A Secure Predictive Framework for Preventing Health Care Data

Bhargavi Konda, Akhila Reddy Yadulla, Vinay Kumar Kasula*, Mounica Yenugula, Supraja Ayyamgari

Department of Information Technology, University of the Cumberlands, KY, USA

Abstract The incorporation of Artificial intelligence (AI) with online resources in the healthcare sector has significantly enhanced medical services. Facilitating accurate diagnoses, tailored treatment strategies, and ongoing patient surveillance, Internet of Things (IoT) devices promote efficient communication, while AI analyzes detailed healthcare data to improve decision-making and reduce costs. Nevertheless, securing data storage and transmission poses a significant challenge, especially as the threat of data breaches and cyber-attacks increases. Protecting patient privacy and securing health records is essential to maintaining public health. To address these challenges, a new approach called Butterfly Optimization Based Modular Neural Network (BObMNN) has been developed, focusing on data security and predictive performance. The healthcare database was first assembled and imported into a Python environment for further analysis. Following this, data security protocols were established using encryption and decryption methods. The encrypted data was then subjected to preprocessing, specifically feature selection, where butterfly optimization (BOA) was utilized to determine the most important attributes for predictive analysis. The constructed model was evaluated using a variety of measures, including Area under the Curve (AUC), accuracy, Recall, F-score, and Precision.

Keywords Data Security, Butterfly Optimization, Preprocessing, Modular Neural Network, Encryption, Decryption.

1. Introduction

The worldwide healthcare sector is experiencing a significant transformation due to the digitalization of patient and health information. This change is driven by a variety of factors, including lifestyle changes, an aging population, advances in software applications, innovative treatments, and the principles of evidence-based medicine [1]. This transformation offers prospects for improved clinical decision-making, healthcare delivery, and disease surveillance [2]. The utilization of data analytics within the healthcare sector presents numerous benefits alongside certain drawbacks [3]. Every year, there has been a growth in security and confidentiality concerns in health data [4]. Healthcare organizations realize that bottom-up, reactive, and technology-focused measures are insufficient to safeguard patients and organizations [5]. Therefore, innovative information systems and prevention-focused measures to curb leaks and enable effective use of massive healthcare data are needed [6]. Healthcare information, for example, patient data, is highly sensitive and personal in the modern digital age [7]. As the healthcare sector more and more employs digital technologies and data-driven products, privacy must be safeguarded [8]. Involuntary access or disclosure of confidential data can have severe issues, such as identity theft, insurance fraud, and medical identity theft [9]. These issues can also obstruct medical research and innovation, diminish public confidence in healthcare organizations, and create legal and financial problems for healthcare organizations [10]. Conventional measures such as encryption, access controls, and guideline compliance have proved effective in protecting

---
*Correspondence to: Vinay Kumar Kasula (Email: vinaykasula.phd@ieee.org). Department of Information Technology, University of the Cumberlands, KY, USA.

data [11]. Nonetheless, the evolving cyber security threat environment and the complexity of healthcare data systems make these measures insufficient [12]. AI can assist in enhancing healthcare data privacy by continuously monitoring who is accessing data, identifying unusual user behavior, and identifying potential threats in real time [13].

AI-based analysis of user behavior can make access controls more robust and enable threats to be identified earlier [14]. AI plays a critical role in maintaining data anonymity, enabling researchers and healthcare professionals to analyze large data sets without compromising patient privacy [15]. Data integration, interoperability, and sharing in healthcare are critical for providing secure services [16]. Cloud computing can help overcome e-health hurdles, resulting in cost savings and less reliance on technical people [17]. Saudi healthcare firms are embracing cloud computing, utilizing an innovative paradigm for mobile e-health multimedia apps [18]. Big data may be used to extract useful patterns, detect fraud, manage risks, and reduce costs, but it also poses issues in data collecting, storage, and security [19]. However, structured data was mostly taken into consideration in those earlier studies. For instance, employing Decision Tree (DT), K-nearest Neighbour (KNN), convolutional neural networks (CNNs), and Naive Bayesian (NB) algorithms have already garnered unstructured data to extract features [20]. Still, as far as we are aware, no prior effort has dealt with medical imaging data. To the best of our understanding, there have been no previous attempts to tackle clinical imaging data [21].The key contribution of the research is given below,

- The health care database is first collected and imported into the Python environment as input data.
- A new BObMNN was developed, focusing on data security and prediction.
- Following this, the process of securing the database is executed through encryption and decryption methods.
- The decrypted information is then processed using BOA to extract relevant features for predictive analysis.
- Ultimately, the proposed method was utilized to predict both affected and non-affected diseases.
- The developed model's success was compared to other models.

The second part of this article presents contemporary, relevant research; Part three deals with the sampling issue; Section four elaborates on the challenges; Section five explores the case study and capability verification of the innovative framework. This work finally concludes in Part Six.

2. Related Works

A recent literature review described it as follows,

Mohammed et al. [22] have developed a new algorithm called "Pattern-Proof Malware Validation" (PoPMV) to improve security in industrial cyber-physical systems (ICPS). The algorithm employs a deep learning framework alongside reinforcement learning methods to improve processing efficiency, identify potential attacks, and optimize the functionality of Integrated Cyber-Physical Systems (ICPS). Simulation results show a 30% improvement in security, but the algorithm requires significant computational resources, potentially leading to longer processing times.

Singh et al. [23] developed a safe data fusion aggregation approach for the Internet of Medical Things (IoMT) ecosystem. This approach to data quality is done by evaluating link quality and choosing active sensors through the Archimedes Optimization Algorithm. When implemented in NS2.35, this method demonstrates superior performance compared to alternative approaches in terms of energy consumption, computational cost, reliability, network connectivity, latency, and communication overhead. However, it still faces challenges in scalability.

Nadhan et al. [24] Researched a network based on cryptographic techniques that are designed for the encryption and decryption processes of images, displaying promising implications in the sharing of medical images with deep learning technologies. ResNet-50 is the main learning structure that is utilized to convert image representations, improve encryption algorithms on a specific domain, and use reconstructive networks in the decryption process. The system is very reliable in assessing the effectiveness of treatments and has a Return on Investment framework. The research utilizes open-source data and findings of security analyses; however, it demands high computational power.

