Artificial Neural Network based Data sensitive Security Model for Healthcare Data

Brajesh Chaturvedi[1]and Harish Patidar[2]
[1and 2]Mandsaur University,Mandsaur, India

Abstract: -Using artificial neural networks, we propose data-sensitive security architecture for healthcare data sets of unprecedented size and complexity. Modern medical gadgets and the digitization of healthcare data have added new layers of complexity to data analytics. A large amount of data is generated by this procedure, which is then usefully analyzed and categorized to provide useful insights. Using this information and the results of the study, medical diagnoses and prognoses can be made. Many countries' healthcare laws specifically require the protection and storage of patient records. Therefore, the healthcare industry places a premium on protecting the confidentiality of this vast trove of information. The same level of security applies anywhere else, too. Conversely, the factors' relative relevance determines the data's sensitivity, which is not constant. Therefore, it has been shown that employing a solitary security architecture for all data is redundant and inefficient. The adaptive intelligent security system proposed in this study will tailor protection to the specifics of the stored data. Big Data, Security, and Machine Learning, the three primary pillars of the proposed architecture, are mapped to accomplish varying degrees of data confidentiality. Due to the unpredictable and unstructured nature of medical data, an artificial neural network is trained using an electronic medical record and a sensitivity level to produce a novel security model that offers the best security in real time. The ANN is trained with attributes from data sets that include patient information, medical history, insurance information, and so on, and accurate categorization is achieved. The proposed method is tested experimentally to determine its efficacy.

Keywords*: Data security, Healthcare, Big data, adaptive security, data sensitivity, Deep learning, Neural network*

1. Introduction

Lorem The digitization of all medical records has had a profound effect on the healthcare sector over the past decade. This process resulted in electronic medical records (EMR), which revolutionized medical study and diagnosis. The EMR makes it possible to retrieve all of the relevant information at any time and from any location. Information about patients and their medical histories, as well as information about doctors and hospitals, as well as financial and insurance details, are typical examples of what is stored. It is also derived from other sources such as genetic and biological data, social media information, and information collected by far-flung sensors [1]. Innovations in data generation in the healthcare sector can be attributed to the widespread use of biomedical machinery, which includes digital health records, medical equipment, tracking devices, and mobile computers. Complex, energy-hungry machines are needed for such massive data processing.

The term "big data" refers to the processing of massive amounts of data that are difficult to manage with traditional computer methods. Data sets of this size and complexity require state-of-the-art methods of management and analysis, hence the term "big data." The heterogeneity of the data compounds processing difficulties. Big data is distinguished by its diversity, volume, speed, and accuracy of data. Volume is the quantity of data, commonly measured in terabytes; variety is the heterogeneous, random, and unstructured data created from numerous information sources. In the context of big data, veracity refers to the accuracy and reliability of the data, while velocity refers to the speed with which decisions may be made based on the massive amount of data available [2, 3].

Big data analytics, with its ability to manage huge volumes of data, is a perfect tool for the healthcare industry, whose data volume is expanding everyday. Big data analytics methods could be used to put healthcare data to good use in R&D. The gathered datasets allow for the generation of clear insight and focused knowledge. In the event of a global health emergency, such as the spread of corona virus (COVID-19), H1N1, Dengue, Ebola, or any of a number of other viruses, predictive analysis can be incredibly useful because it allows for more informed decision-making and, potentially, the saving of lives by facilitating efficient knowledge and data sharing among all nations. It can improve operational efficiency and reduce healthcare expenses. The use of big data in healthcare will aid in a variety of areas, such as disease outbreak prevention, patient care improvement, clinical trial efficiency, pharmaceutical side effect detection, and more. By making appropriate therapeutic choices, it can also prevent deaths that could have been prevented. Citations [4, 5].

However, in recent years, a number of scholars have made protecting the privacy and confidentiality of healthcare data a central research topic. A compromise in data security has the potential to put people's lives in peril and trigger major medical and financial catastrophes. The medical records that healthcare facilities share, archive, and retain are crucial to their operations. This information has the highest risk of being leaked and is the most likely target of a cyberattack. To address this problem, researchers have developed a variety of security models to prevent intrusions and safeguard sensitive information [6]. Information security, data security, and access control are the three pillars of healthcare big data security outlined by Kim et al. [7]. They suggested that companies protect the privacy of big data by investing in the appropriate hardware and software for handling administrative and clinical data. Data preservation, reuse, and auditing must be accomplished from the start of the project to ensure cost effectiveness.

