ACCESS CONTROL AND PRIVACY PRESERVATION OF MEDICAL RECORDS WITH ENHANCED RIVEST-SHAMIR-ADLEMAN ALGORITHM USING COUNTER MODE ENCRYPTION

BY

USMAN, Hassan
MTech/SICT/2019/9846

DEPARTMENT OF COMPUTER SCIENCE
FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA.

JULY, 2023

ACCESS CONTROL AND PRIVACY PRESERVATION OF MEDICAL RECORDS WITH ENHANCED RIVEST-SHAMIR-ADLEMAN ALGORITHM USING COUNTER MODE ENCRYPTION

BY

USMAN, Hassan
MTech/SICT/2019/9846

A THESIS SUBMITTED TO THE POSTGRADUATE SCHOOL FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA, NIGERIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF MASTER OF TECHNOLOGY (M.Tech) IN COMPUTER SCIENCE.

JULY, 2023

ABSTRACT

Cryptography is a broad research area in the field of information technology, which deals with a protocol for securing data, information. Cryptography involves encryption and decryption mechanism in data access control and protection. Due to the large volume of data availability, a suitable environment has been created for illegal or criminal activities. This includes discrimination, stealing or modifying medical files or content for personal or commercial gain. However, this rapid growth in information theft has led to the development of many encryption mechanisms to protect user (patient) privacy and access control of confidential medical information. In order to overcome these problems of discrimination, data theft and medical records modification, this work proposed a robust integration of an Enhanced Rivest-Shamir-Adleman (RSA) Algorithm using Counter (CTR) mode encryption method in developing a faster, and secure techniques for medical record privacy preservation and access control. Python is used as programming language while kivy language is used to develop user interfaces. The hepatitis diagnosis dataset with 155 datasets (records) was downloaded from Kaggle repository, and used in testing and validating the implemented techniques. The encryption techniques were subjected to performance metrics that include; key length, encryption/decryption time, and Throughput. The proposed integrated algorithm (RSA and CTR) of encryption and decryption time gave the following results; (6.448m/s,38.67m/s) for 1KB data, (12.896m/s,77.38m/s) for 2KB data, (32.24m/s,193.9m/s) for 5KB data (64.48m/s,386m/s) for 10KB data respectively. Also, the key length gave (2^{128}) data size, and RSA (2^{1024}) data size, which together show a trillion possible key combinations, and Throughput gave 162.32(kb/s). In conclusion, the developed Technique shows that the encryption time is the fastest compared to the existing techniques. This technique can be used in medical sectors such as laboratory tests for (HIV and AIDS, COVID 19), to free people from discrimination, and other criminal acts in society.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF ABBREVIATIONS

Abbreviations	Meaning
AES	Advance Encryption Standard
CBC	Cipher Block Chaining
CFB	Cipher Feedback
CSV	Comma-Separated Values
CTR	Counter
DES	Data Encryption Standard
DK	Decryption Key
ECB	Electronic Code Block
EF	Encryption File
EHR	Electronic Health Records
EK	Encryption Key
EMR	Electronic Medical Records
IDE	Integrated Development Environment
NHS	National Health Service
OFB	Output Feedback
RSA	Rivest-Shamir-Adleman

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background to the Study

Large amounts of data are generated, processed, and stored every day in today's industry. The collection and handling of this vast amount of data is becoming difficult because of technology and human resource costs. For this vast volume of data, privacy, and protection are the critical concerns (Mohammed, *et al.,* 2020). This massive amount of data can also be called big data. Big Data defines very broad data sets with extra varied and dynamic systems, such as social media, weblogs, email, sensors, and photographs (Goswami, 2017). These vast data are mainly stored for backup and quick retrieval or access on the internet or in the cloud. With this data being online, it is vulnerable to breaches of privacy and security (Koo, *et al.,* 2020). For instance, using the healthcare industry that stores its information in the cloud, a dishonest cloud provider staff can reveal patient sensitive information (like patient personal data) to commercial organizations for some financial benefit (Roehrs, *et al.,* 2019). In the big data domain, data privacy is a problematic field centred on data security itself. Data owners need to encrypt data before saving the data to the cloud or before transmitting the data between systems to secure and retain data privacy and combat unauthorized access to such sensitive data. Encrypted data for analysis and computer activity cannot occur unless a clear text record is put in place that can easily be understood and effectively used. Therefore, for normal operations to be carried out, the encrypted data must be decoded. The encryption, and general analysis tasks on encrypted datasets cannot be executed by a user who has no credential or an unauthorized user (Roehrs, *et al.,* 2019).

The current trend toward digitizing healthcare workflows and moving to electronic patient records has been seen as a paradigm shift in the healthcare industry.

Abouelmehdi, *et al.*, *(*2017), stated that the quantity of clinical data that are available electronically will be dramatically increased in terms of complexity, diversity, and timeliness, resulting in big data. Due to recent technological development, the amount of data generated by the internet, social networking sites, sensor networks, healthcare applications, and many other companies, is drastically increasing day by day. All the enormous measure of data produced from various sources in multiple formats with very high speed is referred to as big data. The term big data is defined as "a new generation of technologies and architectures, designed to economically separate value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and analysis". On the premises of this definition, the property of big data is a term used for very large data sets that have more varied and complex structure (Deepa, *et al.*, 2022).

Big data is most commonly described as a huge amount of unstructured data or we can refer to it also as semi-structured data. The processing, storage, and analysis of such huge sets of data with the help of classical processing or database approach or tools are insufficient. It requires advanced processing tools with real-time analysis. Healthcare (medical) is one of the important sectors that produce big data because today, healthcare switches paper-based medical records into electronic platforms to store, manage, analyze, and process data in the form of Electronic Medical Records (EMR) or Electronic Healthcare Record (EHR) with the help of internet (Pandey and Pandey, 2018).

Cryptography is an approach used in providing protocol and mechanism used in securing the channel of communication by assuming that an unauthorized third party

existed. Key-based cryptographic algorithms are broadly divided into symmetric and asymmetric encryption. In symmetric encryption, a single key is required to encrypt and decrypt information or data needed to be secure. While Asymmetric encryption adopts or uses two keys, the public key for encryption and the private key for decryption, to secure information or data. However, the public key of an asymmetric encryption scheme is known to the public, while the private key is only known by the intended receiver (Dawson, *et al.,* 2022). One of the most popular and extensively used symmetric encryption schemes is the Advance Encryption Standard (AES), which uses a single key to perform the encryption and decryption operation. It is considered safe against all known attacks. Various key sizes are utilized in AES such as 128,192, and 256 bits with a block size of 128 bits. The operation of the Advance Encryption Standard (AES) is performed on bytes unlike Data Encryption Standard (DES) which operate on bits. There are various modes of AES, including; Cipher Block Chaining, counter (CTR), Electronic Code Block, and the likes (Sultan, *et al.,* 2020). Furthermore, the best known widely used public key system is the Rivest-Shamir-Adleman (RSA) algorithm which consists of a key generation phase, an encryption phase, and finally decryption phase (Al-kaabi and Belhaouari, 2019).

1.2 Statement of the Research Problem

The present society demands a degree of connectivity between individuals, businesses, Hospitals, and governments that must cut across political and cultural boundaries. Digital technology with Big Data provides this connectivity and gives its users many benefits. But at the same time, it provides a rich environment for criminal activities such as impersonating personal health records, stealing of identity or classified government/organizational information, hacking, and discrimination. The rate of information theft is rapidly growing and this pose as a serious threat to internet security.

According to Kittur, *et al.*, (2019), they identify Privacy Preservation for e-health Big Data and integrate mobile devices to mitigate the drawbacks of big data in e-health. However, the lack of encryption became a loophole, which motivated the induction of the RSA algorithm and counter (CTR) mode of encryption for the patient to have complete control over health records. This can be achieved through the Enhanced Rivest-Shamir-Adleman Algorithm Using Counter Mode Encryption. However, based on the research of Al-kaabi and Belhaouari, (2019) RSA come with limitation such as constraints on data length that can be encrypted, relatively slow, and computational cost due to dual key generation. This can be resolved using Counter (CTR) encryption mode along with RSA.

1.3 Aim and Objectives

This study aims to develop an Access Control for the Privacy Preservation of Medical Records with Enhanced Rivest-Shamir-Adleman (RSA) Algorithm using Counter Mode Encryption. The objectives are as follows:

i. Design a medical access control for privacy Preservation Techniques.
ii. Enhanced Rivest-Shamir-Adleman Algorithm using Counter Mode Encryption.
iii. Implement the Techniques with Rivest-Shamir-Adleman Algorithm using Counter Mode Encryption.
iv. Evaluate the implemented techniques using metrics like Key length, Encryption/Decryption time, and Throughput.

1.4 Significance of the Study

The significance of this study cannot be overstated owing to the numerous areas in which it can be applied especially in the health sector. The significance of this study is as follows:

To protect the interest of individuals, from revealing the patient's healthcare records to the public and secure the confidentiality, integrity, and control access to individual personal healthcare records, and also reduce data breaches from the system and provide protection against hackers, and also stepping up to meet the current standard and proper hospital management, to reduce cost and prevent the system from going down. Most significantly to improve from manual file to computer base approach in driving through the protection of confidential health records of the patients.

To make available the encryption technique that will be beneficial to the hospital, staff, patients, and health management, and also increase the bond of secret between the consultant and their patients in the hospital. This can be done owning to the oat taking by the consultants, to always protect the confidentiality and integrity of all patient personal health records.

The patient will benefit from facilitating access to new therapies, improved diagnostics, and tests, through easy encryption and decryption using two keys. This private key will be known only by the patients to boost the confidentiality between patients and the consultants.

1.5 Scope and Limitation of the Study

In recent years, several research works have been proposed in the area of big data in health records. However, there are still many limitations and challenges in developing and securing electronic health records. In this work, the scope was mainly focused on

efficiently securing the confidentiality, integrity, and controlling access to patient personal health information using the RSA algorithm and Counter (CTR) mode encryption. However, other types of techniques like ammonisation encryption, attribute-based encryption, and classification mode were considered in this research work.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Emergence of Big Data

Over the last two decades, data is growing exponentially due to the rapid evolution of new technologies, devices, and communication means Galetsi, *et al.,* (2020) this voluminous data with diverse variety generated with high velocity transform data into Big Data. It is extensively used in the marketing field, sales, banking, finance sector, healthcare, social media, tourism, and many more. But due to its 'Big' features in every aspect, it becomes difficult to handle it with traditional data processing applications. There are many challenges while handling big data as difficulties lie in data capture, storage, searching, sharing, analysis, and visualization. As large datasets are usually non-relational or unstructured, thus processing such data sets poses a significant challenge. Therefore, Big Data Analytics becomes a demanding field for researchers. It is not a single technology, but a data-driven approach used to develop and deploy customized solutions as it analyses a large amount of data to uncover hidden patterns, correlations, and other insights (Galetsi, *et al.,* 2020).

2.2 Sources of Medical Big Data

According to Jasim, *et al.*, (2015), Having understood the meaning of medical big data and their dimensions, the Digitization of content by industries is the new source of medical data, the following are some sources of medical data like Administrative claim records, clinical registries, electronic health records, biometric data, patient-reported data, the internet, medical imaging, biomarker data, prospective cohort studies, and large clinical trials.

However, big data sources are available in areas like astronomy, atmospheric science, social networking websites, life sciences, medical science, government data, natural disaster, resource management, weblogs, mobile phones, sensor networks, scientific research, and telecommunications Fields(Gupta, 2014).

2.3 Application Areas of Big Data

Big data has many areas of application such as the banking sector or more generally the financial sectors. It is used in the aspect of securing banking data as well, transportation and a lot more. The following subsection explains more about this area.

2.3.1 Applications in banking and financial industries

Massive amounts of data are being generated by the banking and financial industries through their various service offerings such as checking/savings accounts, mobile banking, credit and debit cards, loans, insurance, and investment services. Most of these data are structured data. Also, most of these organizations have set up their presence online for better serviceability and marketing through which lots of data are collected (Mohanty and Boinepelli, 2015).

Most of the data collected are unused, and the industry is looking to various new technologies in data mining and business analytics to help understand and identify customer needs and offer new services to enhance their business opportunities and increase their margins and profitability. The industry is also looking for solutions in risk management and fraud detection which will help minimize business exposure.

Banking handles massive volumes of data with proper security has not been easy. Data analytics offers potential benefits to industries such as banking by allowing analysis of customer log files and the handling of customer interactions. Combining structured and

unstructured data types in this way can give companies a better view of both their customers and operations (Davenport, and Dyché, 2013)**.**

Analytics offers banks the ability to segment customers depending on their risk profiles, credit usage, and similar markers, offering products tailored to their needs and ability to handle money. Analytics are utilized throughout the industry, such as in retail banking operations, where every customer transaction is tracked and matched to the customer (Banerjeer, *et al*., 2014). Due to the massive number of transactions and activities in financial institutions such as banks, big data development is inevitable, and this directly impacts the management of scarce resources by individuals, groups, and organizations. Big data analytics is thus used by the financial service sector to predict client behaviours and to gain advantages based on understanding customers and employees (Al-Shiakhli, 2019). In the following sections, we have covered the way big data is applied to a few of the most important areas in more detail.

2.3.1.1 Fraud detection

Various surveys and studies by Field Mohanty and Boinepelli, (2015) indicate that the banking and financial services industry is the victim of most fraud cases among various industries. Following are some of the widely known frauds in the banking industry.

i. Online Banking Fraud: Involves fraudsters taking over access to the victim's account and performing transactions to siphon the funds out of the accounts.
ii. Card Fraud: Involves fraudsters stealing card information and performing fraudulent transactions.
iii. Insider Fraud: Involves fraud by the bank's employees.
iv. Money Laundering: Crime involving transactions with mainly foreign banks to conceal the origins of illegally obtained wealth.

The traditional approach of sifting through the reports manually and applying various rules is only useful for the compliance process and not for detecting fraud and stopping losses. The financial industry requires real-time fraud detection to effectively identify fraudulent transactions in real-time and stop them from executing.

2.3.2 Application in transportation

Transportation experts can rely on big data to generate accurate and timely information for traffic flow prediction, a crucial step for improving real-time traffic management and mitigating urban congestion (Deepa, *et al.,* 2022). Traffic flow prediction usually begins with mining vast amounts of unstructured locational data extracted from physical and social sensing technologies. The data are then analyzed through various standard statistical methods and machine learning techniques to detect traffic patterns and build predictive models of those patterns. The performance of the underlying algorithms varies in many ways that reflect in part their ability to account for network-wide heterogeneity and spatiotemporal relations while maximizing in-sample and out-of-sample prediction Fields(Deepa, *et al.,* 2022). Comparative studies between big data analytics and traditional statistical methods to predict traffic flows point to the superiority of the big data approach which can better account for non-linear relations (Ma, *et al.,* 2020).

Big data analytics and predictive modelling are also increasingly influencing urban management and congestion mitigation efforts (Newman, *et al.,* 2017). The city of Tokyo, for instance, has partnered with a private firm to develop a smartphone-compatible app, ZenryokuAnnai that analyses nearly 360 million observations every second to generate real-time information on the shortest and least-congested travel routes.

2.3.3 Application for insurance

In these industries big data helps in analyzing and predicting customer behaviour through the derived data of social media, Global Positioning System (GPS) enabled devices, and Closed Circuit Television (CCTV) footage to provide customer insights for transparent and simpler products. It also allows for better customer retention from insurance companies (Rajeshwari, 2015).

2.3.4 Application in agriculture

As the economy develops, people's demand for food is getting higher and higher. Hence, the control of agricultural products has become more important. The agricultural sector not only needs traditional agricultural production experience and theory but also needs to use modern science and technology and management methods to serve it and promote the continuous improvement of agricultural productivity, to improve the quality and output of agricultural products. Applying big data to agricultural production can achieve timely monitoring of agricultural products and increase the output of agricultural products.

Wang, *et al.* (2018) and Cao, *et al.*, (2018) designed a database server, mobile client application, and data management system to improve the timeliness of data collection and the convenience of uploading for agricultural workers to conveniently upload agricultural data in real-time, ensuring the reliability and timeliness of the data, and increasing the output of agricultural products.

Hai and Zhang, (2017) to deal with agricultural big data, adopted a service-oriented architecture, with British Standard as the main technical framework, integrating Geographic information system (GIS), Web Service, JSON data exchange, and other technical means to realize data services, application services, data exchange services,

and many more, and to quickly establish an agricultural industry business application system, thereby shortening the development cycle of management information systems, improving R and D efficiency and the value of agricultural big data utilization.

Zeng and Ren, (2018) built the agricultural big data management platform with Hadoop technology as the core. The platform can automatically collect information on the growth, reproduction, market demand, marketing, and other aspects of agricultural operations in a certain area, and generate detailed management information reports.

2.3.5 Applications in manufacturing

Manufacturing companies have become highly competitive across the world with the margins of doing business going down every day. Manufacturers are always on the lookout for optimizing costs in running factories thereby increasing the margins, Big data analytics is helping in a couple of areas. (Mohanty and Boinepelli, 2015).

2.3.5.1 Preventative maintenance

In the automated world of manufacturing, sensors are used everywhere in monitoring the assembly line to be able to identify failures quickly and fixed them to minimize downtime. The root cause of plant failure could be due to one or more of the numerous possible parameters spread across different subsystems linking the assembly line. A huge amount of sensor data, all unstructured data, is accumulated over the running of the manufacturing plant. Historical maintenance records for the various subsystems are also gathered in a semi-structured format. And logs related to the productivity relative to the peak capacity are also gathered along with the maintenance records and sensor data. Time series analysis of the various subsystems based on their respective sensor data and performing pattern matching against the failure case is used for catching the potential failures. Also, path analysis and sessionizing techniques are used to capture the critical

events based on correlations between the sensor readings, historical maintenance records, and logs to predict the probable failures. This helps take preventative measure to keep the line running for an extended period without interruptions and also help to improve the safety of running the operations (Mohanty and Boinepelli, 2015)

2.3.5.2 Demand forecasting

The most important factor in businesses that are tied to the manufacturing industry is to optimally use the resources where day-to-day orders keep changing dynamically. Forecasting sales and the time frame when it happens will help plan for the timely acquisition of raw materials, ramping up/down production, manage warehousing, and shipping logistics. In the short term, overestimating demand leaves the manufacturer with unsold inventory which can be a financial drain and underestimating implies missed opportunities. Demand forecasting is required to plan for strategic investments and business growth. Hence, the effective running of a business with maximum profitability requires a solid forecasting system. (Makridakis, *et al,* 2018).

2.3.6 Applications in telecommunications

Big data can improve the quality of management in telecommunications by making use of real-time data analyses and monitoring machine logs. Predictive analytics can also be used to minimize performance variability and prevent quality issues by providing early warning alerts (Al-Shiakhli, 2019). Big data analytics platforms used in the telecommunication field face the major challenge of storing and processing big data; traditional analysis techniques are too expensive in many cases. Big data techniques such as Hadoop can help in reducing storage costs, particularly where storage modules such as the Hadoop Distributed File System (HDFS) and computation modules such as MapReduce are included (Çelebi, 2013). Big data analytics has the power to extract

more information than traditional data analytics, which can help in improving mobile cellular networks. Such mobile cellular networks generate and move massive amounts of data such as calls and mobile application activities that consist of both structured and unstructured data types. Traditional data analytics deals only with structured data, and thus it is almost impossible to handle the data with traditional data analytics (He, *et al.*, 2011).

2.3.7 Applications in social media

Online social media is growing leaps and bounds as witnessed by the growth in the active user base and the amount of data that it generates. Sites such as Facebook, Twitter, Google+, LinkedIn, Reddit, and Interest are some of the most popular online hangout places these days. Even big corporations have started using social media as a business channel by having their presence through Facebook accounts, Twitter, YouTube channels, and company blogs to name a few. The inherent openness of social media to everyone to hear and voice their opinions and build new relationships has paved the way for the creation of a wealth of data. This has caught the attention of data scientists in exploring the use of social media in various areas. Social media analysis involves gathering and analyzing huge data that social media generate to make business decisions. The goals of this analysis include strategies for product marketing, brand promotion, identifying new sales leads, customer care, predicting future events, foster new businesses. (Mohanty and Boinepelli, 2015).

2.3.8 Applications in healthcare

Research by Almeida, (2017) showed that big data might help in reducing waste and improving efficiency in clinical operations, research and development, and public health. Raghupathi and Raghupathi, (2014) described big data analytics in healthcare

and identified several remaining challenges; big data analytics has the power to develop care, save lives, and minimize costs, using the recent data explosion to extract insights to allow healthcare providers to make better decisions. The potential benefits gained from using big data in healthcare are not limited to, discovering diseases quickly, thus making treatment easier and more effective; identifying healthcare fraud quickly to manage specific individuals; and improving population health.

Furthermore, Zhong, *et al.,* (2016) presented big data applications in healthcare and showed the way big data can be embedded into daily life to offer the ability to examine experiences of illness and healthcare. Big data analytics thus have a large impact on the healthcare sector, reducing operational costs and improving patients' quality of life (Al-Shiakhli, 2019 and Galetsi, *et al.,* 2020).

