Abstract—

Index Terms—

I. Department of Computer Science and Engineering, Government College of Engineering, Dharmapuri, 636704, Tamilnadu, India 2 Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, 641004, Tamilnadu, India *Corresponding Author: Kirupa Shankar Komathi Maathavan. Email:kirupaa1991@gmail.com Received: 02 August 2021; Accepted: 03 September 2021

Abstract: The major operation of the blood bank supply chain is to estimate the demand, perform inventory management and distribute adequate blood for the needs. The proliferation of big data in the blood bank supply chain and data management needs an intelligent, automated system to classify the essential data so that the requests can be handled easily with less human intervention. Big data in the blood bank domain refers to the collection, organization, and analysis of large volumes of data to obtain useful information. For this purpose, in this research work we have employed machine learning techniques to find a better classification model for blood bank data. At the same time, it is vital to manage data storage requirements. The Cloud offers wide benefits for data storage and the simple, efficient technology is adapted in various domains. However, the data to be stored in the cloud should be secured in order to avoid data breaches. For this, a data encryption module has been incorporated into this research work. The com- bined model provides secure encrypted classified data to be stored in the cloud, which reduces human intervention and analysis time. Machine learning models such as Support Vector Machine (SVM), Multinomial Naive Bayes (MNB), Deci- sion Tree (DT), Random Forest (RF), Gradient Boosting (GB), K-Nearest Neigh- bor (KNN) are used for classification. For data security, the Advanced Encryption Standard with Galois/Counter Mode (AES–GCM) encryption model is employed, which provides maximum security with minimum encryption time. Experimental results demonstrate the performance of machine learning and encryption techniques by processing blood bank data. Keywords: Electronic health records (EHR); big data; classification; machine learning; data security; encryption; cloud 1 Introduction World Health organization (WHO) reports that on an average around 118.5 million blood donations happened globally in 2018. The report covers that 72% or 123 out of 171 countries had a national blood policy. From 2013 to 2018 the rate of blood donation has increased into 7.8 million which is reported by 156 countries. Handling these large volumes of data essentially needs an efficient processing system.

donors, blood bags inventories and transfusion services. The manual analysis requires more time and chances of errors is large due to large volume of data. These time consuming and manual data management are eradicated in the digital era. Technology development reduces the human efforts and improves the diagnosis precision in the healthcare sector due to digital technologies. Though the healthcare records are digitized still it requires human intervention to analyze the data. Medical data analysis needs high precision and accuracy so that further issues can be eliminated [2]. The blood data management analysis can be categorized into two modules. The first one is pure technical which essentially manages the data related to blood samples after processing the sample. The second one majorly deals the user data such as personal information, sample collection location, data. Analysis of these user data can be helpful to utilize the same person in future case if there is a required of blood. For this purpose, data analyzers are introduced in healthcare domain which classifies the sensitive data into different classes. Machine learning techniques are one among them which is widely used for various classification and clustering approaches in image processing applications [3]. Whereas in healthcare data analysis, machine learning models are employed in recent years. The sensitive user data can be identified and classified using machine learning techniques reduces the human intervention and errors in data management. While machine learning gains more attention in healthcare data analysis, cloud computing transfers the medical data analysis into next level as virtual storage and ease access of healthcare data. The rapid growth of huge amount of data needs an efficient platform to handle and process the data [4]. Cloud offers numerous benefits and the virtual resources can store larger amount of data. Due to these benefits, Electronic Health Records (EHR) are moved into cloud platform. However, the same digital platform introduces numerous security and privacy challenges. Specifically, in healthcare data the user privacy is a major concern and preserving the user privacy from security attacks is a crucial task. Cloud services are categorized into public cloud, private cloud, and hybrid cloud. Most of the healthcare data management systems employs public cloud which cannot fully be trusted by users [5]. The data outsourced in cloud are sensitive so privacy and security becomes major concern while deploying cloud services for EHR. Cloud offers several security measures to ensure the privacy and security of user data. However, from user side there is no such security measure so while transferring data to cloud it can be accessed. To prevent this, the data is encrypted in the user end and then transfer it to cloud is the only

II. solution. Various encryption algorithms are evolved for data encryption however, it is essential an encryption algorithm should provide maximum security with minimum computation and communication cost. The research work objective is framed to analyze electronic healthcare registers through machine learning techniques. Followed by classification, an efficient encryption for healthcare data is obtained to ensure the user data security and privacy. Finally, the encrypted data is moved into public cloud environment. Data collected from blood banks are analyzed through machine learning algorithms and based on the results the better performance machine learning model results are encrypted using Advanced Encryption Standard (AES) encryption with Galois/Counter Mode (GCM) for enhanced security. 766 IASC, 2022, vol.32, no.2 2 Related Works The rate of healthcare data increases rapidly and handling the data manually is a tedious process. It might consume more time also increases the probability of erroneous in the final results. To avoid this issue, healthcare records are converted digitally as EHR and stored in a data repository [6]. Feature selection is an important process in healthcare data analysis. Based on the selected features, classification is performed so that individual risk and preventive measures can be provided [7,8]. Similar to feature selection, feature reduction is also an important process in healthcare data analysis [9]. Processing huge volume of data will increase the computational complexity of the system. Feature reduction in healthcare data analysis reduces the error rate and increases the performance of the system. Numerous ML applications are applied in the healthcare data analysis [10–15] for classifying the medical data. Based on the classification results a suitable decision can be obtained which reduces the extra burden of physicians. However, the healthcare records have various sensitive and user privacy information. It is essential to identify the sensitive data and preserve the user privacy is essential. ML techniques can be utilized to categorize the data into sensitive and non-sensitive data, so that user data security methodologies can be included in the data management process. A multi-source ordered preserving encryption for cloud-based eHealth system reported in [16] identifies the threats like frequency analysis, identical data inference and privacy leakage. An enhanced model of Merkle Hash Tree for multicopy storage of electronic medical records is reported in [17] that prevents data loss, unauthorized access to the sensitive user data. Lower communication and computation cost are considered as the merits of the research work. Sensitive and Energetic Access Control (SE-AC) mechanism presented in [18] ensures the

III. data confidentiality of electronic health records and prevents authorized access. Secure

Depends on the size of the block, key length, the number of rounds will be given as 14 for 256 bits, 12 for 192 bits and 10 for 128 bits. Tab. 1 depicts the key size and its respective number o

References