Data collection, processing, analysis, and storage are all parts of the knowledge creation security architecture created by Yazan et al. [8]. They provided a model for the full spectrum of large data security processes. R. Zhang and L. Liu [9] established an authentication-based security paradigm for internet data transfer using the Transport layer and the Secure Sockets layer. It operates in the same way as other cutting-edge web technologies do, such as Internet faxing, electronic mail, and web browsing. Mutually trustworthy certificates can be used with TLS or SSL to authenticate the server. C. Yang et al. [10] proposed using a one-time pad mechanism for authentication, eliminating the need for shared passwords between servers. In this model, the consumer and the service provider's identities must be verified at each point of access.

To prevent unauthorized access at every point in the security process, various encryption-based security models have been proposed [11-16]. These models include RSA, Rijndael, AES and RC6, DES, 3DES, RC4, IDEA, and Blowfish. Researchers have a significant challenge due to the fact that the size of the encryption key varies with the amount of data. Some researchers have also employed data masking technology to boost the safety of big data in healthcare by substituting anonymous numerical values for personally identifiable information. These data fragments are then recovered with the help of quasi-identifiers and a de-identification procedure. To conceal individual details in large healthcare databases, Swaney and Samrati [17, 18] proposed a k-anonymity data masking technique. Truta et al. [19] enhanced this work by incorporating p-sensitive anonymity into the security architecture by integrating the attributes in addition to the identifying parameters. However, when dealing with anonymity in high-dimensional datasets, all approaches failed miserably.

Researchers in the healthcare sector have also built a security paradigm based on access control. A. Mohan and D.M. Blough [20] introduced an attribute-based approach to authorization. In addition, H. Zhou and Q. Wen [21] presented a dynamic access control system that uses cloud computing and combines encryption and cipher-text techniques. H. Wang et al. [22] established a framework for big data analytics for the healthcare business employing a number of big data methodologies. Hadoop, map-reduce, pig-Cassandra, Hbase, zookeeper, oozie, avro, mahout, and other parallel systems are used to process enormous data sets. The complexity of the algorithm has been reduced because to distributed architecture and parallel computing. Integration and interoperability between processes have always been prioritized in big data analytics. However, the privacy and protection of the healthcare industry's vast stores of personal medical information have received less focus. The researchers have developed universal, unchanging approaches to security and privacy. Big data is inherently unpredictable;thus, an adaptable and dynamic security architecture is required to deal with its changing properties.

Adaptive security is a concurrent security method that monitors activity and events to lessen vulnerability and get ready for attacks in advance. The primary focus of adaptive security is the creation of an analytical framework that yields ever-improving outcomes in terms of the quality of hazard detection, acknowledgement, and avoidance. One of the cornerstones of flexible security is being ready for any framework difficulties. Improved security engineering and constant monitoring are the absolute minimum requirements. Standard operating procedure calls for predicting, identifying, and responding to an issue before it has a chance to damage the framework [23]. This is in contrast to the more pessimistic approach of just expecting an incident to occur.

Adaptive security relies heavily on advances in machine learning and security research. Furthermore, the employment of symptomatic inquiry, distinctive examination, and predictive research with verified evidence to identify suspicious behavior can all help determine the root cause of an unpleasant situation. Machine learning is able to fill a helpful need in this era of endless Big Data verified by cloud data warehouses, dangerous developments cloaked as respectable bearings, and progressively elusive server demands. Through the automation of numerous procedures, including plan affirmation, which is employed in evaluation, it can assist a security in getting information.

This study presents an adaptive security strategy for healthcare big data analytics based on artificial intelligence. The suggested method achieves AI through the use of an artificial neural network that has been trained using a medical dataset (EMR). The network makes it possible to send out security models in real time. Sensitive information such as medical records, insurance policies, and patient histories all play a role in determining the security architecture.

The unique adaptive intelligent security architecture is mostly built on artificial neural networks, which is a significant contribution to the field. The adaptive and data-sensitive security model further distinguishes this study from others in the field. It combines the benefits of big data analytics in dealing with massive amounts of data with the intelligence of artificial neural networks (ANN) for decision making in the face of noise, ambiguity, and unpredictability. This combination strategy increased the security model's robustness without adding complexity. The remaining sections of the document are organized as follows: In the report's second half, the basics of big data analytics in healthcare are presented. In Section III, we present the mathematical model underlying the ANN. In Section IV, we offer an adaptive security architecture for the healthcare sector's enormous data sets. Part V uses a simulation exercise to assess the efficacy of the proposed method, and Part VII finishes things up.

