

ENABLING SECURE DATA TRANSACTION IN BIO MEDICAL ENGINEERING USING CCART APPROACH

¹A. VELMURUGAN, ²T. RAVI

¹Research scholar, Department of Computer Science and Engineering,
Sathyabama University,
Chennai, India.

²Professor, Department of Computer Science and Engineering,
Madanapalle Institute of Technology and Science,
Andhra Pradesh, India.

ABSTRACT

Biomedical engineering, application is growing day by day. Major work carried by it is to manufacture the medical equipment based on the data of patient produced by professionals. The issue faced is the privacy and confidentiality of patient details that are transferred from professionals. In addition, other important factor is to retrieve the data from the database is other challenging process. This paper considers these issues and proposes a method by transacting the data with security in order to provide privacy. CART approach is implemented to retrieve the required data from the database in an efficient way. It works by searching the content in tree structure which saves the searching time and classification supports effective data (required) from such a huge database. This overcomes the misprocessing of data that are done by manual it leads to loss in time and cost for manufacturing the equipment. The proposed work retrieves the exact data in effective and efficient manner.

Keywords: *Data transaction, privacy, CART, Security and confidentiality.*

1.INTRODUCTION

Biomedical engineering is a process of integrating biology and medicine with engineering to solve the issues related to living beings. The biomedical engineers require strong knowledge in traditional engineering discipline, such as electrical, chemical engineering, or mechanical and increasingly, materials science. This is transparently utilized in areas such as healthcare, from analysis and scrutiny for treatment and recovery, and has invited in public conscience by the way of propagation of implantable medical devices, such as pacemakers and artificial hips, to more enhanced technologies such as stem cell engineering and the 3-D printing of biological organs. The Biomedical engineering process has principles issue and remedy techniques are focused as researching areas.

The application of biomedical technology is increasing due to the growth of bioinformatics and information technology. With the growth of biomedical technology it also raises the biomedical database [1]. This creates demand and difficulty for data mining in bio

medical field. Here the challenge faced is transactions of data require confidentiality and privacy. Data mining is a method of discovering knowledge from a huge database and the goal is to detect the hidden and valuable information. In other hand, data mining algorithm does not depend on any predefined classes or prototype and the training examples while classifying the classes can produce the optimal quality as output that's the proposed method will fit to biomedical engineering for better result. A major difficulty for information retrieval in the life science domain is fighting with its complex and inconsistent terminology.

The main process in bio medical engineering is designing the equipment according to the patient details. The other privacy issue is another challenge faced now days in bio medical. From the doctors to bio-medical engineer the detail of patient is transferred to design the equipment for treatment of a patient. Hence it arise the confidentiality and privacy issues. Therefore this paper focuses on designing the secure transaction of data from professionals to engineering professionals.



Therefore security in the transaction is provided by implementing cryptographic method that provide key in transacting the record.

The factor that focuses is how to bridge the gap among two fields, such as data mining and bio informatics, for successive data mining in biomedical data. An optimized analysis support for efficient and effective bio-medical data analysis and result of some research works may boost efficient bio-medical data scrutiny and outcome of some research problems that create path for advance developments of potential data mining tools for bio data analysis.

Because of their nature of high dimensionality, enlarge volume, difficult, various types of data, hierarchical, time-dynamic and requirement for extremely high accuracy therefore biomedical data proposes the certain challenges to bioinformatics. The biomedical applications enhance in the field epidemiology, bioinformatics, and bio surveillance has received attention and challenges to researchers have tried to incorporate data mining in the medical domain.

An efficient retrieval of data from the database is another major problem therefore an optimized algorithm is deployed to retrieve the required data from a huge database. An additional care must need in retrieving the specified data. If any fault is takes place it will cause an enormous loss of time, data, working and cost for construction of equipment. Therefore exact data retrieval overcomes the issues mentioned above for this optimized algorithm is proposed in this paper. Related papers with respect to the issues faced are discussed in section II. The implementation of proposed system is discussed briefly and the method is described in section III. Results and conclusion are discussed in section IV.

