Intelligent Neighborhood Indexing Sequence Model for Healthcare Data Encoding

Ibrahim M. EL-Hasnony[1], Mohamed Elhoseny[1], Mohammed K. Hassan[2]

[1] Faculty of Computers and Information, Mansoura University, Egypt

[2] Faculty of Engineering, Mansoura University, Egypt

Emails: ibrahimhesin2005@mans.edu.eg, Mohamed_elhoseny@mans.edu.eg; eng.mkamal1976@gmail.com

Abstract

Recently, information security in the healthcare sector has become essential to ensure confidentiality in medical data. At the same time, automated disease diagnosis using deep learning (DL) models also gained considerable attention to accomplish enhanced classification performance. This paper designs an intelligent neighborhood indexing sequence based on encoding with a classification model for healthcare information security (INISEC-HIS). The proposed INISEC-HIS technique aims to accomplish security in medical data transmission and diagnosis. The neighborhood indexing sequence (NIS) technique is applied to securely transmit the data, which transforms the medical data into an encoded format. Besides, a novel artificial fish swarm algorithm (AFSA) with deep neural networks (DNN) model is used for the classification process. The design of AFSA to optimally adjust the hyperparameters of the DNN model shows the study's novelty. An extensive simulation analysis takes place to examine the improved outcomes of the INISEC-HIS technique, and the obtained results highlighted the supremacy over the other techniques.

Keywords: Healthcare; Information security; Encryption; Data encoding; deep learning; Disease diagnosis

1. Introduction

Medical organizations have focused previously on disseminating and gathering data for effective treatment of patients' illnesses. Security was frequently constrained by conventional models of locking file cabinets/file rooms. The advent of computers improved the efficacy of medicinal records and improved privacy exposures [1]. Computer management of records takes data from a controllable hand and creates the opportunity for outright theft, leaks, and mismanagement. When the therapeutic facilities' collection of data is considered, the values are comprehensible. Records consist of private data involving; address, birthday, and social security number. Next, financial with the replication of the earlier banking and credit data. Finally, the medicinal data with the chance of embarrassment or even blackmail [2]. The introduction of electronic patient records has inadvertently created opportunities for healthcare frauds involving medicinal individuality thefts. Incorrect medicinal records might result in medicinal mistreatment, which might create catastrophic consequences for the person.

Security threat is the primary concern to healthcare organizations because of the vulnerability and value of medical information being distributed and recorded [3]. The importance of the information comes

DOI: https://doi.org/10.54216/JISIoT.000102

from the truth that it is historical, and it immediately affects the capacity to securely treats patients; it takes a longer period for rebuilding. It has several medical information, like demographic, personal, and financial information, that allows being utilized for wide individuality thefts [4]. It is determined where we might alter the passwords and their credit card, PIN, and account number in case of a breach, but we might change our mothers' maiden name. The susceptibility comes from the truth that there is a revolution in healthcare with cloud computing, communication of systems, mobile devices, and IOHT, and the change in the working practice of physicians, like working from home, remote monitoring, and telemedicine [5]. This revolution hasn't often matched with healthcare organizations' policies, budgets, security awareness, and practices. There is a growing sophistication of attacks with social engineering methods (for example, Phishing), which could beat "conventional" protections like signature, antivirus, and rule-based detection systems [6]. However, before seeing a novel AI-based tool could perform for organizations, we recommend that it is a very sophisticated tool. Without the fundamentals, they would fail to send on their potential.

The significance of secured conversion in healthcare, medical, and public health transport methods was detected by several institutions [7]. The Networking and Information Technology Research and Development (NITRD) programs have recently launched the Federal Health Information Technology Research and Development Strategic architecture [8]. It has explicated the significance of incorporating engineering, computing, statistics and mathematics, social science and behavioral, and public health study fields to explore the vital revolution to increase the service in the health care scheme [9]. Substantial developments in artificial intelligence (AI), higher performance cloud computing, machine learning (ML), availability, and deep learning of novel dataset makes this incorporation attainable.

This paper designs an intelligent neighborhood indexing sequence based on encoding with a classification model for healthcare information security (INISEC-HIS). The proposed INISEC-HIS technique involves the neighborhood indexing sequence (NIS) technique, transforming the medical data into an encoded format. Besides, a novel artificial fish swarm algorithm (AFSA) with deep neural networks (DNN) model is used for the classification process. An extensive simulation analysis takes place to examine the improved outcomes of the INISEC-HIS technique, and the obtained results highlighted the supremacy over the other techniques.

2. Literature Review

Shaikh et al. [10] developed and designed an improved ECG scheme for privacy preservation and security, ECG visualization, and ECG diagnosing. The QRS technique is utilized for diagnosing the attained ECG signal. The results achieved from the QRS system are utilized for displaying unhealthy/healthy patient conditions. The method would be aware of additional diagnoses when the situation is serious. Carpov et al. [11] presented the solution for analyzing user health information straight to the Cloud while conserving user confidentiality. This result utilizes homomorphic encryption for protecting user information at the time of analysis. They designed a mobile application that offloads user information to the Cloud, and a homomorphic encryption method process this information without revealing data to the Cloud providers.

