Semi Supervised Heterogeneous Graph on Health – Based Classification of Medical Records

Rishin Roy, Abhisek Swain, Mrs. MrudulaOruganti

ABSTRACT-Health is an essential part of life, without it a person can't do his/her work properly and also can't pay heed to the other essential responsibilities in life. For a patient to be healthy, there has to be some kind of universal medium where his/her health related issues all over the years will be stated and a system is needed for conjoining all of them. That helps the patient in using any sort of medical institution to get themselves checked without starting the whole check-up process again. If the patient loses the prescriptions by any means and is not able to access it, then the records stored in the database can be used to get the prescriptions back. This eases the efforts of both the doctor as well as the patient to work on the problem easily.

INDEX TERMS – K-anonymity, Medical Records, Semi-Supervised methodology, Secure Mining

— — — — — — — —◆— — — — — — — — — —

I. INTRODUCTION

The concept of accessing health records of patients which are being stored from years and years together, help the doctors to get a track record of the patient's immune system, and the kind of attention he/she needs. Health records also give the patients, a report of all the diseases they had all these years as well as the present records. But besides all of these, our concentration is mainly on storing the data and keeping it safe from the attackers who might try and get access to the system and manipulate the data or steal important information about any patient. We are going to mine the health records in such a way that the data stored is encrypted and is only accessible to the authentic personnel. For e.g. the patient who is opening an account and gives his/her personal details i.e. the username and password in addition to the mobile number and age, is the only one who can get complete access to the information, the doctor however can view only the patient's name and the disease he/she is having.

The admin is another person who can view the data entered by the patient, but still has restrictions of his own, as he cannot view the patient login credentials and the age and zip code which is justified as privacy information, can use the system's classification methodology to view the patient records in different ways, i.e. the original data records, the generalized data records, the bucketized way, or in a distributive manner.

The original records contain all the information except the patient's password details. The generalized records contain all the information, except the private informations such as the age, zip code and the year of birth are hidden from the admin. The bucketized records are classified based on the range of ages of the patients and are organized as males first and then female patients. The distributive records are classified on the basis of age

groups, where same ages are classified one after the other in the order of Males and then Females, and if not, then are arranged one after the other in ascending order.

The patients who are new to the database and have never registered themselves need to register on the portal which includes a desired username, password along with the disease of the patient, mobile number/phone number of 10 digits, date of birth, age etc.

Once registered on the portal, the patient can log back in to view the record which he/she has stored and in future after inspection can check the database again for viewing the records named after him/her.

Same goes for the doctor's part too, where in case the doctor hasn't registered themselves, have to enter their details first, which includes their credentials, their designation, distinction (degree withheld), age, date of birth, mobile number, address etc.

Once the details have been entered, the doctor is ready to use the system by logging in to his/her portal and then can view the patient records based on the type of specialization he/she has done. This will classify all the patients classified under that particular kind of disease.

The patient, doctor and the admin, all of them have also got one more option, which is of changing the password, which adds up to the security of the system from attackers who might have got the consent of entering into the system.

_____

- *Rishin Roy is currently pursuing Bachelor's degree in Computer Science and Engineering from SRM University, Chennai, India,PH-+917092058285. E-mail: rishinroy124@gmail.com*
- *Abhisek Swain is currently pursuing Bachelor's degree in Computer Science and Engineering from SRM University, Chennai, India, PH-+919940389871. E-mail: abhikar400@gmail.com*
- *Mrs. MrudulaOruganti is our guide and is teaching at SRM University, Chennai as Assistant Professor (M.Tech), PH-+919500039386. E-mail: mrudula.sgiet@gmail.com*

IJSER

II.   LITERATURE SURVEY

The algorithms used in this system are association rule mining and k-anonymity which are comprised under one single algorithm, which is named as Semi Supervised Heterogeneous graph on Health (SHG-Health) algorithm.

