Data Sensitive Safety Framework for Big data analysis in Healthcare using Deep Learning

Preeti chaudhary,
Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Abstract

Medical sector generate huge amount of data which need to be analysed and classified in to the information, information is the meaningful data, whereas data is unorganized and have no proper meaning. To analysis this big amount of data deep learning techniques can be used which provide more efficient and accurate analysis of information with privacy. Privacy of the sensitive data is major concern, so if we use single or same data safety architecture then if it will face failure then all the data will be on hack. In this article we discuss a method which uses benefits of big data analysis analysis, safety of data by EMR and distribution of data using ANN machine learning method to adapt intelligent safety level to health care data.

Keywords: Medical, big data analysis, machine learning, safety, classification.
DOI Number: 10.48047/NQ.2022.20.4.NQ22338 **NeuroQuantology2022;20(4): 1136-1142**

I. INTRODUCTION

The adoption of complete medical data digitalization has ushered in a revolution in the medical area during the previous decade. The electronic medical records (EMR) created because of this approach have drastically changed the diagnostic and research paradigms. The ability to access the EMR at any time and from any location has brought all of the information together in one place. It is also derived from various sources such as remote sensors data, biological and genetic records [1]. The amount of data generated in the field of healthcare has expanded due to advancements in biomedical machineries like as medical equipment, digital health records, apprehending devices, and mobile computing.

Big data analysis, in general, is a complicated and large volume of data that requires the most advanced tools to handle and manage for analysis. The varied character of the data adds to the processing difficulties. The diversity, amount, pace, and veracity of the data are all characteristics of big data analysis. Variety denotes the heterogeneous, random, and unstructured data generated from various sources of information; volume represents the amount of data, which is typically in the terabyte range. In big data analysis, velocity refers to the speed with which this massive amount of data is processed, and decisions are made, whereas veracity refers to the data's trustworthiness and appropriateness [2,3].

Big data analysis analytics' ability to cope with large amounts of data makes it an ideal processing and decision-making tool in the field of healthcare, where the volume of data is growing by the day. Healthcare data may be effectively exploited for research and development using big data analysis analytics approaches. The obtained datasets can then be used to generate targeted knowledge and provide clear insight. It can be extremely beneficial in the event of global health emergencies such as the coronavirus (COVID-19), H1N1, Dengue, Ebola virus, and others, by enabling

accurate decision- making and potentially saving lives through efficient knowledge and data sharing among all countries through predictive analysis. It can also decrease the budget of treatments and advance operational efficiency. The use of big data analysis in the medical area will aid in the prevention of epidemics, the improvement of life style, the efficiency of medical trials, the detection of drug side effects, and so on.

However, data safety and privacy of data collected in healthcare has been a major topic tackled by several scholars in recent years. A breach in data safety could jeopardize people's lives and be utilized in a large-scale medical and economic calamity. Many healthcare organizations rely on the medical data they save, preserve, and communicate for their support and delivery. This information is the most vulnerable to cyber-attacks and is the most likely to be made public. To address this problem, a number of researchers have developed several safety models to protect data from malicious assaults and leaks [6].Data safety, access control, and information safety are three components of big data analysis safety for healthcare, according to Kim et al. [7]. They advised that companies secure the safety of big data analysis using appropriate hardware and software for clinical and administrative data. Data retention, reuse, and auditing must all be done at the start of the project to ensure cost effectiveness.

For knowledge generation, Yazan et al. [8] established a safety architecture that includes data collecting, data storage, data processing, and data analysis. They provided a paradigm for the entire big data analysis safety lifecycle.

Swaney et al. [17,18] presented a k-anonymity strategy of data masking. Truta et al. [19] enhanced this work by providing p-sensitive anonymity in the safety framework by adding the attributes alongside the identification parameters. These approaches, on the other hand, were unable to perform well when dealing with high-dimensional dataset anonymity.

Researchers in the field of healthcare have also developed an access control-based safety paradigm. A. Mohan and D.M. Blough [20] discussed and designed an access control system based on attributes.

Adaptive Safety is a concurrent safety approach that continuously monitors behaviour and events to mitigate risk and respond properly to attacks before they materialize. The primary purpose of adaptable safety is to create an analytical framework of hazard detectable quality, acknowledgment, and shirking that consistently results in more effective outcomes. One of the most important principles of flexible safety is to always expect an issue with the framework. The most basic requirements are consistent monitoring and improvements in safety engineering. The standard operating procedure is to not trust that an incident will occur, but to anticipate, recognize, and react to it before it has a chance to rupture the framework [23].

