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), which employs several data mining techniques, is used to help diagnose and treat patients with similar symptoms. The advantages of safe decision support systems include not only privacy preservation but also improves diagnostic accuracy. It also reduces the diagnostic time. To speed up the diagnosis and improve the accuracy of the diagnosis in the current health care system, it is important to make the diagnosis faster and cheaper. This system is called Secure Medical Decision Support System (SMDSS). The data mining techniques used to assist physicians in diagnosing patients with similar symptoms have been very popular in recent years. Past The advantages of medical decision support systems include improved diagnostic accuracy. It also reduces the diagnostic time. In this article, we use the Support Vector Machine (SVM) mining technique, which is used as a classifier that has advantages over traditional mining methods and is a new way of predicting a patient's disease. Because the system was built on the cloud platform, it was necessary to add some features that meet the security requirements. Particularly with the many data related to health care being created daily, the classification can be used to dig up valuable information to improve medical decision support systems. Homomorphic cryptographic techniques are useful for keeping the privacy of the patient in the cloud. Here, using the SVM Classifier and Encryption Techniques, and trying to make the clinical decision support system more useful in providing some important, accurate and effective death information.**

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

I. INTRODUCTION

The healthcare industry has a global scope in providing health services to patients who have never faced such a large number of problems with electronic information or have experienced rapid growth of information today. However, if there is no development of the right techniques to find the most economically feasible value from many healthcare information, these data may not be justified. But it becomes meaning only. It requires a lot of storage and management space. Over the past two decades, the phenomenal evolution of data mining techniques has given a significant impact on the way we revolve our way of life by predicting future behavior and trends about everything that can be stored. The information is meaningful. These techniques are ideal for decision support in healthcare systems. [1] In order to speed up the diagnosis and improve the accuracy of new diagnostic systems in the healthcare industry, it should be possible to diagnose the diagnosis. Cheaper and faster. Secure Medical Decision Support System (SMDSS), which employs various data mining techniques to assist physicians in diagnosing similarly affected patients, has received a lot of attention lately.

Medical decision support systems are defined as The "active knowledge system", which uses two or more patients' information to create specific cases. [2] This shows that SMDSS is the only decision support system that emphasizes the use of knowledge management. Characteristic to achieve medical recommendations for patient care based on multiple patient records. The main goal of the SMDSS is to help physicians in their care. This means that physicians interact with SMDSS to help diagnose and access diagnostic information based on patient information. The naive Bayer classifier, one of the most popular learning tools in the machine, has been used extensively recently to diagnose diseases in SMDSS.

Despite its simplicity However, it is more appropriate for medical diagnostics in health care than some sophisticated techniques. The SMDSS , which provides a naive Bayesian classification, has advantages over traditional medical systems and allows physicians to predict disease. Of patients However, its prosperity continues to be correlated with the understanding and management of data security and privacy, especially in the decision-making of patients with the disease. [3] One of the main challenges is maintaining informative information. The patient's medical record is kept away from unauthorized disclosure. The use of medical information may be of interest to a wide range of medical stakeholders. For example, online service providers to online consumers are forecasting individual risk for a patient's disease. Patients may feel that their medical information is being leaked and abused and refuse to provide their medical information to SMDSS for diagnosis. Therefore, it is important to protect the patient's medical information.

By increasing the amount of data generated by healthcare industry businesses and researchers, it requires a fast, accurate, and efficient algorithm for data analysis. [5] Database technology improvements, computational efficiency, and artificial

intelligence. Contribute to the development of intelligent data analysis. The main goal of data mining is to discover information patterns that lead to a better understanding of the information-generating process and predictions. One recent technique that has been developed to address these problems is the support vector engine. Support vector engine has been developed as a powerful tool for classification and regression in noisy and complex domains. Vector support is a very sophisticated and sophisticated machine learning algorithm, especially when it comes to predictive analytics. In the case of health-care diagnostics, it is as good as the Naive Bayesian classification, and it also supports a mechanism for prioritizing important medical information.