The association rule mining is a data mining technique used in situations involving a huge amount of tabular represented data, which on a certain condition needs to be classified together into a single table, this makes that classification easier to get implemented. It basically associates the data taken from different tuples together as per the condition mentioned as once it is met, the associativity is made and the records are automatically organised into the new table.

This algorithm is used during the generalization, bucketization and distributive representation of data.

The k-anonymity algorithm plays the important part of encrypting and securing the data from any third party authority or an attacker of some kind who wants to get access to any sensitive information about any patient or the doctor. Encryption takes place as soon as there is any record being viewed or written on to the database.

III.   ANALYSIS OF REFERENCE

According to the previous works being taken under this domain, they have been very interesting indeed, working straight towards the health conditions of the patients. Predicting the health conditions have been their main concern and the accuracy levels have been outstanding with a level of 91.33%.

But their concern was completely on the backend of the purpose, not on the foreground operation, i.e. ensuring security of the data while making transactions of the health records.

Our aim is to:-

- Ensure the simplicity of data access and enrollment.
- Ensure the encryption of sensitive data while accessing or enrolling.
- Ensuring security measures to the best possible way.

The previous system lacked the usage of these characteristics and we are implementing these into our system. This step will make the system more secure thanever, because a simple attack on the database can turn into a huge massacrefor others at the same time because they can be accessed too.

IV.   ARCHITECTURE DIAGRAM
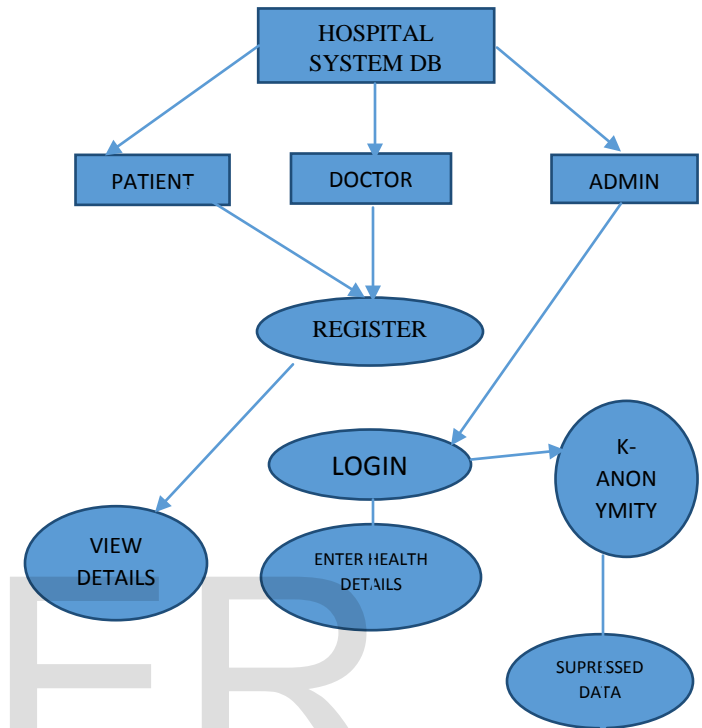
The system architecture consists of three parts, the patient, the doctor and the admin, these are the three types of users who are going to get to use this system, but everyone will have a different instance of access to the system database.

The patient has to first register onto the system and create an account, once done with that, then he/she can login to the system and can update or view the data.

The doctor also has the same procedure to deal with, at first getting registered onto the system, then moving onto to the login screen. But the doctor can view and update the patient health data as well.

The admin however has the only authority of viewing the details of the patient and the doctor, but a certain amount of data, the rest is all worked by the k-anonymity algorithm as is encrypted, even the disease data too.

V.   DATA FLOW DIAGRAM

The flow of data occurs differently for the patient, the doctor and the admin, because each is having a different perspective of accessing the data.

For the patient to access the data, first the registration procedure comes up, once done with that, the next step is to do with the login process, and then the patient can either choose to log out of the system, change the existing password or to view the data enrolled by him/her.

For the doctor, the same procedure as that of the patient occurs, first being the registration process if the doctor is not registered, and if yes, then he/she can simply proceed to the login window. Once logged in to the system database, he/she can view the patient records based on the query made and can also come up with updating the patient disease record . The doctors too can change their password.

