various categories of healthcare professionals and the attributes they are not allowed to view.

Table 2 Healthcare Professionals and the attributes they are not allowed to view

Healthcare Professional	Attribute blocked
Doctors	None of the three
Psychologist	Third person auditory hallucination
Nurse	Third person auditory hallucination and Thought Echo, Insertion or Withdrawal
Social worker	Third person auditory hallucination, Thought Echo, Insertion or Withdrawal and Delusions of Control

III.I.I.III Method of Evaluating of the Prototype

The machine learning approach to data privacy that involved the use of the WEKA version 3.8 software was compared with the application-based approach using the proposed data privacy preserving model.

Both approaches were evaluated based on the quantum of time taken to unblock the attributes that were previously blocked for each category of health workers. Hence, the better approach for data privacy preservation was the one which took a longer time for the blocked attributes to be unblocked given the same conditions for both approaches.

III.I.I.IV Machine Learning Model Design – Algorithms

This study employed three variants of the decision tree algorithm which are Random Forest algorithm, Random Tree algorithm and Decision Stump algorithm. These algorithms were trained with 58% of the data. The 10-fold cross validation was used to determine the accuracy level to measure the validity of the models generated.

IV. RESULTS

IV.I Model for the Preservation of Patient Data Privacy

The proposed model aims at preserving patient data privacy in a HIS that contains health records of persons who at one time or the other have been linked with one or more symptoms of schizophrenia.

Fig 4 displays the Clinical Decision Support System Model depicting internal modules and external Health Information System

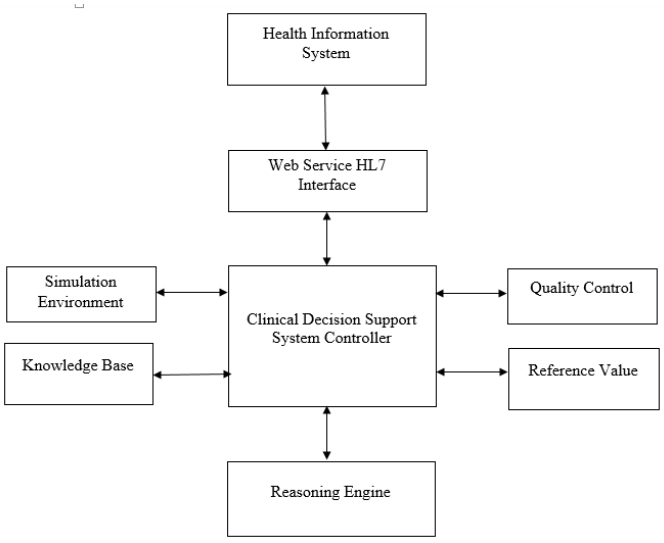

Figure 4 Clinical Decision Support System Model depicting internal modules and external Health Information System

Fig 5 shows the Proposed Model for Data Privacy Preservation in a Health Information System

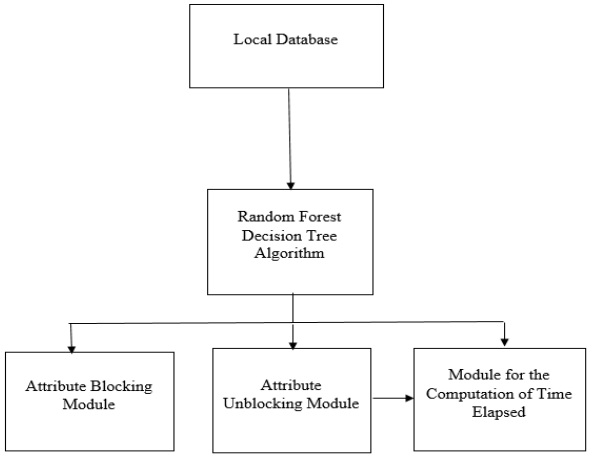

Figure 5 Proposed Model for Data Privacy Preservation in a Health Information System

Fig 6 shows the Clinical Decision Support System Model depicting internal modules and Data Privacy Preserving Section of external Health Information System