2.3.8.1 Finding new treatments

Maintains the database of all the published medical articles on various health topics and has opened up access to all interested researchers. This dataset of documents is huge, and mining meaningful information is a challenge. Researchers have used the semantic searches on this database to uncover new relationships between therapies and outcomes. (William, 2015).

2.3.8.2 Multi-event path to surgery

Applying path and pattern analysis techniques to the data obtained from the patient records with different procedural codes, it is possible to identify the sequence of events leading to expensive surgeries. Using this information, better preventative care can be provided to avoid surgery and help reduce medical costs.

2.3.9 Applications in the retail industry

Big data has a massive impact on retail industries, improving the customer experience and reducing fraud (Wamba, *et al.,* 2017). The retail sector is of major importance in modern society, as almost everyone nowadays must buy their basic needs. Predicting demand for items allows retailers to offer better services to customers (Singh, *et al.,* 2015; Lekhwar, *et al.,* 2019), and retailers can use customers' billing data to gather information for business intelligence. A Hadoop distributed file system (HDFS) tool is used to store, process, and analyze such data to allow the extraction of more information (Singh, *et al.,* 2015). Big data analytics provides these organizations with more information on market decisions and help in segmenting customer based on their characteristics. Social media analytics can also be used to inform companies about their customer's preferences. Applying sentiment analysis to such data provides the organization with early warnings when the customer turns to different products, allowing action to be taken by the organization (Elgendy and Elragal, 2014). Organizations have used the segmentation of customers for many years, but this is now assisted by complex big data techniques such as real-time micro-segmentation which offers better-targeted advertising (Manyika, *et al.*, 2011; Elgendy and Elragal, 2014). Organizations can also gain better targets for social marketing by understanding customer behaviors and predicting market sentiment trends (Russom, 2011; Elgendy and Elragal, 2014). Retailers are thus using data analytics to address new challenges and find opportunities based on increases in market expectations, competition, and volatility. In many companies, additional accuracy, clarity, and insight can be provided by the adoption of data analytics techniques, and such intelligence can be extended to industry supply chains (Hofmann, *et al.,* 2018).

2.3.9.1 Recommendation of products

One of the well-known strategies that retail companies employ to increase their revenues is to recommend products to the customers that they might be interested in, based on the product the customer is currently purchasing.

This is typical of an e-retailer whose back-end systems run product recommendation engines by cross-referencing the items among sales records from various customers that may have purchased the same item earlier, Collaborative filtering technique is used in the recommendation systems by the e-retailers Dash, *et al.,* (2019) such as amazon for recommending products, and the same techniques are used by the movie recommendation engine that Netflix uses.

2.3.9.2 Predicting trends

Retailers collect huge amounts of data about customers including location, gender, and age from their various transactions. Mining of retail data can help identify customer buying patterns and trends which will in turn help identify customer needs to effectively plan for product promotions and attract more customers and increase revenues/profits. Dash, *et al.,* (2019) multi-dimensional analysis and visualization tools of the dataset can be used for the prediction which could help with the company planning of the logistics/transportation of the needed goods.

2.4 Merits and Demerits of Big Data

According to Bonheur, (2019), carrying out an analysis of big data requires an understanding of big data with merit and demerits.

2.4.1 Merits of big data

According to Bonheur, (2019), some of the merits of big data are:

(i) Understanding and Targeting customers

For business organizations, one of the merits of Big Data is to enable them to understand their customers or target market, particularly their behaviours and preferences. And also allow them to provide better product or services, develop new products in consideration of trends in the market, predict consumption patterns and behaviours to provide appropriate marketing responses, and provide better customer experience such as the inclusion of value-added services and after-sales services.

(ii) Optimize and improve Business Processes

Big Data can be essentially considered a source of competition for business organizations. Aside from being able to understand and target customers better, analyzing large datasets can lead to optimization and improvement in specific facets of operations. For instance, retailers can mine Big Data to reveal patterns in production and consumption, as well as in other fulfilment performances to improve the supply chain, promote better inventory management, and optimize distribution channels.

(iii) Support Developments in Artificial Intelligence

Another merit of Big Data involves its critical application in advancing artificial intelligence, particularly in advancing specific fields of AI. For instance, machine learning depends on training data extracted from Big Data to learn from outcomes without being explicitly programmed. Natural language processing requires the collection and analysis of structured and unstructured audio data such as language and dialects, vocabularies, grammar and syntax, and speech patterns to enable human-computer interaction using natural language instead of computer language.

(iv)Empowers Online Businesses and the Digital Ecosystem

It is safe to say that digital communication and Big Data have now become intertwined. Google depends on the analysis of large chunks of web and user data to power its Google search services. The same is true for Facebook and Twitter which use data analytics on a massive scale to deliver targeted content and advertisements. Companies such as Amazon, as well as Netflix regularly crunch data obtained from their customers to improve service delivery, as well as implement a personalized user experience. Nevertheless, the relevance of Big Data online-enabled business stems from the fact that more people are coming dependent on digital communications. Hence, these businesses are utilizing the data generated by these online users to maintain their competitive advantage.

(v) Equipping Organizations with Better Capabilities

Remember that the application of Big Data does not rest alone on businesses. Government agencies have been using methodologies in processing large data sets as part of efforts to promote safety and security, such as in the case of predictive policing, as well as in maintaining national policy and signal intelligence. In science, Big Data expedites the process of data analytics, particularly for continuous experiments such as in the case of particle experiments at the Centre for European Nuclear Research (CERN).

2.4.2 Opportunities of big data in healthcare

Big Data offered many opportunities in healthcare, which are as follows: (Alexandru, *et al.*, 2018)

(i) Upper qualitative care

Big data relies on different sources such as previous meetings with doctors, social media, and exterior activities (Alexandru, *et al.*, 2018). As a result, it creates a

structured picture of the client. The traditional methods of physical charts completed by employees with their online courses brought a little amount of information regarding personal life and charts. Nowadays, there is a huge amount of information leading to the healthcare suppliers' knowledge about helps a person require to lose weight. As an example, when someone is fond of skiing and this is to be seen on social media, a doctor can virtually reach that person's background and indicate the cause of leg pain. Otherwise, with superficial programs, the doctors will not have enough information. The general purpose of Big Data in the healthcare domain is to analyze, identify and solve medical problems before they come to represent serious problems. For example, an individual seeing a doctor to lose weight could be told to take pills for high cholesterol. A person posts on different social media about modification in their life that stresses, and the big data algorithm could make an analysis of that information and indicate the patient with a risk for a heart attack. The doctor can slightly modify the cure to prevent the heart attack, thus solving the problem before it comes to be a real-life risk. Big data can have access to Deoxyribonucleic Acid (DNA) records to understand whether a patient has sick members in the family and whether that disease could be inherited biologically by children.

(ii) Indication of Fraud

Fraud has come to represent a growing issue in the healthcare and insurance domains, but it is a reality that there are patients who make claims hoping to be given money in exchange. Big data is important are useful in solving this aspect because it can use a large amount of information to find differences in written claims and demands, and also identify the deceitful ones for more analysis. By making use of its sophisticated procedures, Big Data can compare a huge number of recordings to identify mistakes more quickly than humans. Medicare saved a lot amount of money from claims via Big

Data. The possible way in which Big Data can detect fraud situations is detailed in articles like “Healthcare Fraud Management using Big Data” published on Trend-wise Analytics.

(iii)Advanced Patient Attention

Health recordings made electronically are of real help in gathering population and medical information such as lab examination, clinical information, medical problems, and conditions, which would further help healthcare providers offer quality procedures.

(iv)Upgrade operational proficiency

Healthcare units make use of Big Data and turn it into a business tactic to analyze historical individual hospitalization and also staff competence. Healthcare companies can lower costs and still offer quality care using predictive statistics. Big Data is also useful in healthcare because it reduces mistakes related to medication by increasing economic and management competence and reducing re-entering. Discovering a remedy for illnesses.

(v) Access to remedies, such as chemotherapy

Analyzing historical and operational cases and conditions, complex procedures can foretell the volume of work and organize appointments without any gaps in between cures. This action increases chair usage, diminishes the time in the waiting room for patients, and reduces the costs of procedures. It is difficult to manage this action without competent information science. For instance, it is hard to order 70 patients for their remedies in a centre that has only 35 chairs to be used. It has no sense trying to solve this issue using a pen, paper, or even an Excel document. Having enough beds for hospitalized patients represents a narrow direction. Every hospital tries to give a solution to this issue by analyzing daily every patient’s health situation, adding some

new patients and thinking about some additional ones, then deciding that the hospital can deal with patients within its capacity of beds. It is a never-ending action, which is repeated daily, with the hope of a plan.

(vi)Follow up Attention

Big Data also tells which patients are going to follow the doctor's recommendation to prevent crowding the hospitals. There has been great interest lately in devices that reveal the way many steps have been taken or the number of many times the heart beats. This action leads to improving a patient's physical results and engagement. Using this kind of technique, health can be monitored and issues such as asthma and blood pressure reduced. There are a lot of applications that can tell when a patient does not feel well, to know whether medication is administered accordingly or the amount of time the patient sleeps, moves, or other actions. The population is rapidly growing in age, and the Japanese combine robotics and remedy and healthcare. Robots are used in different fields, from helping elder people living alone to helping doctors offer medical support to countryside people and helping pets dealing with Alzheimer's patients.

(vii) Reduce costs and times of wait

As in the case of other industries, costs can be dramatically diminished by using Big Data in healthcare. There is also a possibility to decrease wait times which means money for most people. One hospital in France, Paris, uses predictive examination to get employees. By knowing the number of patients in the hospital in a certain period, the hospital can decide the employee's number that must be at work for those patients. There are a lot of means to reduce costs, but only a few hospitals make use of them. Hospital money is intricate, and, even though money invested is money gained, some units are not fully confident in putting their money in Big Data. Money can be saved,

used to replace old things with new technology, or put into other activities. An example of the way big data can benefit hospital budgets is the use of forecasting admission rates. Four Paris hospitals have been trialling machine learning systems together with big data for such outlooks. Data coming from different external and internal sources including records for 10 years is used for predicting the number of patients that are expected to be visiting the hospital in an interval of time. Information from such results can lead to more efficient deployment of doctors and resources. Information from such res ults can lead to more efficient deployment of doctors and resources. Big data usage for predictive analytics has been a subject of discussion in multiple scientific writings. (Saratchandran, 2018).

2.4.3 Benefits of health-related big data

According to Alexandru, *et al.*, (2018), Big Data in healthcare can be used to increase the values in the following fields.

(i) Public Health

Using big data, we can analyze disease patterns and record disease outbreaks, public health issues can be improved with an analytics approach.

(ii) Electronic Medical Record (EMR)

An Electronic Medical Record (EMR) holds the standard (structured and unstructured) medical data that can be assessed with a big data approach to guess patients at risk and deliver effective care.

(iii) Patient Profile Analytics

Applying advanced analytics to patient profiles for identifying individuals that could benefit from a proactive approach.

(iv) Genomic Analytics

The data analytic tactic can be efficiently included in genomic analytics to make this method a part of the regular medical care decision process.

(v) Fraud Analysis

This data analytics approach helps analyze a greater number of claim requests to curtail fraud cases. An effective analysis can help reduce fraud, waste, and abuse.

(vi) Safety Monitoring

Data analytics can also be used to investigate real-time great volumes of risk data in hospitals. The approach may help in the safety monitoring and negative event forecast.

2.4.4 Demerits of big data

Despite the merits or beneficial applications of Big Data, it comes with drawbacks or demerits, as well as challenges that can make its implementation risky or difficult for some organizations. Bonheur, (2019) these issues need to be solved to reap better benefits that come with mining large sets of data.

The following are the Demerits of Big Data

(i) Privacy and Security Concerns

One of the notable disadvantages of Big Data centres emerging over privacy rights and security. Even large business organizations such as Yahoo.com and Facebook have figured out numerous instances of data breaches. With data privacy laws becoming more stringent as exemplified by new policies such as the General Data Protection Regulation (GDRP) of the European Union, organizations seeking to develop and

maintain Big Data capabilities also need to invest in protocols, processes, and infrastructure aiming at protecting data and mitigating security risks.

(ii) Technical Challenges and Requirements

Big Data requires both processing capabilities and technical proficiency. In other words, for an organization to have the capacity to mine large volumes of data, they need to invest in information technology infrastructure composed of large databases, processors with adequate computing power, and other IT capabilities. Furthermore, they need to have a certain degree of competency that would allow them to address more specific issues such as data storage and transportation, database management, data access and sharing, quality and validity assurance, and scalability of the infrastructure, among others.

(iii) Issues Over the Value of Big Data

Another problem with Big Data is the persistence of concerns over its actual value for organizations. As mentioned, resolving the challenges and responding to the requirements of its implementation involve investment. Not all organizations can afford these costs. Large organizations can easily develop Big Data capabilities, thus putting their smaller counterparts at a disadvantage. It appears that due to the costs, as well as the drawbacks and risks of Big Data, its advantageous applications only benefit large organizations while expanding further the competitive gap between them and smaller organizations.

2.4.5 Challenges of big data in healthcare

The emergence of big data in healthcare has many obstacles, which are as follows: (Alexandru, *et al.*, 2018).

(i) Privacy

The lack of privacy and intimacy may be the strongest flaw (Alexandru *et al.*, 2018). To be effective, Big Data must access pretty much everything, from private recordings to social media life. But the price is paid because private information is revealed to solve health problems. But a patient's freedom is not given. However, some regulations state a medical recording's privacy, but they are not taken into consideration since it is thought that information about a person should not be forbidden when it relates to human health. The subject of privacy risks associated with Big Data use cases in healthcare has been tackled in articles like big data security and Privacy in Healthcare. (Harsh and Ravi, 2014)

(ii) Replace Medical Staff

Big data presents the advantage to know about possible future health problems but also has a huge risk, the doctors are replaced. Big Data is not performing to be used without a human touch, but feared that, when use increases, patients would not go to the doctors, but use the technology and undermine doctors' authority. Big Data in the case of healthcare cannot be rejected because more and more units and companies invest in this growing field. But, one should consider its disadvantages and realize a procedure safe for both doctors and patients. The question of whether Big Data can lead to medical staff replacement has appeared as a subject for discussion on major sites like Forbes, Fortune.

(iii) Discussions and conclusions

Leaving in an age where data is produced at every step of the way allows us to achieve many great things, but also, leaves us exposed to many points of failure. As we saw in the results underlined by this research, in healthcare, big data can make a difference and

save real human lives. From upper qualitative care to indication of fraud or from reducing the costs and times of wait to discovering a remedy for illnesses and more others were find it out to be the most important advantages of big data in healthcare. Privacy seems to be an important problem that people cannot overcome, as well as the replacement of medical staff. This problem put the large-scale adoption of big data in health care under observation. This research has shed light on the advantages and disadvantages of the use of big data in healthcare. Throughout analytical critics of publications and journals, this thesis identified the most important points in adopting big data in healthcare as seen by publications and journals. The most important finding of this article is that the advantages always overcome the disadvantages when it comes to saving human life and improving the quality of life (Alexandru, *et al.,* 2018).

Based on the existing studies and frameworks, presents a set of challenges that are related to health big data and health big data analytics. Challenges belong to technologies, collection, storing or storage, aggregation, analysis, sharing, and visualization of health data. (Bizer, *et al.*, 2012; Iwashyna, and Liu, 2014; Bellazzi, 2014; and Kaisler, *et al.,* 2013). None of the previous works merged their existing systems with the big data system or the National Health Information Exchange. The following is a brief description of the challenges addressed in this research work.

(i) Technologies

The challenges include hardware tools (sensors and computers as data collection and storage devices), software tools (data collection and processing software), and infrastructure (for instance, internet access (communications) and power supply). (Chen, and Zhang, 2014; and Jagadish, *et al.*, 2014).

(ii) Collection

The data collection challenges involve the way of collecting and transferring vast amounts and different types of data to the data repository. Collection methods include both devices (sensors and computers) and human resources. Data transferring to the repository is related to network challenges; (Kuo, *et al*, 2014).

(iii)Transforming and Sharing

Data transformation includes different stages, levels, and data structure forms. At the lower level (first stage), the raw data need to be transformed from the collection devices to the data repository, where one of the biggest challenges is due to integration limitations and lack of system interference capabilities, which may require special data adaptation tools. In later stages, the valuable data will be transformed out of the repository based on certain queries. To share such data, a high network bandwidth capacity is needed which is considered a challenge in this phase. (Koo, *et al.,* 2020; and Kuo, *et al.*, 2014)

(iv) Storing

Once the data are collected, another challenge arises, the storage capacity. The huge amount of data needs large space and high-performance I/O devices, which is costly.

(v) Analysis

The analysis is required to convert the raw data into valuable data. In big data analysis, data must be scalable, well-structured, secure, and consistent. The challenges are both data pre-processing and the lack of experts in the analysis tools.

(vi) Visualization

Health big data faces another challenge which is the way to visualize such huge, varied, and different structures of valuable data. (Rehman, *et al.*, 2016; Oliveira and Gerosa, 2011).

(vii) Security

Security is one of the most important challenges. Therefore, cyber security should be applied to protect systems, networks, and all data from digital attacks and to ensure the secure operation of the complex research, and production infrastructures for creating trusted secure environments, and cooperating groups of researchers and technology specialists (Al-Shiakhli, 2019).

2.5 The Concepts of Privacy Preservation

Privacy preservation of data is an important concept, because when the data is transferred or communicated between different parties then it's compulsory to provide security to data such that other parties do not know the data communicated between the original parties. Preserving the data means hiding the output knowledge of the data using several methods when this output knowledge is valuable and private (Google, 2022).

The two types of privacy preservation are as follows:

i. Individual privacy preservation:

The primary goal of data privacy is the protection of personally identifiable information. In general, information is considered personally identifiable when it can be linked, directly or indirectly, to a person. Thus, when personal data are subjected to mining, the

attribute values associated with individuals are private and must be protected from disclosure. Miners are then able to learn from global models rather than from the characteristics of a particular individual (Seliya, *et al.,* 2021).

ii. Collective privacy preservation.

Protecting personal data may not be enough. Sometimes, we may need to protect against learning sensitive knowledge representing the activities of a group. We refer to the protection of sensitive knowledge as collective privacy preservation. The goal here is quite similar to that one for statistical databases, in which security control mechanisms provide aggregate information about groups (population) and, at the same time, should prevent the disclosure of confidential information about individuals. However, unlike the case for statistical databases, another objective of collective privacy preservation is to preserve strategic patterns that are paramount for strategic decisions, rather than minimizing the distortion of all statistics (bias and precision). In other words, the goal here is not only to protect personally identifiable information but also some patterns and trends that are not supposed to be discovered (Turban and Aronson, 2001).

2.5.1 The Concept of big data in healthcare

Healthcare data are more sensitive and centralized than other types of Big Data. Privacy is the privilege to have some control over the way personal information is collected and used. But with the rapid growth of medical data, the frequency of data privacy and security issues is increasing. Kraemer, *et al.,* (2017) showed that with the development of current healthcare devices, more privacy issues and reliability hinder the complete transfer of healthcare monitoring data to the cloud service platform. Iyengar, *et al.,* (2018) showed that the Internet can bring opportunities for the popularization and

intelligent development of medical care, but the privacy issues of patients, doctors, nurses, and healthcare providers are more concerned.

Bachlechner, *et al.,* (2018) investigated the unclassified medical data and domain ontology, inferred that confidential data is prone to privacy invasion, and found that the current problem of data leakage is relatively serious.

He, *et al.,* (2011) showed that the release of user data may also seriously threaten user privacy, and third-party users can use data mining technology to infer sensitive information contained in the released data. It can be found that there are many privacy and security problems in current healthcare data, and it has great significance to find risk indicators that can be quickly identified to improve the privacy protection of medical care patients.

2.6 Records Life Cycle

The term “Records Life Cycle” describes the life of a record from its creation or receipt through the period of its active use, then into a period of inactive retention (such as closed files which may still be referred to occasionally), and finally confidential disposal or archival preservation.

According to Ferguson, (2015), the key components of records management are as follows: Record creation, record keeping and use, record maintenance (including tracking of record movements), access and disclosure, appraisal, retention and archiving, and disposal or archival preservation. Records must be closely monitored and managed throughout their lifecycle.

2.6.1 Purpose of the patient record

The purpose of a clinical record is to facilitate the care, treatment, and support of a particular service user. The record includes clinical notes, letters, summary reports, and assessments (including risk assessments, standardized assessments). Medical record completeness is a key performance indicator that is related to the delivery of healthcare services in the hospital (Tola, *et al.*, 2017).

2.7 The Concepts of Data Encryption

The concept of Data Encryption is simply the translation of data into a secret code, and it is considered the most effective way to ensure data security. Reading an encrypted file, access must be granted to a secret key or password that enables data decryption. Modern encryption is achieved using algorithms with a "key" to encrypt text or other data into digital nonsense and then decrypt it by restoring it to its original form (Olufohunsi, 2019).