Premi et al. [25] have proposed a framework to protect sensitive healthcare information in IoT-based Wireless Sensor Networks (WSNs). The framework includes secure data transmission protocols, encryption methods, access

control mechanisms, and authentication protocols. It considers the impact of GDPR on healthcare systems using IoT technology. The article emphasizes the importance of ethical data management and periodic security audits to ensure the use of healthcare IoT and enhance patient care.

Madavarapu et al. [26] built an Advanced Integrated Data Security (AIDS) architecture particularly developed for cloud-based healthcare systems to solve challenges such as data breaches and illegal access. This framework incorporates various components, including anomaly detection, access control, intrusion prevention, and encryption techniques. It safeguards data confidentiality, regulates user permissions, and identifies irregular patterns. However, it faces challenges in suboptimal detection accuracy.

3. System Model with a Problem Statement

The existing models used for disease prediction and health data security have several challenges, including low prediction accuracy due to improper feature selection, lack of adequate optimization strategies, and inherent security flaws. Furthermore, traditional models often face scalability and generalization issues, which reduce their usefulness when dealing with diverse clinical datasets. Information loss in recorded healthcare data can lead to inaccurate predictions and poor clinical judgment. Furthermore, different techniques can lead to delays in clinical decision-making, which has a negative impact on diagnosis and treatment planning.
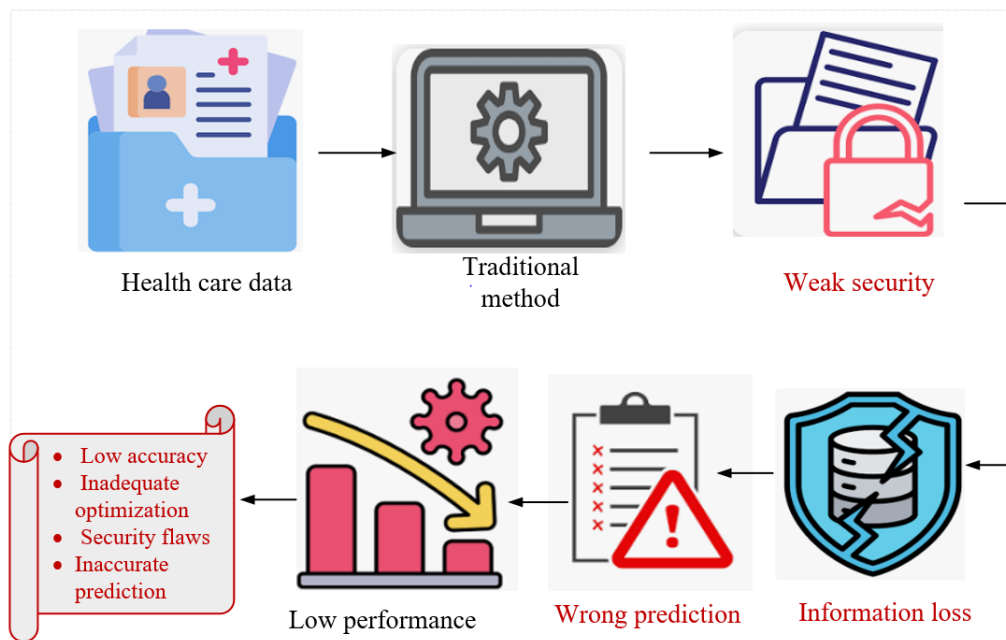

Figure 1. Limitation of the traditional approach

The model for addressing the challenges is depicted in Figure 1. Recent technologies face various limitations, particularly in terms of weak security and information loss. Furthermore, problems such as low accuracy and false predictions persist. In addition, high noise levels prevent the identification of important features required for accurate predictions. As a result, research efforts have focused on developing a new approach to intrusion prediction to overcome these problems.

4. Proposed Methodology

The research presents a novel Butterfly Optimization-based Modular Neural Network (BObMNN) that emphasizes data security and predictive capabilities. Initially, the healthcare database was compiled and imported into the Python environment for processing. Subsequently, the data security measures were implemented through encryption and decryption techniques. The decrypted information underwent preprocessing, specifically feature selection, utilizing BOA to identify the most informative features for prediction purposes.
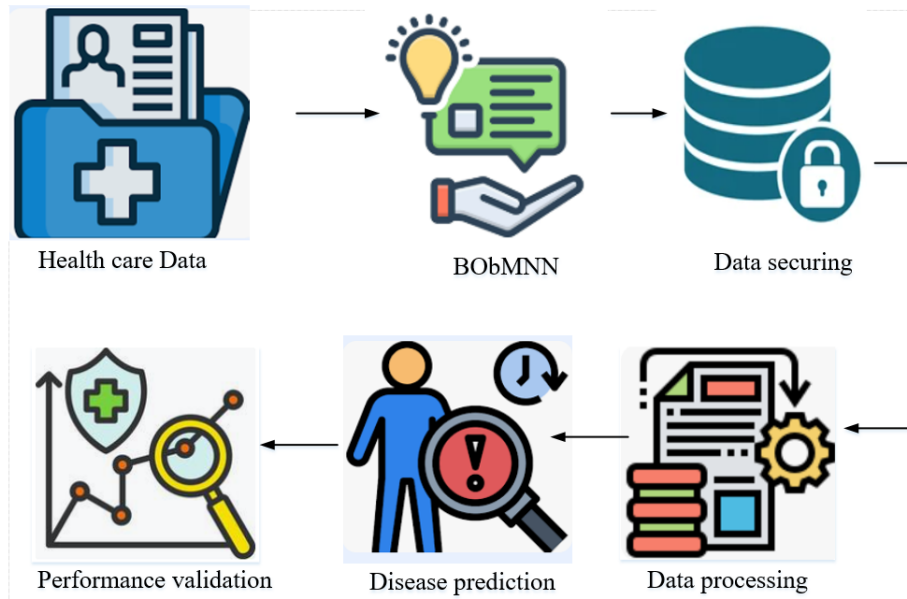

Figure 2. Architecture of developed model

The primary objective of BObMNN is to achieve accurate predictions while ensuring data security. The metrics assessed include accuracy, Precision, recall, F-score, and AUC. Furthermore, an overview of the design methodology is provided. The architecture that has been proposed is depicted in Figure 2.

