ISSN:0975-9646

PPDM and Data Mining Technique Ensures Privacy and Security for Medical Text and Image Feature Extraction in E-Health Care System

M.S Inthumathi[1] , P. Damodharan[2]
Student[1], Associate professor[2], Department of Computer science and engineering
Akshaya College of engineering and technology

Abstract

Objectives: **The main objective of this research is ensuring the privacy as well as security by improving the fully homomorphic data aggregation. It is also focused to achieve the higher classification results in the disease detection.**

Methods: **The existing method of PPDM is used to achieve the privacy in healthcare systems and proposed method of artificial neural network (ANN) algorithm is used to detect the disease more accurately.**

Findings: **The proposed method increases the performance in terms of higher precision, recall and reduction in time complexity.**

Application/Improvements: **The proposed system is done by using gray level co-occurrence matrix feature extraction and ANN approach. ANN method is used for identification of diseased images and improves the classification results significantly.**

Keyword: **Data mining, privacy preservation, security, features extraction and classification.**

1. INTRODUCTION

Data mining is used to extract the most important data from the given database. Data mining can be executed on data signified in quantitative, textual or multimedia forms. Categorization is defined as the recognition of novel patterns such as concurrence among several types. Data mining applications are widely utilized in the medical field to identify and recover the disease also to maintain the information securely. The algorithms and methods are suggested to focus and resolve the patient's issues and keep the sensitive information with higher privacy. Healthcare systems prominently make possible the health constraints observation, disease modeling and evidence based medical treatment. In [1] E. Villalba et.al discussed the topic about heart failure displaying system depends on knowledge expertises. This scenario is focused on the improvement of a heart failure organization system, which depends on daily monitoring of essential body signals along with wearable expertises for the constant evaluation of the long-term infection. However incorporating of the system into their routine is still lacking and it is a major drawback.

In [2], Taeho Jung et.al suggested few approaches how to achieve privacy as well as security in the cloud source more efficiently. This scenario is considered how an outside aggregator study few arithmetic data above participant's secretly owned statistics when preserving the information privacy. An efficient encryption protocols are introduced which guarantee information privacy and it achieves the real time scenario successfully. However it has still an issue with important information leakage. Tiziano et.al [3] discussed about the analysis of computational complexity and cryptographic concepts. The research scenario considered the execution of discrete Fourier transform with the help of cryptographic security systems. It is used to encrypt the most important informations from the document in more securely. However it has issue along with less efficiency as well as unfeasible implementation.

In [4], Claude castelluccia et.al discussed how to provide effective security data in wireless network. This scenario is introduced an easy and provably secure encryption method which permits effective aggregation of secured information. It is used to monitor the aggregation which depends on the pseudorandom function and it is also used to capably analyze the numerical values. However it has problem with communication overhead in this work. Klaus Kursawe et.al [5] suggested efficient protocols to aggregate the huge information along with encryption techniques. The current protocols which can be utilized to confidentially evaluate aggregation meter dimensions and also used to permit determination of fraud and leakage. In few cases the security is still an issue due to complicated conditions.

In [6] Jun Zhou et.al suggested authentication schemes and privacy models for secured healthcare systems. This scenario is introduced the model called as authorized accessible privacy model. To solve the inefficiency issues another scheme is improved named as privacy preserving cooperative authentication method. It is used to avoid the unauthorized persons to access the patient's information from the database. The scenario is achieved the lower computational and communication overhead. However it has issue with huge dimensional database. In [7] the scenario describes novel Private Stream Aggregation (PSA) algorithms which allow users to upload a stream of encrypted data to an untrusted aggregator, and allow the aggregator to decrypt (approximate) aggregate statistics for each time interval with an appropriate capability. We guarantee a strong notion of privacy. However it has issue with reliability factor.