Key objective for proposed system:

Biomedical database are sensitive data and it need confidentiality and security in transaction of data between sources to destination. Implementing efficient data retrieval in addition with providing security avoids the cost, time, manual work in case if any mismatch fault occurs in retrieving information from a huge database. To overcome mentioned disadvantages proposed method implements a cryptographic method for security and optimal data mining algorithm.

This research work completely focuses on enabling privacy for patient reports during transaction between doctors and bio medical engineers. This is achieved clearly through double step security. The related works regarding the proposed work are discussed in section 2. Implementation of projected method is explained clearly in section 3. Results obtained are discussed in section 4 briefly by comparing it with existing available techniques and finally the intention of the proposed work is concluded.

2. LITERATURE SURVEY

In this section, papers are discussed according to the issues faced by the patients regarding their medical records and few efficient methods to overcome those problems. Data mining is the process of extracting knowledge from database with statistics and optimal retrieval algorithm. The functions of data mining is enhancing day by day, therefore mining information from a huge database is a difficult task [1]. The accuracy of information extraction is improved by discovering predictive relationships between different fragments of mined data; data-mining algorithms are used. However, textual dissimilarity due to mistakes, abbreviations, and other sources can prevent the productive discovery and consumption of hard-matching procedures. In textual data recent approaches of soft clustering can exploit predictive relationships. This work discusses an approach by means of soft clustering data mining algorithm to enhance the accuracy of biomedical text extraction.

The innovatory alteration in data mining and bio-informatics, it is very valuable to use data mining techniques to estimate and observe bio-medical records [2]. Based on the various characteristics the bio-databases are framed with the patient affected by diabetes like: Genetic category, Disease type, country, food habits. This work first extracts the correlated, valuable, valid and actionable data from bio databases. Bio databases is frequently independent heterogeneous and spread in nature. The mined data is preprocessed. After preprocessing, data mining techniques are implemented on preprocessed records, local data as well as proprietary data. Then the extracted knowledge is included with expert system knowledge to support researchers and scientists in their research work and crucial decision making process. In bioinformatics it is essential to scrutinize the maximum research problems



and it expand new data mining methods for scalable and effective bio-data analysis.

The biomedical invention process allows scientists to generate novel hypothesis to identify the most serious questions is the main mission for text and data mining is truly influence [3]. The invention of 3-way genetic communications and to the examination of genetic data in the condition of electronic medical records (EMRs), as well as an integrative approach combine data from genetic (SNP) and transcriptomic (microarray) sources for clinical prediction was discussed in this work. The involvement of text mining in the field of biomedical database with genetic interaction method was discussed in this work.

Biomedical informatics (BMI) is the knowledge underlying acquisition, maintenance, retrieval, collecting, manipulating, and analyzing the biomedical knowledge and information to improve medical data analysis, problem solving, and decision making, motivated by efforts in the direction of growth in medical domain [4]. A comprehensive framework will be created which encompasses of numerous data mining methods and assess meaningful information from biomedical data. In bio medical, data mining field is applied to analyze the characteristics, identify patterns of interest, for diagnosing and predicting patient's health.

Automated data gathered with data mining together performs automated decisions. Discrimination can be direct or indirect. Direct discernment utilizes sensitive data for decision making. Indirect discrimination makes decisions on the basis of non-sensitive data [5]. Along with security and privacy, proper discrimination performs vital role in allowing for legal as well as moral point of vision of data mining. This work focuses to develop new preprocessing discrimination prevention methodology which consists of different types of data transformation methods. With the help of that direct discernment, indirect discernment or together of them at the same time get prevented.

