DEVO-DNFN: A Novel Approach for Secure Medical Data Classification in the Cloud

Suruchi Nannaware[1], DR. Hiren Dand[2]
[1]*Research Scholar, Shri JJT University Rajasthan*
[2]*PhD Guide Shri JJT University Rajasthan*

***Abstract*— In cloud Computing, storing and accessing sensitive information is a major concern. The paper tries to address these issues by providing the novel approach for classifying privacy preserved medical data, named DEVO. It integrates the two optimization algorithm namely, Energy Valley Optimizer (EVO) and Dingo Optimizer (DOX). Cloud simulation is the first step in the process, which gathers medical data from the dataset as an input. The privacy of data is then maintained in the cloud environment by using DEVO to create a privacy utility coefficient matrix using deep learning principles. The resulting privacy-preserved data is safely kept in the cloud and can only be retrieved by authorized parties with the same key. In addition, DEVO-DNFN uses Deep Neuro Fuzzy Network (DNFN) to classify medical data, which is then refined with DEVO. Assessment metrics like True Positive Rate (TPR), accuracy, and True Negative Rate (TNR) show encouraging results, having observed values of 0.912, 0.907, and 0.915, respectively. This comprehensive method addresses the important issues of privacy and data security in cloud-based medical data classification.**

Keywords: cloud computing, DNFN, optimization, Privacy preservation

I. INTRODUCTION

Cloud computing revolutionize the way businesses can store, process the data and manage various applications. Fundamentally, cloud computing offers various computing services—such as software, storage, processing power, servers and networking using the internet as opposed to depending on local servers or personal computers. An "pay-per-use" service is used to supply the resources in cloud computing. Its computational architecture is web-based [1]. The healthcare industry makes extensive use of cloud computing, a digital technology [23]. Cloud computing can perform a wide range of tasks, including computation, data storage, sharing, and analysis, without revealing to organizations the details of its implementation and platform [11]. Many companies use the cloud environment to host their model and data from local devices because of these limitless services [9] [2]. Some of the best examples of cloud computing are Gmail, Microsoft Azure, Google App Engine, IBM cloud, Amazon Elastic Compute Cloud and Google Docs. When it comes to storage methods, the majority of cloud areas usually provide three forms of storage mainly block storage for storing data in fixed sized blocks, file storage is used for storing files and folders while object storage treats data as an objects that have corresponding metadata [24].

Cloud computing has the potential to be revolutionary, but it also comes with risks including vendor lock-in, security, compliance, and data privacy. For this reason, strong governance frameworks and risk management techniques are essential. Data security is the major concern when it comes to cloud storage and retrieval processes. Due to this technological issue, clients experience data insecurity when the security of their data is not guaranteed, and they are unable to provide case studies pertaining to the services [1]. Information and data are safe when privacy is maintained, regardless of the type of data. As the number of patients with chronic diseases rises, hospitals are having to test more and more patients' health conditions, which drives up costs for healthcare systems [26]. Decision tree classification is frequently used in the development of health monitoring systems, which assist patients in periodically testing their conditions in an effort to lower healthcare costs and enhance the quality of care [27]. Health monitoring systems operate as follows, in conjunction with recent developments in wearable technology and mobile communication networks: a hospital first creates a clinical decision model by applying the decision tree classification technique, after which it tests the

biomedical information obtained from various wearable devices and using the model it makes decision for the patients [29] [30]. Hospitals often outsource health monitoring services to cloud servers in order to save costs and facilitate practical deployment. This approach has several advantages for both the hospital and its clients, including scalability, ease of management, and ubiquitous access to health information. The hospital also uses this information to make decisions for its clients based on the model [31]. Data encryption is one of the most important security guaranteeing systems; encryption should always be guaranteed, no matter how sensitive the data. Cloud data privacy protection is crucial since user information may be exposed by the service provider to hackers; therefore, privacy measures are essential to ensure security [1]. Various privacy preserving classification schemes have been put forth to safeguard confidentiality of biomedical data and clinical decision models. The majority of currently used convolutional neural network privacy-preserving techniques are based on homomorphic encryption or differential privacy [18] [21]. Homomorphic encryption is used to evaluate the lower order polynomials on encrypted data while differential privacy method adds random noise to data which may cause unavailability of data [26]. Multiple parties can collaboratively and privately classify data using decision trees by employing SMPC-based schemes, but they may result in high communication costs. While traditional cryptographic techniques help solve problems, they are expensive in terms of memory and computation, and their direct implementation is more difficult than that of Deep Neural Networks (DNNs) [20] [22]. Users are allowed to train the global framework without centrally training the data using Federated learning, but it is unable to safeguard privacy after a model is deployed on an unreliable cloud server and test data is being inferred [22].