In [8] Maryam Rajabzadeh et.al suggested an identity based multi proxy signature scheme to provide higher security level in this scenario. The identity based multi proxy signature is a kind of proxy signatures in which allocation of signing amongst several of proxy signers. In this kind of cryptographic primordial, cooperation of every proxy signers in the proxy collection produces the proxy signatures of approximately the identical size as that of standard proxy signatures on behalf of the innovative signer, which is more capable than broadcasting individual proxy signatures. Because identity-based multi-proxy signatures are helpful in disseminated schemes, grid computing, providing a provably secure identity-based multiproxy system is required.

2. Materials And Methods

2.1. Data aggregation with privacy mechanism

In this module, a method introduced named as PPDM which is built based on the efficient privacy-preserving fully homomorphic data aggregation. The main aim of the data aggregation is ensuring the privacy as well as security concept in the existing scenario. The data aggregation is the process of combining unique information of patient details in the dataset and it is in terms of addition and multiplication operation since for privacy with security mechanism. Fully homomorphic data aggregation scheme achieves the information-theoretic security for input privacy if and only if the following event takes place: even the adversary possesses the ciphertext $C_{u,i}$ with respect to the individual data m_i.

$$\text{H}(m_i|C_{u,i}) = \text{H}(m_i) \qquad (1)$$

Where H(·) and H(·|·) is the entropy function and conditioned entropy function in information theory.

2.2. PPDM1 technique

In this module, the scenario is introduced an efficient method named as PPDM1 by outsourcing the correlation function computation to the cloud on the basis of our proposed privacy preserving data aggregation scheme. This correlation extraction function is used to mine the text features from the specified image dataset.

PPDM1 Key generation

It uses the concept of aggregation of key generation procedure steps for privacy. Authorized persons and patients maintain the secret keys privately. N = pq can also be selected by the patient in the next step PPDM1.Enc, and the knowledge of N-factoring is required to be privately kept.

PPDM1 encryption

It uses the aggregation of encryption concept to encrypt each element in the modules. They are respectively performed by the patient's information and the authorized physician in the healthcare centre.

PPDM1 evaluation

It is observed that only multivariate polynomials composed of addition and multiplication operations, are required to compute.

PPDM1 decryption

The authorized people processing the secret key namely the knowledge of factoring the composite *N*, and recover the original correlation between dynamically collected PHI text vector and the template vector. Aggregation decryption fixes some threshold values and based on this value, we can conclude that the patient is suffering or recovering from one specific disease.

2.3. PPDM2 technique

In this module, a secure and efficient privacy-preserving medical image feature extraction scheme is introduced. By exploiting the newly-designed privacy-preserving data aggregation holding the property of full homomorphism, the issue of representing and extracting medical image features can be realized in the encrypted domain while retaining the inherent properties when operating in the plaintext domain. Difference of- Gaussian (DoG) is a technique for extracting the features from the specified images.

The medical image is required to convolve with Gaussian filters assigned with different variance for each scale and the difference between two neighbouring Gaussian-blurred images is taken. Then, feature points are chosen as local extreme of the DoG images occurring at multiple scales. The medical Dog image generated from neighboring scales $\rho_i \text{ and } \rho_j$is symbolized as

$$\text{DoGImg}(x,y,\rho_{i,j}) = Conv_G(x,y,\rho_i) - Conv_G(x,y,\rho_j) \quad (2)$$

Where $1 \leq x \leq X, 1 \leq y \leq Y$, X, Y respectively refer to the horizontal and vertical sizes of the medical image the convolution of the medical image with Gaussian Kernal *G* which is defined as for any *x* and *y*,

$$Conv_G(x,y,\rho_i) = G(u,v,\rho_i) * I(x,y) = \sum_{u,v} G(u,v,\rho_i) I(x-u,y-v) \qquad (3)$$

Where * identifies the convolution process. It uses the aggregation of key generation for achieving the privacy. As usual the processes have been done by encryption and decryption then ensure the patient details with greater privacy.

Achieve security

In this module, perform the feature point detection by using encrypted data comparison. The encrypted feature descriptor matching step compares the descriptors of the medical image collected from the patient and the template medical image from the authorized physicians by computing the similarity value via some certain judging metric. The DoG algorithm performs superior for security against attacks and also achieved higher security for corresponding input privacy as well as output privacy.