2.7.1 Data encryption

It is a process that enables taking the text or data used and converting it into a code also called cipher text that cannot be understood by unauthorized users. For the data to be useable, it must be changed back or decrypted, also it is an efficient means of preventing unauthorized access to sensitive data. Its solutions protect the ownership of data throughout its lifecycle. Encryption is useful to avoid exposure to breaches such as packet sniffing and theft and impersonation (Olufohunsi, 2019).

2.7.1.1 Encryption

ncrpIt is a process of converting information or message which is plaintext into a difficult and unreadable form called ciphertext using an encryption algorithm

Olufohunsi, (2019) Each algorithm uses a string of bits known as a “key” to perform the calculations. The larger the key, the greater the number of potential combinations that can be created, thus making it harder to break the code and unscramble the contents. Encryption diagram Figure 2.1 shows the process.

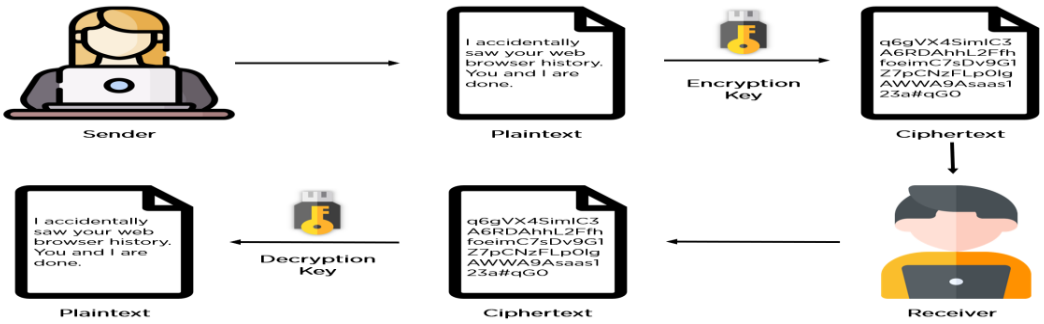

Figure 2.1: Encryption Diagram (Google, 2022)

2.7.1.2 Types of data encryption

According to Ahmed, (2020), there are two types of encryption in widespread use today: Symmetric and Asymmetric encryption.

2.7.1.3 Symmetric encryption

In symmetric key encryption, any user using the encryption system has a copy of the single secret key, this secret key is used for encryption and decryption, it is faster than asymmetric encryption, in symmetric encryption, and the sender shares the same secret key with the receiver to use it in the decryption process. Since the secret key is shared with the sender and the receiver, it becomes risky to use it. In symmetric key encryption, it is recommended to use a 128-bit key length longer to avoid or make it harder for any crack attempt. Some of the most common encryption algorithms are AES, RC4, DES, 3DES, RC5, and RC6. Out of these algorithms, AES is the best. Symmetric encryption is used when the speed process is required over security (Ahmed, 2020). The same key is used for both encryption and decryption. It is therefore critical

that a secured method is considered to transfer data between the sender and receiver (Venkatesh, et *al.*, 2019).

In Addition, the Advance Encryption Standard is known to be a symmetric block cipher containing 128 bits of block length. The algorithm is capable of accommodating three different key lengths which include 128, 192, and 256 bits. In the process of encrypting a 128-bit key 10 execution round is required, while using the 192-bit key encryption 12 execution round, and 14 execution rounds for the 256-bit keys is carried out. Hence, the Advance Encryption Standard algorithm is round-based. However, apart from the final round, the encryption and description contain four functions for each different round. They are various block cipher algorithm modes which include the Electronic Code Block (ECB), Cipher Block Chaining (CBC), Cipher Feedback (CFB), Output feedback (OFB), and finally the Counter (CTR) mode (Pethe, *et al.*, 2017).

(i) Electronic code block (ECB)

Based on the mode of operation the algorithm uses a sequence series of listed message blocks. The first block or section of the plaintext is taken and encrypted with the key in other to produce the first key. This process takes place again in the second block and continuously. In an Electronic code block, the mode of operation is deterministic, thus when the original plain text is encrypted twice using the same key, the outputted cipher block will be the same (Pethe, *et al.*, 2017)..

(ii) Cipher block chaining (CBC)

This mode of operation produces a cipher message independent of the plain text that is resulted in chipper text that is not deterministic. The Cipher block chaining mode load the n-bit initialization vector on the top of the register, then the n-bit plaintext block is XOR with the data value in the top register. The key K is used to encrypt the result

XOR operation using the underlying block cipher, then the cipher text is fed into the top register, and the operation proceeds until all the plaintext blocks are processed (Pethe and Pande 2017). Moreover, the cipher block mode works by merging the current plaintext to the last ciphertext block, and the output is encrypted along with the key.

(iii) Cipher Feedback (CFB)

The mode returns feedback from every ciphertext, in other to use this feedback in the encryption process of the next plaintext. The internal working of the cipher feedback mode differs completely from that of the electronic code block (ECB) mode. In this mode, the given plaintext block does not only depend on the plaintext and the encryption key but also depends on the previous blocks of ciphertext. The message is been depended on by the ciphertext. In this certain strange features are present such as the ciphertext using just only the encryption process of the cipher block. And most importantly the decryption of the background block cipher is never adopted (Pethe *et al,* 2017).

(iv)Output feedback (OFB)

This model includes the successful feeding of the output block derived from the background block cipher back to it. A string of bits is produced by the feedback in other to feed it to the encryption algorithm which is used as a keystream generator in the case of CFB mode. The plaintext and the key steam are XOR-ed together. And IV is required or essential as the initial random n-bit input block. However, the IV is necessary, not secure.

(v) Counter (CTR)

The Counter mode is mostly considered the Counter-based version of the CFB mode without its feedback. Considering this mode, the receiver and the side require access to

a very strong reliable counter. This counter calculates a new shared value shared every time there is an exchange in the ciphertext block. The counter value is not necessarily a secret content, but it is essential to both the receiver and the sender, the counter has to be a synchronized (Pethe, *et al.*, 2017).

2.7.1.4 Asymmetric encryption

It uses a different key for both encryption and decryption processes. One of the keys is typically known as the private key and the other one is called a public key. It is also known as public key cryptography. It involves multiple keys for encryption and decryption. It uses two distinct encryption keys related to each other, the first key known as a public key, and the other key known as a private key. The public key is available for anyone who wants to send a message to the sender. The second private key stays secret, only the sender knows it. A message is encrypted by a public key before it can be decrypted using a private key. A message encrypted by a private key can be decrypted using a public key. Security of public key is not required because it is available and can pass through the internet. Asymmetric Encryption is considered the best choice for information transmitted. Asymmetric Encryption is used in client and server model communication, a certificate is created to locate the server, the certificate contains the organization profile and information, the server and client need a secure encrypted communication, and a query sends over the network to the other party, which send them back the certificate. Secure Sockets Layer/Transport Layer Security (SSL/TLS) uses Asymmetric and Symmetric Encryption.

All communications use only public keys, and there is no private key transmitted. Some of the algorithms that use this technique are Rivest-Shamir-Adlemwn (RSA), which is used for encryption and authentication, and Pretty Good Privacy (PGP), which is used

to secure emails. Asymmetric Encryption is used when security is the priority over speed (Ahmed, 2020).

2.8 Related Studies

According to Mohammed, *et al.,* (2020), the literature review Privacy preservation in Big data Encryption standards, and the least significant bit of steganography that attempts to solve the problem of information theft which is rapidly growing pose a serious threat to internet security. It is important to keep the content of the message hidden and conceal the existing information from criminals inside an image to enhance security by adopting advanced encryption standards and the least significant bit of steganography to solve the problem. The result described the data set used to test and determine the efficiency of the proposed method. System performance has been checked with various images. This application allows the users to open an image, save an image, encrypt a secret image, save the decrypted message, and enter the message. Users can hide information inside an image to produce a stage image and the message recovery can be extracted from the stego-image using a key. Though the limitations are pointed out, such as multiple image operations on the stego-image like rotation, resizing, and cropping all have not been included in this research.

This research contributed to knowledge using the least significance bits approach, integrated with cryptography Advance Encryption Standard (AES) with steganography which can hide information inside an image to protect privacy in big data applications.

Abouelmehdi, *et al.,* (2017) used Attribute base encryption techniques racers control, and homophobic encryption to investigate the security and privacy chattering in big data and give some approaches and techniques for achieving successful security and privacy in healthcare organizations. It shows that privacy and security issues with technical

challenges could be a huge barrier in this research. Limitation shows that the privacy and security issues of the big data lifecycle are not considered in this research along with the advantage and disadvantages of privacy and security technologies in the context of big data in health care.

The data privacy preservation model for health information systems was investigated by Seun, *et al.,* (2019) and they concluded that the application of the privacy protocol used in this work is weak, because, the mechanism provides the unauthorized individual with the opportunity to have access to source code and modify the application. In the unfortunate event of a successful hacking exploit, which has a direct impact on the proposed model used to preserve patient data privacy, that at one time or the other has been linked with one or more symptoms of schizophrenia. It also categorizes healthcare professionals and their access levels to the system database. Hence, it will be difficult to recover from inadequate information availability for Data privacy preservation in health information systems in Nigeria, Based on the studies, this research contributes its quarter by providing a viable alternative to the machine learning approach to data privacy preservation. The consequences of having poor privacy preservation of big data negatively impact the health records, it is supposed to promote.

A study by Gutiérrez, *et al.,* (2020) agreed on the resilience in the integrity of data in the event of connectivity failure, with characteristics of privacy, security, and usability. The result provides a secure and resilient infrastructure for electronic medical records, which maintains the integrity view of the medical records of patients in a hospital system using the central and registration of electronic medical records (EMR) of patient information, and misinformation, to achieve feasibility and implementation of prototypes developed from the Healthy Block architectures in a productive environment.

Various studies have investigated the privacy of big data in an attempt to solve the problem of privacy.

Id, *et al.*, (2020) evaluate a novel framework for the privacy preservation of electronic health records using blockchain technology built on the Hyperledgerfabric framework and were able to achieve a secured framework user-oriented that handles efficient storage and transfer of medical records, ensuring the data ownership of individuals and patients confidentiality and data integrity, the approach and contributes and present a blockchain-based architecture with Hyperledgerfibric to secure health records to preserve patient privacy, public key encryption was not considered which is one of the major limitations.

Ram, *et al.*,(2018) measure the performance of privacy preservation techniques in big data analytics. The approach used was a data lake-based modernistic privacy preservation technique that handles privacy preservation in unstructured data to solve privacy-prone violations of tons of data generated from social media, websites, smartphones, and some-e-commerce sites like Amazon and the threats associated with them. It was observed after a systematic review, that all the existing mechanisms of privacy preservation are concerning structured data and not unstructured data. The review of this work shows more than 80% of data generated today are unstructured, which leads to Convention data mining algorithms that can be applied for classification and clustering problems but cannot be used in privacy preservation, especially when dealing with personal information. Machine learning and soft computing techniques can also be used to solve privacy problems to avoid embarrassment and abuse after many reviews. This research limitation is to develop a concrete solution to protect privacy in both structure and unstructured data. In the field of privacy preservation, robust

techniques need to be developed to handle large-scale heterogeneous data sets. To maximize data utility while ensuring data privacy.

The work of Abouelmehdi, *et al.,* (2018) address security and privacy issues, ensuring a secure and trustworthy big data environment by Preserving security and privacy using the ammonization Encryption method. They proposed a methodology that provides data confidentiality and secure data sharing. This contribute to driving health research, knowledge discovery, clinical care, and personal health management but it increased complexity, and new models proved difficult to interpret.

2.9 Summary of Review

The lack of data integrity and preservation of electronic records in mobile applications has affected big data integration in mobile e-health Kittur, *et al.,* (2019); Al-Kaabi and Belhaouri, (2019) outline certain limitations in securing big data using Enhanced Rivest-Shamir-Adleman (RSA) algorithm, which includes; limitation on data size to be encrypted, its computational cost and relatively slow when performing encryption and decryption, and the dual (public and private) key generation also affect the computational efficiency, and the work is only limited to a comprehensive survey on various ways to improve RSA algorithm. Seun, *et al.,* (2019) also came up with a Data privacy-preserving model for health information systems, and the model was designed for DPP and HIS using iterative design techniques. However, it was limited to only medical records to address the inadequacy. Abouelmehdi, *et al.,* (2018) work on big healthcare data for preserving security and privacy, it focuses on the state of art survey approach to security and privacy challenges, with limitations on anonymization and encryption. Ram, *et al.,*(2018) work on privacy preservation techniques in big data analytics; a survey, using the approach, lake based modernistic privacy preservation techniques, technique was capable of handling privacy preservation in unstructured

data, which is limited to designing a concrete solution in protecting the privacy of both structure and unstructured data. Gutiérrez, *et al.,* (2020) evaluate a Healthy blockchain-based information technology architecture for electronic medical records resilient to connectivity failures. The work used a blockchain network to secure medical record systems, and the result shows high efficiency in keeping the electronic medical records of individual patients unified. It is also limited to medical records. Hence, a hybridized encryption scheme (Counter CTR and Enhanced RSA algorithm) is proposed to solve the limitations identified in the review. Detailed view of table 2.1: A review of related work can be found in the appendix section of the work.

CHAPTER THREE

3.0 RESEARCH METHODOLOGY

3.1 Tools and Materials

The proposed Enhanced RSA Algorithm using Counter mode encryption technique are developed using Python programming language, which is a high-level object-oriented programming language. Hence, a Python interpreter and IDE must be installed for smooth writing of codes. This thesis considered Python version 3.9 and the Visual Studio Code Integrated development environment. The Graphical user interphase considered is Kivy programming language, which is a multi-platform Python framework for designing user interfaces. It is utilized because it integrates well with the Python programming language. MySQlite Relational Database Management is also utilized in the work for keeping patient records.

3.2 Features of the Proposed Techniques

i. The proposed encryption-based medical health record techniques are capable of storing patient medical records, and encrypting patient medical records using Counter mode (CTR) enhanced RSA algorithm. Each record encryption key will be secure using RSA (Private and Public) algorithm for remote distribution.

ii. The Hepatitis medical record of 155 instances is stored locally using the MySQlite database for security purposes and easy access.

iii. A comprehensive usable interface for easy navigation and interaction by users: the interface is developed using Kivy Python User Interface.

3.3 System Requirement

The requirement needed in developing the proposed techniques can be majorly categories into two: Hardware Requirements and Software Requirements.

3.3.1 Hardware requirement

The minimum configuration of the system for the smooth running of the proposed technique includes 4 GB of RAM, 150GB of storage space, and 2.0GHz of processor speed. Hence, the implementation of the techniques used 2.6GHz of processor speed, 8GB of RAM, and 500GB of storage space.

3.3.2 Software requirement

The proposed Enhanced RSA Algorithm using Counter mode encryption techniques were developed using the Visual Studio integrated development environment (IDE), python with version 3.9 (crypto and other necessary module installed), and SQLite dataset required for local storage of data. Finally, the technique is designed on a Windows 10 operating system. The software requirement can further be viewed as a functional and non-functional requirement.

3.3.2.1 Functional requirements

The functional requirement of the proposed techniques includes: accepting input by the developed techniques, the techniques should be capable of verifying doctor or patient, the techniques should also perform encryption or decryption only based on user request and Visual response from the user.

3.3.2.2 Non-functional requirements

The non-functionality requirement of the developed techniques includes; the developed techniques should be easy to interact with. Constant availability of the proposed

techniques for patient or doctor, accurate result is also expected from the proposed techniques, and reliability is provided by the proposed techniques.

3.4 Architecture of the Proposed Techniques

The section describes the overall concept behind the proposed medical record techniques for preserving patient privacy. Concepts such as the doctor, user interface, internet (Kaggle repository), internet (medium in transferring medical records), and encryption box (with two-layer encryption mechanism thus, the CTR encryption and Enhanced RSA Algorithm).

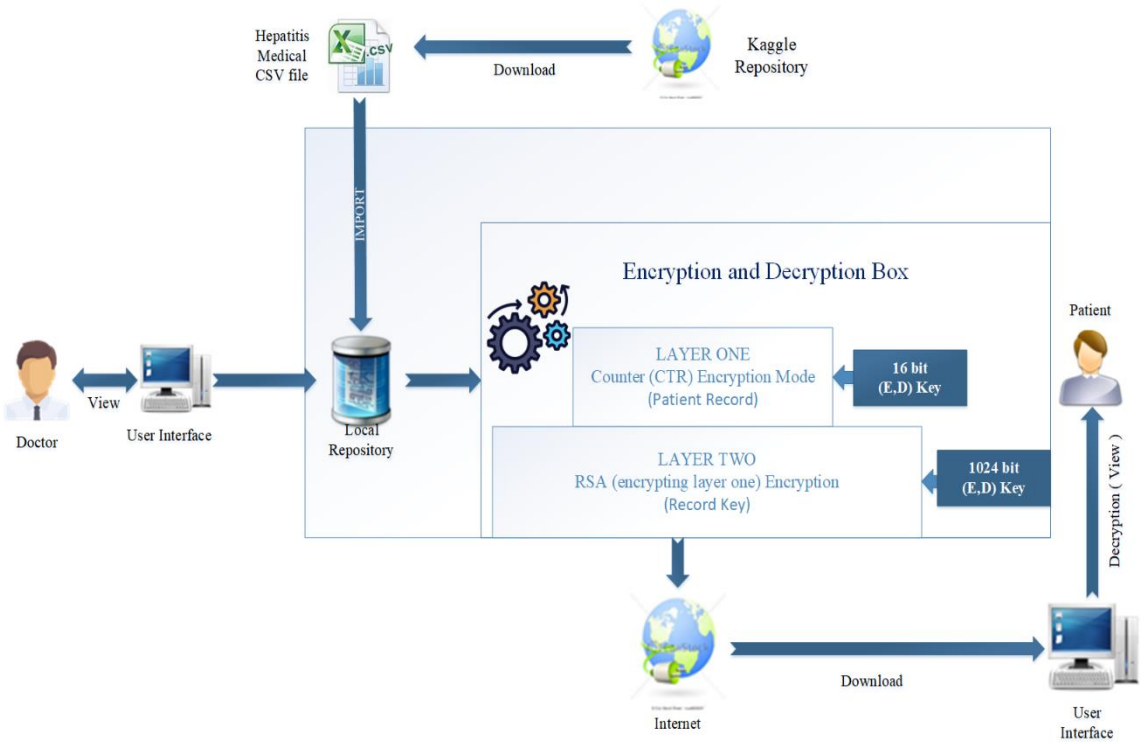

Figure 3.1: Conceptual Depiction of Medical Record Privacy Preservation Techniques

Figure 3.1 shows the internal architecture of the proposed techniques for medical records privacy preservation. However, the entities considered in this conceptual diagram include the doctor and the patient.

3.4.1 The doctor

The diagram in Figure 3.1 shows the entity Doctor is capable of interacting with the encryption techniques via a user interface, developed using Kivy User Interface Framework**.** The doctor is capable of login into the medical encryption techniques via an access code, each medical record can be encrypted independently by the doctor using the Counter (CTR) mode encryption. The key used in encrypting each record can be generated by the doctor via a button on the user interface.

3.4.2 The user interface

The user interface enables communication between the entity (Doctor or Patient) and the internal working of the encryption techniques (the encryption box). The user interface is developed using Kivy Programming Language, which is designed in a manner that can be easily integrated with Python code as back ends. The text field is one of the user interface components to accept information such as username, access code, and encryption key. The Buttons include in the user interface accept commands from the entities, that is the Patient and the Doctor.

3.4.3 Hepatitis CSV file and local database

Through the operation of the encryption techniques, the medical records are stored, saved, and sent within the application. The Hepatitis medical record is downloaded in comma-Separated Values (CSV) format from an online data science repository (Kaggle), and pre-processed for appropriate viewing. The CSV file (containing 155 Patient record instances) is imported and viewed in a tabular manner within the Graphical User Interface.

3.4.4 Encryption and decryption box

The encryption box consists of a two-layer encryption stage. The first layer encryption stage encompasses the encryption of each record using the Counter (CTR) encryption mode and a single 16-bit encryption, decryption key is generated which is forwarded to the second layer encryption stage. The second layer uses the RSA encryption mechanism to secure the key used in encrypting the medical record, and two keys are generated (A Public Key and Private Key). The RSA public 1024-bit key can be used to encrypt the Key used in the Counter (CTR) mode for securing patient records, while the private key is sent to the patient through a medium within the internet infrastructure for decryption purposes.

3.4.5 The patient

The patient is another entity that can interact with the user interface at the other end, which is capable of accessing the techniques via an access code. The patient received an RSA private key (for decrypting the key used in encrypting patient medical records) via the internet, and this can be viewed using the user interface.

3.5 Data Collection (Dataset)

In this research, the proposed medical dataset used was the hepatitis diagnosis medical record. The dataset is collected from an online data science repository, which is Kaggle. The Kaggle repository is a large data science community of machine learning practitioners. It is also considered a subsidiary of the Google Limited Liability Company (LLC), the Kaggle repository is a data science environment for academic research purposes, recruiting, proposing, and solving scientific problems. However, this research work considered using the hepatitis Kaggle medical record dataset for

preserving the privacy of the patient's health records. The downloaded medical record consists of 155 instances, while the attribute utilize from the dataset includes:

Detail view of the table can be found in the appendix section of the work. The hepatitis dataset consists of 8 columns (attribute or features) and 155 rows (Patient records). The feature or attributes include

(i) Id: This denotes the unique attribute associated with each patient, it enables efficient and easy access to any record from the dataset.

