An Optimized Classification Mechanism to Predict Cloud Based Medical Data

Mr. Peddikuppa Siva1, Dr.L.K.Induamthi[2]

[1]Assistant Professor, Department of Computer Science & Engineering, Matrusri Engineering College, Telangana, India.
[2]Associate Professor, Department of Computer Science & Engineering, Matrusri Engineering College, Telangana, India.

Abstract-In India, the average medical expenses for rural patients have grown over 160% and for the urban patient have grown over 175%. Around 85% from rural and 82 % from urban population do not have sufficient health care support. That is why are evolutionary measure is very much needed in medicine than any other field. The existing system involves the physical movement of the patient to the hospital or diagnostic firm to provide the necessary samples. This involves a lot of primary disadvantages like–Travelling expenses, Stress met by the patients, Disruption in the day to day activities of the patient's care-takers. Additional disadvantage is multiple encryptions used for multiple classes of users (with different privileges) which occupies a major space in the cloud (space complexity). Another impediment is the absence of CDA-Clinical Document Architecture in cloud. The system proposes an architecture which overcomes the above mentioned disadvantages by collecting the data remotely from the user and updating the data in the cloud i.e. Real-time data collecting. Advantage of this system is both the doctor and patient can view patient's records as and when required irrespective of climatic changes or geographical location. We also use the sensors to determine the health conditions, viz heart beat, systolic pressure, blood glucose levels etc. The collected data from the sensors are stored in the form of a tree and encrypted by Attribute based Encryption.

***Keywords*:** Attribute based encryption, K-means clustering, Fuzzy logic, R Studio programming, CDA-Clinical Document Architecture, Cloud based medical profile.

I. Introduction

The importance of health in human life has never been a subject of interrogation point. No wonder how much ever the technology and science develops, the arrival of chronic patients continually keeps increasing. Except for a sedentary life style that comes with technology, and the quick paced life, the intake of genetically modified crop varieties has seriously meddled with our systems. The world Health Organization (WHO) report throws focus on a parameter referred to as Disability-Adjusted Life Years (DALY's). The disability for adjusted life suggests the amount of years of his or her lifespan an individual would lose due to the disability caused by any chronic disorder. Apart from maternity and malnutrition, the most damaging health problem is diabetes and cardio vascular disorder. This paper focuses on the issues featured by such patients and helps in the automating the clinical document generation. This paper conjointly addresses the

problem of space complexity for cloud storage. Since the cloud provides numerous services, the user is anticipated to pay for such services. The more he space occupied, the more the charge. The paper additionally focuses on providing an alternative encryption option to improve the space occupied. The complete details and also the attributes are stored as a single tree form at when an encryption (Attribute based encryption).Then relying upon the class of users, the attributes are fetched and decrypted for analysis.

II. Proposed Method

The System and circuit description are as follows,

A. SYSTEM DESCRIPTION

The real time data's of the patients are collected from the sensors. It may be varied such as heart beat, blood pressure or glucose levels. The data is dumped in to the system from the sensors in the form of time stamps. The time stamped data is sampled in to the local ware house. The time stamped data is then encrypted using Attribute Based Encryption algorithm and promoted to the cloud. Another instance of the same time stamped data is queried out of the database and transformed into Excel by an API. The excel document is then forwarded in to the R-studio, where the data is clustered into suitable clusters having similar characteristics using K-means clustering. The clustered data is now classified according to the different fuzzy conditions available. The proposed system collects the data from the users remotely and uplinks it to the access regulated storage infrastructure

