scientific reports

Check for updates

OPEN

Integrating advanced neural network architectures with privacy enhanced encryption for secure and intelligent healthcare analytics

C. Ramesh Babu Durai[1], S. Dhanasekaran[2]✉, M. Jamuna Rani[3] & Sindhu Chandra Sekharan[4]

Healthcare data protection in our mutually connected era has emerged as an issue of serious concern with private patient information, which has been exposed more often due to data violations and cyber-attacks. Network structures CNN and LSTM as part of privacy-based encryption method. Research presents neurosis, a new structure, which combines CNN-LSTM architecture with privacy-secured encryption to provide safe healthcare analytics. Depending on the Kaggle healthcare dataset, the model receives an accuracy of 98.73%, which is better than the current functioning. "NeuroShield" includes characteristic-based access control (ABAC), Advanced Encryption Standard (AES), Multi-Factor Authentication (MFA) and differential privacy-based optimizations that provide strong protection. To increase the interpretation, AI (XAI) is used on the basis of size, making health experts capable of understanding model decisions. Detailed evaluation accepts the performance of structure in maintaining privacy through providing high-demonstration analysis for healthcare data protection. Organized testing and comparative analysis suggest that neuroshield not only improves data security, but also provides excellent accuracy with better performing results in healthcare analytics.

Keywords NeuroShield model, Healthcare data analytics, Convolutional neural networks, Long short-term memory, Advanced encryption standard, Attribute-based access control, Multi-factor authentication

The medical system worldwide becomes more digital and adopts new technology, safety and privacy of medical records of patients is a major concern. In the case of sensitive patient data, medical records, and safety of health care systems from cyber-attacks, data violations and unauthorized access, health safety[1] is a broad array of technology, processes and policies. Digitization of Electronic Health Records (EHRS), connected medical devices, telemedicine systems and patient data has promoted an increase in both healthcare data volume and sophistication. The risk of data violations, identity theft, and unauthorized access to confidential health data highlights the paramount importance of strong health care data security solutions[2].

In relation to data security, unique problems and challenges in the healthcare sector face. Both individually identified information (PII) and protected health information (Phi) are strictly governed, which provide highly sensitive to health data. Data includes patient demographics, medical history and treatment. Protected health information is defined as any individually identified health information under HIPAA that a covered unit or its professional partner[3–5]. Medical history, billing information, insurance coverage, and any other data can be used to identify a person's health status or healthcare provision.

Financial loss, damage to a person's reputation, potential legal consequences, and compromised patient care are some consequences as a result of health care data violations. Similarly, the law for protecting healthcare data has also been applied in other countries and geographical areas, such as the European Union General Data Protection Regulation (GDPR) and Canada's Personal Information Protection and Electronic Documents Act (Pipeda), both Are equal. Hipaa in America. Encryption, access control, audit trails, and brech reporting procedures are among the security of healthcare organizations, providers, insurers and others by these rules that handle the patient's information[6–8]. Emerging cyber threats, changing regulatory requirements, and technology progress all lead to a constant developed landscape in which healthcare data security must be developed. When

[1]Kings Engineering College, Chennai, India. [2]Department of Electronics and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore 641202, India. [3]Department of Electronics and Communication Engineering, Sona College of Technology, Salem, India. [4]Department of Computing Technologies, School of Computing, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu 603203, India. ✉email: dhanselvaraj@gmail.com

it comes to the protection of confidential patient data, Tri-end-Tru classics such as firewalls, anti-virus software and circumference security are not enough[9,10]. During the data cycle, healthcare organizations and stakeholders need to apply preventive, investigative and reactionary solutions to maintain data security. Such a strategy should be multi -level and integrated[11].

Protecting sensitive medical information from hackers and other bad actors is the primary goal of preventative measures. One of the best ways to secure data while it's in use, in transit, or at rest is to utilize encryption. Healthcare organizations can protect the confidentiality and integrity of sensitive data by encrypting it using cryptographic techniques[12,13]. This renders the data unreadable to unauthorized users. The same is true for healthcare data: authentication methods, user permissions, and access controls all work together to ensure that only authorized individuals can see or change sensitive information in accordance with the concept of least privilege[14–16]. Healthcare systems and networks undergo audits and monitoring as a proactive measure to detect any unusual activity, attempted intrusion, or suspicious activity. A lot of healthcare companies utilize security information and event management (SIEM) platforms, intrusion prevention systems (IPS), and intrusion detection systems (IDS) to monitor for and react to security incidents in real-time. This helps them find threats early on and stop them before they get worse. By keeping meticulous records of all operations pertaining to healthcare data, strong logging and auditing procedures also assist with forensic investigations and meeting regulatory obligations. Figure 1 illustrates the healthcare data security challenges.