4.1. Process of the Proposed Bobmnn

The novel BObMNN is a metaheuristic algorithm based on BOA and MNN. BOA mimics the foraging behavior of butterflies, allowing for efficient convergence to a global optimum. MNNs are artificial neural network architectures that decompose complex tasks into manageable blocks, each managed by a separate network. These modules operate autonomously, focus on specific data characteristics, and integrate outputs to produce a definitive conclusion. The healthcare database was imported into Python, and data security protocols were established using encryption and decryption methods, followed by preprocessing using BOA for feature selection for predictive analysis, ensuring the integrity of the data.

4.1.1. Data initialization The data was first collected from the std ga0ggle website and subsequently imported into the Python environment for use as input data. The system then trains on the healthcare data. The initialization process is described in Eqn. (1).

$$H(i) = (H_1, H_2, H_3, \ldots, H_n) \tag{1}$$

Here, i it indicates the initialization process variable, H signifies the quantity of data points in the dataset, n represents the entire amount of information within the dataset.

4.1.2. Data securing Database security is ensured through the application of robust encryption and Decryption methods, which are vital for maintaining the privacy and validity of information related to health. Initially, a public key is generated for data encryption, which ensures that access to the information is limited to authorized personnel. The XOR function is employed in both the encryption and decryption procedures. This encryption procedure converts plaintext data into an unintelligible format by utilizing a cryptographic function. From a mathematical perspective, the encryption process can be represented in Eqn. (2).

$$C = E(k_{\text{Public}} \oplus O_{\text{Data}}) \tag{2}$$

Where C indicates the encrypted data, E indicates the encryption function, k_{Public} indicates the key, O_{Data} indicates the original data. This encryption technique ensures the protection of confidential health data against unauthorized access.

Once the data has been securely stored, a private key is generated to facilitate the decryption of the information, thereby allowing authorized individuals to access the original data exclusively. The decryption process reverses the encryption, transforming the ciphertext back into a comprehensible format. Mathematically, the decryption process can be represented as follows Eqn. (3).

$$O_{\text{Data}} = D(K_{\text{Private}} \oplus C) \tag{3}$$

Here, O_{Data} indicates the decrypted original data, D suggests the decryption function, K_{Private} indicates the private key, C and indicates the encrypted data. This dual-key cryptographic system, which employs both public and private keys, establishes a robust framework for preventing breaches of healthcare data. This ensures the protection of sensitive patient information against unauthorized access and cyber threats.

4.1.3. Data processing The decrypted data undergoes a processing phase to ensure that it is clean, well-structured, and prepared for analysis. The initial stage involves preprocessing, which aims to eliminate irrelevant features, address any missing values, and standardize the dataset for uniformity. This stage is critical for increasing the quality of the data before the process of feature extraction. Mathematically, the preprocessing stage can be represented in Eqn. (4).

$$\text{Pre} = F_{\text{norm}}(D_{\text{raw}}) - F_{\text{noise}}(D_{\text{raw}}) \tag{4}$$

Where Pre refers to the dataset that has undergone preprocessing, D_{raw} represents the original decrypted data, F_{norm} indicates the normalization function utilized to ensure consistency, and F_{noise} denotes the function for noise removal, which is responsible for eliminating unnecessary and redundant features from the dataset.

Following the preprocessing phase, the feature selection process commences, utilizing the BOA to determine the essential features present in the dataset. BOA simulates the foraging behavior of butterflies, using a scent-tracking mechanism to effectively explore and exploit significant features while reducing the dimensionality of the data. The feature selection process can be mathematically articulated in Eqn. (5).

$$Fs = \arg\max\left(\sum_{i=1}^{n} W_i \cdot X_i\right) \tag{5}$$

Here, Fs denotes the chosen optimal features, X_i refers to a specific feature, W_i indicates the weight allocated to that feature according to its significance, and n signifies the overall count of features. Through the application of BOA, the framework adeptly identifies the most essential attributes, thereby reducing computational complexity and enhancing the accuracy of the healthcare data security model.

4.1.4. Disease prediction The suggested approach employs features from MNN to categorize diseases, utilizing deep learning methodologies to evaluate the extracted data. It effectively differentiates between affected and unaffected cases, thereby guaranteeing precise predictions. The procedure is mathematically expressed in Eqn. (6).

$$D_p = \begin{cases} if(Fs = 0) & \text{Non affected} \\ if(Fs = 1) & \text{Affected} \end{cases} \tag{6}$$

Here, The variable Dp denotes the outcome of disease prediction while Fs referring to the optimized set of features obtained through feature selection. The MNN establishes the threshold value to distinguish between affected and non-affected cases. The MNN improves disease prediction accuracy by reducing false positives and negatives, ensuring healthcare data security and patient diagnosis through reliable classification of input data.
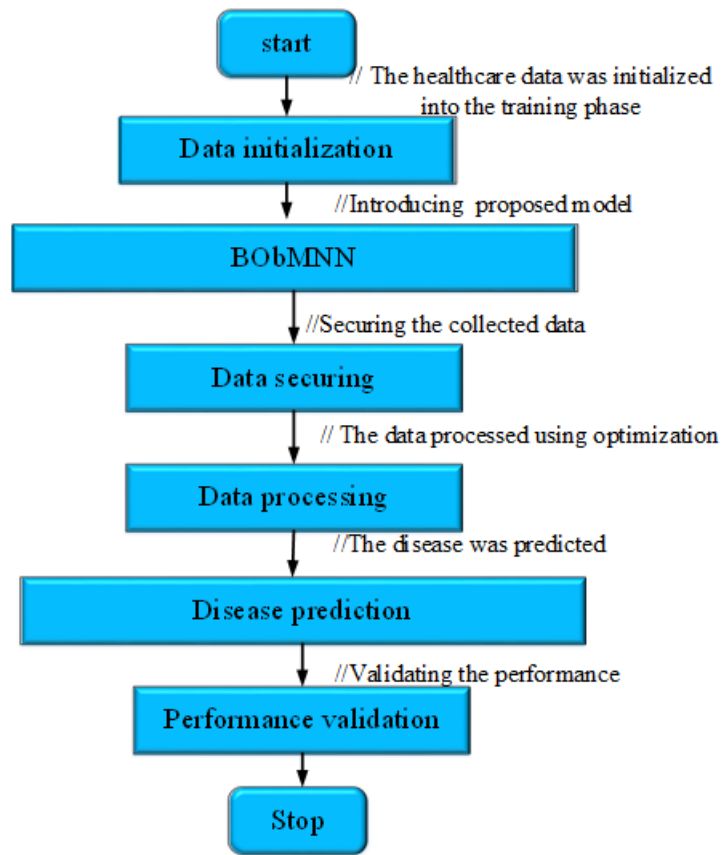
Figure 3. Flow chart of BObMNN