Adaptive safety relies heavily on safety research and machine learning. Furthermore, distinctive examination distinguishes uncommon events, symptomatic investigation assists in determining why an unpleasant event occurred, and predictive investigation can detect suspicious behaviour based on verifiable evidence. Machine learning can meet a supportive demand with unbounded amounts of Big data analysis verified by data warehouses in the cloud and hazardous development covered as genuine bearings, and server requests becoming increasingly difficult to distinguish. It can help a safety obtain information by automating numerous processes, such as plan affirmation, which is employed in evaluation. This study provides an adaptive safety approach for big data analysis analytics in the medical area based on AI.

II. Analyzing the Big Data in Medical sector

The amount of data that has been considered by companies for policy framework and market research has shifted the entire paradigm. Big data analysis analytics is the process of extracting relevant information from a large amount of

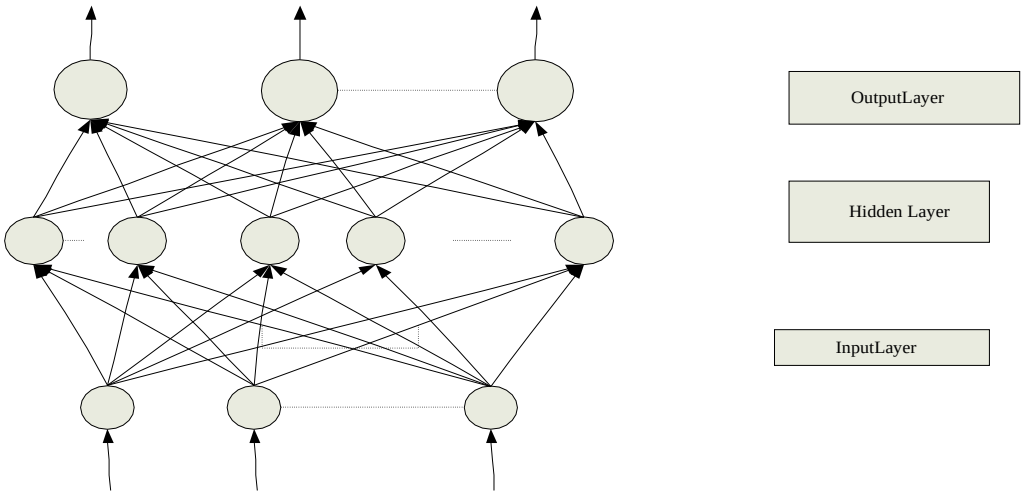
data in the shortest amount of time with the highest level of accuracy. Parallel processing frameworks in computing have shown to be a godsend in big data analysis analytics, allowing for the identification of the data's potential for decision-making. Big data analysis has several advantages, including data mining, which is versatile and generalized. Big data analysis generates a large volume of data at a high rate, and this volume of data is diverse. Big data analysis is a type of data that creates a lot of information. It contains both unstructured and structured data; nevertheless, most of the data is unstructured, which poses the greatest issue. However, the improvements in the IT automation business over the last few years have been nothing short of revolutionary. In the IT sector, huge strides have been made in areas such as electronics, banking, and e-commerce.

In terms of technology's involvement in medical research and diagnosis, the medical area has likewise altered dramatically during the previous two decades. The availability of medical information and treatment histories for millions of patients throughout the world has altered the entire dynamics of the industry. Instead of subjective therapy, this data has been used for novel treatment, focused care quality and value, and evidence-based medical practice.

Big data analysis analytics in the healthcare industry has the potential to dramatically transform medical treatment and research. It has, nonetheless, presented numerous obstacles and limits for the same. The most difficult aspect of implementing big data analysis in healthcare is ensuring the privacy and safety of the EMR. For the ever-increasing threat of cybercrime, the present safety frameworks for the medical area are neither very efficient nor secure. Any breach of the medical record could be used by unscrupulous stakeholders, resulting in a medical blunder over the world. As a result, the researchers are concerned about the safety of the massive volume of medical data.

III . MATHEMATICAL MODEL OF THE ANN

Because of its superior learning skills and capacity to tackle categorization problems, neural networks have gotten a lot of attention in the recent decade. In terms of complexity, speed, and accuracy, it gives the best categorization solution. The most fundamental building element of an ANN is the artificial neuron or perceptron, which is represented by synapses, an adder, and an activation function in a common arrangement. Each neuron is linked by synapses, each of which has its own weights that correspond to the strength of the relevant input link. The adder component adds all of these weighted outputs of neurons. The learning function is also known as the activation function or squashing function.