2. Big Data Analytics and Healthcare Sector

Businesses are now considering vast amounts of data for market research and policy frameworks, which has shifted the entire paradigm. Big data analytics is a method for quickly and accurately collecting useful information from massive datasets. In the realm of big data analytics, computing parallel processing frameworks have shown to be invaluable, as they allow the identification of data's potential value in decision making. Data mining is an invaluable asset to big data because of its adaptability and generalizability. Big Data produces varied volumes of data at a quick rate of volume generation. Big data refers to data sets that contain a large amount of information. It consists of both structured and unstructured information, with the latter being the greater challenge. However, IT automation has seen nothing short of revolutionary changes during the past few years. Electronics, finance, and e-commerce are just a few examples of sub-industries that have benefited greatly from the IT revolution.

The use of technology in healthcare, particularly in the fields of medical research and diagnostics, has seen profound transformations during the past two decades. The accessibility of millions of patients' medical histories on a global scale has drastically altered the competitive landscape. Evidence-based medicine, which puts an emphasis on quality and value, as well as novel treatment in place of subjective therapy, makes use of this data.

Big data analytics has the potential to revolutionize healthcare, particularly in the areas of medical research and treatment. There have been several limitations and obstructions to this, however. The biggest obstacle to adopting big data in healthcare is protecting patients' personal information and medical records. When considering the growing danger of cybercrime and the fact that any vulnerability in a medical record could be exploited by dishonest parties, it is clear that the current security measures for the healthcare business are neither efficient nor secure. As a result, experts are concerned about the safety of an unprecedented volume of sensitive medical information.

3. Methods Mathematical Model of ANN

Lorem Neural networks have garnered a lot of interest in the last ten years due to their superior learning abilities and ability to solve classification difficulties. It provides the most accurate, quick, and complicated classification solution. The artificial neuron, also known as a perceptron, is the most basic component of an artificial neural network (ANN). It is typically composed of synapses, an adder, and an activation function. Synapses, which connect every neuron to every other neuron, are assigned weights based on how strong the corresponding input link is. Together, these weighted neuronal outputs are added by the adder component. The activation function and squashing function are other names for the learning function. A typical activation function for classes that are linearly separable is described by

$$\phi(x) = \begin{cases} 1, & x \geq 0 \\ -1, & x < 0 \end{cases} \qquad (1)$$

A scaled version of the single layer perceptron, the multi-layer perceptron enhances a conventional network's capacity for learning. It is achieved by employing the nonlinear decision region to classify a set of inputs inside an input space in an acceptable manner. The sigmoid function is a commonly used activation function for non-linearly separable classes due to its straightforward derivative. The mathematical definition of the sigmoid function is

$$\varphi(x) = \frac{1}{1+e^{-x}} \qquad (2)$$

The typical architecture of a multi-layer perceptron consists of an input layer, one or more hidden layers, and an output layer. Conversely, the quantity of neurons and layers dictates training accuracy and learning capabilities. An established training set and the corresponding goal values are used to train the neural network. The output layer's results are compared against the intended solution to determine the error signal. This false signal is then carried backwards through the brain network, in the opposite direction of the synaptic connections. The process known as backpropagation learning, which modifies the neural network's weights by utilizing the error signal, can be shortened to

$$\varDelta w_{ji}(n) = \eta \delta_j y_i(n) \qquad (3)$$

where $y_i(n)$ is the neuron output value and η is the learning rate.