Elhoseny et al. [12] present a hybrid security method to secure the analytical text information in medicinal images. The projected method is advanced by incorporating 2D-DWT-2L/2D-DWT-1L steganography techniques with a proposed hybrid encryption system. The presented model is constructed by integration of AES, and Rivest, Shamir, and Adleman methods. Hamza et al. [13] use two-dimensional logistic maps to make the cryptographic key sequence based upon cascading and mixing the orbit of chaotic maps to generate the stream key for the encryption method. The encrypted image developed using this model that shows random behavior, provides a higher level of privacy for the keyframe against several attacks.

Chen et al. [14] developed a verification system according to cloud environments. This system enables us to utilize the digital development method for achieving distributed medicinal information. Incidentally, the digital signature and biometric fingerprint features are utilized for ensuring the privacy of medicinal data in this system. Hua et al. [15] presented a robust and privacy-preserving online medicinal primary diagnoses framework. Within CINEMA architecture, users could enter online medicinal primary diagnoses service precisely without revealing their medicinal information.

DOI: https://doi.org/10.54216/JISIoT.000102

Specifically, depending on the faster secure permutation and comparison method, the encrypted user query has functioned straightaway at the SP without decryptions. The diagnoses results could only be decrypted with the users. In contrast, the diagnosis method in SP could be secured.

3. Proposed Secure Healthcare Framework

In this study, a new INISEC-HIS technique is derived to accomplish security in medical data transmission and diagnosis. The proposed INISEC-HIS technique involves two major processes, namely NIS-based encoding, and AFSA-DNN-based disease diagnosis. These two processes are elaborated on in the following sections.

3.1 Working on NIS Encoding Technique

to improve medical data security, the NIS method is presented to exploit the valuable data amongst nearby bits of the input sequence (inp_seq) for encoding it. NIS method is an individual character encoding system that codes a single character individually. It traverses the input sequence based on zeros and ones with a reference bit. Initially, the series of input characters are mapped to their corresponding ASCII value, and this value is transformed into its equal binary form. This method initiates with the initial bit of the single character and identifies that either the bit is zero/one. Once the initial bit is zero, the process starts by zero and traverses to the neighboring bit to find zero-based and one-based (denoted by 00 for zero-based or 01 for one-based). Once zero is established, the corresponding setting is saved and reinitiated in the procedure by considering the recognized bit as a reference bit. This algorithm repeats until it attains the final bit of the input characters (viz., 7th bit). Likewise, if the initial bit is 1, the process starts with one and traverses to the neighboring bit to find zero-based and one-based (denoted by ten for zeros based or eleven for ones based). Once zero is established, the corresponding position is saved and reinitiates the algorithm by considering the recognized bit as the reference bit. Next, the amount of bits in the zeros-based and based election is related, and the minimal amount of bits is selected as optimum bits [16]. Similarly, the NIS method encodes each character of the input sequence. The possible integration of (00, 01, 11, 10) results in adding eight bits to the optimum amount of bits to the compressed file. The compressed number of bits C_{bits} of the input sequence, inp_seq of length N is denoted as Eq. (1).

$$C_{bits} = \sum_{i=1}^{N} NIS_{opt}(i) + const \qquad (1)$$

In the equation, $const = 8$ denotes the possible four integration of zeros and one.

The typical amount of bits required for representing an individual character NIS_{ch_av} of the input sequence is formulated by

$$NIS_{ch\_av} = \frac{C_{bits}}{N}, \quad 1 \leq NIS_{ch\_av} \leq 5 \qquad (2)$$

The NIS method requires five bits for representing a character only in advanced cases. It maintains a coding table for handling the compression procedure. It follows a symmetrical compression in which the decompression procedure is the precise process of the compression procedure also in the opposite direction. NIS method eliminates the requirement of transferring the coding table alongside the compressed file, i.e., taking into account the main drawbacks of LZW coding and Huffman coding.

3.2 The process involved in the Disease Diagnosis model

To diagnose disease, the DNN model is employed. The presented method makes decisions on health care information. The benefits of this method are the election of essential attributes and the classification of medicinal records based on the time limitation for effective decision-making. The ANN method is a computation intelligence method inspired by the network of biological neurons to resolve predictive challenges, drug detection, NLP, etc. A DNN is a NN method with a definite level of complexity, a NN with several layers. DNN uses a complex arithmetical method to process the information in a complicated way. DNN with several layers combines the classification process and feature extraction to an individual learning body and builds a decision-making process. This kind of NN has succeeded in complicated domains to detect patterns in the last few years. Generally, DNN includes

DOI: https://doi.org/10.54216/JISIoT.000102

the input layer for the new descriptor X_l, L hidden layer, and output layers for predicting information. The DNN is proposed with the help of TensorFlow architecture. The traditional approach constructs an optimum NN using the appropriate amount of layers and neurons for all the layers. Hence, a DNN is created by executing a more comprehensive range of trials.