A. Subramaniyam et al. [1] proposed a TGD based ACNN that provides an efficient way to allow classification of medical data in cloud computing. This approach required less processing time and was simple to maintain. But it didn't take hybrid optimization techniques into account to enhance the classification. Liang, J., et al. [3] developed POMIC scheme for the adaptive cloud-based classification of medical images using a convolutional neural network. The POMIC served as a computationally efficient, storage, and communication system for privacy preserving health monitoring; however, because of the increased complexity of various biomedical features, it had a high time complexity. Block-wise scrambled images and a modified ConvMixer were used by Li, J., et al. [5] to create a privacy-preserving technique that reduced the manipulation of image encryption. This approach worked well with the ciphertexts' deep iterations and private deep learning backgrounds, but it was unable to increase the encryption efficiency. Singh, A.K., Gupta, R., et al. [4] developed PPMD that provides security to sensitive data in the cloud environment. Using this model, the data is divided into sensitive and non-sensitive segments, reducing the impact of added noise. It did not, however, allow the sharing of gathered data with other users upon request. Partially Homomorphic Encryption (PHE) was developed by Du, J. and Bian, F. [7] to process private data without affecting the underlying data. This method had less computational overhead and a shorter run time. However, the technique's security was not improved.

The main objective of this study is to develop and implement a workable method using linear polarized data for preserving privacy of the classified medical data stored in a cloud environment. First, the medical data is taken from the dataset and used as input. Next, using Dingo Energy Valley Optimization, the privacy utility coefficient matrix for cloud data privacy is created. The key matrix's coefficients are then created by generating the linearly polarized data to ascertain the privacy of data. The constructed key matrix is used to generate the cumulative key matrix for finding the second key coefficient. The platform stores data in such a way that when a third party needs to get the information from cloud, they ask for the same key. Finally, the DNFN, trained by the DEVO, is used to carry out the classification of medical data to preserve privacy.

This paper's contribution is to the design a DEVO algorithm which is a combination of Energy Valley Optimizer (EVO) and Dingo Optimizer (DOX) that contributes to protect the privacy of medical data stored in the cloud by generating the privacy utility coefficient matrix.

1.1 Challenges

- The TGD-based ACNN that was created in [1] was unable to enhance classification performance since it did not use any additional hybrid optimization techniques during classifier training. Furthermore, additional datasets were not used to evaluate this model's performance.
- Even though the PPMD method in [4] maintained a high degree of accuracy and guarded against data leakage or theft, it was unable to create an effective privacy-preserving system to secure the data for different owners.

The suggested method is motivated by the problems with the current approaches, and a novel approach to classifying medical data while protecting privacy is developed.

2. PROPOSED ALGORITHM

Many patients' medical records are stored in the cloud computing environment, where they can be retrieved whenever needed. Medical records contain extremely sensitive information about the patient, that must be kept private and not be published. Preserving privacy is essential to keeping the data reliable. Protecting the privacy and usefulness of data, as well as providing accurate classification for disease diagnosis, is the main objective of the work. To safeguard the confidentiality of data stored in cloud, where users' data is kept in an environment protected by a key so that a third party using the key can access the data kept in the cloud. Assume that the input data is given as follows:

$$E = d_{ij}; 1 \leq i \leq l; 1 \leq j \leq m \qquad (1)$$

where, dij represents the ith row data element and jth column data element of the database E, l denotes the overall data records, and m shows the number of features in the database. The database dimension is represented as $l \times m$

An optimization algorithm is used to determine the linear kernel that converts the database's data into privacy-preserved data. The input data E is given as input to linear polarized data and the PU matrix is generated to create key matrix's coefficients and linear polarized matrix is denoted as follows;

$$T = E \otimes P \qquad (2)$$

Here, $\otimes$ is the linear interpolation operator, and P denotes the PU coefficient, which is generated by the proposed DEVO. The PU coefficient is depicted as in figure 1.