Quickly restoring regular operations, limiting downtime, and mitigating the effects of security incidents and breaches are the primary goals of responsive measures. The capacity of a healthcare organization to react promptly to security incidents, natural disasters, and other disruptive situations relies on the incident response plans, disaster recovery processes, and contingency preparations. Incident response teams should be equipped, with exercises and training, to manage security incidents properly and efficiently. This will minimize the potential effect on patient care and corporate reputation[17]. New answers to long-standing issues are emerging in the healthcare data security landscape, the commonality of the latest technology such as blockchain, machine learning (ML), and artificial intelligence (AI). An example is the capacity of AI and ML algorithms to analyze huge amounts of healthcare data in real-time. This enables proactive threat detection and response by recognizing trends, anomalies, and potential vulnerabilities. By enabling secure, tamper-resistant transactions and data exchange between stakeholders, blockchain technology can enhance data integrity, transparency, and trust in healthcare[18,19].

Healthcare data security has evolved a great deal, but there remains much to be tackled. Establishing robust security protocols end-to-end across healthcare firms is difficult owing to various reasons like the intricacy of healthcare ecosystems, interoperability issues, aged systems, and scarce resources. Moreover, it is also important to remain watchful, adapt, and spend money on cybersecurity infrastructure and expertise since cyber-attacks, such as phishing schemes, insider threats, and ransomware attacks, keep evolving. Health care data protection is vital in order to secure patients' private data, sustain trust in the health care systems, and ensure the reliability of the health care services. Healthcare businesses may reduce the likelihood of cyber threats, data breaches, and compliance violations by implementing a thorough, risk-based data security strategy that combines technical controls, regulatory compliance, and organizational governance. The healthcare industry can benefit patients, providers, and stakeholders by embracing innovation and collaborating. By leveraging emerging technology and best practices, they can solve increasing security concerns and develop a robust, secure healthcare ecosystem[20].

The impetus for this activity comes from the growing complexity and sensitivity of healthcare information, in addition to the inability of legacy security solutions to manage changing cyber threats. Modern techniques of healthcare data protection usually fail to provide simultaneous provision of dynamic adaptation of threats, effective data handling, and the preservation of privacy. Past literature has emphasized data encryption or machine learning models separately, with relatively few having used a complete end-to-end holistic approach combining

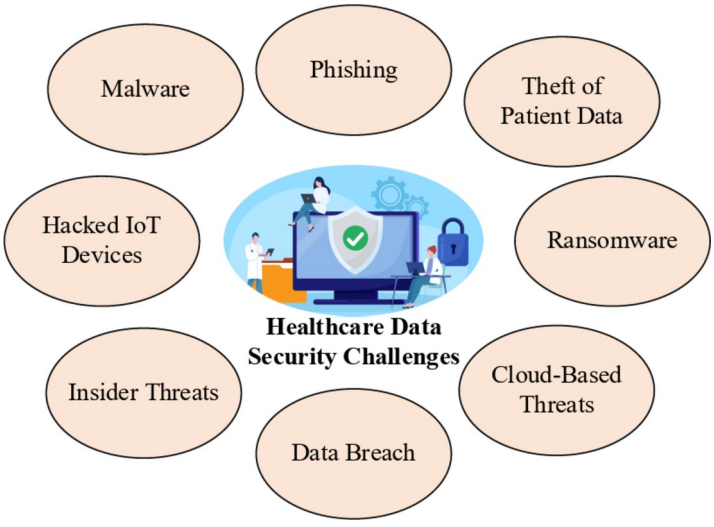

Fig. 1. Healthcare data security challenges.

complex neural networks and optimization algorithms tailored for healthcare analytics. This work seeks to bridge these gaps by introducing a new framework that not only strengthens data security but also improves interpretability and efficiency in healthcare data analysis. Classic models such as Logistic Regression, Decision Trees, and SVMs are usually incapable of capturing the intricate relationships between high-dimensional and sequential healthcare data. Though such models may work well for some tasks, they usually entail a great deal of feature engineering and do not support spatial or temporal data natively. In contrast, more advanced models such as Transformers, though formidable with sequential data, could be computationally infeasible in environments where computational power is scarce, such as in most healthcare environments. In contrast, the integration of CNNS and LSTMs provides uniform-handed approach that is capable of effectively dealing with both spatial and temporary characteristics of health data without consuming too much more computational power.

