# Secured Medical Decision Support System based on SVM

**[1]Akshay Muley, [2]S. A. Kinariwala**

[1]M.E. Student, [2]Asst. Professor
Computer Science & Engineering,
MIT College of Engineering, Aurangabad, India

*Abstract*—**The Secure Medical Decision Support System (SMDSS), with several data mining techniques that are applied to help doctors diagnose diseases of the patient with similar symptoms, has recently received a great deal of attention. The advantages of the Secure Medical Decision Support System include not only improving diagnostic accuracy but also reducing diagnostic time. In this document, we have proposed the classification of the SMDSS with some advanced technologies, such as the Support Vector Machine. The classifier (SVM) offers many advantages over traditional health systems and opens a new way for physicians to forecast the patient's health problems. Specifically, to protect the privacy of historical data of previous patients, a new cryptographic tool called homomorphic additive aggregation scheme (AHPA) was designed. Given that medical care is the field in which the safety of data related to patients' diseases must be preserved, we have used the Pallier Homomorphic encryption technique that substantially fulfills the security objectives. Specifically, with large amounts of clinical data that are generated every day, the classification of support vector machines (SVM) can be used to excavate valuable information to improve the medical decision support system. In this document, we propose the use of the Paillier encryption technique to preserve patient privacy in the cloud. Patient data can be compromised through the cloud. To overcome this scenario, the Homomorphic encryption technique helps. The processing is done in the encrypted data; therefore, there is no possibility of compromising the privacy of the patient's data.**

*IndexTerms*— **Medical Decision Support System, Privacy Preserving, Support Vector Machine, Homomorphic Encryption.**

## I. INTRODUCTION

Today, the health industry is widely distributed around the world to provide health services to patients, has never faced massive amounts of electronic data or has experienced such a sharp growth rate at present. However, if a suitable technique is not developed to find large potential economic values from large health care data, this data may not only be meaningless but also require a large amount of space to store and administer. In the last two decades, the miraculous evolution of the data mining technique has had a great impact on the revolution of human life by predicting future behaviors and trends in everything that can convert stored data into meaningful information. These techniques are adequate to provide support in decision-making in the field of medical care. To accelerate the time of diagnosis and improve diagnostic accuracy, a new system in the healthcare industry must be viable to provide a much cheaper and faster way of diagnosis [1]. The Secure Medical Decision Support System (SMDSS), with several data mining techniques that are applied to help doctors diagnose diseases of the patient with similar symptoms, has recently received a great deal of attention. The Secure Medical Decision Support System has been defined as an "active knowledge system", which uses two or more elements of the patient's data to generate specific advice for each case. This implies that an SMDSS is simply a decision support system that focuses on the use of knowledge management in such a way that clinical advice can be obtained for patient care based on multiple elements of the patient's data. patient. The main objective of modern SMDSS is to help doctors at the point of care. This means that doctors interact with an SMDSS to help with the analyzes and arrive at a diagnosis based on the patient's data. The Naive Bayes classifier, one of the popular automatic data extraction learning tools, has been widely used recently to predict various diseases in SMDSS [1].

Despite its simplicity, it is more appropriate for medical diagnosis in health care than some sophisticated techniques. SMDSS with a naive Bayes classifier has offered many advantages over traditional health systems and opens up a new way for physicians to predict patient illness [2].