For construction the final conclusion there are two phases in which first step include identification of categories and makes clusters of entities whatever it may be, straight indirectly discriminated for making decision. In second step with the support of clustering, conversion of data in definite way such that removes all discrimination. Practical use of healthcare

database systems, Conclusion creation and management equipment like data mining can enormously contribute to increase accuracy in conclusion building in healthcare. Transforming massive, difficult and heterogeneous healthcare data into knowledge can help in monitoring price and preserving high worth of patient care.

Data mining techniques such as grouping, classification, regression, association rule mining, CART (Classification and Regression Tree) are broadly utilized in healthcare domain. Data mining algorithms, when appropriately used, are capable of enhancing the property of calculation, diagnosis and disease classification [6].

Due to the rapid growth of medical data, it has become indispensable to use data mining techniques to help decision support and predication systems in the field of Healthcare. The medical mining yields required business intelligence to support well informed diagnosis and decisions. This paper has provided the summary of data mining techniques used for medical data mining besides the diseases they classified. It also throws light into the importance of locally frequent patterns and the mining techniques used for the purpose. From the analysis a conclusion is made for implementation of proposed system.

Issues identified from existing method:

- Textual dissimilarity due to mistakes, abbreviations, and other sources can prevent the productive discovery and consumption of hard-matching procedures. This leads to mismatching the record of patient and makes the whole work waste.
- Extraction of exact medicine for disease analysis and it needs a crucial decision making process.
- 3 way genetic communications is used in order to provide security in data extraction from a huge database.

3. IMPLEMENTATION

This section clearly describes the implementation of CART algorithm and working of it in retrieval of information in efficient way in biomedical engineering. The gap between data mining and bio medical information is filled by the transaction of medical data sets that transferred between medical professionals to engineering professionals. Therefore issues are

generated in these sector is confidentiality and privacy of user details hence this focuses on secure data transaction between the professionals.

Objective:

The motivation of this proposed work is constructed by clearly analyzing the issues in bio medical data mining concept. There is no privacy regarding patient record transaction between source and destination.

Mining works are not that much efficient in retrieving exact patient record details this cause the engineer to retrieve wrong record and work on it may cause loss of effort, time and cost.

Most of the available works are concentrating in extraction of patient record in efficient way and it does not concern regarding privacy and security issues of patient record transaction.

3.1 Two-Step Verification:

The medical data is transmitted from doctor professionals to engineering professionals in this cryptographic method have been implemented for secure transaction. In order to provide better security one of the emerging double step encryption is followed. Initially the record is encrypted with a key and transmitted from source to destination. In addition to it one time password is also implemented therefore this increases secure and privacy.

Fig 1: Message cryptographic method

From the medical data received the professional's searches the appropriate remedy from already details available in the database is selected. Therefore the data is retrieved from the database should be done in an effective way by an effective algorithm CART. Proposed algorithm is described briefly as follows,

3.2 CART Algorithm:

CART analysis is carried out in a tree-building method that is differing from traditional data analysis methods. From the analysis, CART has been identified to be effective for making decision rules which perform as well or better than rules developed using more basic approaches. With advantage, CART is frequently able to uncover complex interactions between predictors which may be hard or difficult using fundamental

multivariate techniques. It is now possible to perform a CART analysis with a basic knowledge of each of the several procedures involved in its procedure. Classification tree methods such as CART are proper method to produce a prediction law from a set of observations described in terms of a vector of application and a response value. The goal is to describe a general prediction rule which can be used to assign a response value to the situations solely on the origins of their predictor (explanatory) variables. CART can be implements either as a classification tree or as a regressive tree based on whether the response variable is categorical or continuous. Tree based approaches are not depends on any stringent expectations. These methods can handle large number of variables, are resilient to outliers, non-parametric, more flexible, can handle categorical variables, though computationally more intensive.