P 1	P 2	…… ….	P n

Figure 1. The arrangement of PU coefficient

where, P1, P2…… Pn are the indexes of PU coefficient

DEVO was developed to generate a privacy utility coefficient matrix in order to secure medical data kept on the cloud. EVO [12] and DOX [13] are combined to create DEVO. EVO is a metaheuristic algorithm that falls into two classes: exploitation and exploration. It is based on the principles of physics and takes into account various modes and particle decay stability. EVO is used to enhance the optimization process's precision for a variety of problems. But it didn't take into account challenging optimization issues in various domains or actual engineering design problems. DOX is an algorithm inspired by the hunting behaviors of dingos. It can handle a wide range of constraints and offers reliable solutions. The multiobjective issues were not resolved, though. By overcoming the previously mentioned constraints, the novel DEVO handles a wide range of optimization problems as well as produces more optimal solutions. The process begins with initializing the search space by using the EVO algorithm. the particles' fitness within the search space is calculated. Next, Enrichment Bounding (EB) is used to conduct the divergence that occurs between neutron rich and neutron poor particles. Afterwards, the stability level of the particle is determined. Product's stability level is increased by emitting the rays in the physical reaction. γ rays are released to increase the stability level of the excited particles and are described as an extra position updating process of EVO that results in the production of a novel solution. In this stage, a position updating procedure is carried out in order to generate the subsequent particle. As previously stated, in order to advance the algorithm's exploration and exploitation stages, a new position updating procedure is carried out. If the current best solution proves to be more fit than the previous one after all tasks have been completed, it will be kept. If not, the most effective solution from the previous iteration will be applied as the new one. Until the best solution is found; the previously stated steps are followed.

3.METHOD

The cumulative linear polarized data is created by adding the individual columns of linear polarized data, which is done after the linear polarized data is generated to produce the coefficients of the key matrix. A vital function of the key matrix is to determine data privacy. Confidential information can be preserved by combining the key matrix and original data. CKM depends on the generated PU coefficient and also the key coefficient. The key matrix is used to establish the privacy-preserved medical data with the intermediate element.

3.1 Medical data classification for Privacy preservation

To carry out the classification, the data obtained is supplied as input to the DNFN. DEVO, which is generated by combining EVO and DOX, performs the DNFN training. A type of artificial intelligence called DNFN combines fuzzy concepts with neural networks (NNs). Here, the Deep Neural Network's (DNN) learning capacity and the fuzzy systems' reasoning capacity are combined. Here, the reasoning power of fuzzy systems and the learning capacity of deep neural networks (DNNs) are combined. Fuzzy inference and fuzzy pooling are the two operations that make up the layers in this network.

3.2 Training using DEVO algorithm

The DNFN is trained using the DEVO algorithm, which is used to classify medical data while protecting privacy. Fig 1 shows the structural design of the DEVO-DNFN system

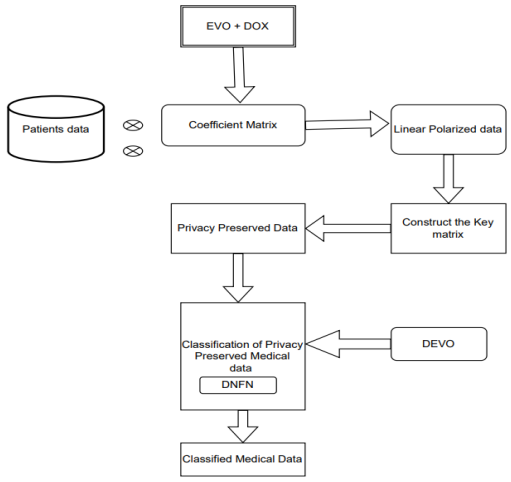

Figure 1. The structural design of the devised DEVO-DNFN

4. RESULTS AND DISCUSSION

The next section goes into detail about the evaluation that was done for this system and compares the DEVO-DNFN with current approaches. The developed DEVO-DNFN is used with a Python tool implementation on a PC.