IV.ANN BASED ADAPTIVE SAFETY MODEL FOR HEALTHCARE BIG DATA ANALYSIS

Big data analysis, safety, and machine learning are three technical verticals covered by the discussed and designed safety architecture for healthcare big data analysis. It uses AI to provide a data-sensitive adaptive safety model. An optimum artificial neural network is used to produce intelligent categorization and decision-making. The most important feature of big data analysis is volume, which refers to vast amounts of data. For example, if there is 3GB of data, only a few MB of it is helpful, while the rest is squandered and useless. The symmetric key cryptography and asymmetric key cryptography algorithms, on the other hand, are safety algorithms. Because key is utilized in this approach, and both the sender and receiver maintain their key to encrypt and decrypt data, it directly jumps on overhead and causes overhead in terms of memory and computation time. The BLOWFISH and RC6 algorithms are the most widely used and appreciated algorithms in existing research; other algorithms include DES, AES, and others. These algorithms work with key lengths ranging from 32 bits through 1024 bits, 2048 bits, 2096 bits, and beyond; as key lengths increase, so does encryption and decryption time. The cypher text we get after data encryption is quite huge.

In healthcare big data analysis, the privacy and safety of the EMR is critical, since it is necessary to encrypt all data with a 2096-bit key and to carry 100 percent computing overhead on the Electronic Health Care System. Medical research relies heavily on medical reports, query files, patient names, surnames, phone numbers, cities, countries, e-mail addresses, and investigation reports stored in electronic health care systems. However, not all aspects of a patient's data are equally sensitive. For example, in the instance of a juvenile crime or a crime against women, such as physical or sexual assault, the patient's personal information in the investigative report is extremely sensitive and cannot be released due to legal and constitutional requirements. However, these facts may not be as sensitive in other cases. Similarly, many patients' medical records contain a variety of features with differing degrees of sensitivity. Applying the same level of protection to all attributes wastes data rate and compromises throughput, latency, computation time, cost, and processing time, among other things.

This study suggests a solution to this challenge by providing different levels of safety for data with varying levels of sensitivity. In this scenario, we use a high level of encryption for data of high sensitivity and a lower level of protection for data of lower sensitivity, such as viral infections, colds and coughs, fever, and so on. The discussed and designed artificial neural network is used to classify the data's sensitivity. The ANN is trained based on sensitivity with target safety level values using the EMR database..

Encryption is a well-known technology for ensuring data secrecy and privacy. Different encryption systems differ in their speed, efficiency, and capacity to protect data from assaults. The encryption algorithms utilized in this project are briefly detailed below:

a) Data Encryption Standard (DES): In 1974, the National Institute of Standards and Technology (NIST) created the Data Encryption Standard (DES), which is the simplest encryption algorithm. The US government later adopted it for military and non-military use. It employs a 64-bit key for a 64-bit plain text with 16 complicated rounds and two transposition boxes, as well as a 64-bit key for a 64- bit plain text. These 16 rounds are iterated with the same cyphers, but the first and last permutations are keyless straight permutations, and the first and last permutations are inverses of each other. The permutation accepts a 64-bit key and processes it appropriately.

b) Advanced Encryption Standard (AES): Vincent Rijmen and Joan Daeman created this encryption algorithm in 2001 to overcome the limitations of DES. AES-128, AES-192, and AES-256 are three block cyphers used in this symmetric encryption technique. Each 128-bit cypher text is

processed with 128-bit, 192-bit, and 256-bit keys, respectively. For 128-bit keys, 192-bit keys, and 256-bit keys, the number of iterations is 10, 12, and 14 accordingly.

c) Blowfish: Bruce Schneir, a well-known cryptologist, devised this encryption approach in 1993. This is the most basic and widely used encryption method in the public domain. It employs a 64-bit block cypher and a key with a configurable length. Because of the ideal hardware design, this technique's implementation is comparatively more practicable.