Decision Trees are usually utilized in data mining with the objective of creating a model that predicts the value of an objective (or dependent variable) depends on the values of several input (or independent variables). The CART or Classification & Regression Trees methodology is used. Classification tree is defined as where the target variable is categorical and the tree is consumed to detect the class within the target variable would likely fall into. Where the Regression tree is discussed as the target variable is continuous and tree is used to predict its value.

The construction of tree is implemented using condition procedures, commonly tree structure support quick search of required data from a huge database in addition extracting the required data exactly is another important factor is done by following mathematical steps. If the condition is true then moves the search in left direction otherwise move it to right side of search. The CONDITION is stated as "attribute $X_i \leq C$ " for successive attributes.

Step 1: The basic key is to select a fragment at each node so that the data in each subset (child node) is "purer" than the information in the parent node. CART evaluates the impurity of the data in the nodes of a split with an impurity measure $i(t)$.

Step 2: If a split sat node t sends a proportion p_L of data to its left child node t_L and a

corresponding proportion p_R of data to its right child node t_R , the decrease in impurity of split

Sat node t is defined as

$$\Delta i(s,t) = i(t) - p_{L_i}(t_L) - p_{R_i}(t_R)$$

$$\dots\dots\dots(1)$$

= impurity in node t

– weighted average of impurities in nodes t_L and t_R

Step 3: A CART tree is raised, beginning from its root node (i.e., the entire training data set) $t=1$, by searching for a split s^* among the set of all available candidates S that give the largest decrease in impurity

$$\Delta i(s^*,1) = \max_{s \in S} \Delta i(s,1) \dots\dots\dots(2)$$

Then node $t=1$ is split in two nodes $t=2$ and $t=3$ using split s^*

Step 4: The above split searching process is repeated for each child node.

Step 5: The tree growing process is stopped when all the stopping criteria are met.

The tree is constructed based on the above steps therefore well-structured result has been obtained from the implementation. Based on record received from doctor professionals data diagnosis is done and required result is obtained by dint of CART.

Working of proposed system:

An algorithmic step of proposed method has been described briefly above. Now it's working in bio medical engineering is explained below. Initially, secure issues are taken into consideration and processed in a way to avoid security and privacy issues among the patient. Symptoms and disease of users are analyzed clearly by the professional and database is maintained.

Fig 2: overall working procedure

Initially the default record collected by doctor contains ID, Name, age, gender, and date of birth, symptoms and disease. Where the complete health detail of the patient is collected and maintained clearly.

Fig 3: Required parameter

For each record a unique ID will be assigned and separated in the format as ID,

symptoms and disease. The first step of privacy is initiated here. This provides a way that the engineering professional does not know the patient detail that is suffering from the disease.

Fig 4: separated parameter

Fig 4 details alone transmitted from the doctor database to engineering professionals in encrypted form. Therefore receiver decrypts it and collects the separated parameter and start the work according to the condition of the patient.

Fig 5: Initial privacy

The engineers receive the separated parameter and design the equipment or suggest solution according to disease and transfer the solution to doctor through security system OTP. Therefore solution to the particular patient record will be retrieved through a mail based decryption method.

Fig 6: second stage security

Finally the doctor retrieves the data using CART for exact identification of user. If any mismatch is done then whole work, cost, time everything is waste and it may cause an issue to the life of patient. Therefore each parameter is checked for exact identification of patient and there solution to it.

4. RESULT AND DISCUSSION

The implementation shows that proposed method have achieved better result and goal of the work is achieved. The issues have been taken into consideration and overcome using the proposed approach. In existing methods treatment procedures were concentrated and privacy issues are not concentrated. The proposed system works in all aspects and output shows better results.

Fig 7: Result of proposed method in various parameters

Proposed method is compared with existing available methods using various parameters and the result is shown below. From the parameters it shown that proposed method generates better results.