Ifj is in the output layer,

$$\delta_j(n) = (d_j(n) - y_j(n)).\frac{b}{a}.(a - y_j(n))(a + y_j(n)), \qquad (4)$$

or if$j$is in a hidden layer,

$$\delta_j(n) = \frac{b}{a}.(a - y_j(n)).(a + y_j(n)).\sum_k \delta_k(n).w_{kj}(n), \qquad (5)$$

a and b are the scaling values from the neuron activation function, and d_j is the intended response. Training keeps going until the neural network's weights yield convergent outputs. When an average error signal, ε_{av}, crosses a threshold, it is said to have reached convergence.

$$e_j(n) = d_j(n) - y_j(n)$$

$$\varepsilon(n) = \frac{1}{2}.\sum_{j\in c} e_j^2(n) \qquad (6)$$

where c is the collection of all output layer neurons and

$$\varepsilon_{av} = \frac{1}{N}.\sum_{n=1}^{N} \varepsilon(n) \qquad (7)$$

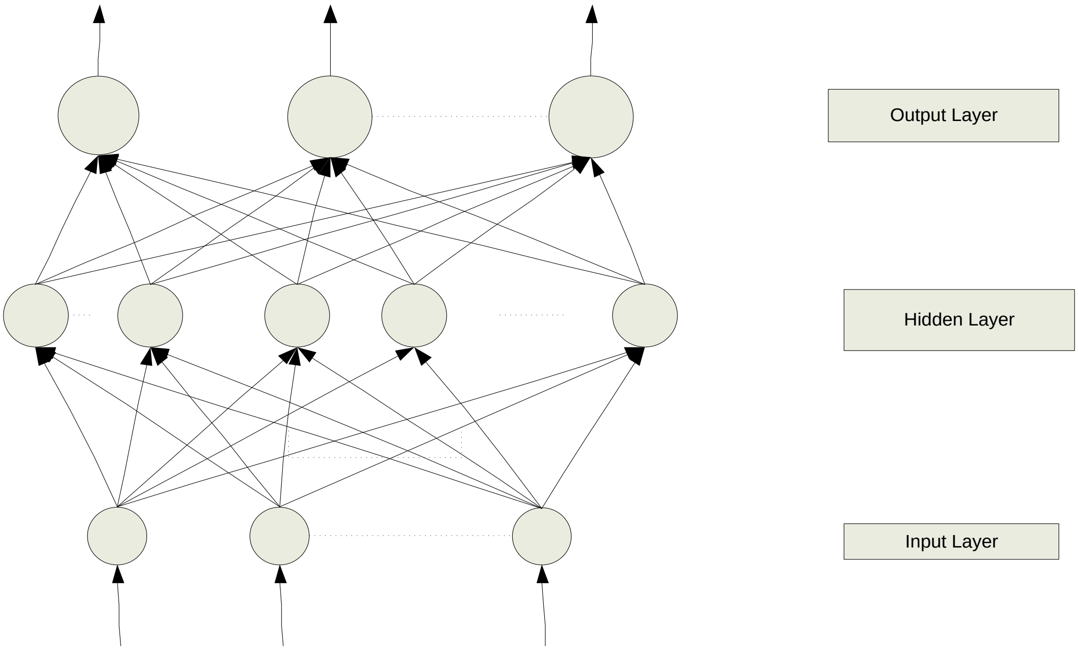

Fig 1: Three-layer Architecture of Neural Network

4. ANN based Adaptive Security Model for Healthcare Big Data

The suggested security architecture for healthcare big data comprises three technology verticals: big data, security, and machine learning. The use of AI allows for a data-sensitive, adaptive security paradigm to be provided. To create insightful categorization and evaluation, a top-notch artificial neural network is used. The most important feature of big data is its volume, or the sheer amount of information being collected. For instance, out of a 3GB data set, only a few megabytes (MB) are actually usable. On the other hand, there are other forms of cryptography that can be used for security, such as symmetric key and asymmetric key cryptography. The use of a shared key between the sender and the receiver to perform the necessary encryption and decryption operations results in an instant increase in overhead, as well as memory and computation expenses. According to the available literature, BLOWFISH and RC6 are the most well-liked and popular algorithms, but DES, AES, and others are also widely used. These methods provide a wide variety of key lengths, from 32 bits all the way up to 1024 bits, 2048 bits, 2096 bits, and beyond, with encryption and decryption durations growing proportionally. A massive cipher text is the end product of data encryption.

Since all healthcare data must be secured with a 2096-bit key, and the EHR carries the whole processing load, protecting patient confidentiality is of the utmost importance. Research in the medical field relies heavily on the data stored in electronic health care systems, including medical records, query files, patient names, surnames, phone numbers, locations, countries, email addresses, and investigative reports. However, not all information about a patient is considered confidential. In the case of a juvenile offense or a crime against women, such as physical or sexual assault, the patient's private information contained in the investigative report is extremely sensitive and cannot be distributed according to legal and constitutional duties. In other contexts, these specifics may not warrant as much caution. In a similar vein, several features of differing sensitivities can be found in the medical records of different people. There is a negative impact on throughput, delay, computation time, cost, and processing time when all characteristics are given the same level of protection. The data transfer is also wasted.