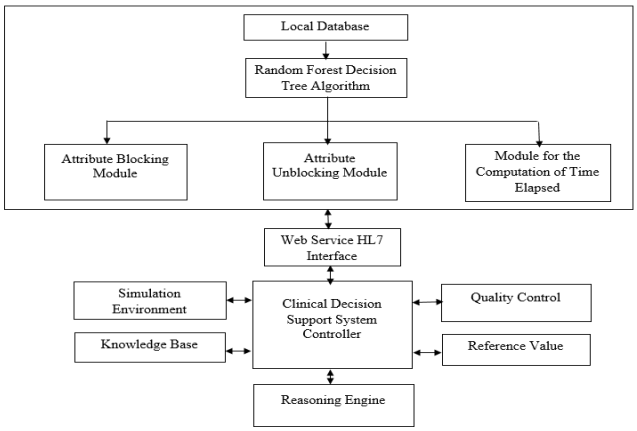

Figure 6 Clinical Decision Support System Model depicting internal modules and Data Privacy Preserving Section of external Health Information System

IV.I.I Prototype Application for Data Privacy Preservation

Fig 7 shows the Architectural Diagram for the Application developed for Data Privacy Preservation (Schizoapp)

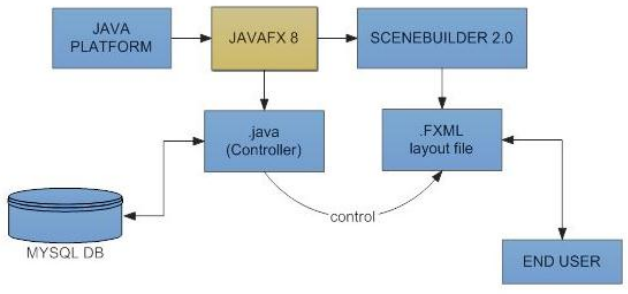

Figure 7 Architectural Diagram for Schizoapp

Fig 8 depicts the Flowchart for the Application developed for Data Privacy Preservation (Schizoapp)

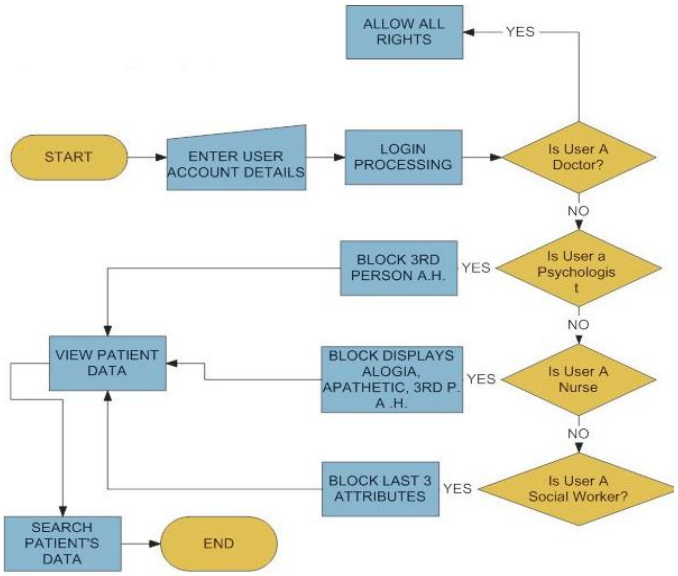

Figure 8 Flowchart for Schizoapp

IV.I.I.I Evaluation of the Prototype

i. Unblocking of Attributes across the Different Health Professional Categories Using the Application (Schizoapp)

The Administrator has access rights to view all attributes of patient data and can unblock attributes that are blocked across the different category of health professionals.