Fig 8: Result compared with existing methods

CONCLUSION:

The discussed methods and their results show knowledge extraction from medical databases is such a complex and enhancing area of research. However, medical

sets and their result searches are growing the security and exact retrieval is issues must concentrate to add an additional advantage to the results. Therefore the proposed method achieves in obtaining better results in addition overcoming the issues that were taken throughout the work. CCART methods provide security in data transaction and implemented exact retrieval of patient for treatment. Finally implementation shows better result when compared to existing methods. In future the work can be extended by encrypting the whole document in order to provide privacy this may create critical issues in encryption and decryption among source and destination.

REFERENCE:

[1] Sumit Vashishta and Dr. Yogendra Kumar Jain, (2011)“Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm” (*IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 2, No. 4.

[2] Hemalatha T., Gowthami B.G., Divya Sri A., Sindhuri Ch., SriDivya Y., BhanuVidhyaKiran B., (2012) “Intelligent Data Mining in AutonomousHeterogeneous Distributed Bio Databases” (*IJCSIT International Journal of Computer Science and Information Technologies*, Vol. 3 (2), 3778-3781.

[3]Graciela Gonzalez, Kevin Bretonnel Cohen, Maricel Kann G., Casey Greene S., Robert Leaman, Udo Hahn, Nigam shah and Jieping ye, (2012) “Text And Data Mining For Biomedical Discovery”.

[4] Madasamy B., Dr.Jebamalar Tamilselvi J., (2013) “General Framework for Biomedical

Knowledge With Data Mining Techniques” *International Journal of Computer Trends and Technology (IJCTT)* – volume 4 Issue 5–May ISSN: 2231.

[5] Ms Nikhita B Ugale, Prof Deshpande S.D., (2016) “Prevention for Direct and Indirect Discrimination in Data Mining” (*IJCSIT International Journal of Computer Science and Information Technologies*, Vol. 7 (2) , 960-963.

[6] Shameem Fathima A., Mani megalai D. and Nisar Hundewale , (2011) “A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue” *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 6, No 3.

[7] Chitra R., and Seenivasagam V., (2013) “Review Of Heart Disease Prediction System Using Data Mining And Hybrid Intelligent Techniques” *ICTACT Journal On Soft Computing*, July, Volume: 03, Issue: 04.

[8] Jiban Pal k., (2011) “Usefulness and applications of data mining in extraction information from different perspectives” *Annals of Library and Information Studies* Vol.58, March, pp.7-16.

[9] Mohammed Abdul Khaleel, Sateesh Kumar Pradham and Dash G.N., (2013) “A Survey of Data Mining Techniques on Medical Data for Finding Locally Frequent Diseases” *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 3, Issue 8.

[10] Radhika, (2008) “Topics in Biomedical Informatics Data Mining and its applications and usage in medicine” *Spring*.