Figure 3 illustrates the flowchart that sequentially depicts the operational process of the proposed model. The procedure for the suggested model is detailed in Algorithm 1 and presented in pseudocode format.

5. Result and Discussion

The BObMNN is a hybrid approach developed in Python, aimed at improving the efficiency and Precision of healthcare data analysis through the integration of optimization techniques and deep learning methodologies. This approach effectively reduces dimensionality while maintaining critical patterns, enabling the MNN to concentrate on pertinent medical features. Evaluation of the model on a healthcare dataset demonstrated

---

Algorithm1: BObMNN

Start

{

      Data initialization ()

      {

          int $H(i), H_1, H_2, H_3, ..., H_n$

             // initialize the input variables

             initialize $\to H_1, H_2, H_3, ..., H_n$

             // The data was initialized

      }

      Data securing ()

      {

          int $C, E, k_{\text{public}}, \oplus, O_{\text{data}}$

             // initialize the encryption process variables

             encrypted data $\to E(k_{\text{public}} \oplus O_{\text{data}})$

             // The data was securely encrypted

          int $O_{\text{data}}, D, K_{\text{private}}, \oplus, C$

             // initialize the decryption process variables

             Original data $\to D(K_{\text{private}} \oplus C)$

             // The data was decrypted

      }

      Data processing ()

      {

          int $Pre, F_{\text{norm}}, D_{\text{raw}}, F_{\text{noise}}$

             // initialized the preprocessing variables

             Remove $\to F_{\text{noise}}(D_{\text{raw}})$

             // The noise characters are removed

          int $Fs, argmax, W_i \cdot X_i, i, N$

             // initialize the feature selection variables

             Selected feature $\to X_i$

             // The needed features are extracted

      }

      Disease prediction ()

      {

          int Dp, Fs

             // initialize the disease prediction process variables

             if $Fs = 0$

             Non affected

             if $Fs = 1$

             affected

             // based on the predicted features, diseases are classified

      }

}

Stop

---

significant improvements in classification accuracy. The proposed method's validity has been confirmed through comprehensive analysis and meticulous assessment of the selected dataset. The operational variables are outlined in Table 1.

Table 1. Parameter specification

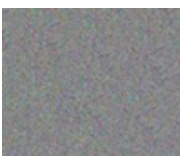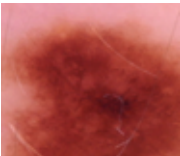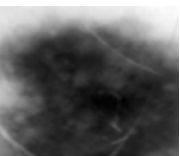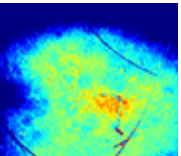| Parameter	Specification
Programming platform	Python
Operating system	Windows 10
Optimization	Butterfly
Network	Modular Neural Network
Dataset	Skin cancer
Dataset format	Image
Total samples	10,015

5.1. Case Study
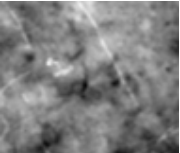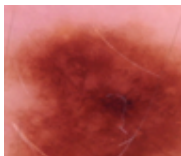
The suggested methodology was subjected to thorough validation tests and a case study to assess its efficacy. The fulfillment of the developed model was evaluated utilizing advanced BobMNN alongside a dataset, specifically a skin cancer dataset sourced from the official Kaggle platform. The outcomes were carefully structured and analyzed to determine their effectiveness. (https://www.kaggle.com/datasets/farjanakabirsamanta/skin-cancer-dataset). The collection of 10,015 dermatoscopic images serves as a significant asset for academic pursuits in machine learning. It encompasses various diagnostic categories related to pigmented skin lesions, such as keratoses, carcinomas, melanomas, and vascular lesions. The dataset's reliability is reinforced by histopathological validation and in-vivo confocal microscopy, establishing a solid basis for the detection and classification of diseases. The dataset details are presented in the Table 2.

Table 2. Dataset description

Total samples	10,015
Training (80%)	8012
Testing (20%)	2003

The dataset contains 10,015 images, including 80% utilized in training and 20% for testing. The model is designed to discover patterns and categorize skin lesions, with 8,012 data used in training and 2,003 for testing.

Table 3. Outcomes

Class	Input image	Encrypted image	Decrypted image	Preprocessing	Feature selection
Non affected					
Affected					

The input image data is subjected to a data protection process that ensures the secure encryption and decryption of the information. Following this, the decrypted data is analyzed to remove noise components and highlight

important features. Advanced techniques are employed to forecast diseases based on the characteristics observed in the imaging. Outcomes are shown in the table 3.

5.2. Comparative Analysis

The assessment of capability scoring entails the analysis of essential metrics, including F-score, AUC, Recall, Accuracy, and Precision, which have been utilized to evaluate the model's success. The evaluation of the suggested model involves the integration of various previously recognized models. These established models comprise. Emotional intelligence-enhanced dynamic Bayesian network (EI-EDBN) [27], CNN [27], Deep CNN (DCNN) [27], Fast Region-based CNN (FRCNN) [27]. Homomorphic [28], Privacy-Preserving Deep Learning (PBDL) [28], Game Theory-Based Spectrum Sharing (GT-BSS) [28], Advanced Encryption Standard (AES) [28], Lionized remora optimization-based serpent (LRO-S) [28].

5.2.1. F1 score The evaluation of the method's effectiveness was conducted utilizing the F1 score, which incorporates both Precision and Recall. Calculating the F1 measure expressed in Equation (7).