(ii) Name: The name of each patient, with a string data type.

(iii) Age: The age attribute of each patient are 10,20,30,40,50,60,70 and 80.

(iv) Sex: This indicates whether the patient is either male or female.

(v) Antivirals: The antiviral indicates whether the patient is antiviral or not, with a (yes or no) value.

(vi) Liver big: The liver big attribute indicates whether the liver of the patient is big or not (Yes or No) value.

(vii) Bilirubin: The bilirubin level is indicated with a floating-point number.

(viii) Albumin: The albumin level is also indicated with a floating-point number.

3.6 The Proposed Data Flowchart

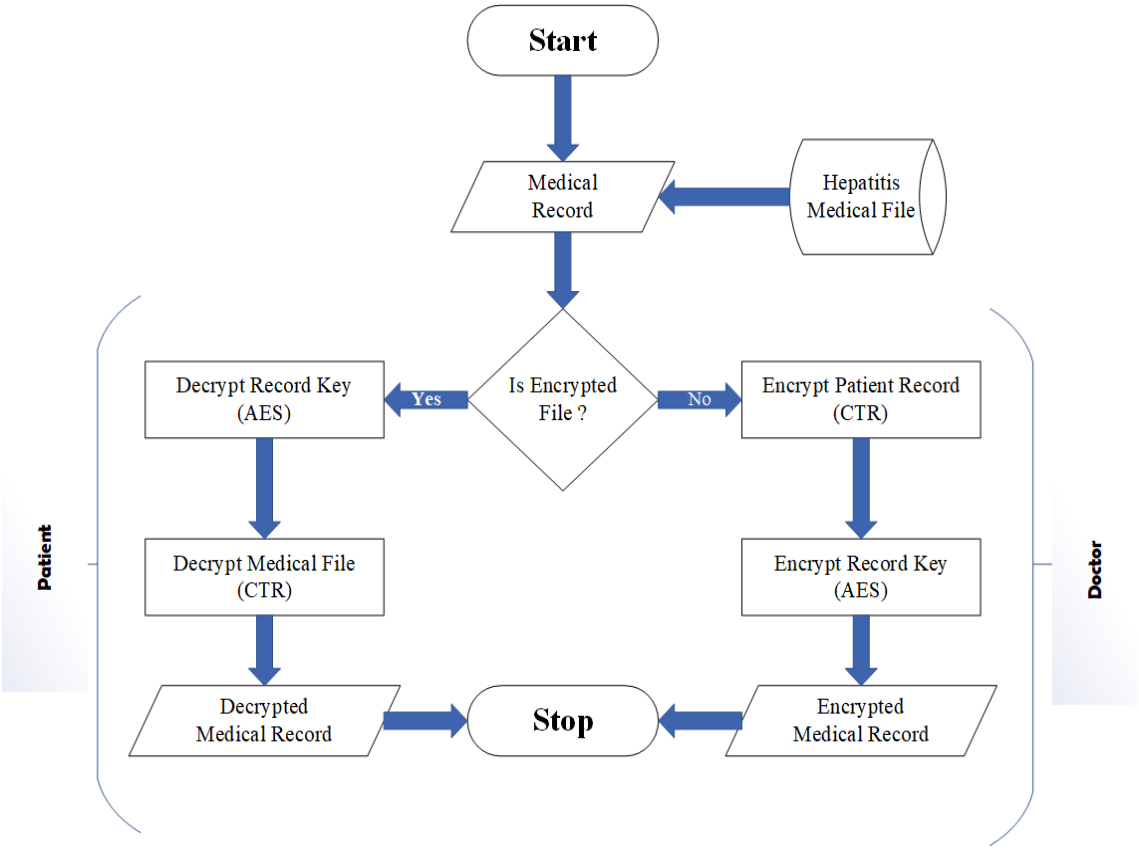

Figure 3.2 Flowchart Diagram of the Medical Record Privacy Preservation Techniques

Figure 3.2 shows the data flow chart of the proposed techniques ranging from the start indicating the beginning of the process and stop indicating the end of the work. The locally stored data is fed into the techniques as input and it is checked by the techniques to understand whether the file is encrypted or not. The No conditional statement indicates that the inputted medical record is not encrypted hence, encryption has to take place using CTR, first follow by securing and encrypting the generated key using an enhanced RSA algorithm then output the encrypted medical record. When the condition is certain then this denotes a decrypted file is supplied to the techniques then are needed to perform the decryption process. The key is decrypted first using the RSA algorithm

then the decrypted CRT key can now decrypt the main file. The result of decryption on encryption is outputted and the flow terminates (stop).

3.7 Use Case Diagram

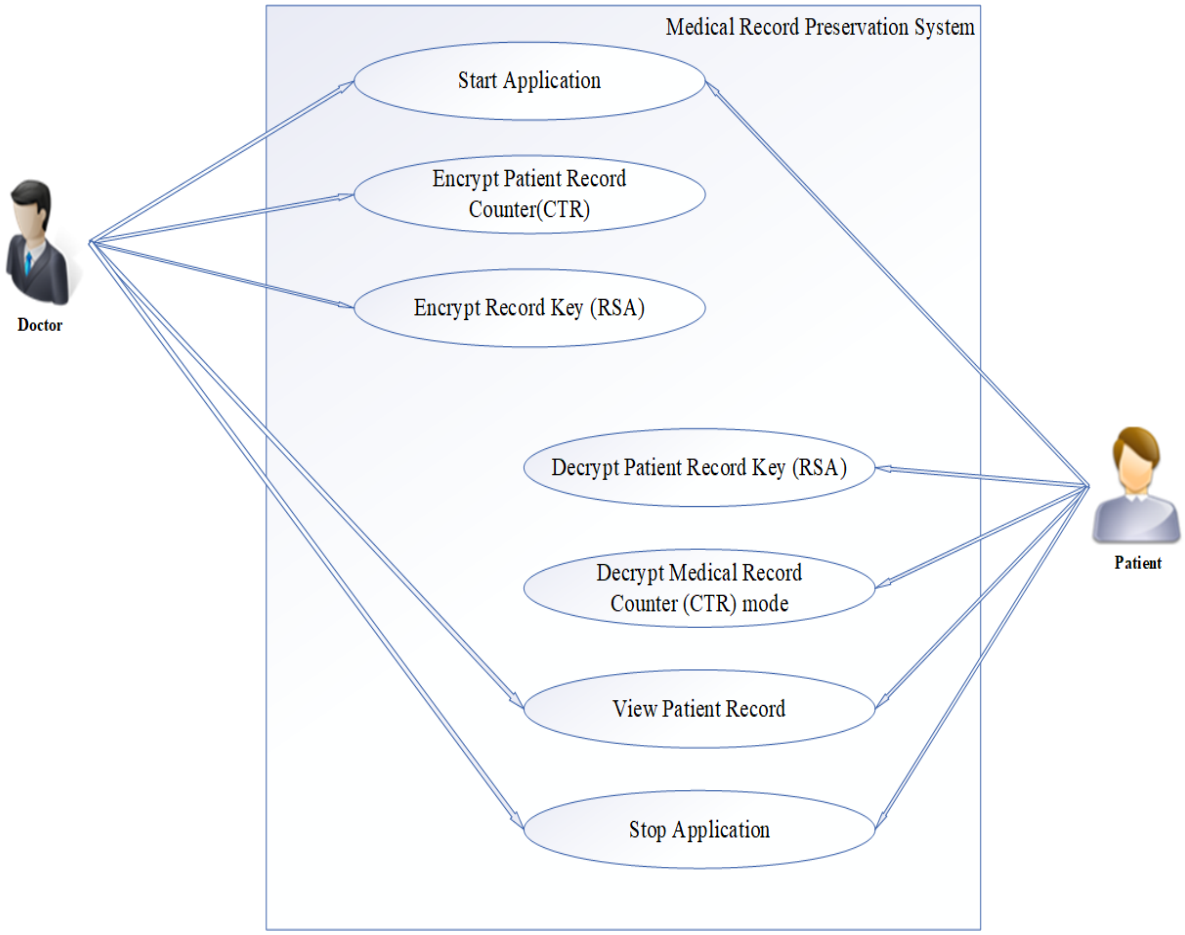

Figure 3.3: Use Case Diagram of the Medical Record Privacy Preservation Techniques

Figure 3.3. Showing the use case of the proposed techniques. The diagram consists of two actors, which include the doctor and the patient. The diagram indicates the use or action performed by both actors.

Doctor

The Doctor can start the application, can encrypt patient records with Count (CTR) mode encryption, can secure the key generated using enhanced RSA algorithm, and can

view all patient records before encryption, and also send the private key to the patients and lastly**,** stop the application or logout.

Patients

The Patient can start the application, can decrypt the patient record with the use of the private key generated by enhanced RSA, can view his medical record, and also stop the Application or log out.

3.8 The Research Approach

This work adopts the integration Counter mode encryption of and enhanced RSA algorithm for preserving medical records by encrypting and decrypting. This approach includes using the Counter (CTR) encryption mode, for encrypting the medical record using a recommended 128-bit key generated by the counter (CTR) mode encryption. The patient information like patient identity or id, name, age, sex, liver big, bilirubin, and antivirals are encrypted using the Counter mode encryption algorithm. The key is generated, and messages are stored somewhere. A counter which is an input block is been adopted. Hence, the length of the counter is equivalent to the size of the block, the counter is encrypted in CTR mode using the key and the result is been X-order with that of the first plaintext block used in generating the first ciphertext block. This research methodology used the CTR mode due to high and fast computational speed in encrypting, and the capability of encrypting large data. The key generated and used in encrypting patient record are now encrypted using the enhanced RSA algorithm. The RSA algorithm keeps the key in the first encryption save by encrypting it and generating a private and public key. The RSA public key which is also referred to as the encryption key is used by the doctor for securing the patient medical record key. Then the RSA private key is used by the patient itself for decrypting the encrypted key. It then used the

decrypted key to unlock the medical record. The enhanced RSA algorithm is an efficient encryption and decryption algorithm that are best suited for data in motion. The following steps denote the encryption and decryption.

i. Initializing data (dataset, key, counter, and many more)
ii. Encrypt Medical Records using Counter (CTR) mode encryption
iii. The Key generated by the Counter (CTR) is encrypted using an enhanced RSA algorithm
iv. RSA generates a public and private key for encrypting and decrypting the counter (CTR) key
v. The private key can be sent through a medium (internet) to the patients for decryption
vi. The public key for encrypting by the sender

3.9 The Proposed Algorithm

Algorithm: The Counter mode Encryption with an enhanced RSA algorithm for medical record encryption and decryption

Medical Record Encryption

Start

Input: *Pr (P_{atient} R_{ecord})*

Input: *E_k ($e_{ncryption}$ K_{ey} CTR)*

Input: *e (p_{ublic} k_{ey} RSA)*

Output: *E_f($e_{ncrypted}$ F_{ile}), $E_{k(e)}$ ($e_{ncrypted}$ K_{ey}), J_f (J_{son} F_{ile})*

Symbols: n ← modulus, n ← counter

1: *E_k ← 128$d_{igit}b_{ytes}$*

2: *Pr ← P_{atient} o_{bject}*

3: ***If*** *p_{atient} o_{bject} = $E_{ncrypted}$*

   a. *E_f= E_k(P_r) (CTR m_{ode})*

   b. *E_f ← E_k^ e Mode n*

   c. *J_f ← E_f*

4: ***end if***

5: ***Stop***

Medical Record Decryption

Start

Input:J_f(J_{son} F_{ile})

Input:d_k(RSA p_{rivate} k_{ey})

Input: $E_{k(e)}$ ($E_{ncrypted}$ S_{ecrete} K_{ey} CTR)

Output: E_k (CTR $e_{ncrypted}$ k_{ey}), P_r(P_{atient} R_{ecord})

Symbols: n ← modulus, n ← counter

1: *Read $E_{k(e)}$ =$E_{ncrypted}$ Secret K_{ey} (J_f)*

   a. *E_k ← $E_{k(e)}$ ^ d_k Mode n*

   b. *E_k= bytes(E_k)*

2: *Read $R_{(e)}$ = P_{atient} R_{ecord} (J_f)*

   a. *P_{atient} R_{ecord} = E_k($R_{(e)}$) (CTR $d_{ecryption}$)*

   b. *P_r ← P_{atient} R_{ecord}*

Stop

Figure 3.4 The Counter mode Encryption and enhanced RSA algorithm for medical record encryption/decryption

The algorithm shows the steps and approach used in medical privacy preservation techniques using integrated Counter (CTR) mode and Enhanced RSA algorithm. The input of the first stage (thus, the medical record encryption) is the Patient Record itself, the Counter (CTR) encryption key (E_k), and the RSA private key generated indicated with (e), while the output of the medical record encryption is the Encrypted file (E_f), Encrypted Key ($E_{k(e)}$), and finally a file in JSON Format. For stage 1, a certain variable has to be calculated such as the modulus, and counter. The Counter (CTR) mode uses 16-digit long bytes key to encrypt any patient record. The patient object is encrypted with the Counter key and the Counter key used is also encrypted using the AES encryption algorithm to generate one private and public key. Furthermore, the outputted Json File in stage one is used as input in stage two along with RSA private key (d_k) which serves as the decryption key to decrypt the CTR mode (encryption and decryption key), and the Encrypted secrete key of Counter (CTR). The output of this stage will be a decrypted Counter (CTR) and plain Patient Record data.

3.10 Performance Evaluation Metrics

Performance is evaluated for the proposed Counter (CTR) mode encryption and Enhanced RSA Algorithm to achieve the objectives outlined measures based on the several metrics which are best suited for this technique. The performance is evaluated for Medical data encryption and decryption. The metrics that are used for the evaluation are Key length which explains the length of the key, Speed, and throughput of encryption and decryption.

3.10.1 Key length

The security of the encryption techniques is a function of the length of the key. The longer the key, the more resistant the algorithm will be to a brute-force attack. For this

reason, key length was chosen as the first parameter for specifying cryptographic algorithms. Key Length is an easy objective, numeric metric to adopt since the key size is universally expressed as several bits. For instance, the standard key length for the Data Encryption Standard (DES) is 56 bits. Assuming there is no better way to break the encryption techniques, other than to try every possible key with a brute force attack, the longer the key, the longer it will take to make the number of attempts necessary to find the correct key. Every extra key bit generally doubles the number of possible keys and therefore increases the effort required for a successful brute-force attack against most symmetric algorithms.

A key length of N bits has 2^N possibilities.

Hence, this work recommends Counter (CTR) mode to 128bit encryption = 2^{128} possible key combination.

The enhanced RSA algorithm also recommends a 1024-bit encryption key of size 2^{1024} possible key combination. The result was about trillions of possible key combinations.

3.10.2 Encryption/decryption time

The encryption and decryption algorithms must be fast enough to meet real-time requirements. Therefore, the speed at which encryption and decryption take place in each of the case study algorithms was also determined. This is carried out by logging the time of the start of the encryption process and logging in the finish time of the encryption/decryption process. Therefore, the time taken to complete the process is given as:

$$Tx = Tx2 - Tx1 \quad (3.1)$$

Start time = Tx1 (m/s)

Finish time = Tx2 (m/s)

3.10.3 Throughput

The throughput of the encryption scheme defines the speed of encryption, when there is an increase in the throughput of the encryption algorithm, there is a decrease in the power consumption algorithm.

The throughput of the decryption scheme defines the speed of decryption, when there is an increase in the throughput of the decryption algorithm, there is a decrease in the power consumption algorithm.

The formula to calculate the average Throughput is given a cap

$$AvgTime = \frac{1}{Nb}\sum_{1=1}^{Nb}\frac{M_i}{t_i} \quad \text{Where} \tag{3.2}$$

AvgTime = Average Data Rate Kbs,

Nb = Number of Messages,

Mi=Message Size (Kb)

Ti=Time taken to Encrypt Message Mi

Encryption time is used to calculate the throughput of an encryption scheme. It indicates the speed of encryption. The throughput of the encryption scheme is calculated as.

$$\boldsymbol{Throughput} = \frac{1}{T} \tag{3.3}$$

The formula for throughput is *TH = I / T.* This is also known as the throughput formula.

TH=Throughput, or the average output of an item for a given period of time

I=Inventory, or the resource that is created when completing the task

T=Time, the time it takes to create the inventory

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Implemented Techniques

This section illustrates the implemented access control techniques for privacy preservation of medical records using Counter CTR and Enhanced RSA algorithms.

```
34
35  def encrypt(self, plain_text, id , passcode):
36
37      # GENERATING RSA KEY
38      self.rsa.generateKeys()
39
40      self.private , self.public = self.rsa.loadKeys()   # retrurn a turple of private, public key
41
42
43      # data = b"secret"
44      data =  bytes(plain_text, 'utf-8')
45      # self.key = get_random_bytes(16)
46
47      # counter =  Counter.new(nbits=16, prefix= unhexlify('f0f1f2f3f4f5f6f7f8f9fafbfcfd'), initial_value=0xfeff)
48
49      self.key = bytes(passcode, 'utf-8')
50      cipher = AES.new(self.key, AES.MODE_CTR)
51      ct_bytes = cipher.encrypt(data)
52      nonce = b64encode(cipher.nonce).decode('utf-8')
53
54      # ENCRYPT  the PASSCODE before placed in the BINARY FILE
55      # USING THE RSA ALGORITHM.....
56      passcode = self.rsa.encrypt(passcode , self.public)
57      print('decrypted ', passcode)
58      pk_file = open(str(id), 'wb')
59      pk_file.write(passcode)
60      pk_file.close()
61
62
63
64      # passcode = base64.b64encode(passcode)
```

Figure 4.1: RSA and Counter CTR Encryption

Figure 4.1. Shows the code snippet for performing Counter (CTR) encryption on the individual patient records, and RSA public system to encrypt the CTR key and generate a cipher version of each record key

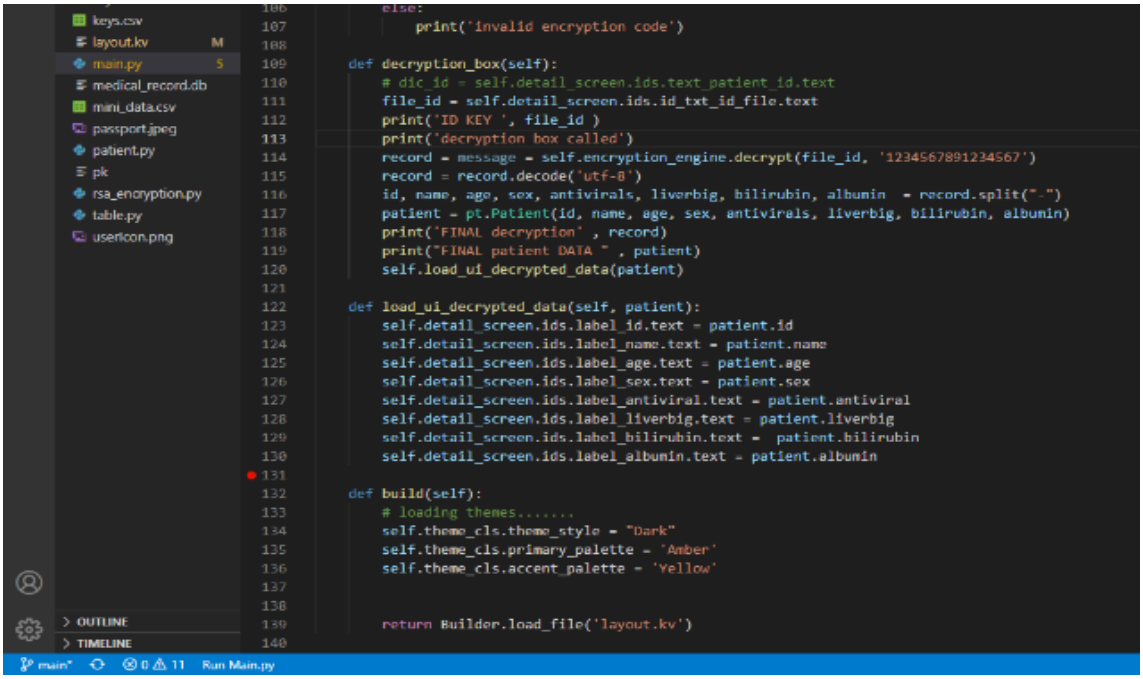

Figure 4.2: RSA and Counter CTR Decryption

Figure 4.2 shows the decryption step of the received record by the patient. Based on the figure the received file is first decrypted using the RSA private key of each patient to decrypt the Counter key (cipher text).