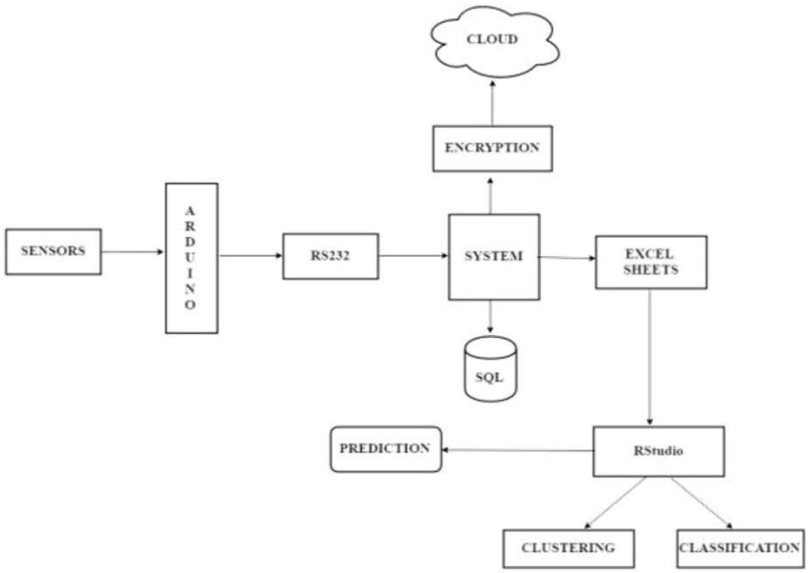

Fig-1: Overall architecture of the proposed system.

B.CIRCUIT DESCRIPTION

1.Sensor oriented data collection

Sensors are electrical devices that collect information. The heart rate sensor is employed for observing the heart rate of the patient. This sensor takes heart beat as an Analog signal and sends it to the database. It senses the real time pulse rate of the observance body. For monitoring the collected data sets, it sends to the management unit to indicate the results. The sensor which is used is connected with a micro-controller named as Arduino. All the data goes from the sensor to the data base through its wireless local area network device via the Arduino. Arduino is an open-source electronics platform which is based on easy-to-use hardware and software. Arduino boards are able to read inputs in the form of light on a sensor, a finger on a button, or a Twitter message etc., and turn it into an output over an activating a motor, turning on a LED, publishing something online. The Arduino components are classified in to two parts viz hardware (pin configuration) and software [1].

2.Components for Data collection and management–The flow and flex of data

Sensor data collected from the observance body through sensor and store and retrieve data from the database. Management unit process the data from and show the result to the user as time stamps. The system should collect an oversized variety of data from the observance body for a period of time to process the data. Currently the system shows the result in a monitor [2].

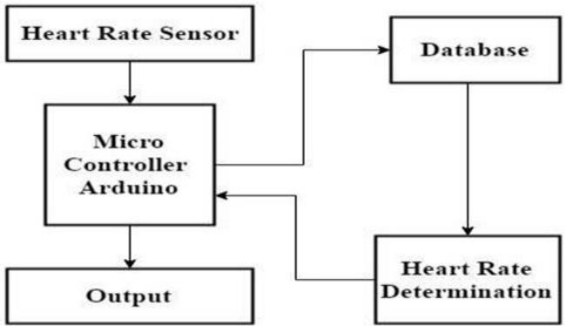

Fig-2:Sensor based and management unit

3. Circuit Diagram

Circuit diagram shows that how the Arduino board is connected with the heart rate sensor. Arduino board has some digital and analog pin. To get the output from the sensor the output pin of the heart rate sensor is connected with the digital pin 2 of the Arduino.The power supply and ground pin is connected with the vcc and gnd pin of the Arduino [1].

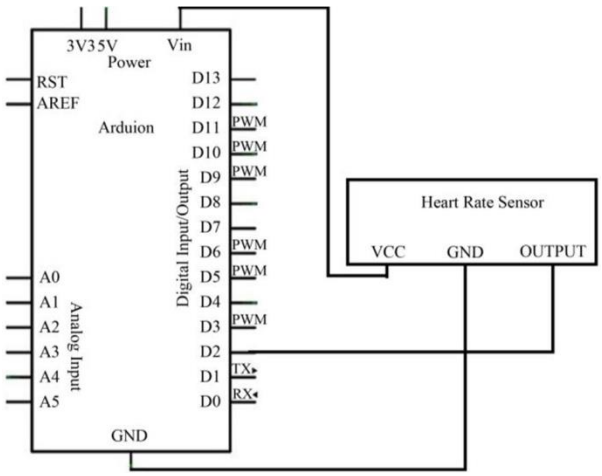

Fig-3: Circuit diagram of Arduino connected to Heart Rate sensor

III. Modules of the Proposed Work

The processed data from RS-232 are sent in to the system and performs the following activities:

A. Local Ware house

The local ware housing is done by My SQL. My SQL is an open source which is available freely and it is a RDBMS which uses SQL queries. One of the important purposes of My SQL is to do Web Database. The main feature of it is to cross platform and store large number of records can be stored. Other applications of My SQL are e-commerce, data ware housing etc. The network connectivity is not always reliable. The presences of no data stored in the local level possess a fatal threat to the efficient functioning of the system. So, the system uses local warehouses where the time stamped reading from the heartbeat sensors reside during the down time of the network.

B. Attribute Based Encryption

The patient's data are encrypted using Attribute based Encryption algorithm. Attribute-based-encryption (ABE) is as public-key cryptography technique in which encryption key and cipher text depends upon attributes. Messages will be encrypted with regard to subsets of attributes or policies outlined over a given set of attributes. The storage is done based on access Trees using n-to- n threshold gates and 1-to- n threshold gates [3]. The user's keys and cipher texts are labeled with a collection of descriptive attributes and a particular key will decipher

Particular cipher text provided that there is a match between the attributes of the cipher text and also the user's key.

C. Cloud Storage and Remote Regulated Access

After encryption is been done the patient's data is being stored in public cloud namely Drop box. Cloud storage and synchronization services let users access their digital content any time, from anywhere, and with any device—Smartphone, tablet, or desktopPC.Manycloudserviceproviders(CSPs)offeralargeamountofspacefor consumer use. Here we focus on Drop box (www.dropbox.com) and other examples are Google Drive (drive.Google.com) and Microsoft Sky Drive (SkyDrive.Live.com). Typically, a user who signs up for a cloud storage service can get a few gigabytes for free or a hundred gigabytes for just several dollars per month. Users can access their data via various interfaces, such as a standard software client, a Web browser, or a Smartphone application. [4]

1.Clinical Document Architecture

The proposed work describe CDA document generation and integration service based on cloud computing, through which hospitals are enabled to conveniently generate CDA documents without having to purchase proprietary software. Physicians and patients can browse the clinical data in chronological order and historical wise. Initially, the Clinical Document Architecture was developed by Health Level Seven. It is an XML based document that is used for the exchange of clinical document. Previously known as the Patient Record Architecture (PRA), was the chronological collection of the health records of the patient. Because of the XML feature, CDA can be processed manually as well as mechanically by electronic processing such as EMR. The latter method involves the decomposition of the document. The structuring facilitates electronic processing. The CDA documents can be transferred using the existent communication protocol i.e., No special protocol is required. The two portions are:

1) Header, which contains meta-data such as the date of creation of the document or the patient information.
2) Body, which consists of a wide variety of data, may be textual or graphical which is encrypted.[5]

Hence once the patient details are integrated into CDA document, the doctor or medical personal can review the patient's clinical history in sequential order per clinical section.

IV Implementation of the proposed work

The real time data are processed from Net beans as an excel file to R programming studio for data analysis. In R programming clustering and classification are processed for normal or abnormal conditions.

A. Clustering

1. k-Means Clustering

K-Means is a clustering approach that belongs to the category of unsupervised statistical learning methods. The general idea of a clustering algorithm is to partition a given dataset into distinct, exclusive clusters so that the data points in each group are

quite similar to each other.