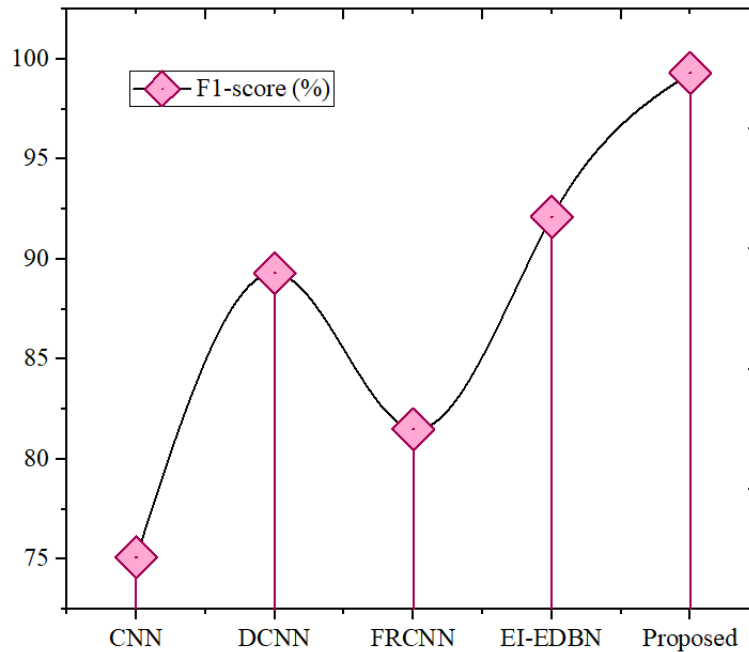$$F1\text{-}score = 2 * \frac{pre * acc}{pre + acc} \tag{7}$$

Figure 4. F1- score correlation

The correlation of F1 scores indicates that CNN obtained a score of 75.1, DCNN achieved 89.3, FRCNN reached 81.5, and EI-EDBN recorded 92.1. Notably, the proposed model demonstrated superior achievements with an F1 score of 99.3, reflecting a well-balanced and highly efficient classification capability. Figure 4 illustrates a comparison of F1 scores.

5.2.2. Precision The prediction for each model was evaluated by calculating the ratio of total true positives (TP) to the sum of TP and FP. The method for calculating Precision is outlined in Equation (8).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

The comparison of Precision indicates that the CNN recorded a score of 78.4, the DCNN achieved 91.4, the FRCNN reached 89.1, and the EI-EDBN obtained 95.4. Notably, the proposed model surpassed all others with a precision of 99.3, thereby ensuring enhanced accuracy in the classification of affected cases while effectively reducing false positives.

5.2.3. Precision Recall is calculated by taking the total number of TP divided by the total of TP and FN. The formula for determining Recall is outlined in Equation (9).

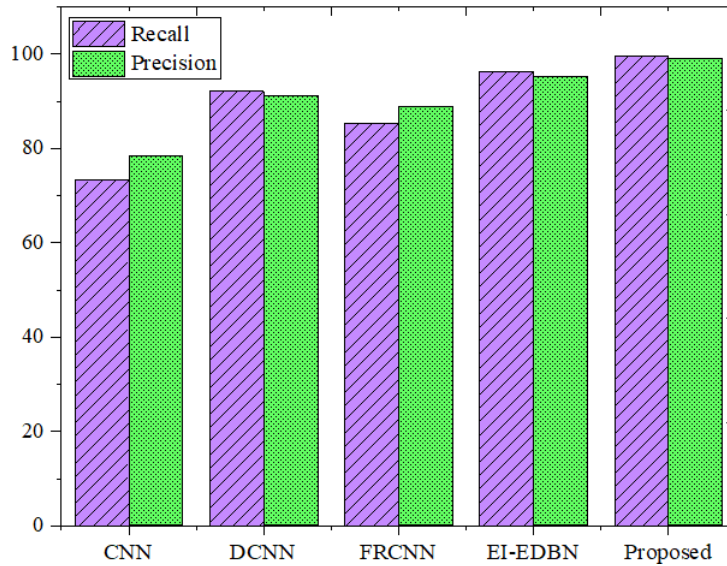$$\text{Recall} = \frac{TP}{FN + TP} \tag{9}$$

Figure 5. Recall and Precision correlation

The suggested model outperformed existing models in disease prediction, with a recall of 99.5. Compared to the previous model, CNN recorded a recall of 73.2. DCNN showed a notable improvement, while the FRCNN achieved 85.3 recalls. The EI-EDBN surpassed with 96.3 recalls. The model's exceptional recall capability, coupled with its ability to identify positive cases and reduce false negatives accurately, makes it a more reliable and effective method in healthcare data analysis. The Recall and Precision correlation is illustrated in Figure 5.

5.2.4. Accuracy Accuracy serves as an indicator of Precision, which is when a model categorizes all instances. This concept is represented in Equation (10).

$$Accuracy = \frac{TP + FP}{FP + TN + FN + TP} \tag{10}$$

The results indicate that CNN obtained an accuracy score of 84.3, while DCNN achieved 87.4. FRCNN recorded a score of 83.2, and EI-EDBN reached an impressive 97.3. Notably, the proposed model exceeded all others with an accuracy of 99.6, showcasing its exceptional reliability in accurate disease classification. The evaluation of accuracy is illustrated in Figure 6.