For the admin, there is a default username and password set, by which he/she can enter into the system and then too can change the password, but the admin however can only view the data and is limited by the k-anonymity algorithm which helps secure the sensitive information about the doctors and the patients.
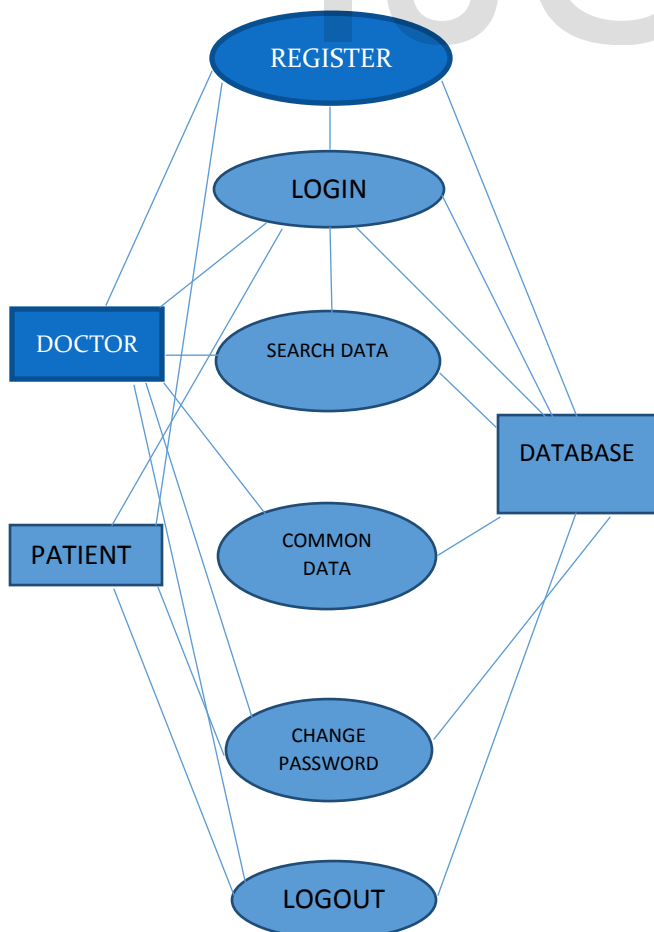
VI.   MODULE DESCRIPTION

1. ADMIN MODULE:

The admin module has two parts, the login part and the home, from where the admin can see all the classified information about the patient and the doctors.

The admin can view the original data, or the generalized data, the bucketized (age customised) data, or the distributive data, but with an exception that the admin cannot seek information completely as it will violate the security norms of the system and thus the algorithm hides sensitive patient and doctor information from the admin by using k-anonymity algorithm.
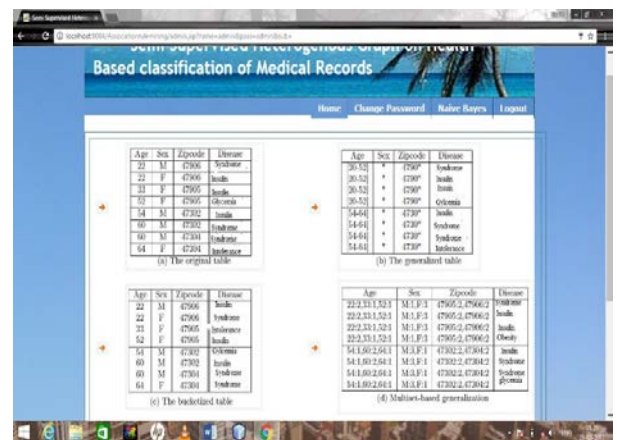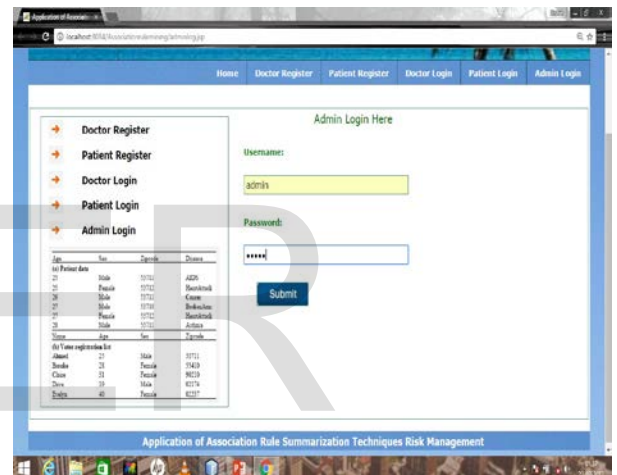
This is the admin home window from where the admin has access to the classified data of the k-anonymity and association rule mined medical records of the patient as well as doctor's information.