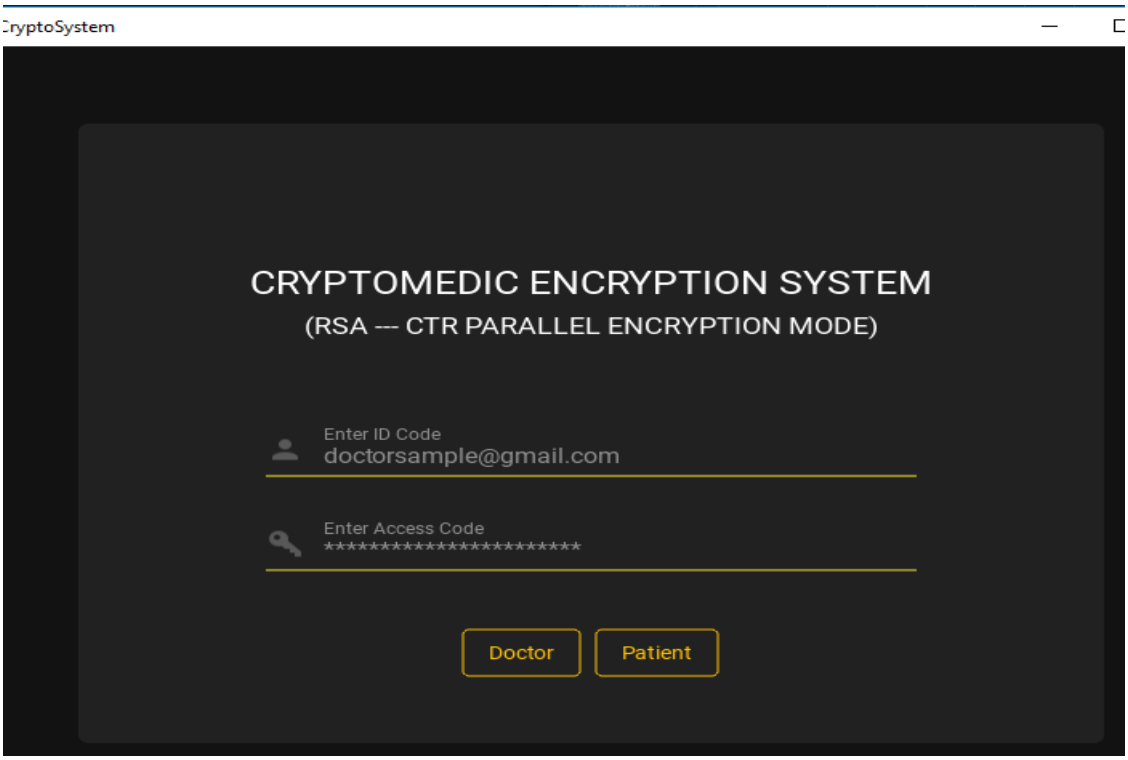

Figure 4.3: Login User Interface Section (doctor or patient

Figure 4.3 shows the login section provided for either the patient or doctors. This enable authentication and accessibility to the techniques by the user (doctor, patient) using valid credential information. The Graphical Interphase provides two text field for ‘id’, ‘access code’ and a button for login.

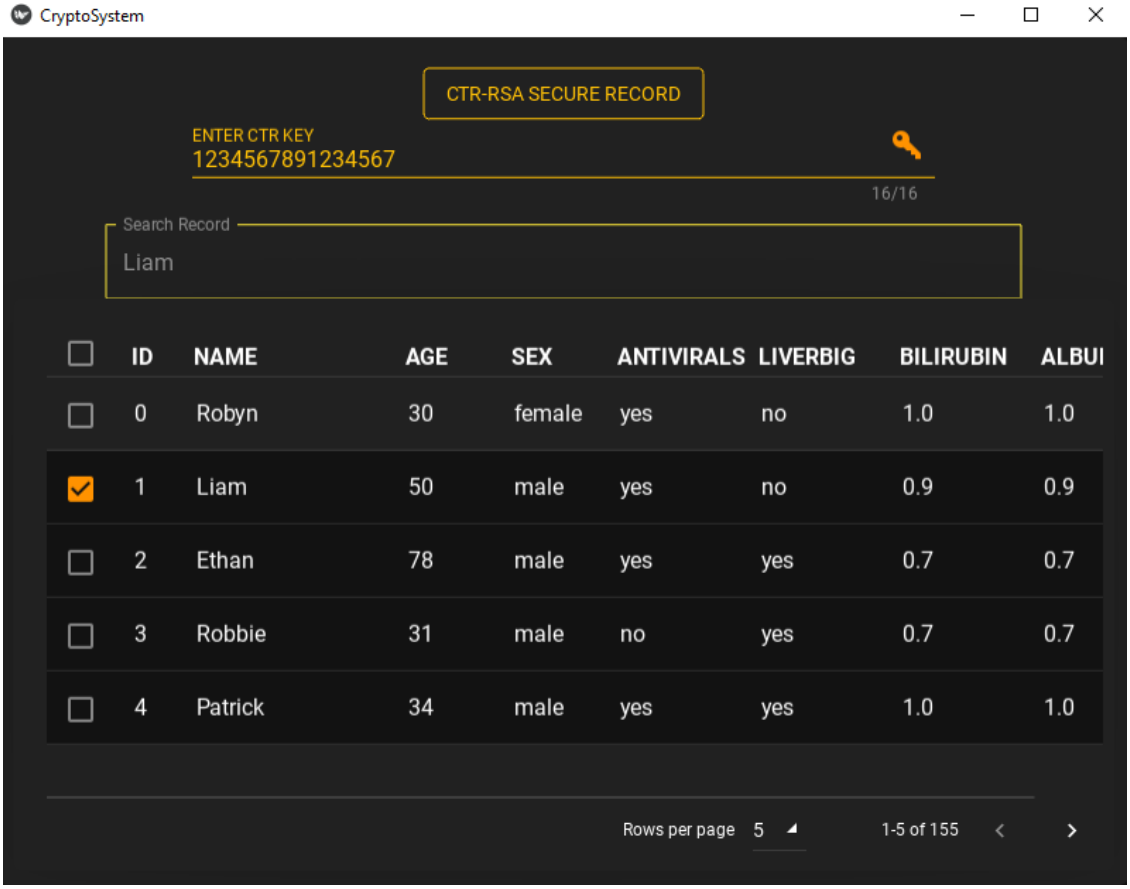

Figure 4.4: Users Record (Doctor View)

Figure 4.4 illustrate the way patient record is shown to doctors after authenticating the doctor. The patient medical information like the unique id, age, sex, antiviral status, liver-big, bilirubin and the like can be view once by the doctors in a tabular format. A section is provided for the doctor to encrypt the patient records using a 16-digit long CTR encryption key, with an activation button to initiate the encryption process.

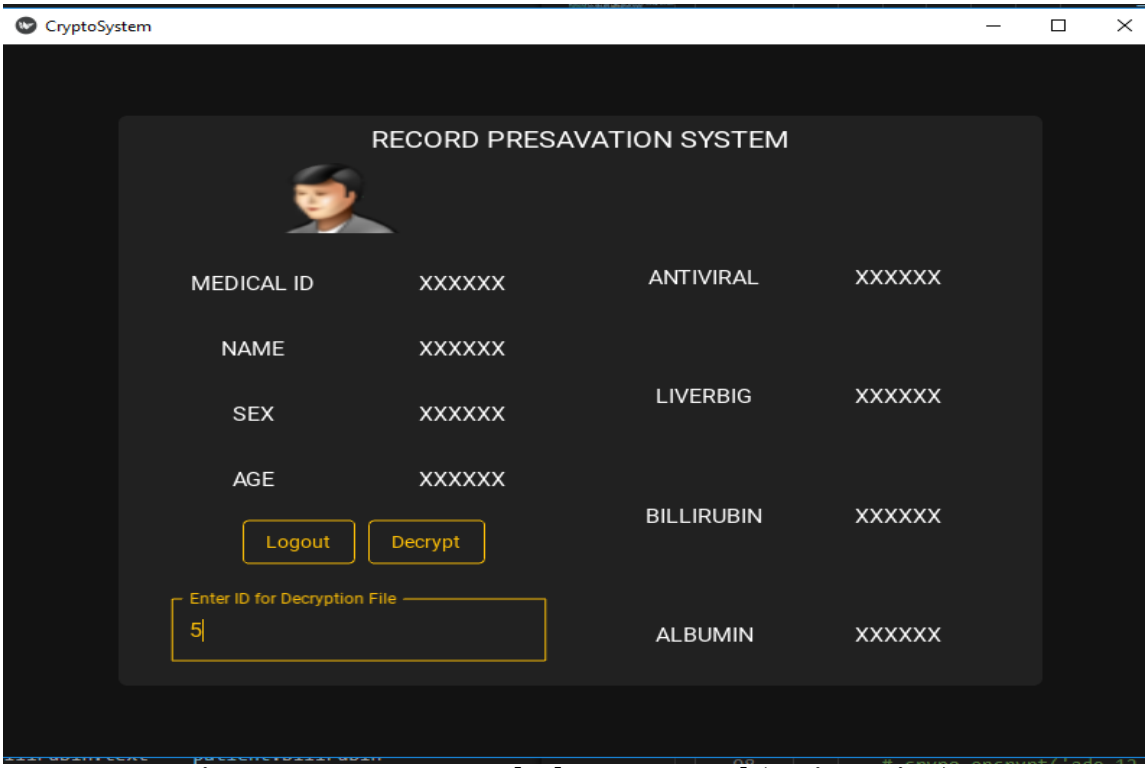

Figure 4.5: Users Record when encrypted (Patient View)

Figure 4.5 illustrating the encrypted view of the patient record for save view purpose. This medical record can only be decrypted by the intended patient using is personal private key.

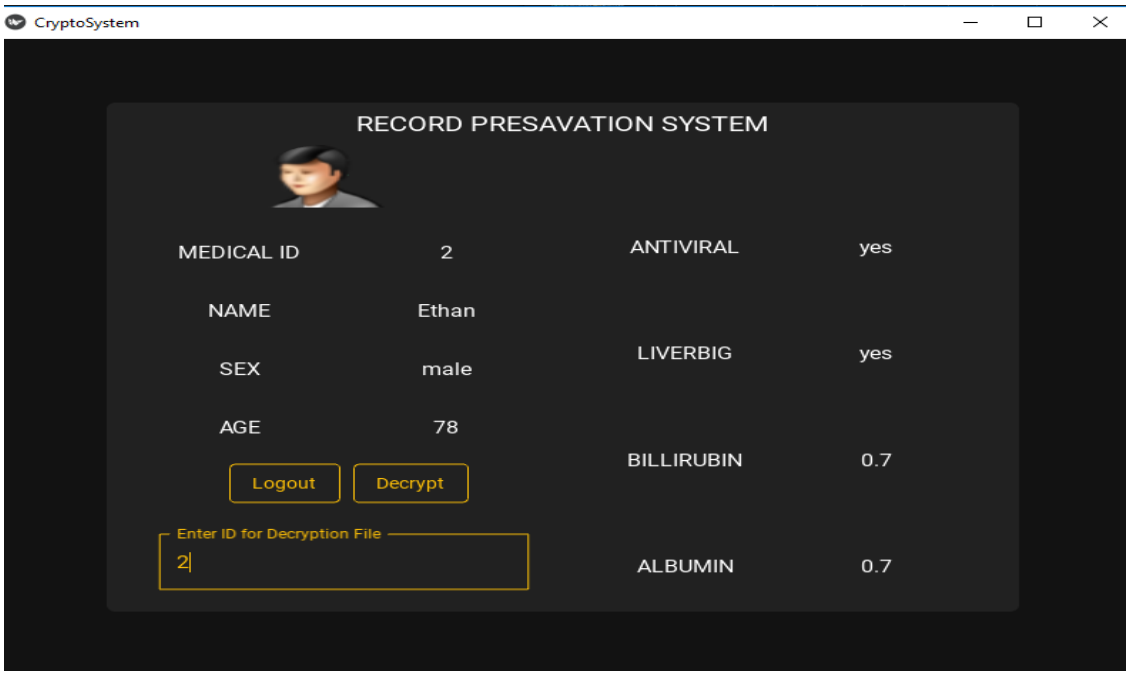

Figure 4.6: Users Record when decrypted (Patient View)

Figure 4.6 indicate the decrypted view of a patient record using a unique id and decryption key (RSA and Counter CTR).

4.2 Data Sample Results

The implementation is sample and tested using 155 hepatitis diagnosis medical data sample. The medical files contain attributes such as patient age, sex, name, antiviral status and bilirubin level. However, the developed techniques is evaluated using standard metrics such as encryption-decryption time, throughput and the key length. Generally, the encryption time of an encryption time denote the complete time taken to convert a plaintext to a cipher text, while the vice visa denotes the decryption time. The throughput metric shows the encryption-decryption speed, and the key size or key length denote the length of key used in encryption and decryption.

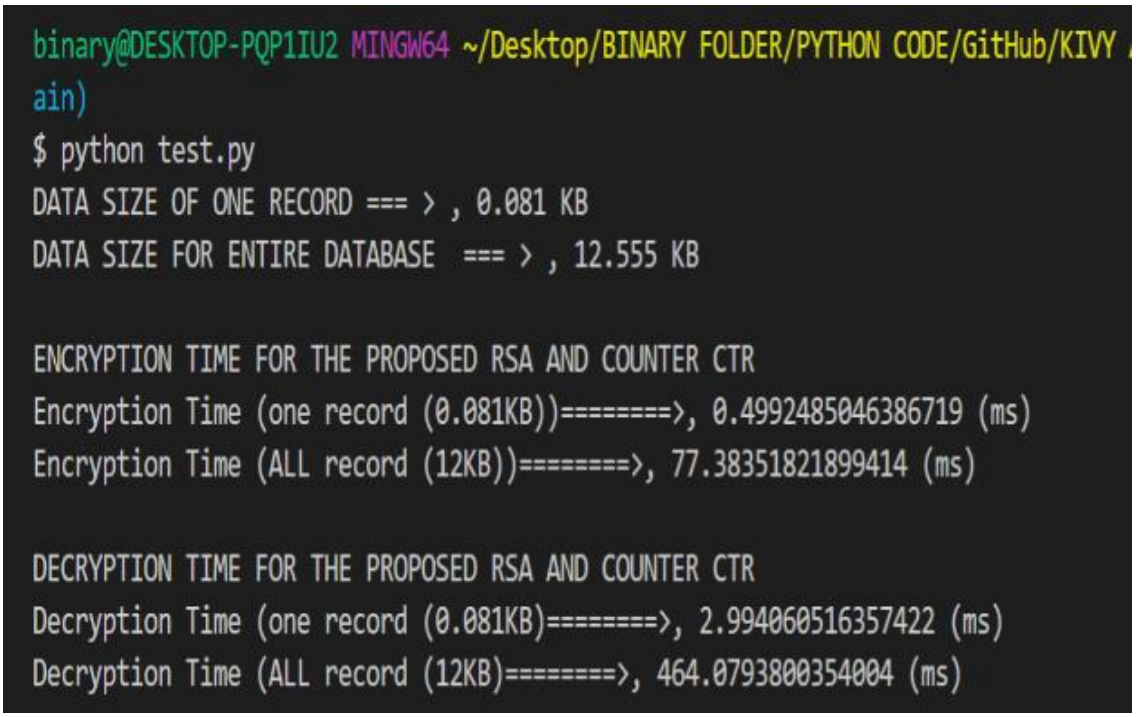

Figure 4.7: Encryption and Decryption time

Figure 4.7 shows each execution time or collective execution time (encryption and decryption speed). The file size of a document is 0.081KB with an encryption time of

0.4992ms and 2.994ms decryption time. Moreover, the entire patient record file is 12KB in size with the encryption time of 464.079 meter/seconds.

Furthermore, the table 4.1 shows the quantitative analysis comparison based on file size with existing research work.

4.3 Encryption/Decryption Time Results Comparison

Table 4.1: The encryption time comparison with other published work Results in meter/seconds (considering same file size)

S/N	Text Size KB	2 key sizes (Bonde, 2017)	SRNN(Bonde, 2017)	RSA with GI(Dawson et al., 2022)	Proposed method
		Encryption time (M/S)			
1	1KB	27m/s	18m/s	11.8m/s	**6.448m/s**
2	2KB	61m/s	24m/s	54m/s	**12.9m/s**
3	5KB	107m/s	55m/s	83m/s	**32.2m/s**
4	10KB	111m/s	71m/s	122m/s	**64.5m/s**

Table 4.1 which denote the encryption speed of the existing system and the proposed system under a certain file size (1kb, 2kb, 5bk, and 10kb). It is clearly indicated that proposed algorithm performs better in comparison to the existing approach.