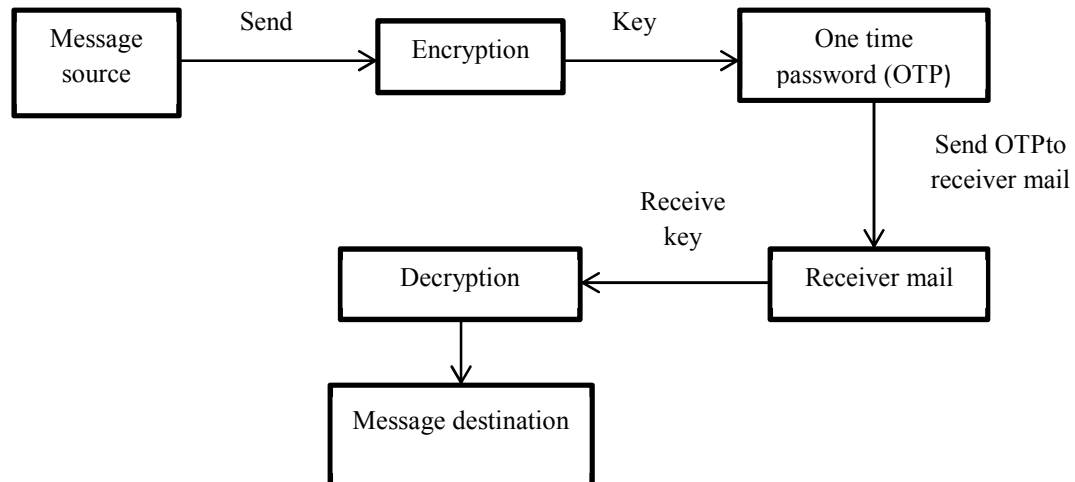


Fig 1: Message cryptographic method

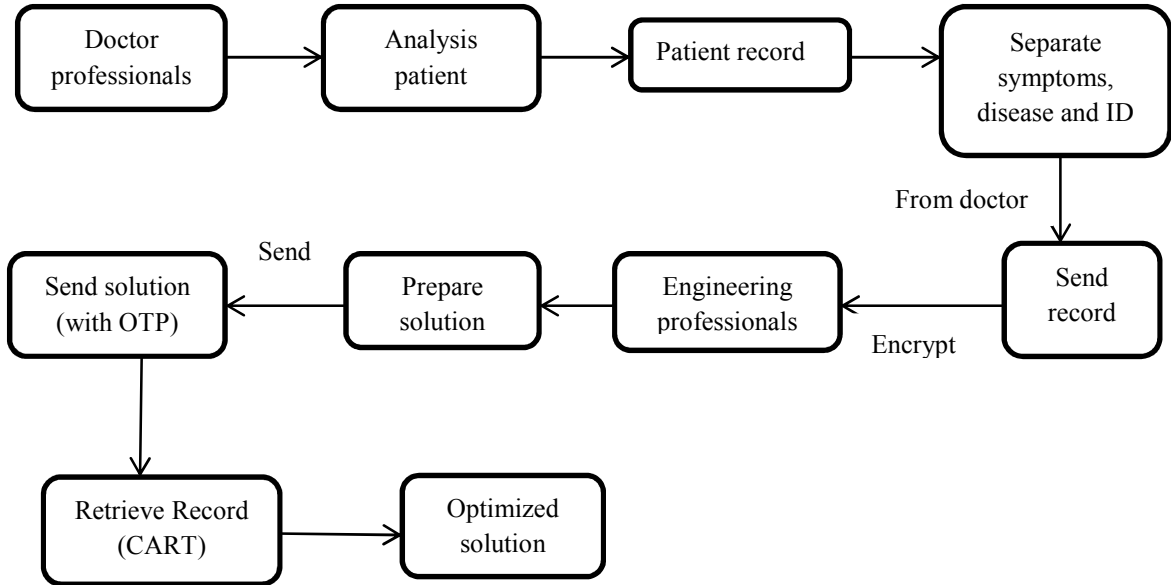


Fig 2: overall working procedure

ID	Name	Age	D.O.B	Gender	Symptoms	Disease
----	------	-----	-------	--------	----------	---------