2.4. Classification technique for disease prediction

In this module, Gray Level Co occurrence Matrix (GLCM) feature method is proposed for efficient feature extraction to improve the classification of disease accuracy for the given dataset. And also introduced classification algorithms such as SVM for predict the patient disease more accurately.

SVM algorithm:

```
Candidate Support Vector (SV) = {closest pair from opposite classes}
While there are violating points do
Find a violator
CandidateSV = candidateSV ∪ violator
If any $\alpha_p$<0 due to addition of c to S then
CandidateSV = candidateSV\p
Repeat till all such points are pruned
End if
End while
```

2.5. Performance evaluation

In this module, we compare the existing scenario and proposed scenario by using efficient methodologies. In existing scenario, we used the methods are named as PPDM1 and PPDM2 for achieving the privacy as well as security successfully. In proposed scenario we introduced feature extraction method and classification method to identify the patient disease more effectively. From the experimental result, we can conclude that our proposed scenario yields higher performance rather than existing scenario.

3. Results And Discussion

The existingPPDM1 and PPDM2 technique is used to discover the privacy as well as security for the specified dataset. In the existing system, the privacy and security is achieved by using the methods of PPDM1 as well as PPDM2 successfully. In proposed system, the diseased parts are identified efficiently by using an ANN classification algorithm. From the experimental result, the conclusion displays that the overall performance is increased in the proposed scenario by using the feature extraction and classification method more effectively. The performance metrics are in terms of precision, recall and time complexity superior in the proposed scenario.

3.1. Precision

Precision is defined as the Percentage of correct predicted results from the set of input terms. The precision value should be more in the proposed methodology than the existing approach for the better system performance.
Precision is calculated by using following equation

Precision = $\frac{\text{True Positive}}{\text{True Positive+False Positive}}$

In this graph, x axis is taken for two methods of and y axis is taken for precision. From the Figure.1 the proposed scenario shows the highest precision rather than existing method. ANN classification method provides higher detection of diseased images in proposed method.

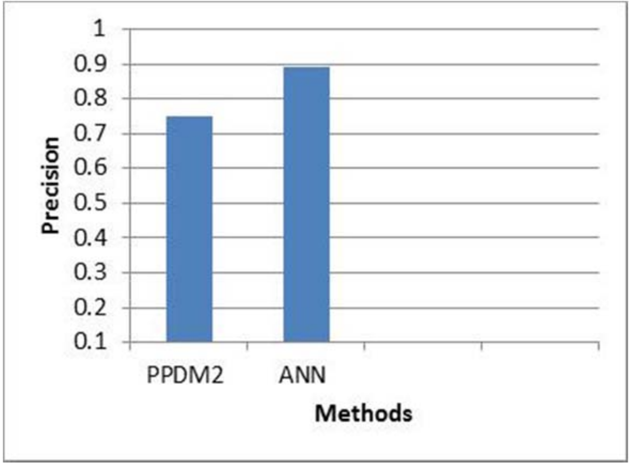

Figure 1. Precision comparison

3.2. Recall

The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation:

Recall = $\frac{\text{True Positive}}{\text{True Positive+True Negative}}$

In this graph, x axis is taken for two methods of and y axis is taken for recall. From the Figure.2 the proposed scenario shows the highest recall rather than existing method. ANN classification method provides higher detection of diseased images in proposed method.