However, its success still depends on the understanding and management of information security challenges and privacy, especially during the decision phase of the patient's illness. One of the main challenges is how to keep patient's medical data away from unauthorized disclosure. The use of medical data may be of interest to a wide variety of people interested in medical care. For example, a direct online consumer service provider offers individual risk prediction for the patient's illness. Without good protection of the patient's medical data, the patient may fear that they leak and abuse their medical data and refuse to provide their medical information to the SMDSS for diagnosis. Therefore, it is crucial to protect the patient's medical data. However, maintaining the privacy of medical data is not enough for the entire SMDSS to thrive. The classifier of the service provider, which is used to predict the patient's illness, cannot be exposed to third parties because the classifier is considered as an asset of the service provider. Otherwise, third parties may abuse the classifier to predict the patient's illness, which could damage the earnings of the service provider. Therefore, in addition to preserving the privacy of the patient's medical data, the way to protect the privacy of the service provider is also crucial for the SMDSS. Preserving the patient-centered Secure Medical Decision Support System called PPCD helps the physician predict the risks of the patient's illness in a way that preserves privacy. A support system for making clinical decisions focused on the privacy of the patient, safe and conservative, which allows the service provider to diagnose the patient's disease without filtering the medical data of any patient. SMDSS provides disease-specific knowledge and information for physicians to improve the effectiveness of diagnosis and improve the quality of medical care. It can elevate patient safety and improve the quality of medical care. With the increase in the amount of data generated by all health industries and researchers, there is a need for fast, accurate and robust algorithms for data analysis [8]. Improvements in database technology, computer performance and artificial intelligence have contributed to the development of intelligent data analysis. The main objective of data mining is to discover patterns in the data that lead to a better understanding of the data generation process and useful predictions. A recent technique that has been developed to address these problems is the support vector machine. The support vector machine has been developed as a robust tool for classification and regression in noisy and complex domains. In this article, we use the Support Vector Machine (SVM) technique, which is one of the most powerful classification techniques that was applied successfully to many real-world problems. This SVM has a greater advantage in case of improving the diagnostic accuracy in the medical decision support system. Along with this, we use the homomorphic encryption technique to provide security to confidential data related to the patient's health information. The remaining document is organized as, Section II provides some Literature Survey that provides brief information of the study carried out in the field of SMDSS Medical Decision Support System.

## II. Machine Learning

The machine itself learns the given inputs and outputs strategy, it also shows the output for new inputs based on given data sets. There are mainly two types of machine learning. Supervised learning: the machine is presented with the input example and its desired output. By using this method, the machine learns the general rule. Unsupervised learning: the machine does not know the input and its desired output. The machine itself learns the algorithm and produces the output.

## III. Support Vector Machine

The vector support machine (SVM) has become an increasingly popular tool in the machine learning task that includes classification, regression, etc. SVM separates the data into two categories and performs the classification and then builds an N-dimensional hyperplanotype. SVM is supervised by the learning model applied mainly for classification. SVM serves as the linear separator between two data points to identify two different classes for the multidimensional environment. The SVM algorithms are in binary format. In the multiple class problem, one must reduce the problem to a set of multiple binary classification problems.

With an approximate set applied for feature selection and SVM for classification, a very high classification accuracy of 99.41% for 50-50% of training test partition, 100% for 70-30% of training test partition and 100 % for 80 - 20% of training - test partition can be obtained. You can also discover a combination of five informative features, which may be important for doctors to diagnose breast. Support Vector Machine is a cutting-edge classification. It works well with real-world applications, such as classifying text, classifying images, etc. SVMs are the standard tools for machine learning and data mining. And with this large number of applications and advantages, we will also use SVM for our classification technique proposed in SMDSS.

## IV. Related Work

Yogachandran Rahulamthavan, et. all have investigated the "Clinical decision support support system for preserving privacy using the Gaussian kernel-based classification". Describes cloud computing technology that is having very rich clinical data. The use of this system improves the decision-making capacity of health professionals. For this purpose, it uses the Paillier cryptosystem, but only encrypts the entire value only depends on the clinical data available locally. The Gaussian kernel only works purely in simple domain and can not be modified to the clinical server.

ErmanAyday, et. al investing "Calculation of the preservation of the privacy of the disease risk through the use of genomic, clinical and environmental data". Describes the privacy to store and process the unit in the system. For this it uses Homomorphic encryption and full comparison of privacy preservation. It uses real patient data and a reliable risk factor for disease. It works efficiently only for genomic data. Specify the disease risk test using genomic data.

H. Monkaresi et. all have presented "An automatic learning approach to improve non-contact heart rate monitoring using a webcam." We have evaluated a method for the remote measurement of human resources in three applications: a controlled laboratory task, a naturalistic HCI and an indoor cycling exercise. This study evaluated the method of Pohet al. And it showed the viability of its methodology to measure HR at rest.