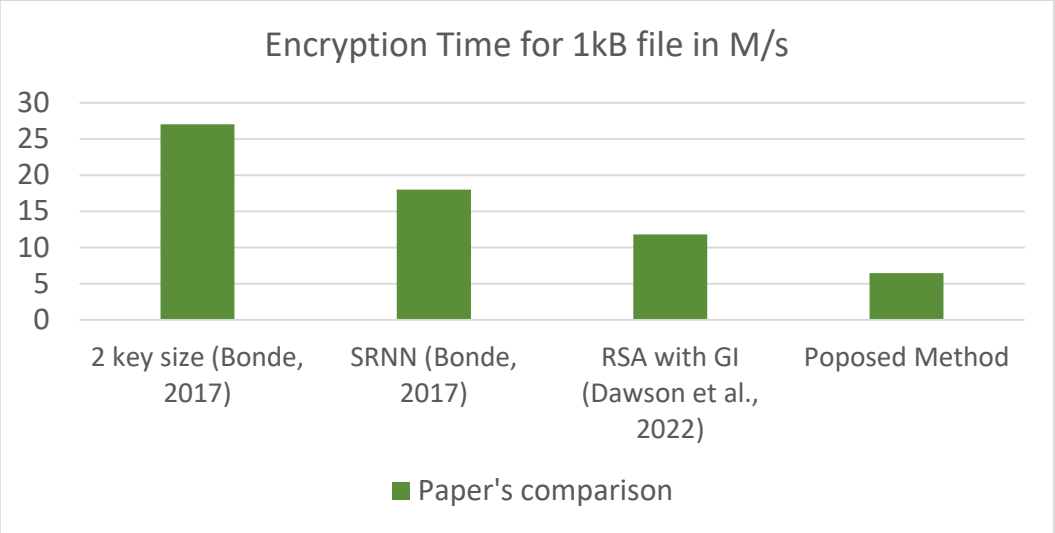

Figure 4.8 Encryption Time for 1kB file in M/s

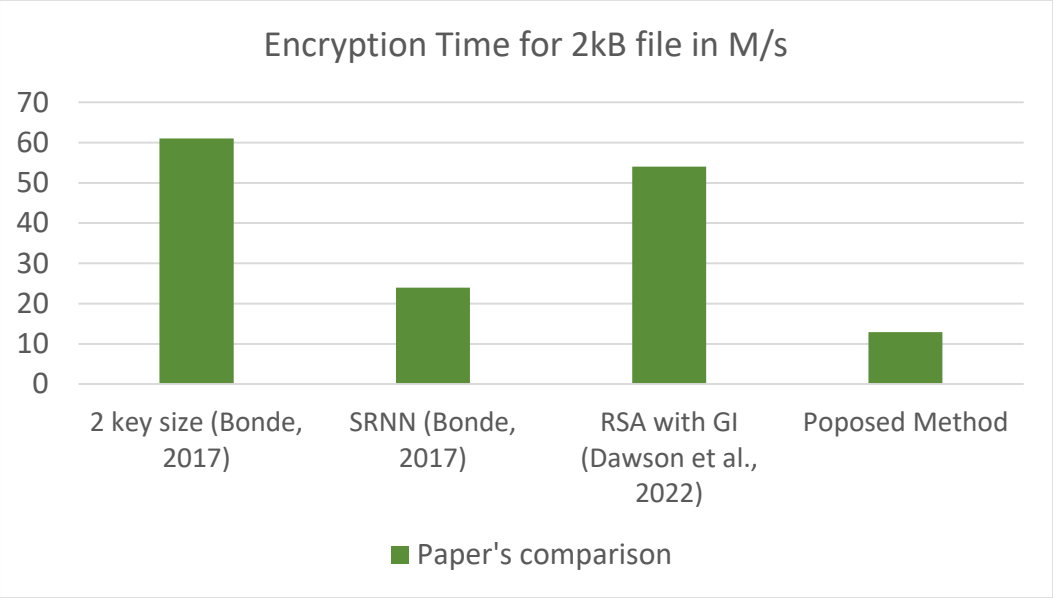

Figure 4.9 Encryption Time for 2kB file in M/s

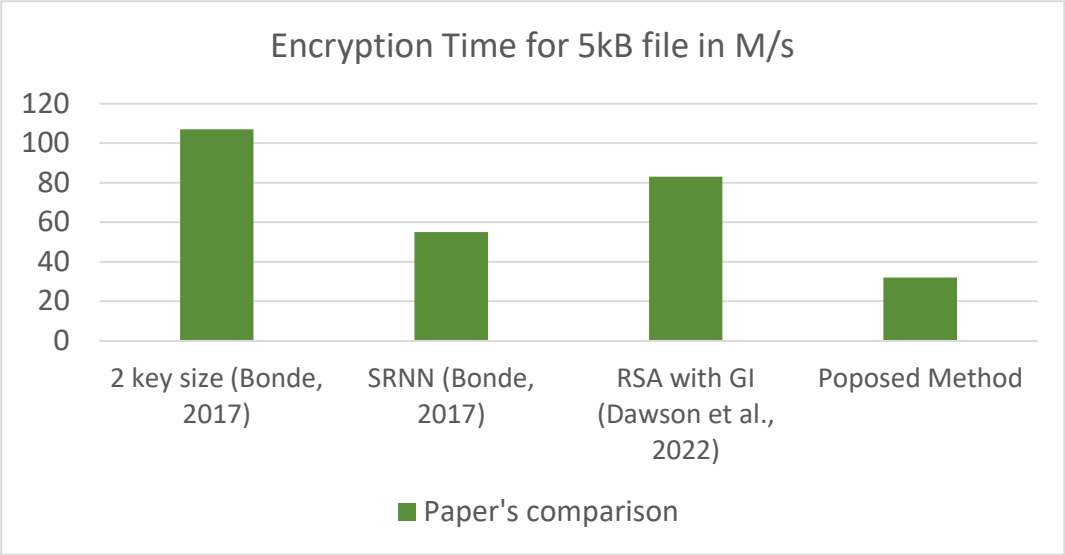

Figure 4.10 Encryption Time for 5kB file in M/s

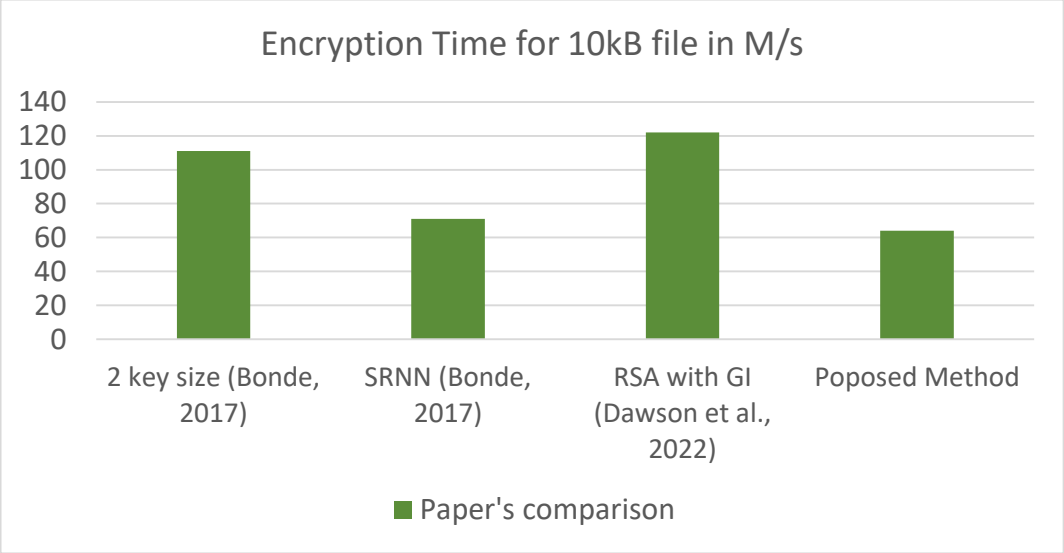

Figure 4.11 Encryption Time for 10kB file in M/s

Figure 4.8, 4.9, 4.10, and 4.11 illustrate the encryption time of various approach using the same encryption time. The algorithms are tested with file size of 1, 2, 5 and 10 kb data respectively. Hence, it shows that the proposed method has the minimum bar across all file size has indicated in red bars in the figure above. It can be concluded that the proposed method takes lesser time in millisecond to perform medical file data encryption.

Table 4.2: The decryption time comparison with other published work results in meter/seconds (considering same file size)

S/N	File size KB	2 key size method (Bonde, 2017)	SRNN (Bonde, 2017)	RSA with GI (Dawson et al., 2022)	Propose method
		Decryption time (M/S)			
1	1kb	81m/s	60m/s	13.7m/s	**38.67m/s**
2	2kb	167m/s	133m/s	83m/s	**77.38m/s**
3	5kb	488m/s	433m/s	342m/s	**193.9m/s**
4	10kb	1037m/s	854m/s	932m/s	**386m/s**

Table 4.2 which denote the decryption time of the existing techniques and the proposed technique under a certain file size (1kb, 2kb, 5bk, and 10kb). It is clearly indicated that proposed algorithm performs better in comparison to the existing approach.

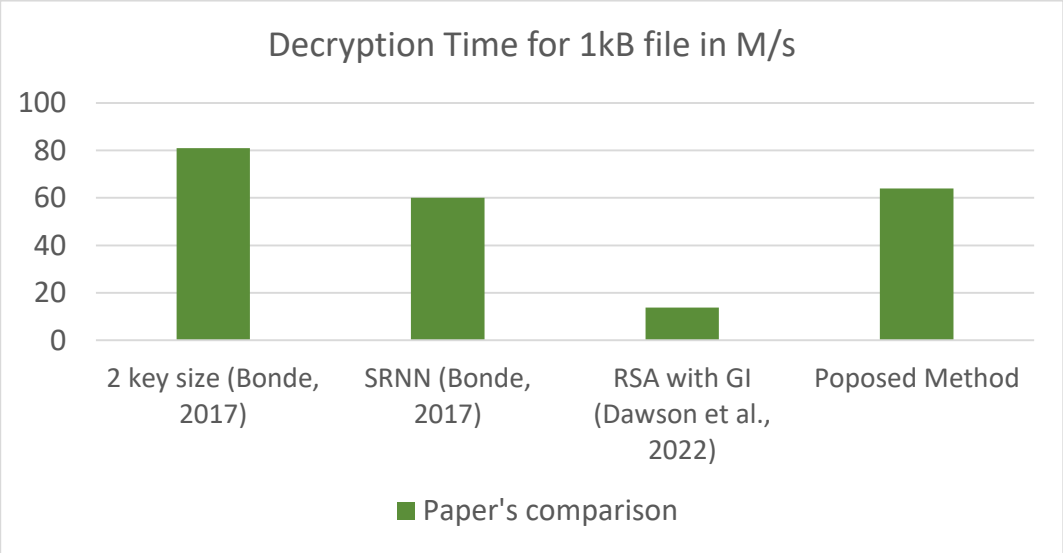

Figure 4.12 Decryption Time for 1kB file in M/s

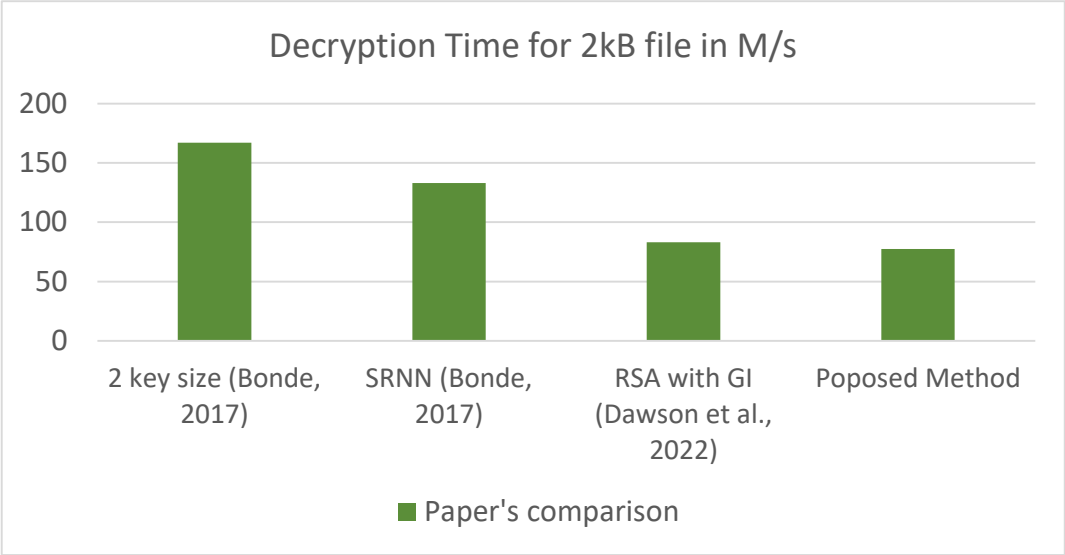

Figure 4.13 Decryption Time for 2kB file in M/s

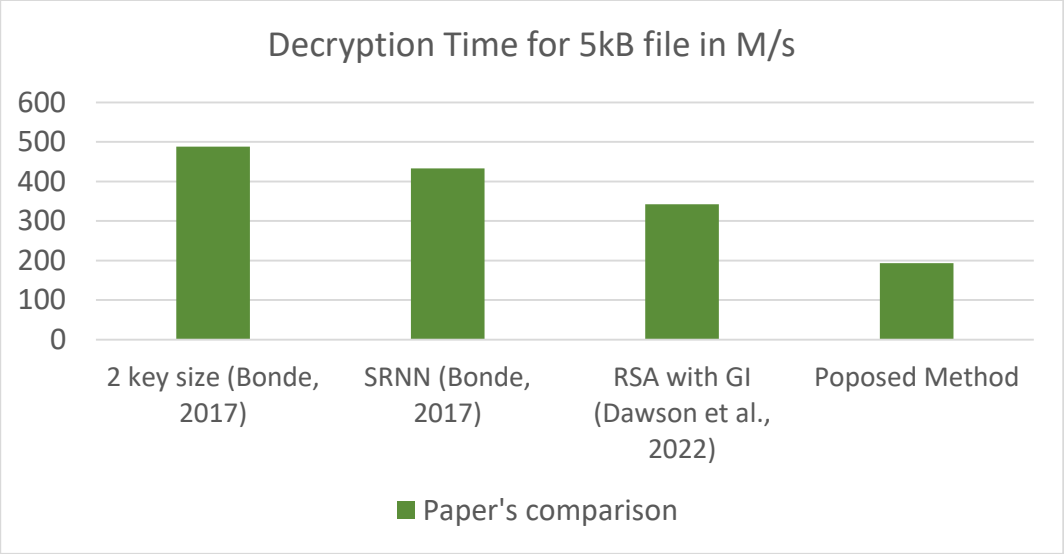

Figure 4.14 Decryption Time for 5kB file in M/s

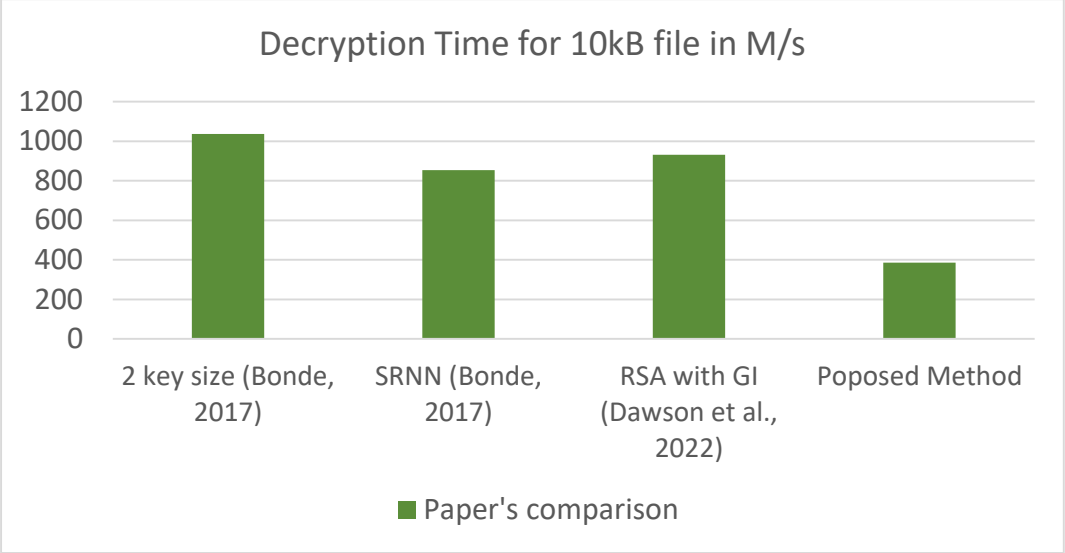

Figure 4.15 Decryption Time for 10kB file in M/s

Figure 4.12, 4.13, 4.14 and 4.15 shows that decryption time performance depends on the file size. The propose algorithm does not perform well small file size (kb) thus 1kb file and 2kb file. Hence, the higher the file size the faster the decryption process.

The efficiency was attained due to the Counter (CTR) performing the actual medical record encryption at high speed. Hence, the throughput of the propose system is computed from the file size and average encryption speed below.

Avg total plain text size (kb) = 12.555 (total file size) / 155 (total records) = 0.081

Average Encryption time (ms) = total encryption time (77.38) / total records (155) = 0.499ms = 0.000499(s)

Throughput = avg total plain text (size) / avg encryption time (s) = 0.081/0.000499 = 162.32 (kb/s)

The key length metrics determine the number of possible key combination for the encryption-decryption key. The proposed approach draws strengths from both counter (CTR) mode encryption and enhanced RSA algorithm.

Counter (CTR) mode 128bit encryption = 2^{128} possible key combination (that is about trillions possible key combination)

The enhanced RSA algorithm 1024bit encryption keyis of size 2^{1024} key length for securing the key generated by CTR mode encryption.

With both key strength, it is practically impossible to decipher or break the encryption mechanism of the proposed approach.

4.4 Results Summary

The summary of the results gotten in table 4.1, 4.2, 4.3 and 4.4 which shows both encryption, decryption, throughput and key length metrics after comparison with other published work.

The developed Techniques perform better using encryption time metrics = 11.8m/s(1kb),54m/s(2kb), 83m/s(5kb), 122m/s(10kb), when using decryption speed metrics it perform lower = 38.67m/s(1kb),77.38m/s(2kb), 193.9m/s(5kb),386m/s(10kb), Throghtput metrics achieved = 162.32(kb/s) and key length metrics = CTR counter (2^{128}) in size, and RSA (2^{1024}) in size. Counter (CTR) mode has 128bit encryption = 2^{128} possible key combination (that is about trillions possible key combination).

Table 4.3 Results Summary

SN	Metrics	Result
1	Encryption Time	11.8m/s(1kb),54m/s(2kb), 83m/s(5kb), 122m/s(10kb)
	Decryption Time	38.67m/s(1kb),77.38m/s(2kb),193.9m/s(5kb),386m/s(10kb)
2	Throughput	162.32(kb/s)
3	Key Length	CTR counter (2^{128}) in size,
		RSA (2^{1024}) in size

4.5 Discussion

The successful implementation of the proposed Techniques shows that efficient privacy preservation of medical records is achieved using the hybrid encryption scheme (Counter CTR and Enhanced RSA algorithm). The proposed encryption base medical health record techniques are capable of storing patient medical records, encrypt patient medical records using the Counter mode (CTR) AES encryption algorithm, each record encryption key will be secure using the Enhanced RSA (Private and Public) algorithm for remote distribution.

CHAPTER FIVE

5.0 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The design of access control for privacy preservation was achieved using the combined techniques like enhanced RSA algorithm with counter mode encryption to protect patient medical records, and the design architecture, flowchart, and use case diagram are cited in chapter three of this thesis work.

An enhanced Rivest-Shamir-Adlemen algorithm with counter mode encryption and decryption of medical records was achieved with the reference in three-point nine in chapter three of this thesis work.

The desktop-based system was implemented using a hybridized enhanced RSA algorithm with counter mode encryption techniques and Python programming language for the back-end business logic and database connection and also used Kivy graphical user interface to interact with the patient and doctor. This is also cited in chapter four of this thesis work.

The performance of the techniques developed was tested using various standard security metrics like encryption and decryption speed or time, throughput, and key length. A comparison was made between the proposed methods' enhanced RSA Algorithm and Counter CTR encryption with two of the existing method RSSN and GI, and the existing method proved to be better in terms of encryption of large quantities of data in a small unit of time, and also made it difficult to crack by unauthorized access.

5.2 Recommendations

The techniques can be completely applied to various sectors like Finances, Medical, Agricultural sector, and many more to preserve user privacy. As long as security and information privacy are essential, the developed techniques are highly recommended.

5.3 Contributions to Knowledge

(i) Integration of enhanced RSA algorithm using Counter Mode Techniques with high-speed encryption and decryption was developed in this thesis. The proposed techniques revealed the way large content is encrypted in a small unit of time.

(ii) A million encryption decryption possible key combination was achieved to overcome a brute force attack.

REFERENCES

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: Preserving security and privacy. *Journal of Big Data*, 5(1),1–18. https://doi.org/10.1186/s40537-017-0110-7.

Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., & Saadi, M. (2017). Big data security and privacy in healthcare: A review. *Procedia Computer Science*,113,73–80. https://doi.org/10.1016/j.procs.2017.08.292.

Ahmed, H., (2020). Encryption. The George Washington University College of Professional Studies. *PSCS 6244 Information Systems Protection*. 08,1-10. Doi:10.13140/RG.2.2.21500.56962..

Ferguson, A., (2015). Introduction to Records Management, Role: Records Manager & Quality Facilitator, V2.5. Retrieved fromhttp://www.shropscommunityhealth.nhs.uk/rte.asp.

Alexandru, A. G., Radu, I. M., & Bizon, M.-L. (2018). Big Data in Healthcare - Opportunities and Challenges. *Informatica Economica*, *22*(2/2018),43–54. https://doi.org/10.12948/issn14531305/22.2.2018.05.

Al-kaabi, E. S. S., & Belhaouari, S. B. (2019). Methods Toward Enhancing RSA Algorithm : A survey. *International Jorunal of Network Security and its Applications*. (IJNSA). 11(3),53-70. https://doi.org/10.5121/ijnsa.2019.11305.

Almeida, F. L. F. (2017). Benefits, Challenges and Tools of Big Data Management.*Journal of Systems Integration*. 8(1), 12-20. Doi: 10.20470/jsi.v8i4.311.

Al-Shiakhli, S. (2019). Big data analytics: A literature review perspective. *Luleå University of Technology*, 1(1), 1–57. Doi10.3390/bdcc6040157.

Bachlechner, D., Fors, K. La, & Sears, A. M. (2018). The Role of privacy-preserving technologies in the age of big data. *Proceedings of the 13th pre-ICIS workshop on information security and privacy (SIGSEC).*13, 2018, 1–15. https://aisel.Gisnet.org/wisp2018/28.

Banerjeer, A., Bandyopadhyay, T. and Acharya, P., (2014). Data analytics: Hyped up aspirations or true potential. *Vikalpa journal*, 38(4), 1-12. http://doi.org/10.1177/025609092013040.

Bellazzi, R. (2014). Big data and biomedical informatics: A challenging opportunity. *Year book of Medical Informatics.* 9(1), 8-13. Doi:10.15265/IY-2014-0024.

Bizer, C. Boncz, P. Brodie, M.L. Erling, O. (2012). The meaningful use of big data: Four perspectives–four challenges. *Computer science SIGMOD Rec*. 40(4), 56-60. Doi:https://doi.org/10.1145/2094114.2094129.

Bonheur, K., (2019). Merit and demerits of Big Data. Posted on March 21, retrieved on 1-12-2021, from google (profqlus).

Cao, W., Liu, Z., Wang, P., Chen, S., Zhu, C., Zheng, S., Wang, Y., & Ma, G. (2018). A rewiew PolarFS. Ultra-Low Latency And Failure Resilient Ditributed File

System For Shared Storage Cloud Database. *Proceedings of the VLDB Endowment*, 11(12), 1–28. https://doi.org/10.14778/3229863.3229872.

Çelebi, Ö.F., Zeydan, E., Kurt, Ö.F., Dedeoğlu, Ö.,İieri, Ö., AykutSungur, B., Akan, A. and Ergüt, S., (2013). On use of big data for enhancing network coverage analysis. *In ICT 2013 IEEE journal,* 1-5. Doi:10.119/ICTEL.2013.6632155.

Chen, C.P.; Zhang, C.Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences.* 275, 314–347. Doi:275.314-347.10.1010/j.ins.2014.01.015.

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data,* 6(1), 2-25. https://doi.org/10.1186/s40537-019-0217-0.

Davenport, T.H., Dyché, J. (2013). Big data in big companies. Baylor Business Review: *International institute for analitics.* (online). 32(1), 1-31. Retrieved from http://www.sas.com/resources/whitepaper/wp.

Dawson, J. K., Benjamin, J., & Acquah, H. (2022). An Enhanced Rsa Algorithm For Data Security Using Gaussian Interpolation Formula. *Reseach square.* 16(4). 1-13. http://doi.org.21203/rs.3.rs-1326669/v1.

Deepa, N., Pham, Q. V., Nguyen, D. C., Bhattacharya, S., Prabadevi, B., Gadekallu, T. R., Maddikunta, P. K. R., Fang, F., & Pathirana, P. N. (2022). A survey on blockchain for big data: Approaches, opportunities, and future directions. *Future Generation Computer Systems*, 131, 209–226. https://doi.org/10.1016/j. future. 2022.01.017.

Elgendy, N. and Elragal, A., (2014). Big data analytics: a literature review paper. *Lecture note on computer science.* 2014 Springer, 214-227. Doi:10.1007/978-3-319-08976-8-16.