4.1Dataset description

Two common databases—the Cleveland and Hungary— has been taken from the UCI repository for the simulation purpose. This model uses the Heart Disease Database [25] to obtain the medical data. Although there are 76 features in this repository, only 14 of them are used in the works that are currently available.

4.2 Evaluation measures

The accuracy, TNR, and TPR metrics are considered to evaluate the performance of DEVO-DNFN model.

a) Accuracy: It represents the ratio of correctly classified health information to all health data fed into the model. It is stated as follows

$$Accuracy=\frac{v_1' + \gamma_1'}{v_1' + \gamma_1' + v_2' + \gamma_2'}$$

where v_1', indicates true positive and v_2' indicates false positive , r_1' terms true negative while false negative is given by r_2' .

b) TNR: It is the ratio of the quantity of health data in the simulation that is classified as true positive data to the total amount of positive data.

$$\text{TNR} = \frac{r_1'}{r_1' + r_2'}$$

c) TPR: It is a ratio of the total amount of medical data in the model that is classified as negative to the amount of true negative data classification.

$$TPR = \frac{v_1'}{r_1' + r_2'}$$

4.3 Comparative methods

The traditional techniques used to evaluate the developed DEVO-DNFN are PPMD [4], POMIC [3], DA-PMLM [2], and ACNN [1].

4.4 Comparative assessment

The study assesses the DEVO-DNFN's ability to classify medical data using two datasets: Cleveland and Hungary. The results are explained in the following sections.

4.5 Using Hungarian dataset

Using the Hungary dataset, the DEVO-DNFN's performance is examined in comparison to training data. The evaluation is depicted in figure 4. Figure 4a presents the accuracy assessment for the DEVO-DNFN. The accuracy attained by the methods ACNN, DA-PMLM, POMIC, PPMD, and the DEVO-DNFN with 90% of the training data is, respectively, 0.741, 0.798, 0.866, 0.870, and 0.912. Compared to the PPMD, DEVO-DNFN has 4.54% more accuracy. The DEVO-DNFN analysis is shown in terms of TPR in Figure 4b. When the training data is 90%, the calculated TPR values of the applied methods ACNN, DA-PMLM, POMIC, PPMD, and the suggested DEVO-DNFN are, respectively, 0.740, 0.844, 0.873, 0.878, and 0.915. The evaluation based on TNR is shown in Figure 4c. With 90% training data, the achieved TNR values of the applied techniques ACNN, DA-PMLM, POMIC, PPMD, and the DEVO-DNFN are, respectively, 0.855, 0.821, 0.778, 0.751, and 0.915. Compared to the ACNN, the TNR values of the proposed DEVO-DNFN are 5.10 percent higher.

Figure 2a. Comparative assessment of a accuracy

Figure 2b. Comparative assessment of TPR of the DEVO-DNFN with other methods based on training data using Hungary dataset.

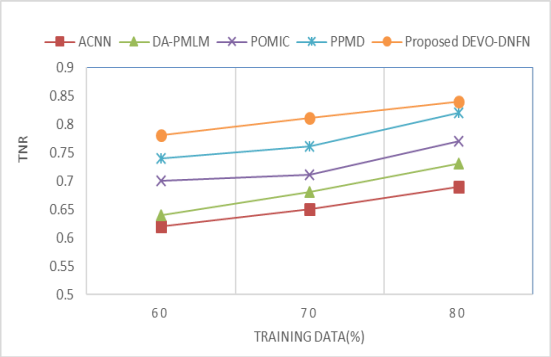

Figure 2c. Comparative assessment of TNR of the DEVO-DNFN with other methods based on training data using Hungary dataset.