One of the primary steps in building a K-Means clustering work is to define the number of clusters to use. Subsequently, the algorithm assigns each specific data point to one of the clusters in an exceeding fashion. The most common way to describe this variation is using the **squared Euclidean distance** [6]. This method of categorizing groups of similar data points can be a comparatively complex task since there is excess number of ways to partition data points into clusters.

$$J(Z) = \sum_{i=1}^{a} \sum_{j=1}^{a_i} (\|y_i - z_j\|)^2$$

'$||y_i – z_j||$' is the Euclidean distance between y_i and z_j. 'a_i' is the number of data points in i^{th} cluster. 'a' is the number of cluster centre's.

2. Algorithmic Steps for k-Means Clustering

Let $Y=\{y_1,y2,y3,……..,y_n\}$ be the set of data points and $Z=\{z_1,z_2,…….,z_c\}$ be the set of centre's.

1) Randomly select 'a' cluster centres.
2) Calculate the distance between each data point and cluster centres.
3) Assign the data point to the cluster centre whose distance from the cluster centre is Minimum of all the cluster centres.
4) Recalculate the new cluster centre's using: Where, 'a_i' represents the number of data points in i^{th} cluster.
5) Recalculate the distance between each data point and new obtained cluster centre's.
6) If no data point was reassigned then stop, otherwise repeat from step3).

B. Classification

1. Fuzzy Logic

The Fuzzy Identity-Based Encryption (FIBE) concept was presented by Sahai and Waters who also presented the concept of Attribute-Based Encryption. This scheme builds upon many concepts from Identity Based encryption. In FIBE, an identity is viewed as a group of attributes. FIBE permits for a private key for an identity,μ, to decrypt to a cipher text encrypted with an identity, μ, if and only if the identities ω and ω are near to each other as measured by the "set overlap" distance metric. In alternative words, if the message is encrypted with a collection of attributes μ, a private key for a collection of attributes ω permits decrypting that message, if and only if $|\mu \cap \mu| \geq d$, where d is mounted throughout the setup time[7]. Here the Fuzzy Logic Systems are used to produce definite and acceptable output with respect to in complete or distorted (fuzzy) input.

The approach of Fuzzy Logic intimates all the possible intermediate values between YES or NO as its output. Let us consider an example of heart rates of different persons and age groups of various categories for10 sec as follows:

Table–I, Heart Rates for Women

	A	B	C	D	E
1	AGE	VERY FIT	FIT	AVERAGE	UNIT
2					
3	30-39	<78	78-99	100-109	>109
4	40-49	<80	80-100	101-112	>112
5	50-59	<86	86-105	106-115	>115
6	60-69	<90	90-108	109-118	>118

TABLE II HEARTRATES FOR MEN

	A	B	C	D	E
1	AGE	VERY FIT	FIT	AVERAGE	UNIT
2					
3	30-39	<84	84-105	106-122	>122
4	40-49	<88	88-108	109-118	>118
5	50-59	<92	92-113	114-123	>123
6	60-69	<95	95-117	118-127	>127

For the above mentioned data we build a set of rules into the knowledgebase in the form of IF-THEN-ELSE structures.

Table–III, Fuzzy Rules–Examples

S.No.	Condition	Action
1	IF Male(Age=30-49)AND (If heart beat=(106-122)) AND target=AVERAGE THEN	No Change
2	IF Female(Age=60-69) AND (If heart beat=(90-108)) AND target=FIT THEN	No Change

Conclusion

The proposed system effectively addresses the limitations of existing healthcare practices by integrating advanced automation, streamlined management, and improved accessibility. By maintaining patient medical history in a standardized CDA (Clinical Document Architecture) format, it ensures structured and interoperable data exchange across healthcare platforms. Furthermore, by eliminating the need for additional layers of encryption, the system enhances security while reducing complexity, thereby enabling faster, safer, and more reliable healthcare delivery. Overall, this solution provides a comprehensive and efficient approach to modernizing healthcare services for both patients and providers.