This research proposes a solution to this issue in the form of many tiers of protection for data with varying degrees of sensitivity. Here, we use a stronger level of encryption for sensitive information, such as names and addresses, and a weaker level of encryption for less sensitive information, such as symptoms of the common cold, cough, and fever. The proposed artificial neural network is used to classify the data's level of sensitivity. In the proposed system, we classified a wide range of sensitivities as follows, for example: Level 5: Victims of rape; Level 4: Infertility and sexual disorders; Level 3: Extreme Cancer; Level 2: Various cancer cases (breast & blood cancer); Level 1: Virus-related fever, seasonal flu, cold, cough, etc. Using the EMR database, the ANN is trained on the basis of sensitivity with target security level values.

The sensitivity scale runs from 1 (least sensitive) to 5 (most sensitive). Due to the fact that there are N types of diseases and N categories of disease, such as more than 150 subtypes of cancer, we have broken down disease sensitivity into a number of subcategories. The proposed method's level of sensitivity can be set by the user. The output is the degree of sensitivity, which can take on values between 1 and 5. The domain, the list of sensitive words, and the sensitivity level of the term are all considered inputs. The input is processed sequentially, and the resulting data is shown in rows. To determine the sensitivity and compute the overhead, the row is matched against the list of sensitive phrases. In this study, we use DES, AES-128, AES-256, Blowfish, and RSA as our security measures, one for each of the five levels of data sensitivity.

Encryption is a well-known method for protecting the privacy and confidentiality of one's data. Different encryption methods have different speeds, levels of efficacy, and levels of protection against attackers. A brief overview of the encryption algorithms employed in this research is provided below:

a) Data Encryption Standard (DES): In 1974, the Data Encryption Standard (DES) was created by the National Institute of Standards and Technology (NIST) as the bare minimum for secure data transmission. The U.S. government eventually approved it for use in military and civilian contexts. For both the 64-bit plain text and the plain text with 16 complicated rounds and two transposition boxes, a 64-bit key is used. Each of the 16 cycles uses the same set of cyphers, with the exception of the first and last permutations, which are both keyless straight permutations but are their inverses. The permutation works with a 64-bit key without any problems.

b) Advanced Encryption Standard (AES): In 2001, Vincent Rijmen and Joan Daeman devised a replacement encryption algorithm for DES that was able to circumvent its flaws. This symmetric encryption method uses AES-128, AES-192, and AES-256, three block cyphers. Each 128-bit cipher text is deciphered using a key that is 128 bits, 192 bits, or 256 bits in length. For 128-bit, 192-bit, and 256-bit keys, the appropriate iteration counts are 10, 12, and 14.

c) Blowfish: This encryption system was created by world-famous cryptographer Bruce Schneir in 1993. This method is the most basic and widely used one available to the public. It combines a 64-bit block cypher with a key whose length can be modified. The most efficient hardware design makes this method far more practical to apply.

d) RSA: Researchers Rivest, Shamir, and Adleman (thus the RSA acronym) developed the method in 1978. Due to its universal exponentiation in a finite field over integers, including prime values, it is considered to be the safest encryption method. This is a public-key and private-key cryptography algorithm, making it asymmetric.

The proposed artificial neural network selects the optimal strategy for real-time data encryption based on its findings. Incorporating artificial intelligence into decision-making, this work uses the best-trained neural network design using the current EMR feature dataset. These features are used to make quick decisions by controlling how sensitive the data is. The three-layered neural network is trained to deliver the optimal encryption solution for real-time data. DES, AES-128, AES-256, Blowfish, and RSA, representing levels 1 through 5, are used as training variables for the neural network. In Figure 2, we see a flowchart depicting the suggested security architecture's full functionality.