In this regard, we are working to improve the existing system using medical decision support systems. Using the Support Vector Machine (SVM), we use a vector-based classification technique to support medical decision support. The system will work faster and more efficiently using SVM. [7] It is widely used in real life due to its simplicity and efficiency, both in theory and practice. However, in large-scale problems that require a lot of training and data, such as logging, logging, training, and system testing, may be very necessary in terms of computing. So for a larger problem, simplifying the calculation is important. We use encryption techniques to maintain the privacy of patient information. And to maintain the privacy of tha information over the network, we are using a homomorphic encryption technique to encrypt the data again. All processing is done at the server side and encrypted data.

This means that the doctor interacts with the SMDSS to assist in the diagnosis and diagnosis of the patient. The Naive Bayes Extract, a popular automated data mining tool, has been widely used in the prediction of SMDSS [1].

Despite its simplicity But it is more appropriate for medical diagnostics in health care than some sophisticated techniques. The SMDSS, with Bayesian naive indicator, offers advantages over traditional health systems and opens new avenues for doctors to predict illness. Of patients [2].

However, success depends on understanding and managing information security and privacy challenges, especially at the discretion of the patient. One of the main challenges is keeping the patient's medical information away from unauthorized disclosure. The use of medical information may be of interest to a large number of people interested in medical treatment. For example, an online provider for online service providers will offer individual risk forecasts for patient illnesses. Patients may fear that such information will leak and use medical information and refuse to provide medical information to the SMDSS for diagnosis. Therefore, it is important to protect the patient's medical information. However, maintaining the privacy of medical information is not enough for all SMDSS to thrive. The provider classification used to predict patient illness can not be disclosed to third parties because the classifier is the provider's asset. Otherwise, third parties may use the classifier to predict the illness of the patient, which could damage the provider's income. Therefore, in addition to maintaining the privacy of patient medical information, how to protect the privacy of providers is also important for SMDSS patient-centered treatment, a systematic medical decision support support. The PPCD allows doctors to predict the risk of illness in a manner that preserves privacy. The system supports medical decision-oriented, patient-safe and conservative patient privacy, which allows providers to diagnose patients without the need to filter any patient's medical information. SMDSS provides knowledge. Specialized disease and information for physicians to improve diagnostic performance and improve the quality of care. It can elevate patient safety and improve the quality of medical care. By increasing the amount of data generated by every healthcare industry, researchers need a fast, accurate, and efficient algorithm for data analysis. [8] Improvements in database technology, computer efficiency, and artificial intelligence help. In the development of intelligent data analysis. The primary purpose of data mining is to discover patterns in information that lead to an understanding of the information-generating process and useful forecasting. The latest techniques that have been developed to address these problems are the support vector machines. Support vector engine has been developed as a powerful tool for classification and regression in noisy and complex domains. In this article, we use the Support Vector Machine (SVM) technique, one of the most effective classification techniques applied to real-world problems. This SVM has more advantages in terms of improving the Accuracy in diagnosis in decision support systems. We also use homomorphic encryption techniques to provide confidential information relating to patient health information. The rest of the material is in Section II. A partial bibliography is provided that provides brief information about the study conducted in the field of medical decision support systems of SMDSS.

II. Machine Learning

The machine will learn the inputs and inputs and then output the results for the new inputs according to the set data. There are two types of learning machine. Supervised Learning: Machines are presented with sample data and required output. When using this method, the machine learns the general rule. Inadvertent learning: The machine does not know the input and output information. The machine learns the steps and produces results.

III. Support Vector Machine

Vector support (SVM) has become a more popular tool in machine learning, including classification, regression, etc. SVM divides data into two categories and classifies and generates. The N-dimensional SVM hyperplanotype is maintained by the learning model used to classify the SVM. It acts as a line separator between two data points to identify two distinct layers for the multidimensional environment. SVM schema is in binary format. In one multi-layered problem, one must minimize the problem to a set of multiple binary classification problems.

Accuracy of classification is very high, up to 99.41% for 50-50% of the test partition, 100% training for 70-30% of the training partitions, and 100% for 80 - 20% Training - can find the test partition. You can also discover a combination of five data attributes, which may be important for a doctor to diagnose breast. Support vector machine is a modern categorization. Works well with real-world applications such as image classification, classification, etc. SVM is a standard tool for machine learning and data mining. And with so many applications and advantages, we use the SVM for our classification techniques offered in SMDSS.