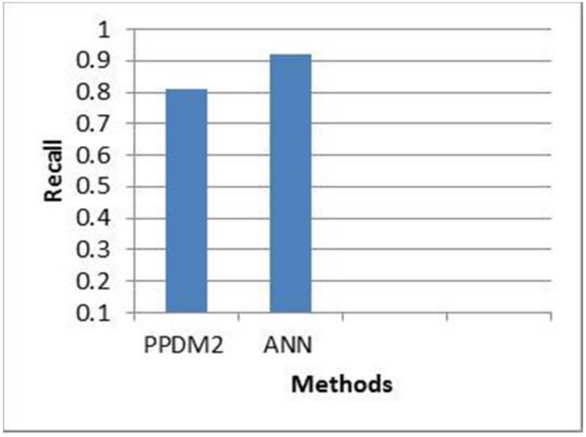

Figure 2. Recall comparison

3.3. Time complexity

In computation, the algorithms are estimated to reduce the time complexity. For number of files the existing and proposed algorithms are executed in various time factor values. The less time execution values called higher performance in the scenario which is provided by using proposed algorithm.
In this graph, x axis is taken for two methods of and y axis is taken for time complexity. From the Figure.3 the proposed scenario shows the lower time complexity rather than existing method. ANN classification method provides higher detection of diseased images in proposed method.

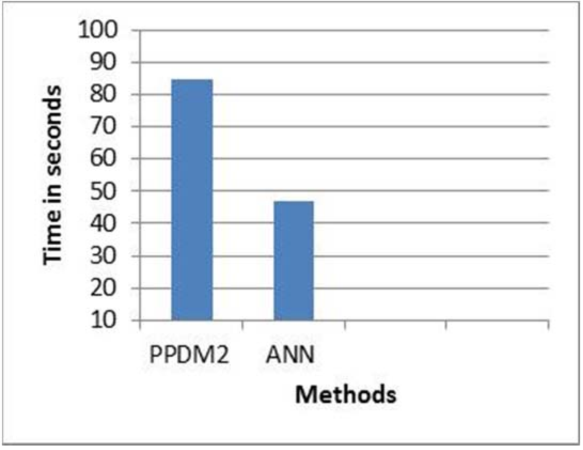

Figure 3. Time comparison

4. Conclusion

In this section, the conclusion decides that the proposed scenario produces superior performance rather than existing scenario. An efficient privacy preserving fully homomorphic data aggregation is introduced and disease modeling is ensured by using method of correlation matching PPDM1. And also it is ensured through privacy preserving medical image feature extraction PPDM2 method. The proposed classification method named as artificial neural network which is used to discover and classify the diseased images more effectively in the given dataset. The experimental result has shown that the proposed system is better in terms of precision, recall and time metrics.

ACKNOWLEDGEMENT

We the authors assure you that, this is our own work and also assure you there is no conflict of interest.

REFERENCES

1. Villalba, Elena, et al, Heart Failure monitoring system based on Wearable and Information Technologies, Journal of Communications 2.2 (2007): 10-21.
2. Jung, Taeho, et al, Privacy-preserving data aggregation without secure channel: Multivariate polynomial evaluation, INFOCOM, 2013 Proceedings IEEE. IEEE, 2013.
3. Bianchi, Tiziano, Alessandro Piva, and Mauro Barni, On the implementation of the discrete Fourier transform in the encrypted domain, Information Forensics and Security, IEEE Transactions on 4.1 (2009): 86-97.
4. Castelluccia, Claude, et al, Efficient and provably secure aggregation of encrypted data in wireless sensor networks, ACM Transactions on Sensor Networks (TOSN) 5.3 (2009): 20.
5. Kursawe, Klaus, George Danezis, and Markulf Kohlweiss, Privacy-friendly aggregation for the smart-grid, Privacy Enhancing Technologies, Springer Berlin Heidelberg, 2011.
6. Zhou, Jun, and Zhenfu Cao, PSCPA: Patient Self-controllable Privacy-preserving Cooperative Authentication in Distributed m-Healthcare Systems, IACR Cryptology ePrint Archive 2012 (2012): 44.
7. Erkin, Zekeriya, and Gene Tsudik, Private computation of spatial and temporal power consumption with smart meters, Applied Cryptography and Network Securit, Springer Berlin Heidelberg, 2012.
8. Asaar, Maryam Rajabzadeh, Mahmoud Salmasizadeh, and Willy Susilo, Security Pitfalls of a Provably Secure Identity-based Multi-Proxy Signature Scheme.