Galetsi, P., Katsaliaki, K., & Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management*, *50*(May 2019), 206–216. https://doi.org/10.1016/j.ijinfomgt.2019.05.003.

Google, (2022). Privacy Preservation Technic (PPT). Retrieved 20-1-22. www.igi-global.com.

Goswami, P. (2017). A survey on Big Data and Privacy Preserving Publishing Techniques. *A journal advances in computational science and Technology*. *10*(3), pp. 395–408. ISSN: 0973-6107.

Gupta, R. (2014). Journey from Data Mining to Web Mining to Big Data. *International Journal of Computer Trends and Technology*, 10(1), 18–20. https://doi.org/10.14445/ 22312803/ijctt-v10p104.

Gutiérrez, O., Romero, G., Pérez, L., Salazar, A., Wightman, P., & Charris, M. (2020). Healthyblock: Blockchain-based it architecture for electronic medical records resilient to connectivity failures. *International Journal of Environmental*

Research and Public Health, 17(19), 1–38. https://doi.org/10.3390/ ijerph17197132.

Hai, Y.R. and Zhang X.Y. (2017). Research and design of big data sharing and comprehensive application platform for agricultural internet of Ningxia. *Ningxia Agriculture for Science and Technology*. 58(9), 51–53.Retrieved from http://www.Googlescholar.com.

Harsh, K. P. and Ravi, S. (2014). Big data security and privacy issues in healthcare, in proceedings of *2014IEEE Computer Society International Congress on Big Data, Dallas*. 2014, 762-765. Doi:10.1109/BigData.congress.2014.122.

He, Y., Lee, R., Huai, Y., Shao, Z., Jain, N., Zhang, X. and Xu, Z., (2011). RCFile: A fast and space efficient data placement structure in MapReduce-based warehouse systems. *IEEE International conference 0f Data Engineering (ICDE) conference*, © 2011k.IEEE, 2011, 1199- 1208. Doi:10.1109/ICDE.2011.5767933.

Hofmann, E. and Rutschmann, E., (2018). Big data analytics and demand forecasting in supply chains: a conceptual analysis. *The International Journal of Logistics Management*, 29, 739-766. Doi: 10.1108/IJLM-04-2017-088.

Id, S. C., Ahmed, K., Wang, H., Whittaker, F., & Chen, Z. (2020). *Healthchain*:A novel framework on privacy preservation of electronic health records using blockchain technology. 15(12), 1-35. https://doi.org/10.1371/journal.pone.0243043.

Iwashyna, T.J. Liu, V. (2014). What's so different about big data? A primer for clinicians trained to think epidemiologically. *Annals of the American Thoracic Society*. 11(7), 1130–1135. Doi:10.1513/AnnalsATS.201405-185AS.

Iyengar, A. Kundu, A. Pallis, G. (2018). Healthcare informatics and privacy, *IEEE Internet Computing*. 22(2), 29–31. Doi10.1109/MIC.2018.022021660.

Jagadish, H.V. Gehrke, J. Labrinidis, A.; Papakonstantinou, Y. Patel, J.M. Ramakrishnan, R.; Shahabi, C. (2014). Big data and its technical challenges. *Communication of the ACM*. 5(7), 86–94. Doi:57.86-94.10.1145/2611567.

Jasim, H., Hameed, S., A., Hadishaheed, S., & Ahmad, A. H., (2015). Big Data and Five V'S Characteristics. *International Journal of Advances in Electronics and Computer Science*,*2*,(1), 16-23.ISSN:2393–2835. https://www.researchgate.net/publication/ 332230305.*Journal of Systems Integration*, 8(4), 12–20.https://doi.org/10.20470/jsi.v8i4.311.

Kaisler, S., Armour, F., Espinosa, J.A., (2013). Money, W. Big data: Issues and challenges moving forward. In Proceedings of the System Sciences (HICSS), *46th Hawaii International Conference on IEEE system science*. 7–10, January 2013; 995–1004. Doi:10.1109/HICSS.2013.645.

Kittur, P. K., Zhang, F., Kimeli, V. K., Omala, A. Anyembe, and Eugene, O. M., (2019). Privacy Preservation for eHealth Big Data in Cloud Accessed Using Resource-Constrained Devices: Survey. *International journal of Network security*. 21, No.2, 312-325, Mar. 2019. Doi:10.6633/IJNS.201903_21(2).16.

Koo, J., Kang, G., & Kim, Y. G. (2020). Security and privacy in big data life cycle: A survey and open challenges. *Sustainability (Switzerland)*, 12(24), 1–32. https://doi.org/10.3390/su122410571.

Kraemer, F.A. Braten, A.E. Tamkittikhun, N. (2017). Fog computing in healthcare-a review and discussion, *IEEE Access*. 5, 2017, 9207-9222. Doi:10.1109/ACCESS.2017.2704100.

Kuo, M.H.; Sahama, T.; Kushniruk, A.W.; Borycki, E.M.; Grunwell, D.K. (2014). Health big data analytics: Current perspectives, challenges and potential solutions. *International Journal of Big Data Intelligence.* 1, (1), 114–126. Doi:10.1504/IJBDI.2014.063835.

Lekhwar, S., Yadav, S. and Singh, A., (2019). Lekhwar, S., Yadav, S. and Singh, A., 2019. Big Data Analytics in Retail. *Proceedings of ICTIS 2018 Conference Singapore, Springer*. 2, 469-477. Doi:10.1007/978_13_1747_7_45.

Ma, S. David, K., A. P., & Stepchenkova, S. (2020). Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse.*Tourism Management*, *77*(October 2019), 335-339. https://doi.org/10.1016/j.tourman. 2019.

Makridakis, S. Spiliotis E, Assimakopoulos, V. (2018) Statistical and Machine Learning forecasting methods: Concerns and ways forward. *Journal of plos one Statistical and ML forecasting methods.* 13(3), 1-26. https://doi.org/ 10.1371/journal.pone.0194889.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H., (2011). Big data: The next frontier for innovation, competition, and productivity. *Journal of Computer and Communication.* 5(3), 40-48. Doi:10.4236/jcc.2017.53005.

Mohammed, D. A., Ojerinde, O., Folorunsho V. M., & Ogbuka K. M., (2020). Privacy Preservation in Big Data Application Using Advanced Encryption Standard and Least Significant Bit Steganography. *I-Manager's Journal on Information Technology*, 9(2), 1-10. https://doi.org/10.26634/jit.9.2.17550.

Mohanty, H. Boinepelli, H. (2015). Applications of Big Data. *Studies in big data II*. 161–179. https://doi.org/10.1007/978-81-322-2494-5_7. (c) springer 2015.

Newman, P., Stantic, A. B., Systems, I., Ho, D., Conley, D., Hojem, A., Pyke, O., Wood, J., Aird, R., Nguyen, K., & Grant, G. (2017). Technologies , and Transportation . *Journal of Transportation and Technology*. (3,). ISSN:2160-0473.

Oliveira, L.S.; Gerosa, M. (2011). Collaborative features in content sharing Web 2.0 social networks: A domain engineering based on the 3C collaboration model. *In Proceedings of the Collaboration Researchers'International Workshop on Groupware (CRIWG'11);* 6969, 142–157.Retrieved from http://www.Googlescholar.com.;

Olufohunsi, T., (2019). Data Encryption, cyber security, threat intelligence and forensics.*A Research Gate Article.* University of Salford, Manchester UK. 11 December, 2019, 1-4. Doi:10.13140/DE.2.2.10215.47529.

Pandey, S., & Pandey, R. (2018). Medical (Healthcare) Big Data Security and Privacy. International journal of scientific & engineering reseach. Vol.*9*(2), pp. 180–182.ISSN:2229-5518.

Pethe, P. R., Harshala, B. & D. Subhash (2017). Comparative study and Analysis of Cryptographic Algorithms DES and RSA. *International of Advance Research in Computer Science and Management Studies*. 5(1), 48-56. ISSN:2321-7782.

Raghupathi, W. and Raghupathi, V., (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems,* 2(3), 1-10. Doi:https://doi.org/10.1186/2047-2501-2-3.

Rajeshwari, D., (2015), State of the art of big data analytics: A survey, *International Journal of Computer Applications*, 120(22), 39-46. Doi: 10.5120/21395-4456.

Ram, P. Mohan Rao, S. Murali Krishna and A. P. Siva Kumar (2018). Privacy preservation techniques in big data analytics: a survey. *Journal of big data.* 5(33), 1-12. https://doi.org/10.1186/s40537-018-0141-8.

Rehman, Ur. M.H. Chang, V. Batool, A. Wah, T.Y., (2016) Big data reduction framework for value creation in sustainable enterprises. *International Journal Information Management*. 36(16), 917–928. https://doi.org/10.101016/j.ijinfomgt.2016.50.013.

Roehrs, A., da Costa, C. A., da Rosa R., da Silva, V. F., Goldim, J. R., & Schmidt, D. C. (2019). Analyzing the performance of a blockchain-based personal health record implementation. *Journal of Biomedical Informatics*, *92*, 1-8. https://doi.org/10.1016/j.jbi.2019.103140.

Russom, P., (2011). Big data analytics, *TDWI Research best practices report, fourth quarter*. 19, 1-34.Retrieved from www.tdwi.org/blogs/philip-russom.

Saratchandran, V. (2018), Big Data in Healthcare - Opportunities and Challenges"https://www.fingent.com," 9 January. [Online]. Available: https://www.fingent.com/blog/5-ways big-data-is-changing-the-healthcare industry. 1-6. Doi:10.13140/RG.2.2.19964.16000.

Seliya, N., Abdollah Z. A., & Khoshgoftaar, T. M. (2021). A literature review on one-class classification and its potential applications in big data. *InJournal of Big Data.* 8(1), 1-31. Springer International Publishing.https://doi.org/10.1186/s40537-021-00514-x.

Seun, E., Adekunle, Y. A., Omotosho, O. J., Adebayo, A. O., & Omolara, T. (2019). Data privacy preserving model for health information system. *International Journal of Engineering Research and Technology*, 12(6), 745–752. ISSN: 0974-3154.

Singh, J. and Singla, V., (2015). Big data: tools and technologies in big data. *International Journal of Computer Applications.* 112(15), 6-10. Doi:10.5120/19740-0029.

Sultan, I., Mir, B. J., Banday, M. T., & Member, S. (2020). Analysis and Optimization of Advanced Encryption Standard for the Internet of Things.*In 7th International Confrence on Signal Processing and Integrated Networks (SPIN).* 1, 571–575. Doi:10.1109/SPIN48934.2020.9071380.

Tola, K., Abebe, H., Gebremariam, Y., & Jikamo, B. (2017). Improving Completeness of Inpatient Medical Records in Menelik II Referral Hospital, Addis Ababa, Ethiopia. *Advances in Public Health*, 1-5. Doi:10.1155/2017/8389414.

Turban, E. and Aronson, J. E. (2001). Decision Support Systems and Intelligent Systems. Prentice-Hall, New Jersey, USA. *Journal of Computer Sciences and Applications.* 6, 1-41. ISBN:13:978-0130894656, ISBN: 10:0130894656.

Venkatesh, M. Satish, G. K. Ram S. and Sudarshan, M. (2019). Secure Data Encryption and Decryption Using Crypto-Stego.A Project: *International journal of Research and analytical Reviews (IJRAR).* 7(1), 1-55. E-ISSN:2348-1269, P-ISSN:2349-5138.

Wamba, S.F., Gunasekaran, A., Akter, S., Ren, S.J.F., Dubey, R. and Childe, S.J., (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*. 70, 356-365. http://dx.doi.org/10.10/6/j.jbusres.2016.08.009.

Wang, Y.Q, Chen, Y.Q. and Wang, Z.W. (2018). Design of agricultural statistical data collection platform under big data environment. 26(24), 111–115. Retrieved from http://www.Googlescholar.com.

William, J., (2015), Applications of Big Data. Big Data Disease Breakthroughs: Information Week. Retrieved from http://www. Informationweek.com. Springer India 2015.

Zeng, L. and Ren, Y. (2018). Design of agricultural information management platform based on big data technology. *South Agricultural Mechanism*. 49(1), 57-67. Retrieved from http://www.Googlescholar.com.

Zhong, R.Y., Newman, S.T., Huang, G.Q. and Lan, S., (2016). Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives. *Computers & Industrial Engineering journal*, 10(11),572-591. http://dx.doi.org/10.1016/j.cie.2016.07.013.

APPENDICES

Aes_ctrmode.py file

```
import base64
fromctypes import sizeof
fromCrypto.Cipher import AES
fromnumpy import byte
import time
data = b'ade-12-yes-1995'
# https://pycryptodome.readthedocs.io/en/latest/src/cipher/classic.html#ctr-
# https://pycryptodome.readthedocs.io/en/latest/src/cipher/aes.html
importjson
from base64 import b64encode
fromCrypto.Cipher import AES
fromCrypto.Util import Counter
fromCrypto.Random import get_random_bytes
frompathlib import Path
importjson
from base64 import b64decode
fromCrypto.Cipher import AES
importcsv as cv
importrsa_encryption as rsa
frombinascii import hexlify, unhexlify
classAESCTRXEncryption:
def __init__(self) -> None:
self.key = None
    # generate and RSA key for each encryption
self.rsa = rsa.RSAEncryption()
self.private ,self.public = ('', '')
def encrypt(self, plain_text, id , passcode):
    # GENERATING RSA KEY
self.rsa.generateKeys()
self.private ,self.public = self.rsa.loadKeys()   # retrurn a turple of private, public key
    # data = b"secret"
data =  bytes(plain_text, 'utf-8')
    # self.key = get_random_bytes(16)
    # counter =  Counter.new(nbits=16, prefix=
unhexlify('f0f1f2f3f4f5f6f7f8f9fafbfcfd'), initial_value=0xfeff)
self.key = bytes(passcode, 'utf-8')
    # print("KEY SIZE" , len(self.key))
cipher = AES.new(self.key, AES.MODE_CTR)
start_time = time.time()
ct_bytes = cipher.encrypt(data)
nonce = b64encode(cipher.nonce).decode('utf-8')
    # ENCRYPT  the PASSCODE before placed in the BINARY FILE
    # USING THE RSA ALGORITHM.....
passcode = self.rsa.encrypt(passcode , self.public)
```

```
end_time = time.time()
        # print('Encryption Time ', (end_time-start_time)*1000)
        # print('decrypted ', passcode)
pk_file = open(str(id), 'wb')
pk_file.write(passcode)
pk_file.close()
        # passcode = base64.b64encode(passcode)
        # print(passcode)
        # self.cerfertext = self.encrypt(self.key, self.public)
ct = b64encode(ct_bytes).decode('utf-8')
result = json.dumps({'nonce':nonce, 'ciphertext':ct})
Path(str(id)+".json").write_text(result)
        # print(result)
return (end_time-start_time)*1000
def decrypt(self, id_file , passcode):
         # We assume that the key was securely shared beforehand
data = Path(str(id_file)+'.json').read_text()try:
            b64 = json.loads(data)
nonce = b64decode(b64['nonce'])
ct = b64decode(b64['ciphertext'])
            # key_text = b64['Code']
            # decrypting the AES KEY  with RSA encryption
            # key_text = key_text.decode("ascii")
            # rsa_key_decrypt = self.r.decrypt(self.key , self.private)
self.private ,self.public = self.rsa.loadKeys()
             # retrurn a turple of private, public key
        # DECRYPT  the CTR KEY.. LOAD BINARY FILE..
        # USING THE RSA ALGORITHM.....
pk_read = open(str(id_file), 'rb')
pass_key = pk_read.read()
pk_read.close()
            # print('AES KEY SIZE ' , len(pass_key))
            # print(f'RSA key SIZE private:{sizeof(self.private)}
public:{sizeof(self.public)}')
            start_time2 = time.time()
pass_key = self.rsa.decrypt(pass_key, self.private)
            # print('decrypted ', pass_key)
            # temp key movement
self.key = bytes(pass_key, 'utf-8')
            # key = bytes(rsa_key_decrypt, 'utf-8')
cipher = AES.new(self.key, AES.MODE_CTR, nonce=nonce)
pt = cipher.decrypt(ct)
            end_time2 = time.time()
            # print("The message was: ", pt)
returnpt , (end_time2-start_time2)*1000
except (ValueError, KeyError):
print("Incorrect decryption")
```

database.py file

```
import sqlite3  as database
importcsv as cv
import patient as pt
class Database:
defload_csv_data(self, filename)-> list:
with open(filename) as file:
dataset = cv.reader(file)
header = [next(dataset)]
datalist = []
for row in list(dataset):
print('row==>:' , row)
datalist.append(row)
returndatalist
def __init__(self) -> None:
self.connection = None
withdatabase.connect('medical_record.db') as db :
self.connection = db
print('connection established')
defcreate_medical_record_table(self):
self.connection.execute("""CREATE TABLE IF NOT EXISTS medical_table (
id INT NOT NULL,
name TEXT NOT NULL,
age INT NOT NULL,
sex TEXT NOT NULL,
antivirals TEXT NOT NULL,
liverbig TEXT NOT NULL,
bilirubin TEXT NOT NULL,
        Albumin TEXT NOT NULL,
        PRIMARY KEY (id)
    )
    """)
self.connection.commit()
print('medical record table created.....')
defadd_record(self, patient):
id = patient.id
name = patient.name
age = patient.age
sex = patient.sex
antivirals = patient.antiviral
liverbig = patient.liverbig
bilirubin = patient.bilirubin
albumin = patient.albumin
    # print('ID .....====> ', id)
self.connection.execute(
```

```
        'INSERT INTO medical_tableVALUES(?,?,?,?,?,?,?,?)', (id, name, age,
sex,antivirals, liverbig,bilirubin,albumin) )
self.connection.commit()
print(patient.id  , '  .... record inserted.....')
deffetch_all_record(self):
record_list = []
cr = self.connection.cursor().execute('SELECT * FROM medical_table')
data = cr.fetchall()
for patient in data:
#  id , name , age, sex, antivirals, liverbig, bilirubin, albumin
          us = pt.Patient(patient[0], patient[1], patient[2], patient[3], patient[4], patient[5],
patient[6], patient[7])
record_list.append(us)
returnrecord_list
defcsv_to_database(self, f):
csv_record = self.load_csv_data(f)
patient_list = []
for record in csv_record:
id,  age, sex, antivirals, liverbig, bilirubin, albumin , name = record
patient = pt.Patient(id, name, age, sex, antivirals, liverbig, bilirubin, albumin)
print(f"{patient}")
self.add_record(patient)else:
print('all record inserted completely')
defdatabase_to_csv(self):
patient_records = self.fetch_all_record()
      # write to csv
with open("datas.csv", "w") as file:
writer = cv.writer(file)
writer.writerow(['ID', 'NAME', 'AGE', 'SEX', 'ANTIVIRALS' , 'LIVERBIG',
'BILIRUBIN', 'ALBUMIN'])
for patient in patient_records:
id, name, age, sex, antivirals, liverbig, bilirubin, albumin = patient.id, patient.name,
patient.age, patient.sex, patient.antiviral, patient.liverbig, patient.bilirubin,
patient.albumin
writer.writerow([id, name, age, sex, antivirals, liverbig, bilirubin, albumin])
             # print(f"{id} {name} {age}")
print('program completed....')
# db = Database()
# # db.create_medical_record_table()
# patient = db.fetch_all_record()
# # db.load_csv_data('final_medical_record.csv')
# # db.csv_to_database('final_medical_record.csv')
# # for p in patient:
# #     print('record : ' , p)
# # print([str(p) for p in patient])
# db.database_to_csv()
```