IV. Related Work

Yogachandran Rahulamthavan, et. Everybody has examined the support system for medical decision support for privacy by using Gaussian classification. Describe cloud computing technology with a wealth of medical data. Using this system improves the decision-making ability of health professionals. For this purpose, the Paillier cryptographic system is used, but all encodings are based only on the medical data available on the machine. Only Gaussian kernels work in a simple domain and can not be modified to a medical server.

ErmanAyday, et al., Investing "Calculating the Privacy Treatment of Disease Risk through the Use of Genetic Information, Medical and Environmental Information," describes privacy in storing and processing machines in the system. For this, Homomorphic coding and privacy privacy metrics are used. Uses real patient information and is a reliable risk factor for disease. Works effectively for genomic data only. Identify disease risk tests using genomic data.

H. Monkaresi et. All have presented. "How to automatically learn to improve non-contact heart rate monitoring using a webcam." We have evaluated three methods of human-scale remote sensing: HCI-controlled laboratory work. And indoor exercises. This study evaluates the methodology of Pohet al., And illustrates the feasibility of measuring the residual human resource.

Ximeng Liu et. "Effective processing and privacy protection in the era of large data." This document examines the privacy of large data-era challenges by identifying major privacy needs. First, let's talk about whether private storage techniques are sufficient for large data processing. It also offers robust cosine analog processing protocols that help maintain privacy in response to the performance and personal needs of data mining in the era of large data.

Y. Tong et. Launches "Mobile Phone Access with Helpful, Private and Verifiable Health Information." The authors propose to create privacy in the mobile health system with the help of Private cloud We have a privacy-centric storage solution that integrates the key management of PRF into an unimaginable combination of concealment, search patterns, and redundant access. Create a privacy index for secure keyword searches. We also investigate techniques that help control access. (Both in normal and emergency) and authorized parties to avoid misconduct by incorporating the siren signatures controlled by the ABE with role-based encryption.

Tien Tuan et al. Presents "Sky Stream: Access Control for External Data Sources for Cloud Data Transfer." In this paper, we present a system that enables data access control. Precise flow of money through unreliable information. This allows the owner to encrypt data before forwarding it to the cloud. Encryption guarantees cloud confidentiality and fraudulent user access control. Current strength uses three common coding schemes: ABE Proxy Assignment Project and Window Slider Project. We have shown that the cloud can control access to encrypted conversations by authorized users without having to read clear text.

B. K. Samanthula et. All have entered "A k neighbor's secure query near the encrypted data in an external environment." This document provides two new SkNN protocols for cloud-encrypted data. The first protocol, which acts as a basic solution, filters some data to the cloud. On the other hand, our second protocol is completely secure, that is, it protects the confidentiality of user input and stealth access.

Y. Rahulamathavan et.al has introduced a "clinical decision support system for privacy using Gaussian Kernel-based classification." This document provides a decision support system that maintains privacy using the support vector engine. Based on Gaussian algorithms, this proposed algorithm introduces new techniques such as cloud computing. New York City (Or health knowledge) that exists in remote locations over the Internet without compromising privacy, thereby improving decision-making. Of health professionals.

J. Chen, H. Huang and et.al presented two performance evaluations (CDMs and MORs) for Naïve Bayes classifiers that are used in different types of text collections. They compared CDM and MOR with EOR, CMO and MC-OR. Three variants of opportunity ratios for multiple data sets.

R. Bellazzi et.al has introduced "clinical predictive data mining: current issues and approaches." The purpose of this review is to analyze the scope and role of countermeasures.

X. Yi and Y. Zhang have introduced "Bayesian innocence classification to maintain privacy in distributed data through heterogeneous mixers". Several protocols are proposed, depending on the mixer model in which each site The data will be sent to the two-part mixer, which will use two protocols and deliver the classification results. Because our protocols do not involve collusion between the two mixers, and there is no need for inter-site communication, it facilitates both the management and use of trust.

V. Proposed Architecture

We are using a vector-based classification technique to support clinical decision support. The system will work faster and more efficiently using SVM. [7] It is widely used in real life due to its simplicity and efficiency, both theoretical and practical. However, in large problems, which have large training data and should be used, such as traffic signal detection, training and testing procedures of this method may be necessary in terms of calculation. So for a larger problem, simplifying the calculation is important. We use encryption techniques to maintain the privacy of patient information. And to maintain the privacy of data sent over the network, we use a homomorphic encryption technique to encrypt new data. All processing is done on the server side and in

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).