Fig 9 displays the Nurse's Page showing the Third Person Auditory Hallucination and Thought Echo, Insertion or Withdrawal attributes blocked

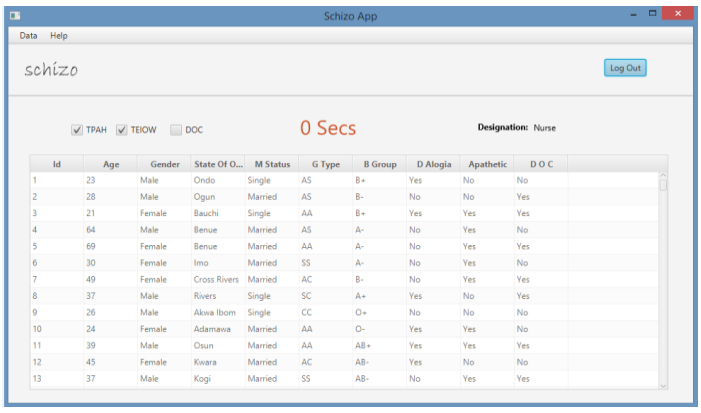

Figure 9 The Nurse's Page showing Third Person Auditory Hallucination and Thought Echo, Insertion or Withdrawal attributes blocked

Fig 10 shows the Nurse's Page showing the amount of time taken by the Administrator to unblock the Third Person Auditory Hallucination attribute

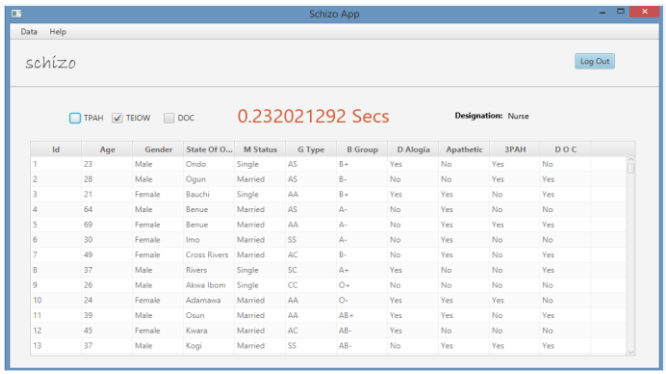

Figure 10 The Nurse's Page showing that Third Person Auditory Hallucination attribute has been unblocked by the Administrator within 0.232021292 seconds

Fig 11 depicts the Nurse's Page showing the amount of time taken by the Administrator to unblock that the Thought Echo, Insertion or Withdrawal attribute

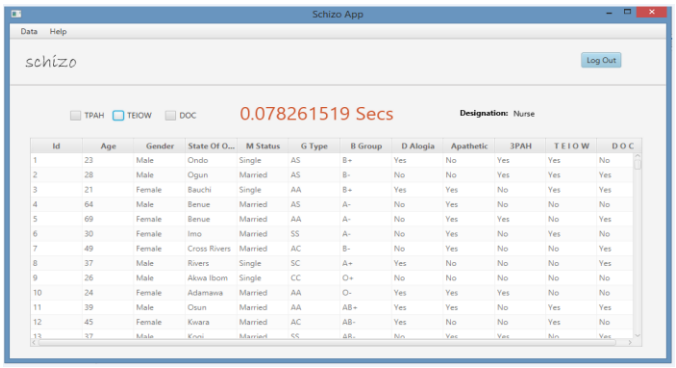

Figure 11 The Nurse's Page showing that the Thought Echo, Insertion or Withdrawal attribute has been unblocked by the Administrator within 0.078261519 seconds

Fig 12 displays the Social Worker's Page showing the Third Person Auditory Hallucination, Thought Echo, Insertion or Withdrawal and Delusions of Control attributes blocked

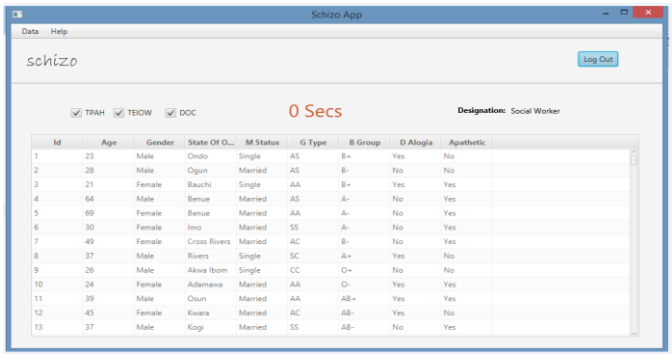

Figure 12 The Social Worker's Page showing Third Person Auditory Hallucination, Thought Echo, Insertion or Withdrawal and Delusions of Control attributes blocked