The DNN Classifiers are applied to generate each neuron layer, with the ReLU activation function. From Eq. (3), it is understood that the DNN is effective and more straightforward.

$$f(x) = x^{+} = \max(0, x) \tag{3}$$

Whereas x represents the input to the neuron. Also, it is named ramp function i.e., equivalent to the half-wave rectification method in electrical engineering [17]. A component using a rectifier is known as ReLU method:

$$f(x) = \ln[1 + \exp(x) \tag{4}$$

That is termed as soft plus function. In the predictive method, a novel representation of the raw descriptor is filtered from the hidden layer by:

$$X_{t+1} = H(W_l X_l + B_l), \qquad l = 1, \ldots, L \tag{5}$$

In the equation, the bias for the l^{th} hidden layer, W_l & B_l represent the weight matrix, and H denotes the similar activation functions, i.e., chosen as ReLu.

For tuning the hyperparameters of the DNN model, the AFSA is applied. AFSA has been a novel intelligence-optimized algorithm that generates a set of artificial fish and imitates natural fish swimming in the water. Using the local interaction and lower-level behaviors of an individual, the method illustrates the higher-level behaviors of AI of the group on the macro level. Typical AFSA is an arbitrary search model that depends on population, i.e., generally starts with a group of arbitrarily made early populations and next searches for the optimal solution through iteration.

Assuming that the early population of fish is N, the issue in discussion is a D-dimension issue. Initiate the state vectors of the fish swarm so that the condition of single fish could be expressed by:

$$X_i = (x_1, x_2, \cdots, x_D) \tag{6}$$

Food satisfaction (fitness concentration) is denoted as $Y_{\text{i}} = f(X_{\text{i}})$. The Euclidean distance that represents the relationships among any two fishes is represented as $D_{\text{i},j} = \|X_{\text{i}} - X_j\|$. For $i, j = 1, 2, \ldots, N$, X_{i}, and X_j denote the distinct state of the fish. The parameters that affect the performances are the degree of crowdedness [18], the visual area of fish, and step size denoted as *Crowded factor* δ, *Visual*, and *Step* correspondingly.

The exploration phase of the sensors in the sensor network toward a large network coverage rate has been the same as the procedure of artificial fish toward high food satisfaction. Each artificial fish mainly searches for the optimal food location with the Following, Random, Preying, and Swarming behaviors.

Random Behavior has been the performance of fish that moves arbitrarily from its visual fields. Assume the present state of fish i is X_i; it elects a state arbitrarily in their visible areas. The moving can be expressed by:

$$X_{i|next} = X_i + Visual \times Rand(\ ) \tag{7}$$

Visual represents the visual area of fish, and $Rand(\ )$ denotes the arbitrary value made between zero & one. Fig. 1 shows the flowchart of AFSA.