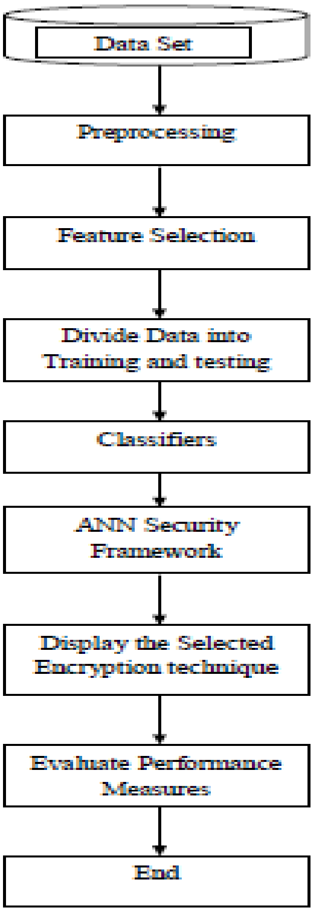
Data Set
Preprocessing
Feature Selection
Divide Data into Training and testing
Classifiers
ANN Security Framework
Display the Selected Encryption technique
Evaluate Performance Measures
End

Fig 2: Flow graph of the proposed security framework

5. Result Analysis

This paper makes use of the MIMIC-III dataset, a sizable database of patients hospitalized to a major tertiary care hospital's critical care units. This dataset contains medical records for patients treated in the intensive care units at Beth Israel Deaconess Medical Center between 2001 and 2012 [25]. This work aims to extract meaningful semantic information from unstructured data. The full dataset was utilized; however, only the free-text clinic notes and the noteevents table were examined. Discharge summaries distinguished themselves from other forms of summaries by including free text and real information. Discharge reports were prepared after the diagnosis was made, therefore any reference to the patient's classification (ICD-9 codes) was removed before being included. Table 1 displays the overall patient count, number of hospitalizations, individual ICD-9 codes, and ICD-9 categories. MIMIC-III has the full dataset, whereas noteevents and discharge summaries only cover the relevant subgroups.

Table 1. Analytics of MIMIC-III dataset

Dataset File	**Hospital Admissions**	**Number of patients**	**ICD-9 Codes**	**Categories of ICD-9**
Complete MIMIC-III	58976	46520	6984	943
Discharge	52726	41127	6918	942

Summaries				
Noteevents	58726	41127	6918	942

The top ten ICD-9 codes and top ten ICD-9 categories are shown in Table 2.For training, validation, and testing, the filtered datasets is divided into 50-25-25 groups.

Table 2. Top 10 ICD-9 codes and categories and respective admissions

ICD-9 Codes	Admissions		ICD-9 Category	Admissions
4019	20046		401	20646
4280	12842		427	16774
42731	12589		276	14712
41401	12178		272	14212
5849	8906		414	14081
25000	8783		250	13818
2724	8503		428	13330
51881	7249		518	12997
5990	6442		285	12404
53081	6154		584	11147

The effectiveness of a suggested adaptive security paradigm is measured through simulated testing. The proposed method is evaluated on an Electronic Health Care System with data sizes ranging from 1 MB to 5 GB, at which point single and multiple clusters are created.

Single-cluster requirements:

• Computer with an I3 processor

• 4 GB of RAM

•| NetBeans

• UBUNTU 18

• HADOOP 2.7

They will be used to create a java-based desktop application.

Multi-cluster requirements:

• System that is heterogeneous

• A cluster of four machines will be created.

The results of the testing and training procedures are used to evaluate the efficacy of the ANN-based decision-making system. The testing error has a root mean square value of less than 1x10-4. Using a dataset of 5000 input and target samples, the suggested neural networks are trained and tested in order to determine the efficacy of the intelligent security model. The sensitivity of the data is determined by the features obtained from the database using keywords such disease name, patient attributes, and EMR data as training inputs for the ANN. The proposed ANN is supposed to generate a decision for the assigned encryption techniques for each data sample utilizing DES, AES-128, AES-256, Blowfish, and RSA, with sensitivity levels 1 through 5. Figure 3 depicts the effectiveness of the proposed framework in making decisions. Each medical data packet is encrypted using one of five different methods, which are represented by the five different categories.