Future Scope

The proposed system can be further enhanced by integrating emerging technologies such as Artificial Intelligence and Machine Learning to enable predictive diagnosis and personalized treatment recommendations. Incorporating IoT-enabled medical devices will allow real-time patient monitoring and automatic updates to the CDA records. Block chain technology can be adopted to ensure transparent, tamper-proof medical data exchange across multiple healthcare providers. Additionally, expanding the system into cloud-based platforms will support scalability, remote accessibility, and interoperability with global healthcare standards. In the future, the solution can evolve into a comprehensive digital healthcare ecosystem that not only stores and secures patient data but also actively supports clinical decision-making and improves overall patient outcomes.

References

[1] S. S. Priya, A. Kumar, and P. R. Sharma, “Hybrid Attribute-Based Encryption to Secure Data in Wireless Sensor Networks,” *International Journal of Engineering Research & Technology (IJERT)*, vol. 13, no. 11, pp. 520–526, Nov. 2023. [Online]. Available: https://www.ijert.org/research/hybrid-attribute-based-encryption-to-secure-data-in-wireless-sensor-networks-IJERTV13IS110130.pdf

[2] M. Alam, M. M. Islam, N. M. Nayan, and J. Uddin, “An IoT-Based Real-Time Environmental Monitoring System for Developing Areas,” *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 52, no. 1, pp. 106–121, Oct. 2024. doi: 10.37934/araset.52.1.106121

[3] H. M. Ken and M. Behjati, “Advancing Air Quality Monitoring: TinyML-Based Real-Time Ozone Prediction with Cost-Effective Edge Devices,” *arXiv preprint*, arXiv:2504.03776, Apr. 2024. [Online]. Available: https://arxiv.org/abs/2504.03776

[4] A. Bhattacharjee, S. Samanta, J. Bhattacharya, and M. K. Singh, “GreenShield: CNN-Based Real-Time Forest Monitoring and Response,” *arXiv preprint*, arXiv:2406.16917, Jun. 2024. [Online]. Available: https://arxiv.org/abs/2406.16917

[5] “Recent progress in Arduino- and smartphone-based sensors for chemical and biological detection,” *Sensors and Actuators B: Chemical*, vol. 400, p. 135146, Mar. 2024. doi: 10.1016/j.snb.2024.135146

[6] N. Gao, “Optimal Deployment of Large-Scale Wireless Sensor Networks Based on Graph Clustering and Matrix Factorization,” *EURASIP Journal on Advances in*

Signal Processing, vol. 2023, no. 65, pp. 1–12, Mar. 2023. doi: 10.1186/s13634-023-00989-5

[7] A. Haque, M. Hassan, and S. Islam, "Wireless Sensor Networks Anomaly Detection Using Machine Learning: A Survey," *arXiv preprint*, arXiv:2303.08823, Mar. 2023. [Online]. Available: https://arxiv.org/abs/2303.08823

[8] V. K. H. Prasad and S. Periyasamy, "Energy Optimization-Based Clustering Protocols in WSN and IoT—Survey," *International Journal of Distributed Sensor Networks (IJDSN)*, vol. 19, no. 1, pp. 1–19, Jan. 2023. doi: 10.1177/15501477231152145

[9] "A Secure Routing Protocol for Improving the Energy Efficiency in Wireless Sensor Networks," *Ecological Information Systems*, vol. 16, no. 2, pp. 98–110, Sep. 2024. doi: 10.1016/j.ecoinf.2024.101248

[10] "Exploring Coverage and Security Challenges in Wireless Sensor Networks: Toward AI-Driven Resilient Protocols," *Computer Networks*, vol. 240, pp. 25–38, Jan. 2024. doi: 10.1016/j.comnet.2023.110123

[11] R. Gupta and S. Das, "Privacy-Preserving Attribute-Based Access Control Using ABAC, Homomorphic Encryption, and Zero-Knowledge Proofs," *Cybersecurity*, vol. 7, no. 12, pp. 1–19, Jan. 2024. doi: 10.1186/s42400-024-00323-8