4.6 Comparative Discussion

Table 1 gives the comparative analysis of DEVO-DNFN. In this case, the effectiveness of DEVO-DNFN is evaluated using metrics such as TPR, TNR, and accuracy by contrasting it with the most popular methods for classifying medical data. Because DEVO is used to structurally optimize the DNFN used for data classification, the established DEVO-DNFN exhibits high accuracy, TPR, and TNR. Moreover, the privacy preservation was improved by converting the input data into a linearly polarized form.

Table 1. A comparative analysis of DEVO-DNFN

Metric	*Accuracy*	*TPR*	*TNR*
ACNN	0.741	0.74	0.736
DA-PMLM	0.798	0.844	0.791
POMIC	0.866	0.866	0.866
PPMD	0.87	0.87	0.87
Proposed DEVO-DNFN	0.912	0.915	0.907

5. CONCLUSION

This study created a DNFN model that was DEVO-optimized for classifying medical data that protects patient privacy. The initial step involves obtaining the input medical data from the dataset. Next, by creating the PU coefficient matrix using the suggested DEVO, the privacy of the cloud data is preserved. Subsequently, the linearly polarized data is produced in order to produce the key matrix coefficients. After that, a key matrix is constructed to ascertain the data's privacy. The second key coefficient is found using the cumulative key matrix that is created from the built key matrix. Data privacy is maintained and the data is stored on cloud servers. Finally, the medical data classification is completed by the DNFN. In this particular case, the developed DEVO trains DNFN. With accuracy values of 0.912, TPR values of 0.915, and TNR values of 0.909, the suggested DEVO-DNFN performed better than alternative approaches. Further goals include using large datasets and incorporating cutting-edge deep learning as well as optimization techniques as part of a future research direction.

REFERENCES

[1] A. M. R. a. S. P. Subramaniyam, ""“Taylor and gradient descent-based actor critic neural network for the classification of privacy preserved medical data”," *Journal og Big data,* vol. vol.7, no. no.3, pp. pp.176-191, 2019.

[2] R. a. S. A. Gupta, ""“A differential approach for data and classification service-based privacy-preserving machine learning model in cloud environment”," *New Generation Computing,* vol. vol.40, no. no.3, pp. pp.737-764, 2022..

[3] Q. Z. H. X. H. a. K. F. Yu, ""“POMIC: Privacy-Preserving Outsourcing Medical Image Classification Based on Convolutional Neural Network to Cloud”," *Applied Sciences,* vol. vol.13, no. no.6, p. pp.3439, 2023.

[4] A. a. G. R. Singh, ""“A privacy-preserving model based on differential approach for sensitive data in cloud environment”," *Multimedia Tools and Applications,* vol. vol.81, no. no.23, pp. pp.33127-33150, 2022..

[5] J. K. X. L. S. M. X. a. T. Y. Li, ""“Privacy preservation for machine learning training and classification based on homomorphic encryption schemes”,," *Information Science,* vol. vol.526, pp. pp.166-179, 2020.

[6] Y. K. H. L. H. a. C. J. Kim, ""“Privacy-preserving parallel kNN classification algorithm using index-based filtering in cloud computing”," *Plos one,* vol. vol.17, no. no.5, p. pp.e0267908, 2022.

[7] J. a. B. F. Du, ""“A privacy-preserving and efficient k-nearest neighbor query and classification scheme based on k-dimensional tree for outsourced data”," *IEEE Access,* vol. vol.8, pp. pp.69333-69345, , 2020.

[8] J. a. L. D. Park, ""“Parallelly Running and Privacy-Preserving k-Nearest Neighbor Classification in Outsourced Cloud Computing Environments”," *Electronics,* vol. vol.11, no. no.24, p. pp.4132, 2022.

[9] S. Y. G. M. Y. a. D. R. Xu, ""“Secure fine-grained access control and data sharing for dynamic groups in the cloud”,," *IEEE Transactions on Information Forensics and Security,* vol. vol.13, no. no.8, pp. pp.2101-2113, 2018.