Ximeng Liu et. all have studied "Towards efficient computing and preserving privacy in the era of big data". This document has investigated the privacy challenges in the big data era by first identifying the big data privacy requirements and then discussing whether existing privacy preservation techniques are sufficient for big data processing. They have also introduced an efficient cosine similarity computing protocol that preserves privacy in response to the efficiency and privacy requirements of data mining in the era of big data.

Y. Tong et. to introduced "Mobile access assisted by health data cloud with privacy and audit capability". The author has proposed to create privacy in mobile health systems with the help of the private cloud. We provide a data storage solution that preserves privacy by integrating PRF-based key management for combination impossibility, a pattern of search pattern concealment and access based on redundancy and a secure indexing method for the Search for keywords that preserves privacy. We also investigate techniques that provide access control (both in normal and emergency cases) and auditability of authorized parties to avoid misconduct, by combining the threshold signature controlled by ABE with role-based encryption.

Tien Tuan et. Al have presented "Stream on the Sky: Outsourcing access control for streaming data to the cloud." In this document, we presented a system that provides fine grain access control for flow data over unreliable clouds. This system allows owners to encrypt data before retransmitting it to the cloud. Encryption guarantees confidentiality against the cloud and access control against dishonest users. The current strength uses combinations of three encryption schemes: a deterministic scheme, an ABE proxy scheme and a sliding window scheme. We have shown how the cloud can impose access control on encryption texts by transforming them for the authorized user, without learning the clear texts.

B. K. Samanthula et. all have entered "Secure query of nearest k neighbor over encrypted data in an outsourced environment". This document has proposed two new SkNN protocols on data encrypted in the cloud. The first protocol, which acts as a basic solution, filters certain information to the cloud. On the other hand, our second protocol is completely secure, that is, it protects the confidentiality of the data, the user's input query and also hides the access patterns to the data.

Y. Rahulamathavan et.al have introduced the "Clinical decision support system for preserving privacy using the Gaussian Kernel-based classification". This document has proposed a decision support system that preserves privacy using a support vector machine based on Gauss cores. Since the proposed algorithm is a potential application of emerging techniques such as cloud computing technology, clinicians can use clinical data sets (or health knowledge) available in remote locations over the Internet without compromising privacy, thereby improving decision-making capacity. of health professionals.

J. Chen, H. Huang et.al presented two function evaluation metrics (CDM and MOR) for the Naïve Bayes classifier applied to text collections of various kinds. They have compared CDM and MOR with EOR, WOR and MC-OR, three variations of Odds Ratio for data sets of various kinds.

R. Bellazzi et.al have introduced "Predictive data mining in clinical medicine: current issues and guidelines". The objective of this review is to analyze the scope and role of the research area of predictive data mining and to propose a framework to deal with the problems of construction, evaluation and exploitation of data mining models in clinical medicine.

X. Yi and Y. Zhang have introduced "the naive classification of Bayes to preserve the privacy in data distributed through semiconsistent mixers". The proposed multiparty protocol is based on the semiconsistent mixer model, in which each data site sends messages to two semiconsistent mixers, respectively, that execute our two-part protocol and then transmit the result of the classification. Because our protocol does not involve any collusion between the two mixers and does not require communication channels between data sites, it facilitates both the administration and the implementation of trust.

## V. Proposed Architecture

We are using the Support Vector Machine Data Mining classification technique for the Clinical Decision Support System. The system will work faster and more efficiently using SVM [7]. It is widely used in real life applications due to its simplicity and good performance both in theory and in practice. However, in large-scale problems, where large training data are available and should be used, such as the detection of traffic signals, the training and testing phases of the method can be prohibitively demanding in terms of calculations. Therefore, for large-scale problems, the reduction of computational complexity is essential. We are using encryption techniques to preserve the privacy of patient data. And to preserve the privacy of the data that passes through the network, we are using the homomorphic encryption technique to re-encrypt the data. All processing will be done on the server side and in the encrypted data. We have defined the SMDSS system model in Figure 1, which includes Trusted Authority (TA), Cloud Platform (CP), Data Provider (DP), Processing Unit (PU) and Undiagnosed Patient (PA).