DOI: https://doi.org/10.54216/JISIoT.000102

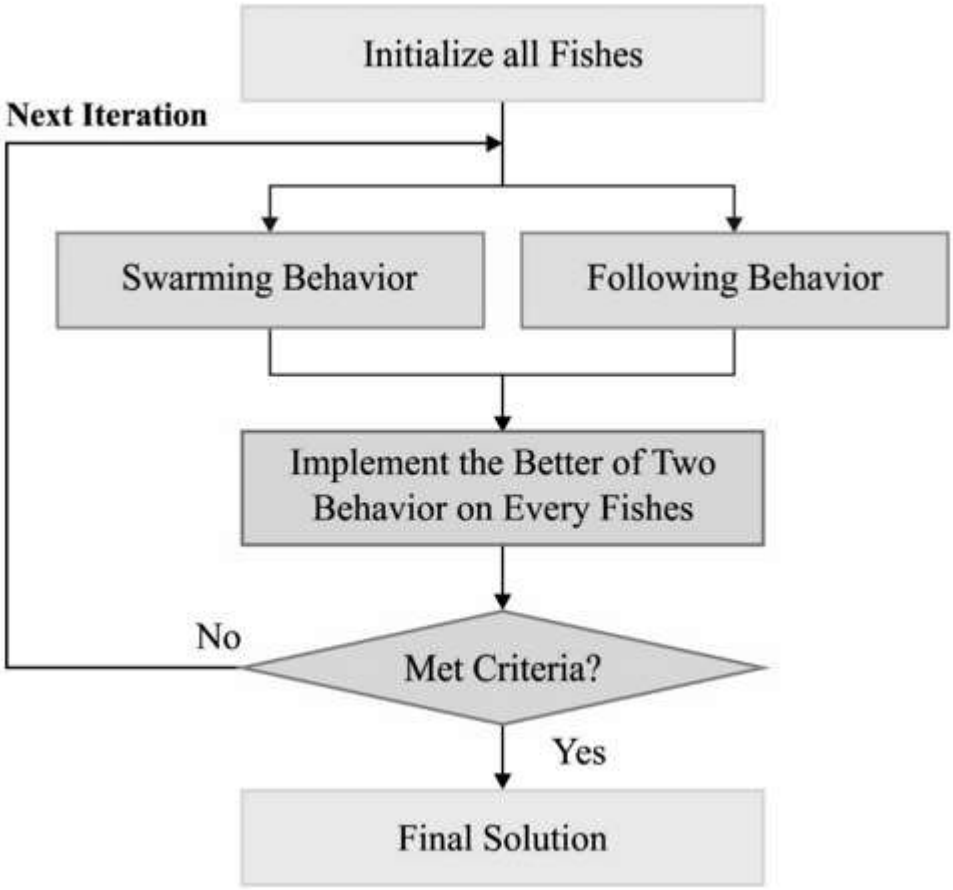

Figure 1: Flowchart of AFSA

Preying performance has been the performance of fish that moves arbitrarily from the water, search for feeding, and moves toward the position with higher food satisfaction. Assume the present state of fish i is X_i, and it selects a novel state X_j arbitrarily within its visual fields as:

$$X_j = X_i + Visual \times Rand(\quad) \tag{8}$$

Assume $f(X_i)$ & $f(X_j)$ represent the food satisfaction of the state X_i and the state X_j, correspondingly.

Case 1: $f(X_i) < f(X_j)$ In the maximal, a fish proceeds a step toward a succeeding way, and the present state $X_{i|next}$ of the fish is formulated by:

$$X_{i|next} = X_j + \frac{X_j - X_i}{\|X_j - X_i\|} \times Step \times Rand\ (\quad) \tag{9}$$

In which $\|\cdot\|$ represents the Euclidean distance among the artificial fish j and artificial fish i.

Case 2: $f(X_i) \geq f(X_j)$ The artificial fish reselect another state arbitrarily. When the fish could not encounter the given time requirement, it moved one step arbitrarily as

Swarming performance has been a performance that the artificial fish is similar to collecting in sets and moving to the center of its fellow. The central position state can be evaluated by:

$$X_c = \left(\sum_{j=1}^{n_f} X_j / n_f\right), D_{i,j} < Visual \tag{10}$$

DOI: https://doi.org/10.54216/JISIoT.000102

Whereas $D_{i,j}$ represents the Euclidean distance between the artificial fish j and artificial fish i.

The food satisfaction of central position state X_c is $f(X_c)$. If $f(X_c)/n_f > \delta f(X_i)$, this area isn't crowded. δ represents the *crowded factor*. When $f(X_c) > f(X_i)$, the fish move one step toward the fellow central location:

$$X_{i|next} = X_i + \frac{X_c - X_i}{\|X_c - X_i\|} \times Step \times rand(\quad) \tag{11}$$

The present state of artificial fishes is X_i, n_f represents the number of its fellows within the visual areas. Assume X_j means the optimal fish (in the optimal feed location) from the visual field of X_i. The food satisfaction of location state X_j is $f(X_j)$ when $f(X_j)/n_f > \delta f(X_i)$ and $f(X_j) > f(X_i)$, the fish move one step toward X_j.

$$X_{i|next} = X_i + \frac{X_j - X_i}{\|X_j - X_i\|} \times Step \times rand\ (\quad) \tag{12}$$

When this behavior isn't beneficial, the fish execute the preying behavior.

4. Performance Validation

The results analysis of the INISEC-HIS technique takes place in terms of different measures.

Table 1 and Fig. 2 investigate the INISEC-HIS technique's performance with existing CR, BPC, and SS techniques. The results portrayed that the INISEC-HIS technique has gained better performance over the other techniques. On examining the results in terms of CR, the INISEC-HIS technique has obtained a CR of 0.47, whereas the BWT, LZMA, and LZW techniques have obtained a CR of 0.62, 0.69, and 0.71, respectively. Moreover, on investigative the results concerning BPC, the INISEC-HIS algorithm has obtained a BPC of 4.67, whereas the BWT, LZMA, and LZW techniques have gained a BPC of 5.12, 5.43, and 5.90 correspondingly. Furthermore, on inspecting the outcomes in terms of SS, the INISEC-HIS system has gained a SS of 48.39%, whereas the BWT, LZMA, and LZW approaches have reached a SS of 60.37%, 62.66%, and 67.90% correspondingly.