Fig 13 displays the Social Worker's Page showing the amount of time taken by the Administrator to unblock the Third Person Auditory Hallucination attribute

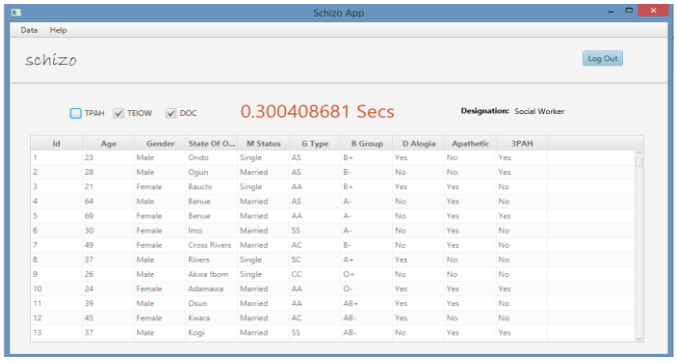

Figure 13 The Social Worker's Page showing that the Third Person Auditory Hallucination attribute has been unblocked by the Administrator within 0.300408681 seconds

Fig 14 displays the Social Worker's Page showing the amount of time taken by the Administrator to unblock the Thought Echo, Insertion or Withdrawal attribute

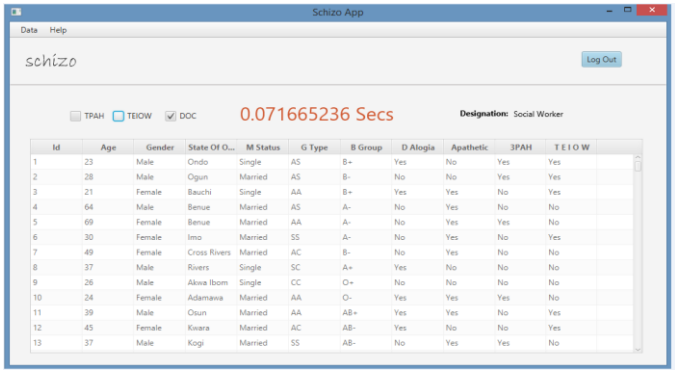

Figure 14 The Social Worker's Page showing that the Thought Echo, Insertion or Withdrawal attribute has been unblocked by the Administrator within 0.071665236 seconds

Fig 15 displays the Social Worker's Page showing the amount of time taken by the Administrator to unblock the Delusions of control attribute

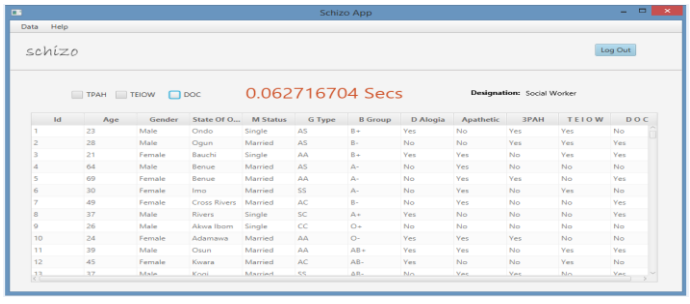

Figure 15 The Social Worker's Page showing that the Delusions of control attribute has been unblocked by the Administrator within 0.062716704 seconds

Fig 16 displays the Psychologist's Page showing the Third Person Auditory Hallucination attribute blocked

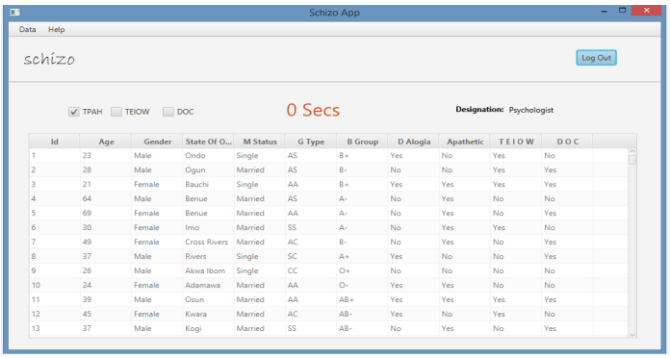

Figure 16. The Psychologist's Page showing that Third Person Auditory Hallucination, attribute is blocked