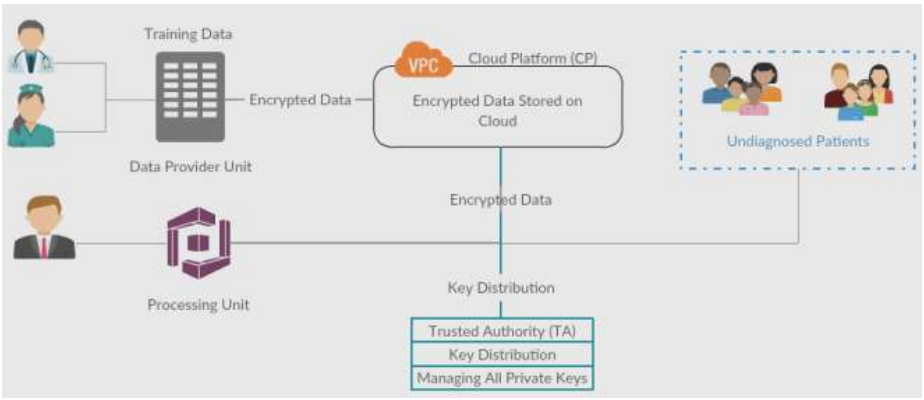


Figure 1) System Architecture

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.

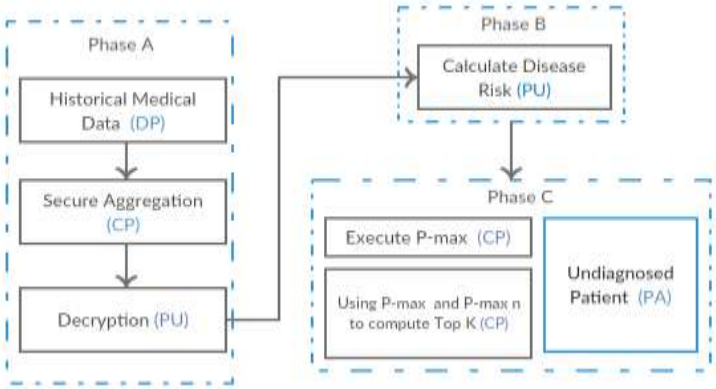


Figure 2) Process Flow Diagram

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

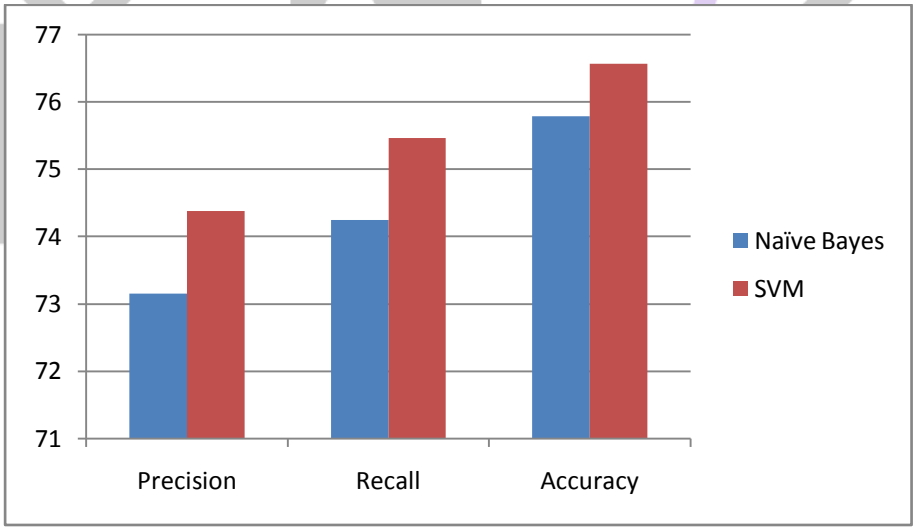
VI. Experimental setup Result and discussion

The basic flow of operation of the proposed method is shown in Figure 2 above. The step-by-step process of our system is described above. The results show that the medical decision support system using vector vectors provides better predictive results than the sibilifier. We have taken samples of the four most important diseases, namely cancer (lung cancer, breast cancer), heart disease, diabetes and chronic kidney disease. In terms of accuracy, when we compare the accuracy of the results generated using the SVM classifier, it is found that the predictions generated by the SVM classifier are 5% to 10% more accurate than the bias algorithm. The Naive Bayes predictive validity comparison graph and our SVM classification machine are shown in Figure 1. The SVM classifier provides predictions for all types of diseases. But 5 to 10 are more accurate than the Naive Bayes algorithm.