Table 1: Performance Evaluation of Proposed method with various existing techniques

Methods	Compression Ratio	BPC	Space Savings (%)
INISEC-HIS	0.47	4.67	48.39
BWT	0.62	5.12	60.37
LZMA	0.69	5.43	62.66
LZW	0.71	5.90	67.90

DOI: https://doi.org/10.54216/JISIoT.000102

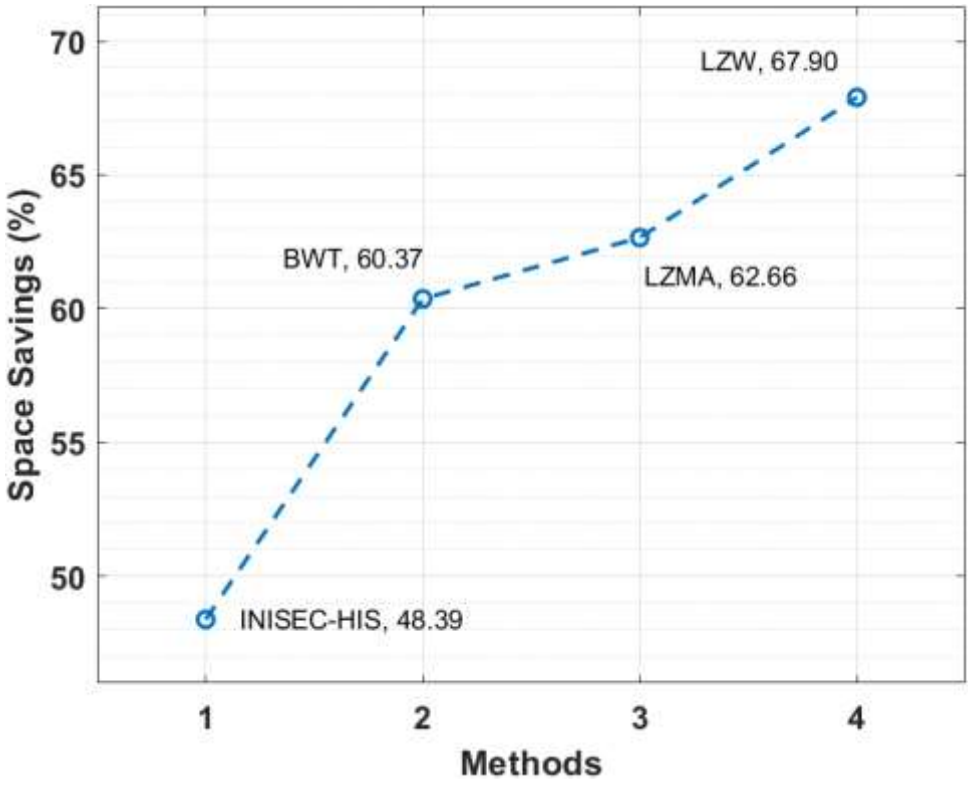

Figure 2: SS analysis of INISEC-HIS model

Table 2 provides a comparative results analysis of the INISEC-HIS technique with existing techniques. The table values denoted that the INISEC-HIS technique has gained effective outcomes with higher diagnostic performance.

Fig. 3 inspects the sensitivity analysis of the INISEC-HIS technique with other methods. The figure has shown that the INISEC-HIS technique has resulted in an increased sensitivity of 99.65%. In contrast, the DNN, MLP, and SVM techniques have obtained a reduced sensitivity of 98.53%, 89.65%, and 81.65%, respectively.

Table 2: Performance Evaluation of Chronic Kidney Disease using various methods

Methods	Sensitivity	Specificity	Accuracy	F-score	Kappa
INISEC-HIS	99.65	94.98	98.90	97.65	90.98
DNN	98.53	86.26	97.54	92.21	73.71
MLP	89.65	81.65	86.65	89.65	69.65
SVM	81.65	68.31	76.65	81.65	48.31

Fig. 4 examines the specificity analysis of the INISEC-HIS algorithm with other techniques. The figure demonstrated that the INISEC-HIS technique has an increased specificity of 94.98%, whereas the DNN, MLP, and SVM techniques have obtained a reduced specificity of 86.26%, 81.65%, and 68.31%, respectively.

Fig. 5 studies the accuracy analysis of the INISEC-HIS manner with other methods. The figure exhibited that the INISEC-HIS method has resulted in a superior accuracy of 98.90%, whereas the DNN, MLP, and SVM systems have obtained a reduced accuracy of 97.54%, 86.65%, and 76.65%, respectively.

DOI: https://doi.org/10.54216/JISIoT.000102

Fig. 6 considers the F-score analysis of the INISEC-HIS technique with other techniques. The figure portrayed that the INISEC-HIS system has resulted in a maximum F-score of 97.65%, whereas the DNN, MLP, and SVM approach has reached a lower F-score of 92.21%, 89.65%, and 81.65% respectively.