The neuroshield models combine the CNNS and LSTM network as they complement each other's strength in dealing with the complex nature of healthcare data. This architecture was preferred for other potential models because healthcare data contains both spatial and cosmic characteristics that demand a strong analytical method. CNNs were chosen due to their established capacity in extracting and processing spatial features from data such as medical images, sensor data and structured patient records. For example, in medical imaging, the edges, textures and complex structures such as spatial patterns may be important for diagnosis. CNNS employs conversional filters to directly learn these features from data without requirement for manual extraction of features. Unlike other machine learning models such as support vector machine (SVM) or Random Forest (RF), which depend on pre-specified features, CNN can learn the most suitable features for work by itself, increase the accuracy and credibility of analysis. The LSTM network was chosen to identify the temporary dependence inherent in health care data. Most healthcare dataset consists of time-series data, e.g., patient significant signs, treatment history and pharmaceutical regime. LSTMs are specially engineer for sequence data and have the ability to learn long -term dependence using their special gating mechanisms. This ability enables them to remember longer information, which is important in healthcare applications where it is important.

This model is designed to catch the spatial and cosmic characteristics inherent in health care data efficiently, which facilitates more accurate and meaningful analysis of patient records, medical images and time-series data. In addition, the framework includes advanced data preprocessing methods, such as the K-nearest neighbor (KNN) copy, to manage the missing values in the healthcare dataset without compromising on data integrity. Furthermore, robust encryption techniques like AES are employed for securing sensitive health data and maintaining patient privacy as per privacy laws. Furthermore, the framework proposes that access controls should be regulated using ABAC policies and MFAs to govern data access as a function of user role and privilege. This integrates data protection through the prevention of unauthorized access to health data and protecting against data violations. Finaly framework employs differential privacy-based adaptation techniques to ensure the privacy of individuals by adapting machine learning models for sensitive health care data. Through such privacy-protection techniques, framework guarantees analytical insight can be derived from health care data without cheating the patient's privacy or breaching privacy policies. The neurocardiac is separated from others in that it integrates a hybrid CNN-LSTM model to facilitate enhanced spatial-temporal analysis and privacy-protection adaptation for safeguarding sensitive health care data. This involves additional elucidation of AI (XAI) to enable additional interpretation, which provides more transparent model decisions for health care professionals.

Main contributions of the work

- Introduction of the NeuroShield Model, a novel framework integrating LSTM networks with Cascaded CNNs, designed to enhance the analysis of healthcare data by effectively capturing spatial and temporal features.
- Implementation of KNN imputation technique to handle missing values in the healthcare dataset, ensuring data completeness and reliability for subsequent analysis.
- Deployment of AES for data encryption and access control, ensuring robust protection of sensitive healthcare information during storage and transmission, thereby upholding patient privacy and compliance with regulatory standards.
- Implementation of ABAC Policy and MFA mechanisms to regulate data access based on user roles and permissions, thereby strengthening data security measures and preventing unauthorized access to healthcare data.
- NeuroShield incorporates a hybrid CNN-LSTM structure for improved spatial-temporal analysis, employs privacy-preserving optimization for the protection of healthcare data, and uses Explainable AI (XAI) with SHAP to enhance interpretability and transparency for clinicians.

The structure of the paper unfolds as follows: Sect. 2 delves into prior research on Healthcare Data Analytics and Classification. In Sect. 3, we provide a detailed exposition of the proposed NeuroShield Model, which leverages the synergies between LSTM networks and Cascaded CNNs. Subsequently, Sect. 4 showcases the results gleaned from extensive testing and comparative analyses. Lastly, Sect. 5 encapsulates our findings and delineates potential avenues for future research endeavors.