Fig 17 displays the Social Worker's Page showing the amount of time taken by the Administrator to unblock the Third Person Auditory Hallucination attribute

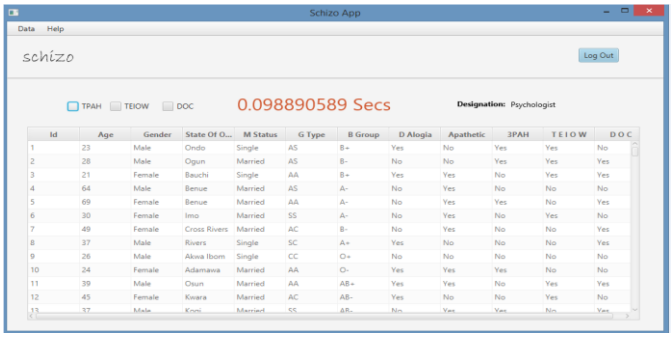

Figure 17 The Psychologist's Page showing that the Third Person Auditory Hallucination attribute has been unblocked by the Administrator within 0.098890589 seconds

Fig 18 displays the Administrators' Page showing the quantum of time taken to unblock attributes across each category of Healthcare Professional as well as the total time elapsed in unblocking all previously blocked attributes using both WEKA and the Application (Schizoapp)

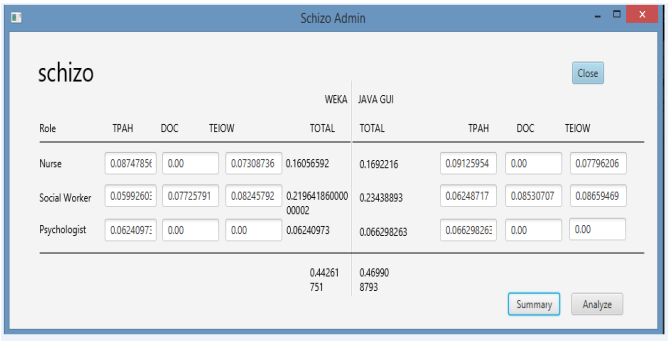

Figure 18 Administrators' Page showing the time elapsed in unblocking attributes across each category of Healthcare Professional as well as the total time elapsed in unblocking all previously blocked attributes using both WEKA and the Application (Schizoapp)

Fig 19 depicts the Administrators' Page showing the efficiency of the WEKA approach and the Application approach to data Privacy Preserving

Figure 19 Administrators' Page showing the efficiency of the WEKA approach and the Application approach to data Privacy Preserving with information on the more accurate one after testing both approaches with the same data.

The efficiency of each approach to data privacy preserving was computed using the formula:

Efficiency = (Total Time Elapsed in unblocking attributes)/(Total Time Allowed for unblocking attributes) X 100%

V. CONCLUSION

V.I Findings from Comparison Between the WEKA Approach and the Application approach to data Privacy Preserving

The efficiency of each of the two approaches was computed as a function of the time elapsed in unblocking the attributes. Thus, a conclusion was reached as to the more efficient of the two approaches to data privacy preserving. The WEKA approach gave an efficiency of 73.77% while the Application (Schizoapp) yielded an efficiency of 78.32%. Hence, the Application (Schizoapp) approach to data privacy preserving proved to more efficient than the WEKA approach.