5.2.5. AUC curve AUC is a statistical metric used to evaluate the performance of binary classification models. The initial phase in calculating the area beneath the curve involves plotting the ROC curve.
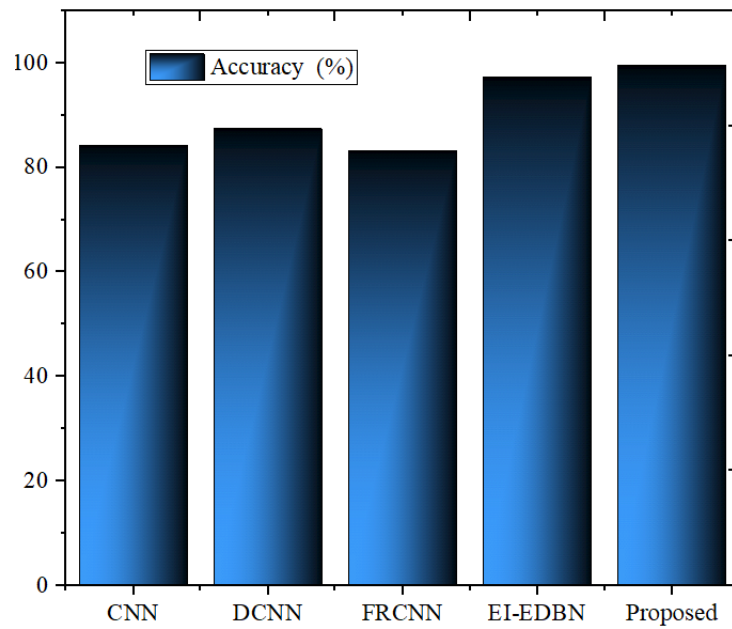
Figure 6. Correlation of accuracy
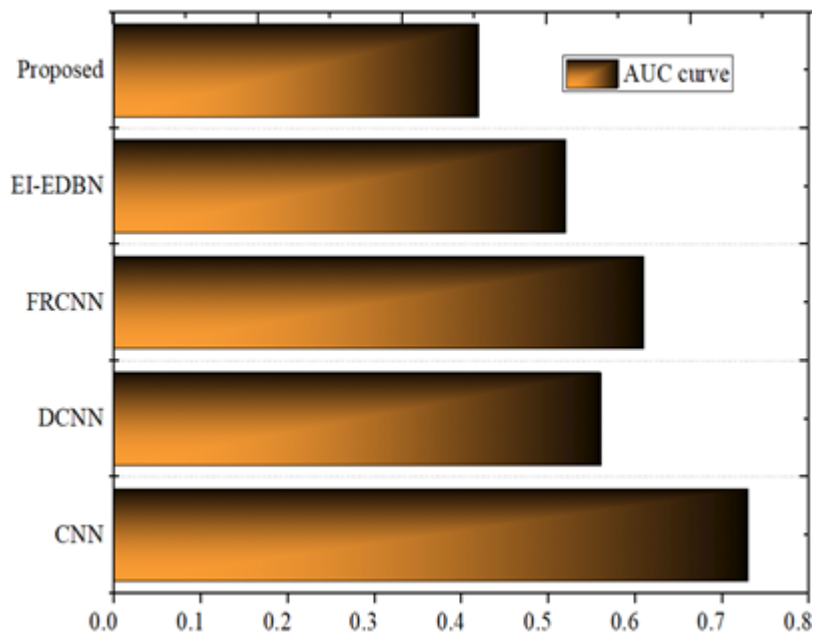
Figure 7. Correlation of AUC

The illustration of the AUC correlation is presented in Figure 7. The AUC study shows that the CNN model scored 0.73, whereas the DCNN model scored 0.56. The FRCNN model scored 0.61, whereas the EI-EDBN model has an AUC of 0.52. In contrast, the proposed model demonstrated an AUC of 0.42, highlighting its unique classification attributes in comparison to the other models. Table 4 presents the overall comparison values.

Table 4. Overall comparison

Methods	AUC	Recall	Accuracy	F1 score	Precision
CNN	0.73	73.2	84.3	75.1	78.4
DCNN	0.56	92.1	87.4	89.3	91.4
FRCNN	0.61	85.3	83.2	81.5	89.1
EI-EDBN	0.52	96.3	97.3	92.1	95.4
Proposed	0.42	99.5	99.6	99.3	99.3

5.2.6. Encryption time The duration required to transform data from its original plain format to an encrypted cipher format is referred to as encryption duration. The calculation of encryption time is represented in Equation (11).

$$ET = \frac{ED}{T} \tag{11}$$

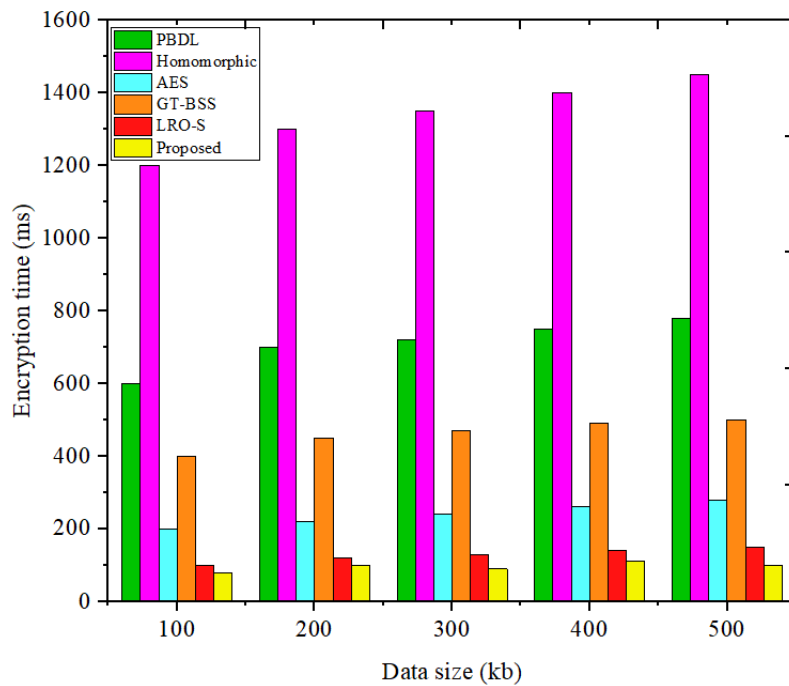In this context, ED signifies the size of the encrypted data, while T represents the time.

Figure 8. Comparison of encryption time

The research examined the encryption time metrics associated with various data sizes measured in kilobytes. For a data size of 100 kb, the encryption times varied from 1200 ms for the Homomorphic method to 120 ms for the developed method. In the case of 200 kb, the encryption times were recorded as 1300 ms for Homomorphic, 700 ms for PBDL, 450 ms for GT-BSS, 220 ms for AES, and 120 ms for the created method. Figure 8 shows a correlation of encryption times.