Related work

AI technology offers the greatest answer for enhancing data security and dependability, making it the ideal choice for healthcare applications. This is why traditional constructions for the IoT-cloud architecture incorporate a number of security measures based on artificial intelligence. Nevertheless, it faces major challenges such as increased time consumption, higher costs associated with Internet of Things (IoT) sensors, inefficient data handling, and an increase in algorithm design complexity. It is also not ideal for processing unstructured data. For that reason, a probabilistic super learning–RH intelligent feature learning method is presented based on artificial

intelligence to strengthen the safety of healthcare data kept in the IoT cloud[21]. This article's suggested learning model is also an attempt to lower the price of Internet of Things (IoT) sensors. In this case, the training model is kept running to detect assaults early, updating the reported attack attributes to learn their characteristics. Elliptic Curve Cryptography (ECC) serves as a common method for securing data, employing the hash value of the data matrix to generate a random key. Subsequently, the improved ECC-RH method encrypts and decrypts the data using the newly-generated random hash key. Various performance indicators are utilized during performance evaluation to validate and compare the outcomes of both current and suggested methodologies.

Connecting medical devices and their accompanying software to the computer networks utilized in healthcare 5.0 is the groundbreaking Internet of Medical Things (IoMT). The rapid development of smart medical devices on IoMT platforms has greatly enhanced the adoption of important technologies, modernizing healthcare practices, illness management, and patient treatment standards. Data screening, data interchange, patient monitoring, data collection and analysis, and sanitary hospital attention are just a few of the cloud-based services offered by the IoMT. It is the job of wireless sensor networks (WSNs) to collect and transmit data. The healthcare industry places a premium on patient safety and respects their right to privacy. The wireless transmission of data from these smart devices through the airways allows anyone to access and edit the patient's medical records. An innovative protocol was developed, ECC-EERP, based on elliptic curve cryptography, to meet the need for a secure and energy-efficient system in healthcare[22]. Data was encrypted using the ECC-EERP key-based approach. By encrypting and decrypting web traffic using pairs of public and private keys, a WSN's overall energy consumption is reduced. The proposed method's efficiency was compared with a number of current approaches. The proposed approach was assessed using a wide range of metrics, including safety, encryption speed, power consumption, network lifetime, communication congestion, processing time, and implementation expense. The findings show that the suggested method improves both safety and efficiency in terms of energy usage.

Particularly difficult is the task of protecting sensitive healthcare data. Due to their high number of patients and direct access to patient records, nursing staff play a vital role in ensuring the security of sensitive medical information. Although the connection between healthcare data protection and information security culture (ISC) is not yet fully understood, the former plays a significant role in the latter. Two additional aspects of organizational ISC pertaining to privacy and security are initially defined and made practical[23]. A survey was conducted among 527 nursing staff in Slovenia to verify the assessment instrument and also investigated any connections between the newly created ISC characteristics and nursing employees' illicit access to healthcare data, using the theory of planned behavior (TPB) as an explanation. A confirmatory factor analysis followed an exploratory one to ensure the reliability of the measuring tool. The newly constructed ISC dimensions are reliable and have sufficient validity, according to both evaluations. According to the PLS-SEM analysis findings, there is a negative correlation between privacy-oriented ISC and attitude towards conduct, and security-oriented ISC with subjective norms and normative views. They also show that TPB provides a good explanation for why certain people get illegal access to patients' medical records. Thus, our study's findings suggest a tangential relationship between ISC and healthcare data breaches. Nursing personnel can only achieve proper practical implementation of ethical principles, such as privacy-preserving conduct, through awareness training. Awareness therapies that target nursing staff members' values and principles have the potential to improve their outlook[23].

A revolutionary change in healthcare is underway, propelled by technology advancements, with the emergence of Patient-Generated Health Data (PGHD). With the rise of PGHD and innovations like home monitoring systems and wearable devices, data gathering may now happen outside of traditional healthcare settings, allowing for constant tracking and patient participation in their own healthcare management. Despite its increasing use, stakeholders are confused about what PGHD means and have concerns about the accuracy, privacy, and security of their data. By looking at its history, different types, technical underpinnings, and problems, particularly with regard to privacy and security rules, this provides a comprehensive overview and explanation of PGHD[24]. This review provides a holistic view of PGHD's present status and future prospects by highlighting the field's contributions to healthcare reform via patient-centric approaches, their comprehension, and individualized treatment; it also delves into new technology and tackles data privacy and security concerns. This methodical approach covers all the existing literature on PGHD in a thorough and organized manner, paying close attention to the many areas mentioned in the aim. In addition, existing articles were used from the area of PGHD to answer the fourth RQ which is focused on the future and was not addressed in the previous study.