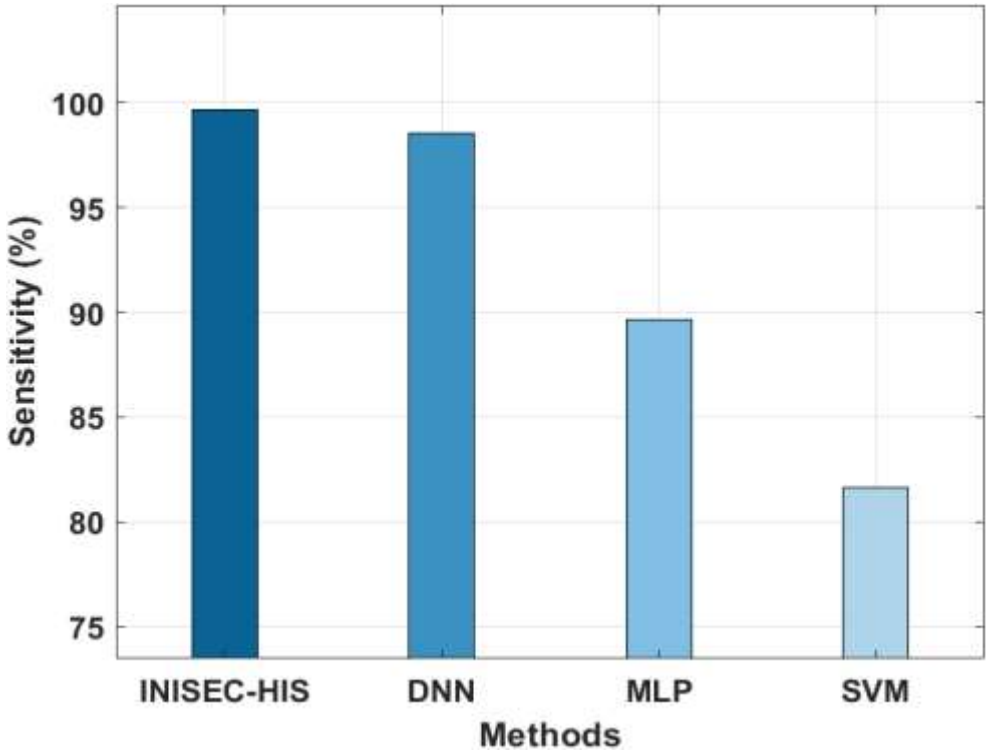

Figure 3: Sensitivity analysis of INISEC-HIS model with existing techniques

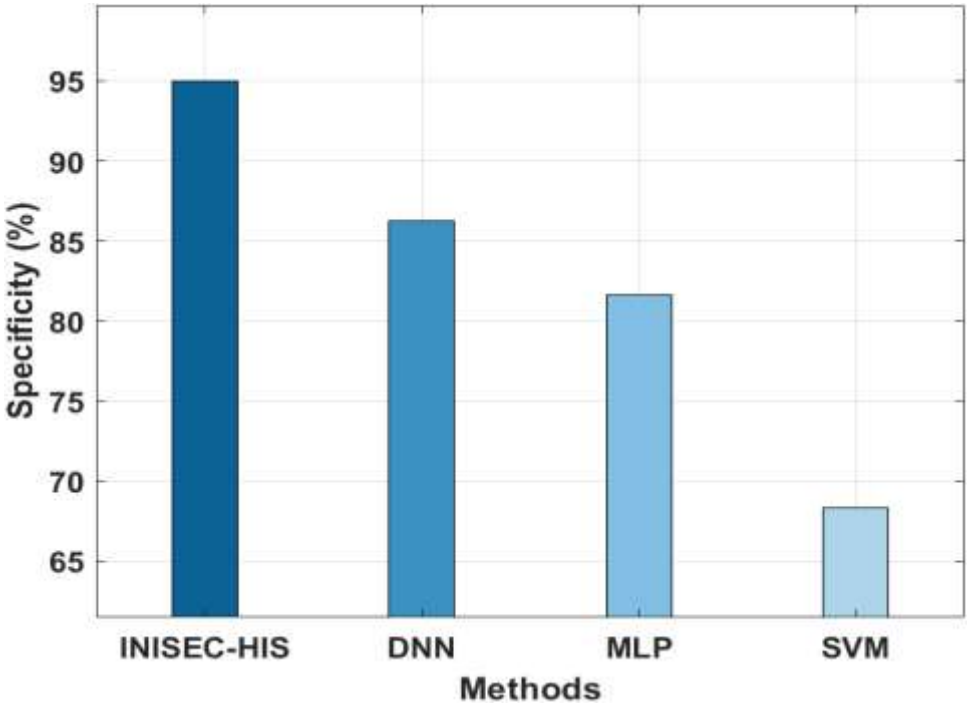

Figure 4: Specificity analysis of INISEC-HIS model with existing techniques

DOI: https://doi.org/10.54216/JISIoT.000102

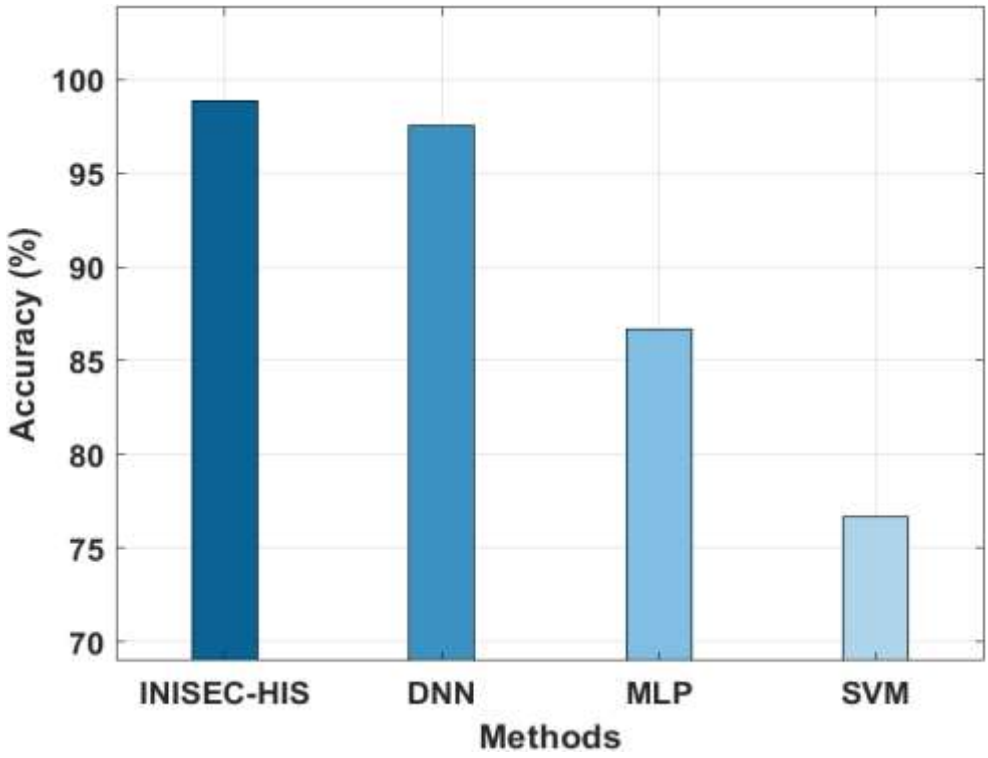

Figure 5: Accuracy analysis of INISEC-HIS model with existing techniques

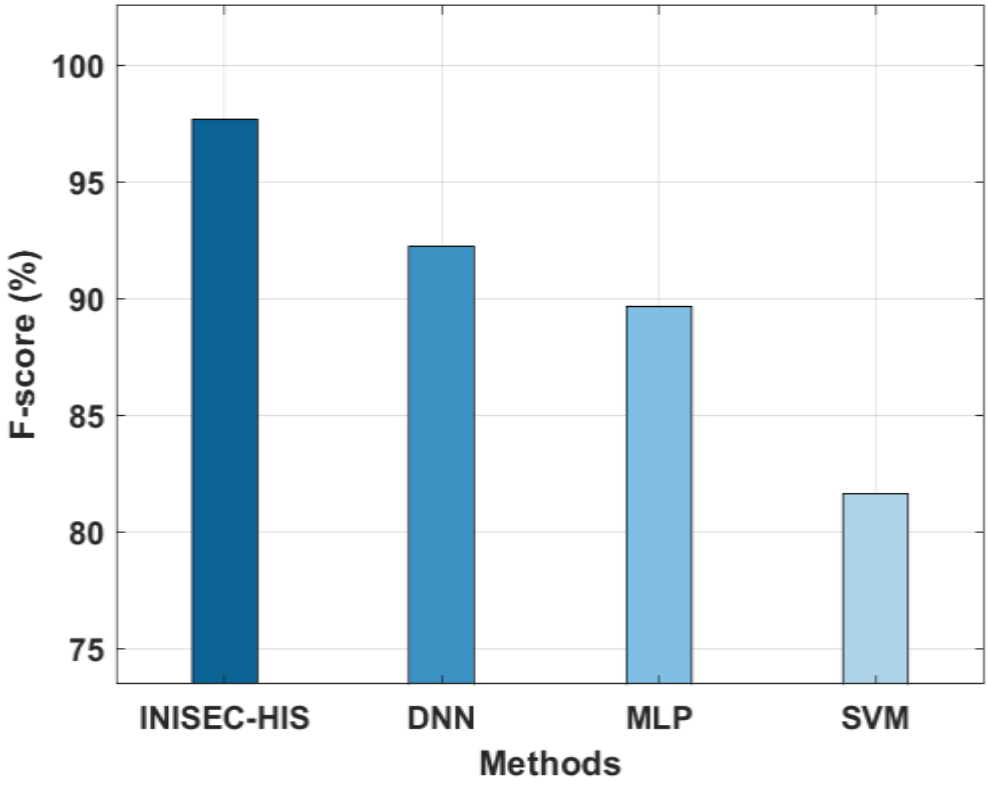

Figure 6: F-score analysis of INISEC-HIS model with existing techniques

DOI: https://doi.org/10.54216/JISIoT.000102

5. Conclusion

In this study, a new INISEC-HIS technique is derived to accomplish security in medical data transmission and diagnosis. The NIS encoding technique is applied to securely transmit the data, which transforms the medical data into an encoded format. Moreover, a novel AFSA with a DNN model is used for the classification process. The design of AFSA to optimally adjust the hyperparameters of the DNN model shows the study's novelty. An extensive simulation analysis takes place to examine the improved outcomes of the INISEC-HIS technique, and the obtained results highlighted the supremacy over the other techniques. Therefore, the INISEC-HIS technique can be employed as an effective tool for secure medical data transmission.