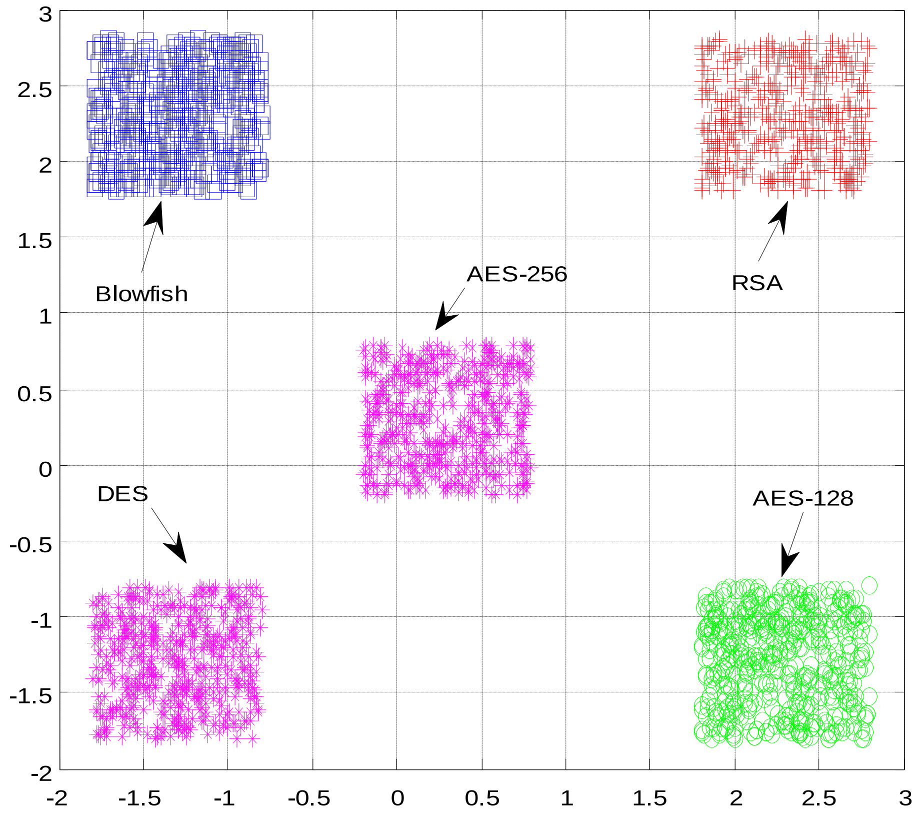

Fig 3:ANN based Encryption technique Allotment

6. Conclusion

This research recommends an ANN-based flexible intelligent security framework for protecting sensitive data. The suggested solution protects a healthcare dataset that includes information about patients, disease features, symptoms, medical reports, and electronic medical records (EMR). There are five levels of data sensitivity for the characteristics of healthcare data, ranging from level 1 to level 5. In this analysis, five distinct encryption methods are utilized as the target security model, with the right choice being made based on the level of sensitivity. The ANN has been trained on 5000 samples of medical data to determine the most effective encryption method to use for each individual packet. When compared to the standard static security architecture, the proposed solution achieves superior results in performance evaluations. Simulation results reveal that the proposed approach is more effective than the standard static security architecture. Low-dimensional security minimizes algorithmic complexity and enhances system efficiency when it is deployed to low-sensitivity data. Protecting the most sensitive data using the strongest security method also improves the overall security of the system. However, the study concludes that multi-cluster processing can accommodate the bigger data quantities, despite the fact that encryption times increase with data volume. Time spent at low sensitivity is reduced, while time spent at medium sensitivity is increased, and time spent at high sensitivity is increased for a considerable period. Because data encrypted at Level 3 does not require decryption, less processing time is needed for mining purposes. Finally, the amount of protection afforded to sensitive information is classified. Classification using low and medium levels of processing reduces data volume along with diversity. When diversity drops, data naturally splits into subsets, simplifying data mining.

Refrences

[1] Land, K.J.; Boeras, D.I.; Chen, X.-S.; Ramsay, A.R.; Peeling, R.W. REASSURED diagnostics to inform disease control strategies, strengthen health systems and improve patient outcomes. Nat. Microbiol. 2019, 4, 46–54.

[2] Marques, G.; Pitarma, R.M.; Garcia, N.; Pombo, N. Internet of things architectures, technologies, applications, challenges, and future directions for enhanced living environments and healthcare systems: A review. Electronics 2019, 8, 1081.

[3] Shafique, K.; Khawaja, B.A.; Sabir, F.; Qazi, S.; Mustaqim, M. Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios. IEEE Access 2020, 8, 23022–23040.