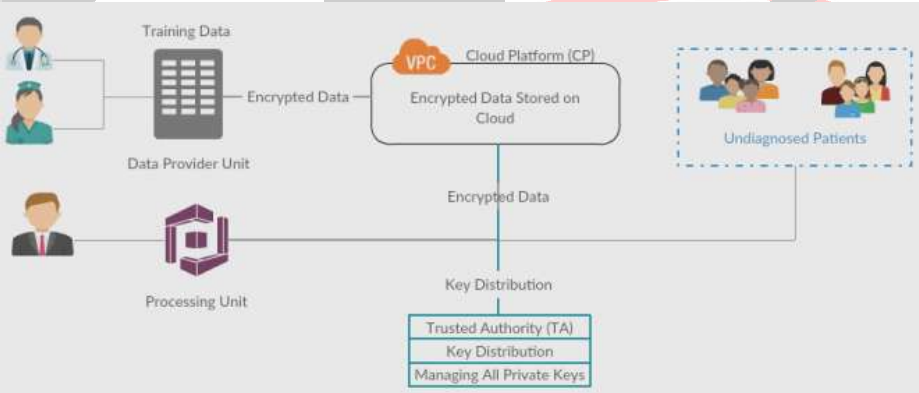

Figure 1) System Architecture

***A. Trusted authority (TA):*** TA is the indispensable entity trusted by all the entities involved in the system, which is responsible for distributing and managing all the private keys involved in the system.

***B. Cloud Platform (CP):*** CP contains an unlimited storage space that can store and manage all data in the system. Other parties that have limited storage space can subcontract their data to CP for storage.

***C. Data Provider (DP):*** DP can provide historical medical data containing the patient's symptoms and confirmed diseases, which are used for the training of the SVM classifier. All this data is subcontracted to CP for storage.

***D. Processing Unit (UP):*** PU can be a company or hospital that can provide online service directly to the client and offer individual risk prediction for various diseases according to the client's symptoms. PU uses medical data to construct the SVM classifier and then uses the model to predict the disease risk of undiagnosed patients.

***E. Undiagnosed patient (PA):*** the PA has some information about the symptoms that are collected during the visits to the doctor or directly provided by the patient. (for example, blood pressure, heart rate, weight, etc.). Symptoms can be sent to UP for diagnosis of the disease

***System Flow Description***

A. Step-by-step system workflow:

Step 1: the undiagnosed patient will send his symptoms to the Platform in the cloud (CP) in the encrypted format, using his public key.

Step 2: the data provider will provide the historical medical data to the CP in an encrypted format using the Homomorphic encryption technique.

Step 3: The CP will decrypt this data and send it to the SVM classifier for training. Once the training is done, the risk of illness will be calculated based on the symptoms provided by the undiagnosed patient and the result of the training. All processing is done in encrypted data, which preserves the privacy of the patient's data.

Step 4: Once the risk of illness is calculated, the expected result will be sent to the next level. At this level, the predicted risk of disease risk will be calculated and, according to the patients' preferences, the results will be sent to the patient in an encrypted form.

Step 5: If the patient wants the names of the predicted top-k diseases, then they can give their own preferences accordingly. For this, on the server side we will use the top-k algorithm. In this algorithm, the risk of maximum likelihood disease will be calculated. And the top-k results will be sent to the patients according to their preferences. Once the result of the encrypted diagnosis is obtained on the client side, the undiagnosed patient will decipher these results using his private key.