The admin can access the original table, or the generalized table or the bucketized table or the distributive table.
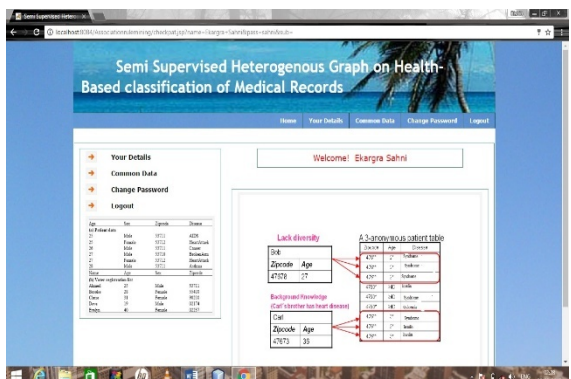
2. PATIENT MODULE

The patient module consists of the registration part and the login part, and can be used to input data into the database.

The patient also can change the password and can alter his/her information from time to time.

At first the patient must be registered into the database which is done from the registration page, once done with the registration procedure, the patient can choose to close the system or can simply login to the profile to view his/her records.
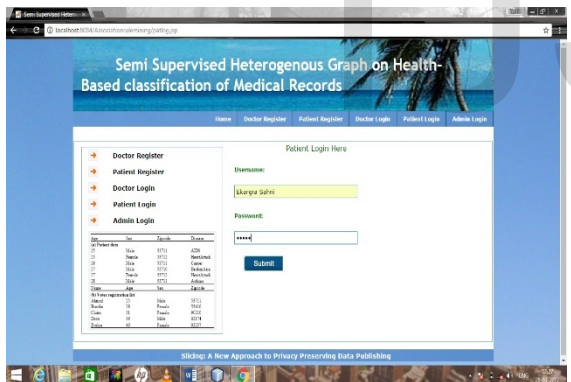
The patient logs in to the system from this section of the software.
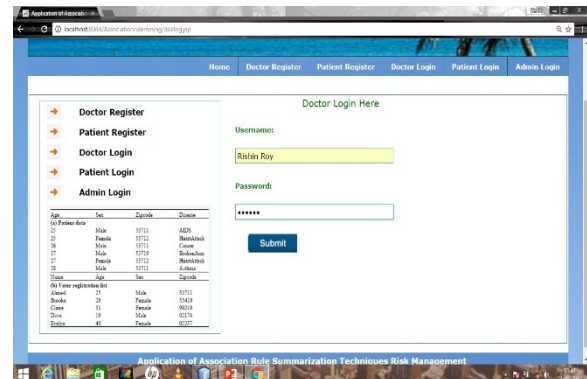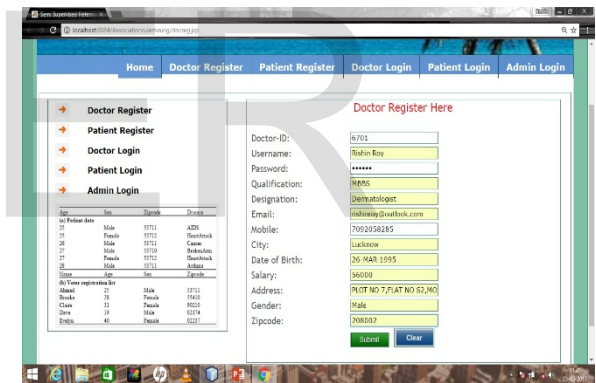

This is the patient's console from where the patient can view the entered record, can view the common records, can change the existing password with a new one, and once done with all of them can finally log out of the system console.