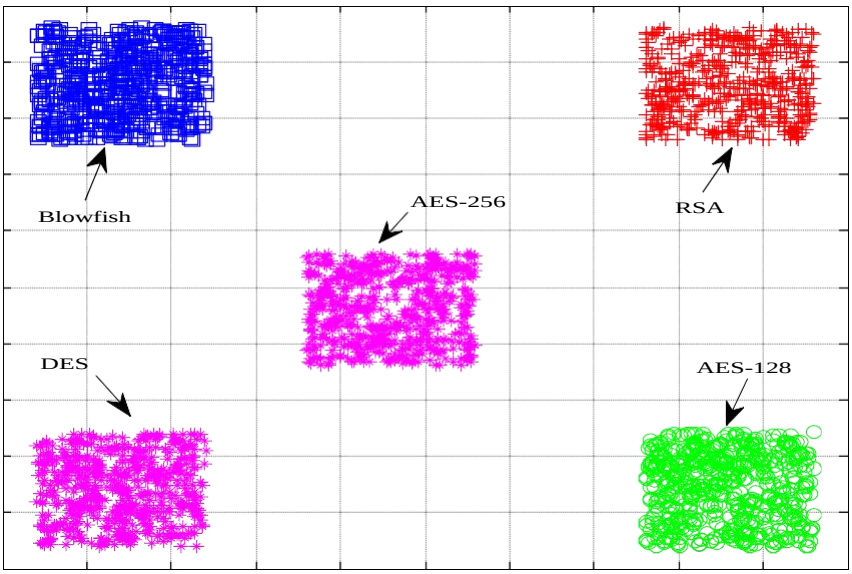
d) RSA: The Rivest, Shamir, and Adleman researchers who devised the RSA technique in 1978 are called after it. Because of its widespread exponentiation in a finite field over integers, including prime numbers, it is regarded as the most secure method of encryption. Because it uses two separate keys, it is an asymmetric cryptographic algorithm (public and private)

The discussed and designed artificial neural network derives and decides on the optimum encryption scheme for immediate data. This work incorporates intelligence into the decision-making process by utilizing the ideal design of a neural network that is trained using the available dataset of EMR features. These features control the data's sensitivity and are utilized to make quick decisions. The three-layered neural network is trained to generate the optimum encryption solution for the instantaneous data in real time. The whole operation of the discussed and designed safety architecture is depicted in fig. 2 as a flow graph.

V. RESULT ANALYSIS

The training and testing results are used to evaluate the performance of the ANN-based decision- making system. To assess the performance of the intelligent safety model, the discussed and designed neural networks are trained and tested on dataset samples of input and targets. The data sensitivity level is determined by features retrieved from the data base using keywords such as illness name, patient traits, and EMR information as training inputs for the ANN. Figure 3 depicts the suggested framework's decision-making efficiency. Every medical data packet corresponds to one of five encryption algorithms represented by the five classes.

Fig2: ANN based Encryption technique Allotment



VI. CONCLUSION

This research proposes an adaptive intelligent safety framework that employs ANN for data- sensitive safety. The suggested technique is utilized to secure a healthcare dataset that includes patient information, illness attributes, symptoms, medical reports, and electronic medical records (EMR). The discussed and designed technique outperforms the traditional static safety framework in terms of performance evaluation. The accuracy obtained through simulation analysis demonstrates the superiority of the discussed and designed technique over the traditional static safety framework. By applying low-dimensional safety to data with low sensitivity, the algorithmic complexity is reduced, and the system's performance is improved. In addition, using the highest safety approach for the most sensitive data improves the system's safety.

REFERENCES

[1] S. Sharathkumar and G. Jagadamba, "Adaptive content-aware access control of EPR resource in a healthcare system," 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, 2017, pp. 205-210.

[2] M. Jayabalan and T. O'Daniel, "Continuous and transparent access control framework for electronic health records: A preliminary study," 2017 2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2017, pp. 165-170.

[3] R. Sánchez-Guerrero, F. A. Mendoza, D. Díaz-Sánchez, P. A. Cabarcos and A. M. López, "Collaborative eHealth Meets Safety: Privacy-Enhancing Patient Profile Management," in IEEE Journal of Biomedical and Health Informatics, vol. 21, no. 6, pp. 1741-1749, Nov. 2017.

[4] A. Dev Mishra and Y. Beer Singh, "Big data analysis analytics for safety and privacy challenges," 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, 2016, pp. 50-53.

[5] Beyer, Mark "Gartner Says Solving 'Big data analysis' Challenge Involves More Than Just Managing Volumes of Data".. Archived from the original on July 2011. Retrieved 13 July 2011

[6] Large-Scale Adaptive Machine Learning for Safety Analytics, Ling Huang, Joint work with ISTC and McAfee Labs ISTC Summer Retreat, OS/3112013.

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

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

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

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

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

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

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

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

[20] 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.

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

[22] 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.

[23] 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.

[24] Yang C, Lin W, Liu M. A novel triple encryption scheme for hadoop-based cloud data safety. In: Emerging intelligent data and web technologies (EIDWT), 2013 fourth international conference on. 2013. p. 437–42.

[25] Federal Information Processing Standards Publication 197. Specification for the advanced encryption standards (AES). 2001.