5.2.7. Decryption Time The duration required to convert encoded data back to its original form is known as decryption time. The calculation of decryption time is also detailed in Equation (12).

$$DT = \frac{DD}{T} \tag{12}$$

In this context, DD represents the decrypted data, while T signifies the time.
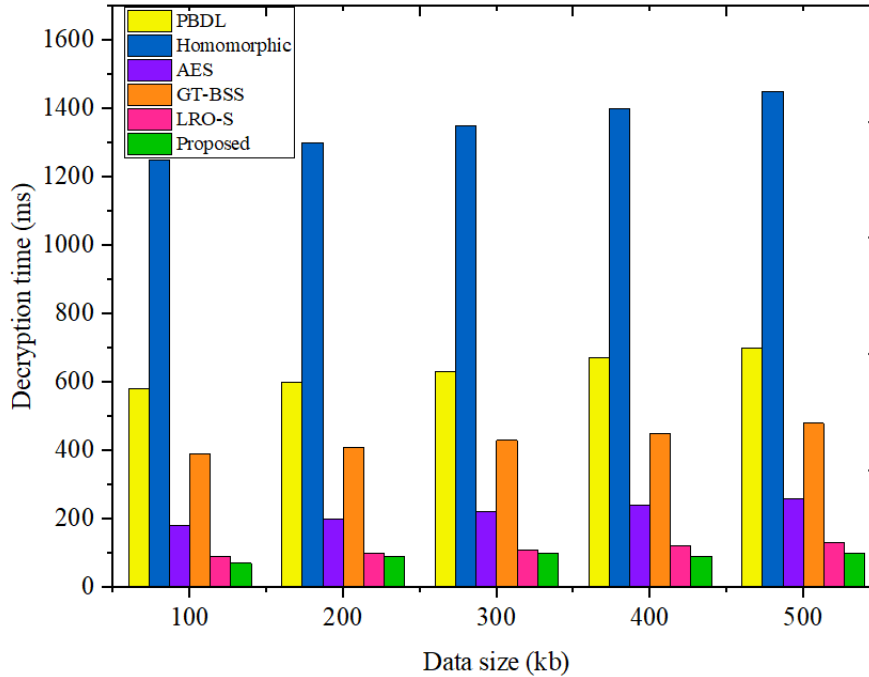

Figure 9. Correlation of decryption time

Figure 9 shows a correlation of decryption times. The proposed method has different decryption times for various data sizes in kilobytes. The complexity of 1250 ms for Homomorphic, 580 ms for PBDL, 390 ms for GT-BSS, 180 ms for AES, and 130 ms for the developed method. The times for encryption and decryption are compared in Table 5.

Table 5. Encryption and decryption time comparison

Methods	PBDL (ms)		Homomorphic (ms)		AES (ms)		GT-BSS (ms)		LRO-S (ms)		Proposed (ms)	
Data Size (kb)	ET	DT	ET	DT	ET	DT	ET	DT	ET	DT	ET	DT
100	600	580	1200	1250	200	180	400	390	100	90	80	70
200	700	600	1300	1300	220	200	450	410	120	100	100	90
300	720	630	1350	1350	240	220	470	430	130	110	90	100
400	750	670	1400	1400	260	240	490	450	140	120	110	90
500	780	700	1450	1450	280	260	500	480	150	130	100	100

5.3. Discussion

The evaluation of fulfillment indicated that the BObMNN model attained outstanding predictive outcomes, exceeding those of earlier research studies. This advancement considerably improves the Precision of skin cancer predictions and bolsters data security. Additionally, the findings of the study confirmed the enhancements in

the model's accomplishment through an extensive comparative analysis. Detailed achievement metrics for the BObMNN model are provided in Table 6.

Table 6. Achievements of the developed BObMNN

Metrics	Percentages
Precision	99.3
Recall	99.5
AUC curve	0.42
F1 score	99.3
Accuracy	99.6
Encryption time	96 ms
Decryption time	90 ms

This model's F1 score is 99.3%, with a recall of 99.5%, an accuracy of 99.3%, and a precision of 99.6%, showing accurate forecasting capabilities. Furthermore, the AUC curve value was 0.42, showing a moderate degree of separation between the groups. In terms of data security, the generated model had an encryption time of 96 ms and an encryption time of 90 ms, indicating efficient and secure data processing.

6. Conclusion

The research concentrated on the creation and execution of BObMNN, a system specifically developed to secure the data and predict the disease utilizing a trained dataset. Additionally, the expected results from the developed model were evaluated and compared with other methodologies. The evaluation of the suggested approach will be based on several assessment criteria, including Precision, accuracy, F-score, Recall, and AUC. The model achieved remarkable achievement indicators, including 99.5% recall, 99.3% precision, and 99.6% accuracy, as well as an F1 score of 99.3%, indicating exceptional predictive ability. Furthermore, the AUC curve value was reported to be 0.42. This model showed encoding times of 96 and 90 ms, which ensured efficient and secure data processing. Future studies should focus on privacy and data safety in the healthcare sector. Integrating AI methods has the possibility of enhancing ways of making decisions, boosting scalability, and improving the prediction of multiple diseases. Furthermore, the application of hybrid optimization techniques may bolster encryption security and enhance the accuracy of predictions.

REFERENCES

1. S. M. Alhashmi, *Knowledge management, health data, and advanced data analytics to expedite solutions in the healthcare industry*. In *Digital Healthcare, Digital Transformation and Citizen Empowerment in Asia-Pacific and Europe for a Healthier Society* (pp. 231-247). Academic Press, 2025.
2. F. Hussain, M. Nauman, A. Alghuried, and N. Akhtar, *Leveraging big data analytics for enhanced clinical decision-making in healthcare*. *IEEE Access*, 11, 127817-127836, 2023.
3. C. Guo, and J. Chen, *Big data analytics in healthcare*. In *Knowledge technology and systems: Toward establishing knowledge systems science* (pp. 27-70). Singapore: Springer Nature Singapore, 2023.
4. M. Paul, L. Maglaras, M. A. Ferrag, and I. Almomani, *Digitization of healthcare sector: A study on privacy and security concerns*. *ICT Express*, 9(4), 571-588, 2023.
5. S. Gupta, M. Kapoor, and S. K. Debnath, *Impact of AI-Enabled Healthcare Security on Patient Outcomes*. In *Artificial Intelligence-Enabled Security for Healthcare Systems: Safeguarding Patient Data and Improving Services* (pp. 137-156). Cham: Springer Nature Switzerland, 2025.
6. E. V. N. Jyothi, M. Kranthi, D. Gowda, and R. C. Tanguturi, *Design of intelligent medical integrity authentication and secure information for public cloud in hospital administration*. In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)* (pp. 256-261). IEEE, 2023.
7. N. Khalid, A. Qayyum, M. Bilal, A. Al-Fuqaha, and J. Qadir, *Privacy-preserving artificial intelligence in healthcare: Techniques and applications*. *Computers in Biology and Medicine*, 158, 106848, 2023.