3. DOCTOR MODULE

This module is for the doctors to view the patient's information which can be helpful for diagnosing the patient and if any changes are required in the patient report, the doctor can update the report too, but can't get access to the other irrelevant details which is not supposedly for the doctor to get access to, such as mobile number, age, date of birth and zip code of the patient.

The image shown below depicts the doctor's profile registration screen wherein the doctor fills in his/her information which is only accessible by the doctor and no one else can view the sensitive/private information except the doctor's name and contact number.

This registration process are for those doctors who are not registered from before and are new to the system.

Once the registration process is done, the doctor can move on to the login window to view the patient records based on the query given by him/her, and can also change the profile password anytime.
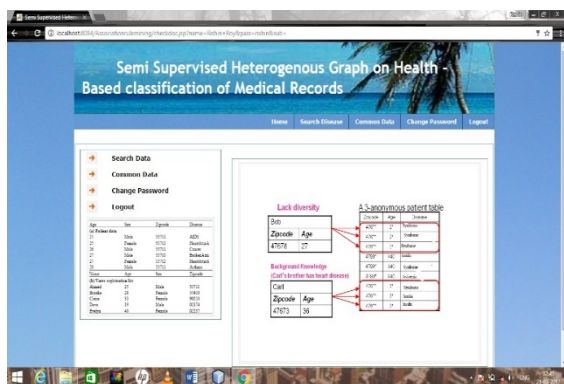
This is the doctor's profile window from where the doctor can get access to the patient's record, can view the common patient records at the same time for faster review, change the password to the profile, and once done with all the work, can log off from the system console.

VII.   CONCLUSION

The security of any kind of data, let it be a stock market report, a electricity bill or the medical record, each of them requires a privacy concern, which turns the whole table onto the point of security, this system mainly focuses on the security benefits of the management system rather than going to the backend and predicting the health condition of the patient based on the similar characteristics of other patients.

VIII.   ACKNOWLEDGMENT

IX.   REFERENCES

[1]   M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Extraction of interpretable multivariate patterns for early diagnostics," IEEE International Conference on Data Mining, pp. 201–210, 2013.

[2]   T. Tran, D. Phung, W. Luo, and S. Venkatesh, "Stabilized sparse ordinal regression for medical risk stratification," Knowledge and Information Systems, pp. 1–28, Mar. 2014.

[3]   M. S. Mohktar, S. J. Redmond, N. C. Antoniades, P. D. Rochford,

[4]   Artificial Intelligence in Medicine, vol. 63, no. 1, pp. 51–59, 2015.

[5]   J. M. Wei, S. Q. Wang, and X. J. Yuan, "Ensemble rough hy-percuboid approach for classifying cancers," IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 3, pp. 381–391, 2010.

[6]   E. Kontio, A. Airola, T. Pahikkala, H. Lundgren-Laine, K. Junttila,

[7]   Q. Nguyen, H. Valizadegan, and M. Hauskrecht, "Learning classi-fication models with soft-label information." Journal of the Ameri-can Medical Informatics Association : JAMIA, vol. 21, no. 3, pp. 501–8, 2014.

[8]   L. Chen, X. Li, S. Wang, H.-Y. Hu, N. Huang, Q. Z. Sheng, and

[9]   C. Y. Wu, Y. C. Chou, N. Huang, Y. J. Chou, H. Y. Hu, and

[10] P. Li, "Cognitive impairment assessed at annual geriatric health examinations predicts mortality among the elderly," Preventive Medicine, vol. 67, pp. 28–34, 2014.