Fig 3: Required parameter

ID	Symptoms	Disease
----	----------	---------

Fig 4: separated parameter

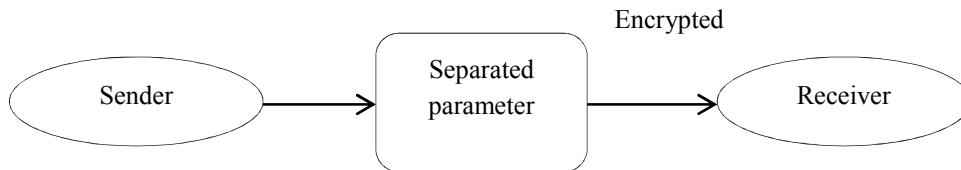


Fig 5: Initial privacy

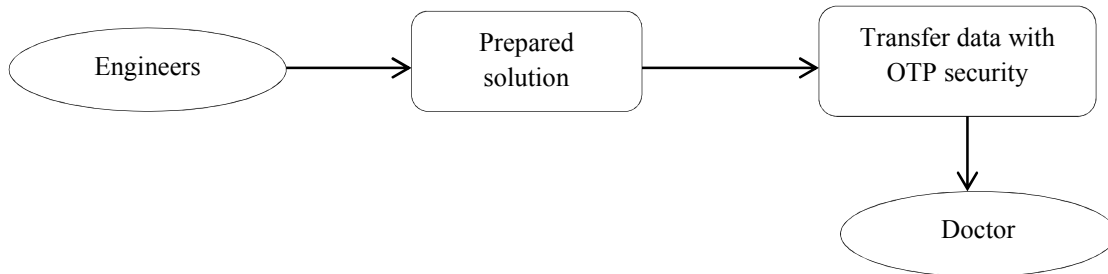


Fig 6: second stage security

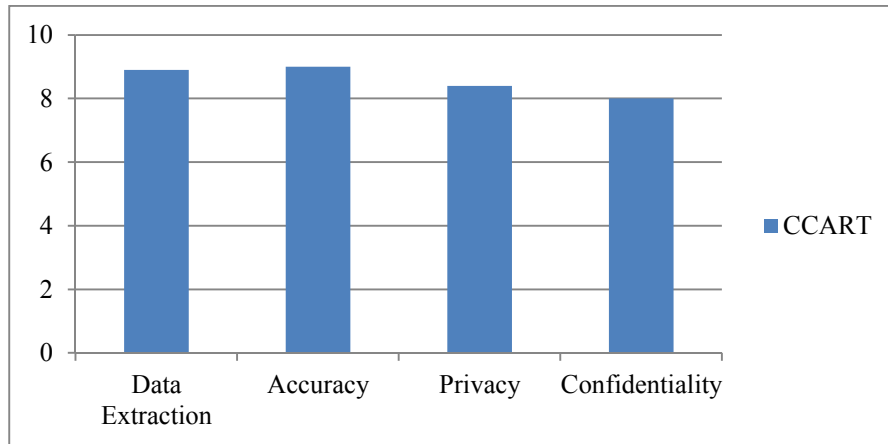


Fig 7: Result of proposed method in various parameters

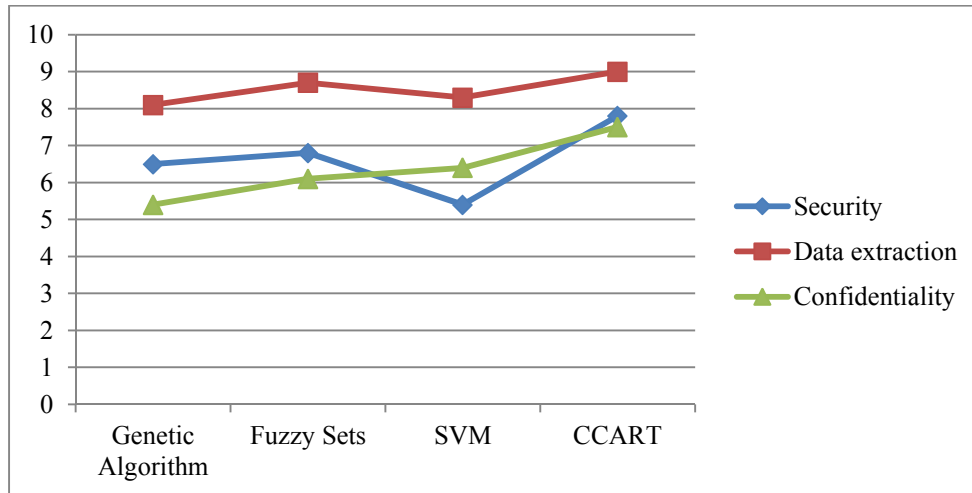


Fig 8: Result compared with existing methods