A developing technology in cyberspace, the metaverse has recently come to the forefront of global attention. There is much promise in the metaverse for delivering a wide range of health services in an immersive environment, both for patients and doctors. In order to create more efficient, safer, and more lifelike virtual healthcare facilities in the metaverse, provided combined AI with blockchain technology[25]. There are three distinct settings: the one in which the doctor works, the one in which the patient is located, and the metaverse. Blockchain technology facilitates interactions between physicians and patients in a metaverse setting while also guaranteeing the confidentiality, integrity, and availability of patient data. This design is mostly based on the metaverse environment. Avatars serve as representations in the metaverse, and physicians, patients, and nurses can access this environment by registering on the blockchain. The doctor-patient consultation was documented and collect, send, and store all relevant data on the blockchain. Models powered by explainable artificial intelligence (XAI) utilize these datasets to forecast and diagnose diseases. The XAI GradCAM and LIME methods ensure trustworthiness, explainability, interpretability, and transparency in illness prediction and diagnosis by providing a logical rationale for the process. Blockchain technology protects patients' data, ensuring its transparency, traceability, and immutability. Because of these blockchain capabilities, patients may be certain that their data is secure.

The Internet of Medical Things (IOMT) has a fairly advanced healthcare, especially in the decentralized communication systems for collecting and monitoring patient data. The machine learning algorithm is employed to assess the patient risk scores based on various factors, which supports healthcare providers in Kovid – 19 Care and follow -up, where data privacy is a major concern. Dasharatha et al.[26] study discovers a federated learning (FL) to integrate blockchain technology (BT) to increase safety and decentralization. The discovery of a delivery of a distributed reinforcement within the multi-disciplinary system. Data is collected from IOMT applications, which ensures the monitoring of the patient without relying on intermediate dependence. The proposed blockchain-competent reinforcement FL model improves clinical monitoring, facilitates safe communication, and strengthens data confidentiality by maintaining efficiency and scalability in distributed environment. Results show that the approach acquires high reliability, improving the existing model in future stating accuracy and safety measures. However, borders include computational complexity to integrate FL with blockchain, real -time applications include challenges in ensuring spontaneous differences in diverse IOMT platforms, and need to adopt widespread adoption in healthcare and require adaptation.

While the studies discussed offer valuable insights and propose innovative solutions to address various challenges in healthcare data security and privacy, they also exhibit certain limitations. To begin with, most of the suggested solutions are based on sophisticated technologies like artificial intelligence (AI), blockchain, and Internet of Things (IoT), which can be challenging to implement in actual healthcare environments because of resource limitations and technical complexities. Moreover, the research tends to concentrate on isolated areas of healthcare data security, like encryption and access control, without addressing the larger picture of regulatory compliance and ethical issues. Additionally, scalability and efficacy of suggested solutions may vary in various healthcare settings and infrastructure and thus requires further tests and fittings in different contexts. Additionally, research focuses mainly on technology-based solutions without considering aspects of organizational culture, training and governance that play an important role in ensuring overall data security measures. Finally, technology and medical practice development requires continuous updates and modifications to combat new threats and weaknesses, underlining the faster speed of development and medical practice development.

Methodology

The approach used in this research is a multi-dimensional that focuses on improving healthcare data analytics, ensuring safety and privacy. First, the neuroshield model is presented, a combination of LSTM network with cascade CNN to effectively catch spatial and temporary features in health care data. Missing data treatment is controlled by KNN copy of KNN to complete data. Strong encryption techniques like the AES are employed to protect sensitive healthcare data, supported by access control processes like ABAC policies and MFA to manage data access. Additionally, Differential Privacy-based Optimization algorithms are used to maintain individual privacy during model training over sensitive healthcare data. The overall methodology supports strong data analysis with the protection of patient confidentiality and compliance with privacy laws. Figure 2 illustrates the architecture of the proposed model.