References

[1] Engelhardt, M.A., 2017. Hitching healthcare to the chain: An introduction to blockchain technology in the healthcare sector. Technology Innovation Management Review, 7(10).

[2] Javed, S.A. and Ilyas, F., 2018. Service quality and satisfaction in the healthcare sector of Pakistan—the patients' expectations. International journal of health care quality assurance.

[3] Collyer, F.M., Willis, K.F. and Lewis, S., 2017. Gatekeepers in the healthcare sector: Knowledge and Bourdieu's concept of field. Social Science & Medicine, 186, pp.96-103.

[4] Almajali, D.A. and Tarhini, A., 2016. Antecedents of ERP systems implementation success: a study on the Jordanian healthcare sector. Journal of Enterprise Information Management.

[5] Ghazisaeidi, M., Safdari, R., Torabi, M., Mirzaee, M., Farzi, J. and Goodini, A., 2015. Development of performance dashboards in the healthcare sector: key practical issues. Acta Informatica Medica, 23(5), p.317.

[6] Ahsan, K. and Rahman, S., 2017. Green public procurement implementation challenges in the Australian public healthcare sector. Journal of Cleaner Production, 152, pp.181-197.

[7] Manyazewal, T. and Matlakala, M.C., 2017. Beyond patient care: the impact of healthcare reform on job satisfaction in the Ethiopian public healthcare sector. Human resources for health, 15(1), pp.1-9.

[8] Mascia, D., Dello Russo, S. and Morandi, F., 2015. Exploring professionals' motivation to lead: a cross-level study in the healthcare sector. The International Journal of Human Resource Management, 26(12), pp.1622-1644.

[9] Evans, J.M., Brown, A. and Baker, G.R., 2015. Intellectual capital in the healthcare sector: a systematic review and critique of the literature. BMC health services research, 15(1), pp.1-14.

[10] Shaikh, M.U., Ahmad, S.A. and Adnan, W.A.W., 2018, December. Investigation of data encryption algorithm for secured transmission of electrocardiograph (ECG) signal. In 2018 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES) (pp. 274-278). IEEE.

[11] Carpov, S., Nguyen, T.H., Sirdey, R., Constantino, G. and Martinelli, F., 2016, June. Practical privacy-preserving medical diagnosis using homomorphic encryption. In 2016 IEEE 9th International Conference on Cloud Computing (CLOUD) (pp. 593-599). IEEE.

[12] Elhoseny, M., Ramírez-González, G., Abu-Elnasr, O.M., Shawkat, S.A., Arunkumar, N. and Farouk, A., 2018. Secure medical data transmission model for IoT-based healthcare systems. Ieee Access, 6, pp.20596-20608.

[13] Hamza, R., Muhammad, K., Arunkumar, N. and Ramirez-Gonzalez, G., 2017. Hash-based encryption for keyframes of diagnostic hysteroscopy. IEEE Access, 6, pp.60160-60170.

[14] Chen, C.L., Hu, J.X., Fan, C.L. and Wang, K.H., 2016, October. Design of a secure medical data sharing system via an authorized mechanism. In 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 002478-002482). IEEE.

[15] Hua, J., Zhu, H., Wang, F., Liu, X., Lu, R., Li, H., and Zhang, Y., 2018. CINEMA: Efficient and privacy-preserving online medical primary diagnosis with skyline query. IEEE Internet of Things Journal, 6(2), pp.1450-1461.

[16] Le, S.T., Prilepsky, J.E. and Turitsyn, S.K., 2015. Nonlinear inverse synthesis technique for optical links with lumped amplification. Optics express, 23(7), pp.8317-8328.

[17] Yu, H., Tan, Z.H., Ma, Z., Martin, R., and Guo, J., 2017. Spoofing detection in automatic speaker verification systems using DNN classifiers and dynamic acoustic features. IEEE transactions on neural networks and learning systems, 29(10), pp.4633-4644.

DOI: https://doi.org/10.54216/JISIoT.000102

[18] Zhang, C., Zhang, F.M., Li, F. and Wu, H.S., 2014, June. Improved artificial fish swarm algorithm. In 2014 9th IEEE Conference on Industrial Electronics and Applications (pp. 748-753). IEEE.

DOI: https://doi.org/10.54216/JISIoT.000102