Rsa_encryption.py file

```
importrsa
classRSAEncryption():
defgenerateKeys(self):
        (publicKey, privateKey) = rsa.newkeys(1024)
with open('keys/publicKey.pem', 'wb') as p:
p.write(publicKey.save_pkcs1('PEM'))
with open('keys/privateKey.pem', 'wb') as p:
p.write(privateKey.save_pkcs1('PEM'))
defloadKeys(self):
with open('keys/publicKey.pem', 'rb') as p:
publicKey = rsa.PublicKey.load_pkcs1(p.read())
with open('keys/privateKey.pem', 'rb') as p:
privateKey = rsa.PrivateKey.load_pkcs1(p.read())
returnprivateKey, publicKey
def encrypt(self, message, key):
returnrsa.encrypt(message.encode('ascii'), key)
def decrypt(self, ciphertext, key):try:
returnrsa.decrypt(ciphertext, key).decode('ascii')except:
return False
def decrypt(self, ciphertext, key):try:
returnrsa.decrypt(ciphertext, key).decode('ascii')except:
return False
# rs = RSAEncryption()
# # rs.generateKeys()
# private, public = rs.loadKeys()
# print(f'private {private} \n public {public}')
# cipher_text = rs.encrypt('1995' , public)
# print('message encrypted succesfully')
# message = rs.decrypt(cipher_text, private)
# print(f'original_message {message}')
```

layout.kivy file

```
ScreenController:
LoginScreen:
MainScreen:
MedicalDetailScreen:
<LoginScreen>
name:'login_screen'
MDScreen:
MDCard:
size_hint: .80, .8
pos_hint:{'center_x':0.5 , 'center_y':0.5}
orientation: 'vertical'
elevation:10
padding:30
```

```
spacing:8
MDLabel:
text: "CRYPTOMEDIC ENCRYPTION SYSTEM"
            font_size:25
            # text_color:app.theme_cls.accent_color
            theme_text_color:"Primary"
halign:'center'
size_hint_y: None
height:self.texture_size[1]
MDLabel:
text: "(RSA --- CTR PARALLEL ENCRYPTION MODE)"
            font_size:18
text_color:app.theme_cls.accent_color
            # theme_text_color:"Primary"
halign:'center'
size_hint_y: None
height:self.texture_size[1]
                Widget:
size_hint_y:None
height:40
MDTextField:
hint_text:'Enter ID Code'
icon_left:"account"
padding:15
normal_color:app.theme_cls.accent_color
line_color_normal:app.theme_cls.accent_color
color_mode:"custom"
            size_hint_x:.70
pos_hint:{'center_x':.50}
MDTextField:
hint_text:'Enter Access Code'
icon_left:"key-variant"
hint_color:app.theme_cls.accent_color
password:True
padding:15
normal_color:app.theme_cls.accent_color
line_color_normal:app.theme_cls.accent_color
            size_hint_x:.70
pos_hint:{'center_x':.50}
MDCard:
orientation:'horizontal'
spacing:10
padding:10
pos_hint:{'center_x':.50}
size_hint:None, None
height:80
width: 180
```

```
MDRectangleFlatButton:
text: 'Doctor'
                # text_color:app.theme_cls.accent_color
                # theme_text_color:'Primary'
                # md_bg_color: app.theme_cls.primary_light
pos_hint: {'center_x': 0.5,'center_y': 0.5}
on_release:app.root.current='main_screen'
MDRectangleFlatButton:
text: 'Patient'
                # text_color:app.theme_cls.accent_color
                # theme_text_color:'Primary'
                # md_bg_color: app.theme_cls.primary_light
pos_hint: {'center_x': 0.5,'center_y': 0.5}
on_release:app.root.current='detail_screen'
            # MDRectangleFlatButton:
            #     text: 'Cancel'
            #     # text_color:app.theme_cls.accent_color
            #     # theme_text_color:'Primary'
            #     # md_bg_color: app.theme_cls.primary_light
            #     pos_hint: {'center_x': 0.5,'center_y': 0.5}
<MainScreen>
name:"main_screen"
MDCard:
id:layout
orientation:'vertical'
pos_hint:{"center_x":.5 , "center_y":.5}
      size_hint:1,1
padding:10
      # MDDataTable:
      #     id:table
MDCard:
orientation:'vertical'
pos_hint:{"center_x":.5}
         # md_bg_color:app.theme_cls.primary_light size_hint:.7,1
spacing:50
padding:10
MDRectangleFlatButton:
text: 'CTR-RSA SECURE RECORD'
            # text_color:app.theme_cls.accent_color
            # theme_text_color:'Primary'
            # md_bg_color: app.theme_cls.primary_light
pos_hint: {'center_x': 0.5}
on_release:app.encryption_box()
MDCard:
orientation:'horizontal'
pos_hint:{"center_x":.5}
         # md_bg_color:app.theme_cls.primary_light size_hint:.7,1
```

```
spacing:20
padding:10
MDTextField:
hint_text:'ENTER CTR KEY'
icon_right:"key-variant"
            # mode:"rectangle"
id:text_encryption_key
hint_color:app.theme_cls.accent_color
password:False
padding:15
            max_text_length:16
normal_color:app.theme_cls.accent_color
line_color_normal:app.theme_cls.accent_color
            size_hint_x:.20
pos_hint:{'center_x':.60}
MDTextField:
hint_text:'Search Record'
icon_right:"magnifier"
mode:'rectangle'
hint_color:app.theme_cls.accent_color
password:False
padding:15
normal_color:app.theme_cls.accent_color
line_color_normal:app.theme_cls.accent_color size_hint_x:.80
pos_hint:{'center_x':.50}
<MedicalDetailScreen>
name:'detail_screen'
MDCard:
orientation:'vertical'
pos_hint:{"center_x":.5 , "center_y":.5} size_hint:.8,.8
padding:8
MDLabel:
text:"RECORD PRESAVATION SYSTEM"
size_hint_y: None
height:self.texture_size[1]
        theme_text_color:"Primary"
halign:'center'
        font_size:20
        text_align:'center'
#  ROW 1 MEDICAL ID DATA
MDCard:
id:layout
orientation:'horizontal'
padding:10
spacing:10
#  THE DETATIL VERTICAL SPLITTING.....
MDCard:
```

```
id:layout
orientation:'vertical'
padding:2
spacing:1
            # LEFT SIDE WITH  ID, NAME, AGE AND SEX
            # ============================
FloatLayout:
canvas:
Color:
rgb: 1, 1, 1
                    Ellipse:
pos: 180, 400
size: 100 , 100
source: 'userIcon.png'
angle_start: 0
angle_end: 360
MDCard:
id:layout
orientation:'horizontal'
padding:2
spacing:2
MDLabel:
text:"MEDICAL ID"
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
halign:'center'
                    text_align:'center'
MDLabel:
text:"XXXXXX"
id:label_id
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
halign:'center'
                    text_align:'center'
#  ROW 2 MEDICAL NAME DATA
MDCard:
id:layout
orientation:'horizontal'
padding:2
spacing:2
MDLabel:
text:"NAME"
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
```

```
halign:'center'
                text_align:'center'
MDLabel:
text:"XXXXXX"
id:label_name
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
        text_align:'center'
#  ROW 3 MEDICAL ID DATA
MDCard:
id:layout
orientation:'horizontal'
padding:2
spacing:2
MDLabel:
text:"SEX"
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
                text_align:'center'
MDLabel:
text:"XXXXXX"
size_hint_y: None
id:label_sex
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
        text_align:'center'
#  ROW 4 MEDICAL ID DATA
MDCard:
id:layout
orientation:'horizontal'
padding:2
spacing:2
MDLabel:
text:"AGE"
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
                text_align:'center'
MDLabel:
text:"XXXXXX"
id:label_age
```

```
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
                text_align:'center'
MDCard:
orientation:'horizontal'
spacing:10
padding:10
pos_hint:{'center_x':.50}
size_hint:None, None
height:80
width: 180
MDRectangleFlatButton:
text: 'Logout'
                # text_color:app.theme_cls.accent_color
                # theme_text_color:'Primary'
                # md_bg_color: app.theme_cls.primary_light
pos_hint: {'center_x': 0.5,'center_y': 0.5}
                size_hint:.5, .6
on_release:app.root.current='login_screen'
MDRectangleFlatButton:
text: 'Decrypt'
                # text_color:app.theme_cls.accent_color
                # theme_text_color:'Primary'
                # md_bg_color: app.theme_cls.primary_light
pos_hint: {'center_x': 0.5,'center_y': 0.5}
                size_hint:.5, .6
                # on_release:app.root.current='detail_screen'
on_release: app.decryption_box()
MDTextField:
hint_text:'Enter ID for Decryption File'
icon_right:"magnifier"
id:id_txt_id_file
mode:'rectangle'
hint_color:app.theme_cls.accent_color
password:False
padding:5
normal_color:app.theme_cls.accent_color
line_color_normal:app.theme_cls.accent_color
            size_hint_x:.80
pos_hint:{'center_x':.50}

        Widget:
size_hint_y:None
height:0
    # RIGHT SIDE WITH  ID, NAME, AGE AND SEX
```

```
            # ============================
MDCard:
id:layout
orientation:'vertical'
padding:10
spacing:2
MDCard:
id:layout
orientation:'horizontal'
padding:5
spacing:2
MDLabel:
text:"ANTIVIRAL"
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
halign:'center'
                    text_align:'center'
MDLabel:
text:"XXXXXX"
id:label_antiviral
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
halign:'center'
          text_align:'center'
#  ROW 6 MEDICAL ID DATA
MDCard:
id:layout
orientation:'horizontal'
padding:5
spacing:2
MDLabel:
text:"LIVERBIG"
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
halign:'center'
                    text_align:'center'
MDLabel:
text:"XXXXXX"
id:label_liverbig
size_hint_y: None
height:self.texture_size[1]
                    theme_text_color:"Primary"
halign:'center'
          text_align:'center'
```

```
#  ROW 7 MEDICAL ID DATA
MDCard:
id:layout
orientation:'horizontal'
padding:5
spacing:2
MDLabel:
text:"BILLIRUBIN"
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
                text_align:'center'
MDLabel:
text:"XXXXXX"
id:label_bilirubin
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
                text_align:'center'
MDCard:
id:layout
orientation:'horizontal'
padding:5
spacing:2
MDLabel:
text:"ALBUMIN"
size_hint_y: None
height:self.texture_size[1]
                theme_text_color:"Primary"
halign:'center'
                text_align:'center'
MDLabel:
text:"XXXXXXX
```

Table 2.1: Review of related work

S/N	Author/Year	Methodology	Result	Limitation
1	(Mohammed, *et al*., 2020)	the use of Least Significant Bit and Advance Encryption	A merge approached used in securing or concealing privacy in big Data application was achieved as a results	Multiple image operation on stego image like rotation, resizing and cropping.
2.	(Abouelmehdi, *et al.*, 2017)	An Attribute Based encryption techniques racers controls, and homophobic encryption to investigate security and privacy	The techniques achieve great success in security and privacy in healthcare organization as a results	It shows that the privacy and security issues, and technical challenges was a huge barrier in this research
3.	(Gutiérrez, *et al.*, 2020)	The use of block chain network to secure medical record system	The result shows high efficiency in keeping the EMR'S of individual patients unified.	Medical record are only considered in the paper work
4.	(Ram, *et al.*, 2018)	The use of lake based modernistic privacy preservation techniques	The results shows technique capable of handling privacy preservation in unstructured data	The paper limitation is to design a concreate solution in protecting the privacy of both structure and unstructured data
5	(Abouelmehdi, *et al.*, 2018)	A State of art survey approach on security and privacy challenges	A comprehensive survey security and privacy application in healthcare was achieved.	The researcher only focus on recent proposed method based on anonymization and encryption.
6	(Al-kaabi & Belhaouari, 2019)	A survey approach	A comprehensive survey on various ways to improve RSA algorithm was achieved.	RSA algorithm is only considered in this paper work
7	(Seun, *et al.*, 2019)	A model was design for DPP and HIS using iterative design techniques	The results shows Model for DPP and HIS to address the inadequacy	Medical record is only considered in this paper
8	(Alexandru, *et al*, 2018)	Investigation method was proposed.	The results shows the way big data is noticed by the public and the way it is used in saving human lives	Health record are only considered in the paper work
9	(Seliya, *et al*, 2021)	One class classification, for outliers and novelty detection & deep learning was use in this review.	The result use one- class classification to detect abnormal data points, and can serve to address issues related to severely	Area of application context and its inheritance association problems are been

			imbalanced datasets in Big Data.	omitted.
10	(Abouelmehdi *et al.*, 2018)	Ammonization Encryption method was proposed	The results posed methodology provides data confidentiality and secure data sharing.	It increased complexity, make new models more difficult to interpret.

Table 3.1 Hepatitis Dataset (Pre-processed and cleaned).

ID	NAME	AGE	SEX	ANTIVIRALS	LIVERBIG	BILIRUBIN	ALBUMIN
0	Robyn	30	Female	Yes	No	1.0	1.0
1	Liam	50	Male	Yes	No	0.9	0.9
2	Ethan	78	Male	Yes	Yes	0.7	0.7
3	Robbie	31	Male	No	Yes	0.7	0.7
4	Patrick	34	Male	Yes	Yes	1.0	1.0
5	Charlie	34	Male	Yes	Yes	0.9	0.9
6	Noah	51	Male	Yes	Yes	0.9	1.3722580645161286
7	Robert	23	Male	Yes	yes	1.0	1.0
8	Oscar	39	Male	Yes	yes	0.7	0.7
9	Ross	30	Male	Yes	yes	1.0	1.0
10	Jon	39	Male	No	No	1.3	1.3
11	Nathan	32	Male	No	yes	1.0	1.0
12	Stuart	41	Male	No	yes	0.9	0.9
13	Cooper	30	Male	Yes	yes	2.2	2.2
14	Stuart	47	Male	No	yes	2.2	1.3722580645161286
15	Nathan	38	Male	Yes	yes	2.0	2.0
16	Sam	66	Male	Yes	yes	1.2	1.2
17	Declan	40	Male	Yes	yes	0.6	0.6
18	Stewart	38	Male	Yes	yes	0.7	0.7
19	Kyle	38	Male	No	No	0.7	0.7
20	Anna	22	Female	No	yes	0.9	0.9
21	Michael	27	Male	Yes	No	1.2	1.2
22	Cameron	31	Male	Yes	yes	1.0	1.0
23	Calum	42	Male	Yes	yes	0.9	0.9
24	Yvonne	25	Female	No	yes	0.4	0.4
25	Martin	27	Male	Yes	yes	0.8	0.8

26	Alistair	49	Male	No	yes	0.6	0.6
27	Courtney	58	Female	Yes	yes	1.4	1.4
28	Alistair	61	Male	Yes	No	1.3	1.3
29	Brandon	51	Male	No	yes	1.0	1.0
30	Kevin	39	Male	No	yes	2.3	2.3
31	Rory	62	Male	Yes	yes	1.0	1.0
32	Rachel	41	Female	No	yes	0.7	0.7
33	Ruth	26	Female	Yes	yes	0.5	0.5
34	Alexander	35	Male	Yes	yes	0.9	0.9
35	Mark	37	Male	Yes	yes	0.6	0.6
36	Daniel	23	Male	Yes	yes	1.3	1.3
37	Victoria	20	Female	Yes	No	2.3	2.3
38	Owen	42	Male	Yes	yes	1.0	1.0
39	John	65	Male	Yes	yes	0.3	0.3
40	Calum	52	Male	No	yes	0.7	0.7
41	Cameron	23	Male	Yes	yes	4.6	4.6
42	Nicholas	33	Male	Yes	yes	1.0	1.0
43	Jake	56	Male	Yes	yes	0.7	0.7
44	Fraser	34	Male	Yes	yes	0.7	1.3722580645161286
45	Harry	28	Male	Yes	yes	0.7	0.7
46	Andrew	37	Male	Yes	yes	0.6	0.6
47	Elizabeth	28	Female	Yes	yes	1.8	1.8
48	Adam	36	Male	Yes	yes	0.8	0.8
49	Owen	38	Male	No	yes	0.7	0.7
50	Charlie	39	Male	Yes	yes	0.9	0.9
51	David	39	Male	Yes	yes	1.0	1.0
52	Daniel	44	Male	Yes	yes	0.6	0.6
53	Jonathan	40	Male	No	yes	1.2	1.2
54	Max	30	Male	Yes	yes	0.7	0.7

55	Leo	37	Male	Yes	yes	0.8	0.8
56	Mark	34	Male	Yes	yes	0.8	1.3722580645161286
57	Christopher	30	Male	No	yes	0.7	0.7
58	Luke	64	Male	No	No	1.0	1.0
59	Beth	45	Female	Yes	yes	1.0	1.0
60	Angus	37	Male	Yes	yes	0.7	0.7
61	Mohammed	32	Male	Yes	yes	0.7	0.7
62	Alex	32	Male	Yes	yes	3.5	3.5
63	Gordon	36	Male	Yes	No	0.7	0.7
64	Kyle	49	Male	Yes	yes	0.8	0.8
65	Blair	27	Male	Yes	yes	0.8	0.8
66	Gerald	56	Male	Yes	yes	0.7	0.7
67	Callum	57	Male	Yes	yes	4.1	4.1
68	Jude	39	Male	Yes	yes	1.0	1.0
69	Evan	44	Male	Yes	yes	1.6	1.6
70	Dale	24	Male	Yes	yes	0.8	0.8
71	Gregor	34	Male	Yes	No	2.8	2.8
72	William	51	Male	Yes	No	0.9	0.9
73	Robbie	36	Male	Yes	yes	1.0	1.0
74	Bradley	50	Male	Yes	yes	1.5	1.5
75	George	32	Male	No	yes	1.0	1.0
76	Roy	58	Male	Yes	No	2.0	2.0
77	Jacqueline	34	Female	No	yes	0.6	0.6
78	Grayson	34	Male	Yes	No	1.0	1.0
79	Murray	28	Male	Yes	yes	0.7	0.7
80	Charles	23	Male	Yes	yes	0.8	0.8
81	Cameron	36	Male	Yes	yes	0.7	0.7
82	Leo	30	Male	Yes	yes	0.7	0.7
83	Karen	67	Female	Yes	yes	1.5	1.5

84	Thea	62	Female	Yes	yes	1.3	1.3
85	Luca	28	Male	Yes	yes	1.6	1.6
86	Ryan	44	Male	Yes	yes	0.9	0.9
87	Samuel	30	Male	Yes	yes	2.5	2.5
88	Murray	38	Male	Yes	yes	1.2	1.2
89	Liam	38	Male	Yes	No	0.6	0.6
90	Heather	50	Female	Yes	No	0.9	0.9
91	Douglas	42	Male	Yes	yes	4.6	4.6
92	Mitchell	33	Male	Yes	yes	1.0	1.0
93	Justin	52	Male	Yes	yes	1.5	1.5
94	Alistair	59	Male	Yes	yes	1.5	1.5
95	Jake	40	Male	No	No	0.6	0.6
96	Nathan	30	Male	Yes	yes	0.8	0.8
97	Steven	44	Male	Yes	No	3.0	3.0
98	Jonathan	47	Male	Yes	yes	2.0	2.0
99	Christopher	60	Male	Yes	No	2.0	1.3722580645161286
100	John	48	Male	Yes	yes	4.8	4.8
101	Dean	22	Male	Yes	yes	0.7	0.7
102	Malcolm	27	Male	Yes	yes	2.4	2.4
103	Euan	51	Male	Yes	yes	4.6	4.6
104	Kai	47	Male	Yes	yes	1.7	1.7
105	Fraser	25	Male	Yes	yes	0.6	0.6
106	Samuel	35	Male	Yes	yes	1.5	1.5
107	Dylan	45	Male	Yes	yes	2.3	2.3
108	Rory	54	Male	No	No	1.0	1.0
109	Lucas	33	Male	Yes	yes	0.7	0.7
110	Francis	7	Male	Yes	yes	0.7	0.7
111	Robbie	42	Male	No	yes	0.5	0.5
112	Charlie	52	Male	Yes	yes	1.0	1.0

113	Francis	45	Male	Yes	yes	1.2	1.2
114	Barry	36	Male	Yes	yes	1.1	1.1
115	Katie	69	Female	Yes	yes	3.2	3.2
116	Donald	24	Male	Yes	yes	1.0	1.0
117	Euan	50	Male	Yes	yes	1.0	1.0
118	Wayne	61	Male	Yes	yes	1.0	1.3722580645161286
119	Cameron	54	Male	Yes	No	3.2	3.2
120	Lee	56	Male	Yes	No	2.9	2.9
121	Cole	20	Male	Yes	yes	1.0	1.0
122	Matthew	42	Male	Yes	yes	1.5	1.5
123	Shaun	37	Male	Yes	yes	0.9	0.9
124	Robert	50	Male	Yes	yes	1.0	1.0
125	Taylor	34	Female	Yes	No	0.7	0.7
126	Ben	28	Male	Yes	No	1.0	1.0
127	Calvin	50	Male	Yes	yes	2.8	2.8
128	John	54	Male	Yes	yes	1.2	1.2
129	Calum	57	Male	Yes	yes	4.6	4.6
130	Derek	54	Male	Yes	yes	1.0	1.0
131	James	31	Male	Yes	yes	8.0	8.0
132	Joshua	48	Male	Yes	yes	2.0	2.0
133	Luke	72	Male	No	yes	1.0	1.0
134	Jackson	38	Male	Yes	yes	0.4	0.4
135	Charles	25	Male	Yes	No	1.3	1.3
136	Mason	51	Male	Yes	No	0.8	0.8
137	Ross	38	Male	Yes	yes	1.6	1.6
138	Sam	47	Male	Yes	yes	1.0	1.0
139	Kai	45	Male	No	yes	1.3	1.3
140	Stuart	36	Male	Yes	No	1.7	1.7
141	Thomas	54	Male	Yes	No	3.9	3.9

142	Finn	51	Male	Yes	yes	1.0	1.0
143	Owen	49	Male	Yes	yes	1.4	1.4
144	Jay	45	Male	Yes	yes	1.9	1.9
145	Cole	31	Male	Yes	yes	1.2	1.2
146	Graeme	41	Male	Yes	yes	4.2	4.2
147	Greig	70	Male	Yes	yes	1.7	1.7
148	Jude	20	Male	Yes	yes	0.9	0.9
149	Martin	36	Male	Yes	yes	0.6	0.6
150	Mason	46	Male	Yes	yes	7.6	7.6
151	Connor	44	Male	Yes	yes	0.9	0.9
152	George	61	Male	Yes	No	0.8	0.8
153	Jennifer	53	Female	Yes	yes	1.5	1.5
154	Bruce	43	Male	Yes	yes	1.2	1.2