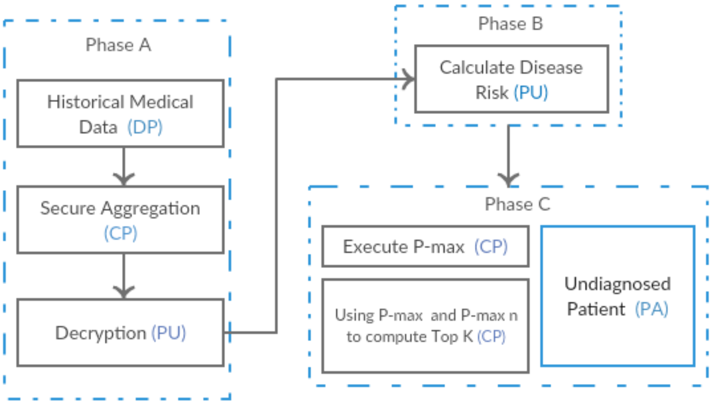

Figure 2) Process Flow Diagram

The proposed system should achieve requirements for the preservation of privacy. As indicated above, if SMDSS does not consider the privacy requirements, highly confidential patient information (information on symptoms and diseases) will be disclosed to PU, PC and unauthorized parties in the patient's medical decision. It will allow the patient to involuntarily provide their own data to SMDSS. In addition, PU is always a profitable company that prevents its own data from leaking to other parts of the system. Therefore, the proposed system should achieve privacy of PA and PU simultaneously.

The proposed system should achieve calculation efficiency. The patient always has limited computational resources that can not support overload calculations. To support the recovery of patient-centered diagnosis results from the PC on time, the proposed system should consider the efficiency of computing. Therefore, it is important to allow PA to retrieve the results of the diagnosis in real time

## VI. CONCLUSION

In this document, we have proposed a Secure Medical Decision Support System using the data extraction classification technique called Support Vector Machine. Using SVM, the computational time and the diagnostic rate in our system can be improved. SVM has excellent performance in generalization, so it can produce high accuracy in classification for diagnosis. The patient can safely recover the top-k diagnosis result according to their own preferences. With the advantage of the homomorphic encryption technique, the patient's privacy over the cloud will be achieved. The processing is performed on the encrypted data, so that there is no loss in the privacy of patient data while training the SVM classifier. These results will evidently prove the proposed method.

## REFERENCES

[1] Ximeng Liu, Rongxing Lu, Jianfeng Ma, Le Chen, and Baodong Qin, "Privacy- Preserving Patient-Centric Clinical Decision Support System on Naïve Bayesian Classification", IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS, VOL. XX, NO. XX, DECEMBER 2014.

[2] R. S. Ledley and L. B. Lusted, "Reasoning foundations of medical diagnosis," Science, vol. 130, no. 3366, pp. 9–21, 1959.

[3] H. R. Warner, A. F. Toronto, L. G. Veasey, and R. Stephenson, "A mathematical approach to medical diagnosis: application to congenital heart disease," Jama, vol. 177, no. 3, pp. 177–183, 1961.

[4] C. Schurink, P. Lucas, I. Hoepelman, and M. Bonten, “Computer- assisted decision support for the diagnosis and treatment of infectious diseases in intensive care units,” The Lancet infectious diseases, vol. 5, no. 5, pp. 305–312, 2005.

[5] M. Kantarcıoglu, J. Vaidya, and C. Clifton, “Privacy preserving naive bayes classifier for horizontally partitioned data,” in IEEE ICDM workshop on privacy preserving data mining, 2003, pp. 3–9.

[6] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, “Tools for privacy preserving distributed data mining,” ACM SIGKDD Explorations Newsletter, vol. 4, no. 2, pp. 28–34, 2002.

[7] X. Yi and Y. Zhang, “Privacy-preserving naive bayes classification on distributed data via semi-trusted mixers,” Information Systems, vol. 34, no. 3, pp. 371–380, 2009.

[8] A. Amirbekyan and V. Estivill-Castro, “A new efficient privacy- preserving scalar product protocol,” in Proceedings of the sixth Aus- tralasian conference on Data mining and analytics-Volume 70. Aus- tralian Computer Society, Inc., 2007, pp. 209–214.

[9] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, “Toward efficient and privacy-preserving computing in big data era,” IEEE Network, vol. 28, no. 4, pp. 46–50, 2014.