[4] Mansour, R.F.; El Amraoui, A.; Nouaouri, I.; Diaz, V.G.; Gupta, D.; Kumar, S. Artificial intelligence and internet of things enabled disease diagnosis model for smart healthcare systems. IEEE Access 2021, 9, 45137–45146.

[5] Zeadally, S.; Siddiqui, F.; Baig, Z.; Ibrahim, A. Smart healthcare: Challenges and potential solutions using internet of things (IoT) and big data analytics. PSU Res. Rev. 2020, 4, 149–168.

[6] Kim S-H, Kim N-U, Chung T-M. Attribute relationship evaluation methodology for big data security. In: 2013 international conference on IT convergence and security (ICITCS), IEEE. p. 1–4.

[7] "Data-driven healthcare organizations use big data analytics for big gains" IBM white paper February. 2013.

[8] Yazan A, Yong W, Raj Kumar N. Big data life cycle: threats and security model. In: 21st Americas conference on information systems. 2015.

[9] Zhang R, Liu L. Security models and requirements for healthcare application clouds. In: IEEE 3rd international conference on cloud computing. 2010.

[10] Xindong WU, Gong Qing WU and Wei Ding, "Data Mining with Big Data," IEEE Transaction on Knowledge and Data Engineering, vol. 26, no. 1, pp. 97- 107, Dec. 2012.

[11] G. Ghinita, "Privacy for location-based services synthesis," Lectures on Information Security, Privacy, and Trust, University of Massachusetts, Boston, Tech. Rep., April 2013.

[12] X. Liang, R. Lu, L. Chen, X. Lin, and X. Shen, "Pec: A privacy preserving emergency call scheme for mobile healthcare social networks," Communications and Networks, Journal of, vol. 13, no. 2, pp. 102–112, April 2011.

[13] M. A. D. Mashima, D. Bauer and D. Blough, "User-centric identity management architecture using credential-holding identity agents," Digital Identity and Access Management: Technologies and Frameworks, IGI Global, December 2012.

[14] F. Paci, R. Ferrini, A. Musci, K. Steuer, and E. Bertino, "An interoperable approach to multifactor identity verification," IEEE Computer, vol. 42, no. 5, pp. 50–57, 2009.

[15] R. Lu, X. Lin, and X. Shen, "Spoc: A secure and privacy preserving opportunistic computing framework for mobile-healthcare emergency," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 3, pp. 614–624, March 2013.

[16] Sweeney L. Achieving k-anonymity privacy protection using generalization and suppression. Int J Uncertain Fuzziness Knowl Based Syst. 2002;10:571–88.

[17] Samrati P. Protecting respondents identities in microdata release. IEEE Trans Knowl Data Eng. 2001;13:1010–27.

[18] Truta TM, Vinay B. Privacy protection: p-sensitive k-anonymity property. In: Proceedings of 22nd international conference on data engineering workshops. 2006. p. 94.

[19] Mohan A, Blough DM. An attribute-based authorization policy framework with dynamic conflict resolution. In: Proceedings of the 9th symposium on identity and trust on the internet. 2010.

[20] Zhou H, Wen Q. Data security accessing for HDFS based on attribute-group in cloud computing. In: International conference on logistics engineering, management and computer science (LEMCS 2014). 2014.

[21] Wang, H., Yin, J., Perng, C. S., & Yu, P. S. (2008, October). Dual encryption for query integrity assurance. In Proceedings of the 17th ACM conference on Information and knowledge management (pp. 863-872). ACM.

[22] Shafer J, Rixner S, Cox AL. The hadoop distributed filesystem: balancing portability and performance. In: Proceedings of 2010 IEEE international symposium on performance analysis of systems & software (ISPASS), March 2010, White Plain, NY. p. 122–33.

[23] Othman, S.B.; Almalki, F.A.; Chakraborty, C.; Sakli, H. Privacy-preserving aware data aggregation for IoT-based healthcare with green computing technologies. Comput. Electr. Eng. 2022, 101, 108025.

[24] Rawat, R.; Mahor, V.; Garg, B.; Chouhan, M.; Pachlasiya, K.; Telang, S. Modeling of cyber threat analysis and vulnerability in IoT-based healthcare systems during COVID. In Lessons from COVID-19; Academic Press: Cambridge, MA, USA, 2022; pp. 405–425.

[25] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. Mimic-III, a freely accessible critical care database. Scientific data, 3:160035, 2016.