8. S. Arunprasath, and S. Annamalai, *Improving patient centric data retrieval and cyber security in healthcare: privacy preserving solutions for a secure future. Multimedia Tools and Applications*, 83(27), 70289-70319, 2024.

9. S. Gupta, M. Kapoor, and S. K. Debnath, *Cybersecurity Risks and Threats in Healthcare*. In *Artificial Intelligence-Enabled Security for Healthcare Systems: Safeguarding Patient Data and Improving Services* (pp. 39-64). Cham: Springer Nature Switzerland, 2025.

10. P. Esmaeilzadeh, *Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations. Artificial Intelligence in Medicine*, 151, 102861, 2024.

11. A. Garcia-Perez, J. G. Cegarra-Navarro, M. P. Sallos, E. Martinez-Caro, and A. Chinnaswamy, *Resilience in healthcare systems: Cyber security and digital transformation. Technovation*, 121, 102583, 2023.

12. S. Silvestri, S. Islam, D. Amelin, G. Weiler, S. Papastergiou, and M. Ciampi, *Cyber threat assessment and management for securing healthcare ecosystems using natural language processing. International Journal of Information Security*, 23(1), 31-50, 2024.

13. S. Aminizadeh, A. Heidari, M. Dehghan, S. Toumaj, M. Rezaei, N. J. Navimipour, and M. Unal, *Opportunities and challenges of artificial intelligence and distributed systems to improve the quality of healthcare service. Artificial Intelligence in Medicine*, 149, 102779, 2024.

14. D. Kavitha, and S. Thejas, *Ai enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation. IEEE Access*, 2024.

15. N. J. Herzog, D. Celik, and R. B. Sulaiman, *Artificial intelligence in healthcare and medical records security*. In *Cybersecurity and Artificial Intelligence: Transformational Strategies and Disruptive Innovation* (pp. 35-57). Cham: Springer Nature Switzerland, 2024.

16. M. Kumar, H. Raj, N. Chaurasia, and S. S. Gill, *Blockchain inspired secure and reliable data exchange architecture for cyber-physical healthcare system 4.0. Internet of Things and Cyber-Physical Systems*, 3, 309-322, 2023.

17. S. Harnal, G. Sharma, S. Malik, G. Kaur, S. Simaiya, S. Khurana, and D. Bagga, *Current and future trends of cloud-based solutions for healthcare. Image Based Computing for Food and Health Analytics: Requirements, Challenges, Solutions and Practices: IBCFHA*, 115-136, 2023.

18. A. Ayari, H. Hamdi, and K. A. Alsulbi, *E-health Application In IoMT Environment Deployed in An Edge And Cloud Computing Platforms. Procedia Computer Science*, 246, 1019-1028, 2024.

19. A. Jain, S. Mittal, A. Bhagat, and D. K. Sharma, *Big Data Analytics and Security Over the Cloud: Characteristics, Analytics, Integration and Security*. In *Security and Risk Analysis for Intelligent Edge Computing* (pp. 35-66). Cham: Springer International Publishing, 2023.

20. S. Asif, Y. Wenhui, S. ur-Rehman, Q. ul-ain, K. Amjad, Y. Yueyang, and M. Awais, *Advancements and prospects of machine learning in medical diagnostics: unveiling the future of diagnostic precision. Archives of Computational Methods in Engineering*, 1-31, 2024.

21. T. Dhar, N. Dey, S. Borra, and R. S. Sherratt, *Challenges of deep learning in medical image analysis—improving explainability and trust. IEEE Transactions on Technology and Society*, 4(1), 68-75, 2023.

22. M. A. Mohammed, A. Lakhan, D. A. Zebari, M. K. Abd Ghani, H. A. Marhoon, K. H. Abdulkareem, and R. Martinek, *Securing healthcare data in industrial cyber-physical systems using combining deep learning and blockchain technology. Engineering Applications of Artificial Intelligence*, 129, 107612, 2024.

23. S. Singh, and D. Kumar, *Energy-efficient secure data fusion scheme for IoT based healthcare system. Future Generation Computer Systems*, 143, 15-29, 2023.

24. A. S. Nadhan, and I. J. Jacob, *Enhancing healthcare security in the digital era: Safeguarding medical images with lightweight cryptographic techniques in IoT healthcare applications. Biomedical Signal Processing and Control*, 88, 105511, 2024.

25. G. Premi, P. Solainayagi, C. Srinivasan, and P. G. Kuppusamy, *Data Privacy and Confidentiality in Healthcare Applications of IoT-Enabled Wireless Sensor Networks*. In *2023 Second International Conference On Smart Technologies For Smart Nation (Smart Tech Con)* (pp. 610-614). IEEE, 2023.

26. J. B. Madavarapu, R. K. Yalamanchili, and V. N. Mandhala, *An ensemble data security on cloud healthcare systems*. In *2023 4th International Conference on Smart Electronics and Communication (ICOSEC)* (pp. 680-686). IEEE, 2023.

27. M. A. Almaiah, S. Yelisetti, L. Arya, N. K. Babu Christopher, K. Kaliappan, P. Vellaisamy, and T. Alkdour, *A novel approach for improving the security of IoT–medical data systems using an enhanced dynamic Bayesian network. Electronics*, 12(20), 4316, 2023.

28. A. Almalawi, A. I. Khan, F. Alsolami, Y. B. Abushark, and A. S. Alfakeeh, *Managing security of healthcare data for a modern healthcare system. Sensors*, 23(7), 3612, 2023.