Dataset collection

The model is trained with a healthcare dataset containing synthetic patient records, consisting of variables ranging from demographics and medical history to treatment information. With 10,000 observations amounting to synthetic patient healthcare records, the Healthcare Dataset available on Kaggle is a dataset paradise[27]. Patient demographic information, medical history, and admission details and so on, are some of the variables used. For the purpose of health-related analysis and modeling, the large dataset proved to be useful. Researchers can enhance health care delivery and patient care procedures by delving into patient characteristics, diseases, interventions, and results. Medical care organizations employ the dataset for augmented strategic planning and quality enhancement programs by learning from admission patterns, insurance cover, and practitioner performance, to mention a few. Predictions of patient admissions, billing levels estimation, and disease outcomes projections become simpler to advanced statistical instruments such as predictive modeling. This subsequently results in improved provision of healthcare services. To ensure the privacy and confidentiality of patients during analysis, it is essential to responsibly manage data by conducting adequate preparation and adhering to all the relevant laws.

Data preprocessing and cleaning

Anonymize personally identifiable information (PII)

To ensure patient privacy, individually identifying information (PII) is unknown through methods such as pseudo naming and data normalization. For example, the patient's names are replaced with unique identifiers, and the exact date of birth is normalized in age.

When scrubbing the healthcare database of such information, one must take into account the sensitive nature of individually identified information (PII). Personally identifiable information (PII) has a wide range, which can be employed to identify a person, such as names, addresses, social security numbers and medical records. To meet the requirements such as HIPAA patient privacy by organizations must be maintained, which ask for the complete or effective neutrality of individual identified data. To achieve this balance between preserving data utility and security of privacy, methods of approval are paramount. Such methods preserve the analytical utility of the dataset by using pseudonym or by normalizing the data. One of the ways to protect the privacy of patients allowing effective data analysis is to give them a special identity rather than using their name directly. Similarly, the exact dates of birth can be replaced with more common age groups to hide the identity of people without abandoning the utility of the dataset. This systematic process of preserving healthcare data in healthcare, beyond patient privacy, regulatory in healthcare and research in analysis, is meaningful for modeling and analysis.