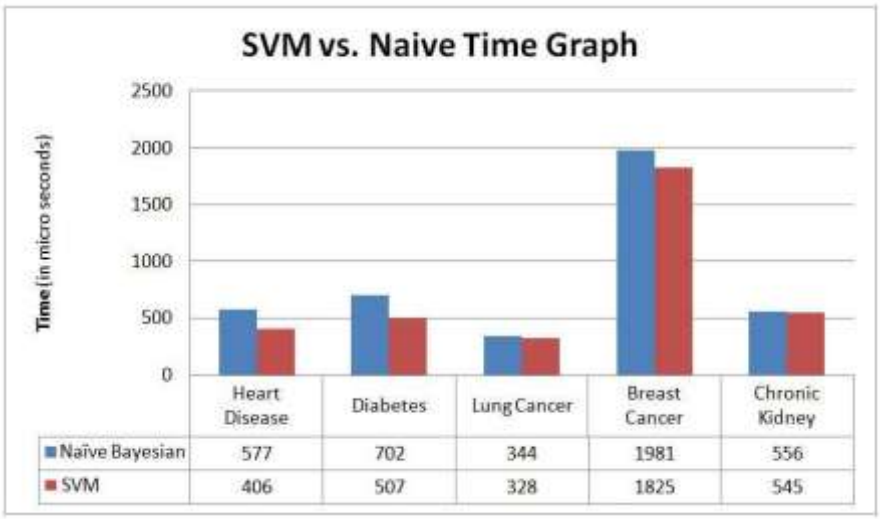
The results show that supportive financial decision support systems using vector-assisted machines provide better predictors of the sutures. We have taken samples of the four most important diseases, namely cancer (lung cancer, breast cancer), heart disease, diabetes and chronic kidney disease. In terms of accuracy, when we compare the accuracy of the results generated using the SVM classifiers, the predictions generated by the SVM classifiers are 5% to 10% more accurate than the algorithms. Naive bias.

Table.4.1.7 Comparisons among various parameters between Naïve bayes and Support Vector Machine algorithm

	Naive Bayes	SVM
Accuracy	75.78 %	76.56 %
Precision	73.15 %	74.38 %
Recall	74.24 %	75.46 %



Graph 1: Comparing accuracy of prediction between SVM and Naïve Bayes



Graph 2: Comparing Time Efficiency of prediction between SVM and Naïve Bayes

In terms of efficiency, the time to find the prediction of most chronic diseases, the traditional Bayesian algorithm will take more time compared to our algorithm in supporting vector vectors (SVM). The efficiency of time spent between SVM and Naïve Bayes is shown below (Figure 2). We will find that the SVM has more time and produces more accurate results than traditional naive algorithms.

Advantages:

By designing this system, we can do this.

- Significant improvement in diagnostic accuracy.
- Reduced diagnostic time allows for accurate prescriptions in less time.
- High rate of prediction without any burden.
- Reduce communication costs.
- Maintain patient privacy.

VII. Conclusion

In this paper, we have proposed Medical decision support system using the classification technique o data mining called Support Vector Machine. Using SVM, the computational time and diagnosis rate in our system gets improved. SVM has excellent performance in generalization so it can produce high accuracy in classification for diagnosis. The patient can securely retrieve top-k diagnosis result according to their own preferences. With the advantage of Homomorphic encryption technique, the patient's privacy over the cloud will achieved. The processing is done on the encrypted data, so that there is no loss in the privacy of patient's data while training the SVM classifier. These results evidentially proved the proposed method and showing the nice performance of classification accuracy based on dataset of Diabetes. From the implementation system we can conclude that the results of proposed work are better than result of previous work as the classification accuracy of SVM classifier is 76.56% which is more than the accuracy of Naïve bayes that is 75.78%.

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.