[10] N. a. K. N. Gomathi, ""“Ontology and hybrid optimization based SVNN for privacy preserved medical data classification in cloud”," *International Journal on Artificial Intelligence Tools,* vol. vol.28, no. no.03, p. pp.1950009, 2019.

[11] Xu, L., Tian, C., Zhang, G., Li, L., Tian, W. and Zhang, Y., , ""“PPDRM: Privacy-Preserving DRM Training and Classification on the Cloud”," *Available at SSRN 4460902, ,* 2023..

[12] Azizi, M., Aickelin, U., A. Khorshidi, H. and Baghalzadeh Shishehgarkhaneh, M., " “Energy valley optimizer: a novel metaheuristic algorithm for global and engineering optimization”," *Scientific Reports,* vol. vol.13, no. no.1, p. pp.226, 2023.

[13] Bairwa, A.K., Joshi, S. and Singh, D., ""“Dingo optimizer: a nature-inspired metaheuristic approach for engineering problems”," *Mathematical Problems in Engineering,* pp. pp.1-12,, 2021.

[14] Javaid, S., Abdullah, M., Javaid, N., Sultana, T., Ahmed, J. and Sattar,, "“Towards buildings energy management: using seasonal schedules under time of use pricing tariff via deep neuro-fuzzy optimizer”,," in *In Proceedings of 15th international wireless communications & mobile computing conference*, WCMC, June 2019..

[15] Gayathri, S. and Gowri, S., "“CUNA: A privacy preserving medical records storage in cloud environment using deep encryption”," *Measurement: Sensors*, vol. vol.24, p. pp.100528, 2022..

[16] Wei, L., Zhu, H., Cao, Z., Dong, X., Jia, W., Chen, Y. and Vasilakos, A.V., "“Security and privacy for storage and computation in cloud computing”," *Information sciences*, Vols. vol.258,, pp. pp.371-386, 2014.

[17] Stergiou, C. and Psannis, K.E., "“Efficient and secure big data delivery in cloud computing”," *Multimedia Tools and Applications*, vol. vol.76, pp. pp.22803-22822, 2017.

[18] Pulido-Gaytan, B., Tchernykh, A., Cortés-Mendoza, J.M., Babenko, M., Radchenko, G., Avetisyan, A. and Drozdov, A.Y., "“Privacy-preserving neural networks with homomorphic encryption: C hallenges and opportunities”," *Peer-to-Peer Networking and Applications*, vol. vol.14, no. no.3, pp. pp.1666-1691, 2021.

[19] H. L. X. L. R. a. L. H. Zhu, "“Efficient and privacy-preserving online medical prediagnosis framework using nonlinear SVM”," *IEEE journal of biomedical and health informatics*, vol. vol.21, no. no.3, pp. pp.838-850, 2016.

[20] R. a. S. V. Shokri, "“Privacy-preserving deep learning”," in *In Proceedings of the 22nd ACM SIGSAC conference on computer and communications security,*, October, 2015.

[21] Liang, J., Qin, Z., Xue, L., Lin, X. and Shen, X., " “Efficient and privacy-preserving decision tree classification for health monitoring systems”," *IEEE Internet of Things Journal*, vol. vol.8, no. no.16, pp. pp.12528-12539, 2021.

[22] Qi, Z., MaungMaung, A. and Kiya, H., " “Privacy-Preserving Image Classification Using ConvMixer with Adaptative Permutation Matrix and Block-Wise Scrambled Image Encryption”," *Journal of Imaging*, vol. vol.9, no. no.4, p. pp.85, 2023..

[23] S. A. K. W. H. a. W. F. Chenthara, "“Security and privacy-preserving challenges of e-health solutions in cloud computing”," *IEEE access,*, vol. vol.7, pp. pp.74361-74382, 2019.

[24] Yang, P., Xiong, N. and Ren, J., "“Data security and privacy protection for cloud storage: A survey”," *EEE Access*, vol. vol.8, pp. pp.131723-131740, 2020.