V.I.I Recommendations

This research has brought to bear the need for digitization of health records particularly in developing countries like Nigeria. The study therefore recommends that the Federal Ministry of Health in Nigeria enforces the adoption of Electronic Health Records and HIS in Nigerian hospitals to facilitate better DPP of patients' data and in the long run improve the health status of Nigerians that will in turn lead to a betterment of the Nigerian health care sector. This study also recommends a concerted effort on the part of the administration of the Healthcare parastatals at the various levels in Nigeria to ensure that healthcare personnel are adequately trained on the use of the HIS technology and subsequently monitored to ascertain that they use the technology. The need for patient data privacy preserving will then be very expedient. Thus, applications such as the Schizoapp developed in this study can be employed to preserve the privacy of healthcare patients and this will go a long way in improving the quality of healthcare delivery which the Nigerian health sector provides to its citizenry.

V.I.I.I Suggestions for Further Studies

Future studies can employ the data privacy preserving model presented in this study to effect data privacy using a larger dataset in terms of both the records and the attributes to further extend the use of the model. Future studies can also employ the model in other domains outside the health sector to help preserve the privacy of persons involved in such study.

In many related literature, the decision tree algorithm showed more superiority over most other classification algorithms in the WEKA software. A future research can also employ other algorithms other than the decision tree algorithm to implement data privacy preserving using machine learning.

V.I.I.II Contribution to Knowledge

This research contributes to knowledge by providing a viable alternative to the machine learning approach to data privacy preservation. The application built provides interested parties with the opportunity to have access to the source codes for the application through which they can modify the application viz-a-viz their data privacy need. This becomes worthy of mentioning since the WEKA software has its source codes inaccessible to prospective users but limit users only to the functionalities available within the software by the developer's design thus making it almost impossible for a user to modify the software to meet evolving needs with respect to the specification of the user. The study goes a step further to evaluate the application developed by juxtaposing it to the WEKA software in terms of its efficiency measured in percentage. The study also provided a model for DPP that can be adopted by other researchers to tackle the issue of data privacy with respect to their peculiar need.

REFERENCES

[1] E. Berner, Ethical and Legal Issues in the use of clinical decision support systems, *J Healthc Inf Manag, 4(2)*, 2010, 34-37

[2] M. S. B. Desale, and Javheri, S. B, Implementation of Privacy Preserving Data Classification System Using Machine Learning, *International Journal of Engineering Development and Research, 4(3)*, 2016, 397-403

[3] M. M. V. Deshmukh, P. A. Tijare, and S. N. Sawalkar, A Survey on Privacy Preserving Data Mining Techniques for Clinical Decision Support System, *International Research Journal of Engineering and Technology, 3(5)*, 2016, 2064-2069

[4] A. Dubovitskaya, V. Urovi, M. Vasirani, K. Aberer, and M. I. Schumacher, A cloud-based eHealth architecture for privacy preserving data integration, in IFIP International Information Security Conference (Hamburg: Springer International Publishing, 2015) 585-598.

[5] R. Englander, and C. Carraccio, Domain of competence: medical knowledge. *Academic pediatrics, 14(2S)*, 2014, S36-S37

[6] B. T. Karsh, S. Guerlain, D. Metcalf, J. L. Marquard, E. H. Lazzara, S. J. Weaver, and P. Gorman, Just What the Doctor Ordered?: The Role of Cognitive Decision Support Systems in Clinical Decision-Making & Patient Safety, in Proceedings of the Human Factors and Ergonomics Society Annual Meeting (SAGE Publications, 2010) 826-829.

[7] Y. Lindell, B. Pinkas, N. P. Smart, and A. Yanai, Efficient constant round multi-party computation combining BMR and SPDZ, in Annual Cryptology Conference (California: Springer Berlin Heidelberg, 2015) 319-328

[8] J.A. Osheroff, J.M. Teich, and B.F. Middleton, A roadmap for national action on clinical decision support, *American Medical Informatics Association, 5(3)*, 2011, 217-237

[9] Y. Sun, N. Wang, X. L. Shen, and J. X. Zhang, Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behaviour, 6(2)*, 2015, 278-292

[10] V. Vaishnavi, and W. Kuechler, Design research in information systems. Retrieved from: www.isworld.org/researchdesign/drisISworld.htm. Accessed 15 June 2011.

[11] World Health Organization. Toolkit on monitoring health systems strengthening. Medical products, vaccines and technologies: a toolkit for countries (Geneve: WHO, 2009)