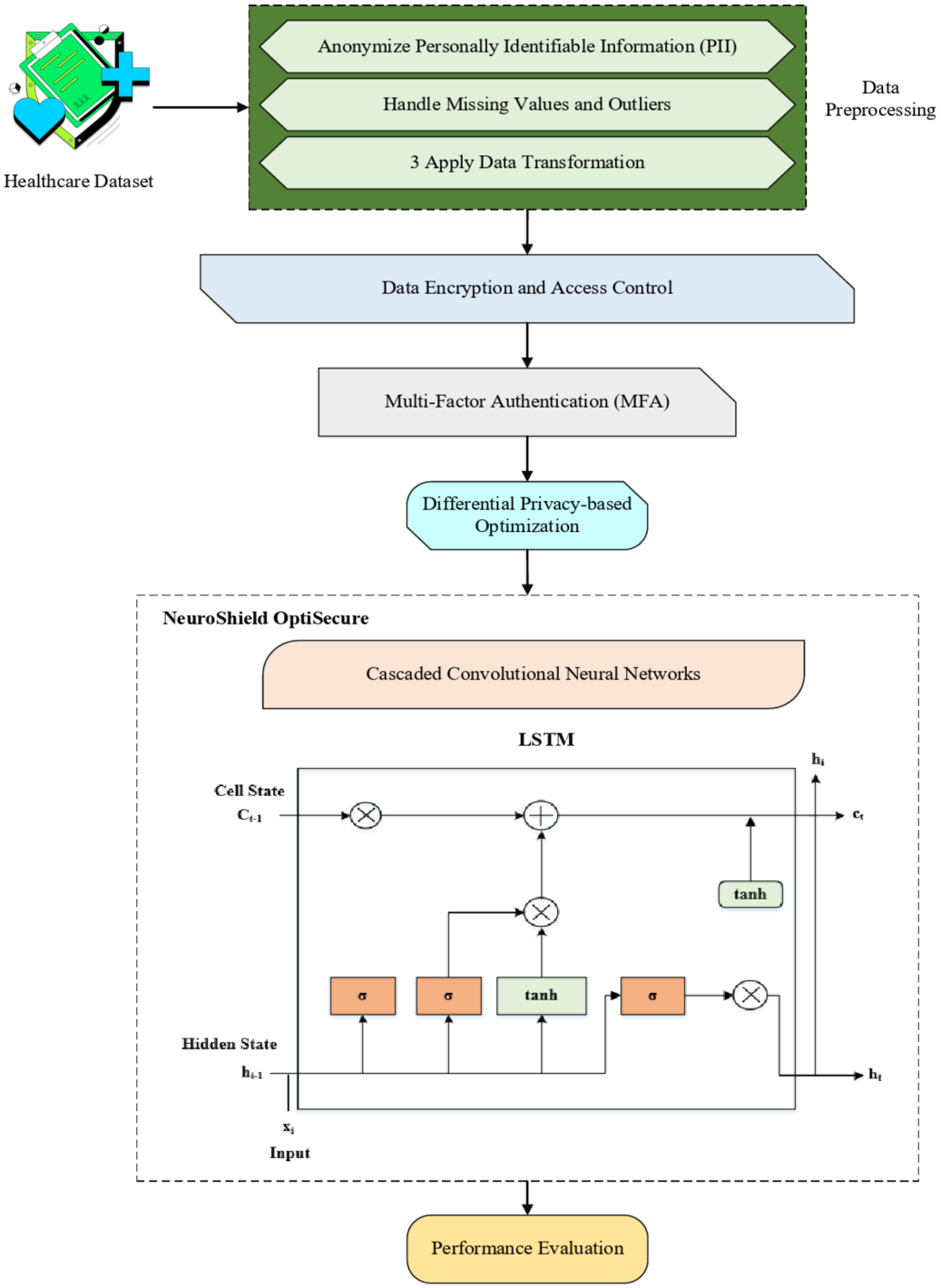

Fig. 2. Architecture of proposed model.

Handle missing values and outliers

In healthcare data analysis, the accuracy and reliability of analytical models depends on the outlier and elimination of missing values. The absence of action on the outlier, most comments and observation away from missing values (usually as a result of data collection or errors in incomplete data) can have a major impact on the performance and accuracy of analytical models. The KNN imperfection method is applied to change the missing values. This technique estimates the missing data points based on the values of the nearest neighbors in the dataset, ensuring data perfection without loss of integrity. To apply missing values and preserve dataset, KNN employs similarities between data examples. Similarly, the outlier must be treated with excessive attention to not slant the results of the analysis. Not treating these unusual data points properly can slant statistical projections and confuse the model interpretations. Various techniques can be employed in outliers handling. Statistical techniques identify the outliers and solve them using functionalization such as historicization to reduce their

impact on the model. Data is done to map numerical characteristics for a uniform range, which increases the stability of nerve network training.

In the management of the healthcare dataset, which contain outliers, historicization is a significant method to minimize their influence while preserving sample distribution. There are outlier observations that are very dissimilar from other remarks, and they can bias figures and model outcomes. Winsorization offers an effective solution by minimizing the effect of the outlier without entirely removing the extreme values at a certain percentage. Winsorization keeps outliers from adding uneven influence on overall distribution by setting the upper and lower boundaries in terms of percentage requirements. By tapering the effect of the outliers, this process keeps the dataset representative of the vast population. Through Winsorization, analysts can keep the accuracy and reliability of the dataset intact while reaching meaningful insights and while making well-informed with healthcare data.

Apply data transformation

Raw data must be subjected to data modification procedures in order to analyse and model healthcare data effectively. These procedures are crucial for normalising data structures, improving the quality of analytical models, and guaranteeing the operation of machine learning algorithms. Obtaining numerical properties within a broad range, like 0 and 1 or between 0 and standard deviation 1, is the goal of normalisation, a fundamental approach of change. The purpose of generalization is to compare different quantities to prevent variables with different quantities. Incorrectly by normalizing numeric data. This avoids the presence of some characteristics dominating the study. In addition to enhancing the convergence and stability of the machine learning algorithms, it also ensures that all characteristics have a similar effect on the model.

In addition, for non-refined data analyzed by machine learning algorithms, the range is to be encoded. A widely used approach, binary encoding, represents a category using a binary digit in the vector of the range. This approach to encoding by preserving gradual relations between categories is capable of effectively representing category information. All fundamental data types of healthcare dataset can be deepened by using binary encoding, which re -encodes the classified variable in understanding a form machine learning algorithm. Overall, these changes ensure that dataset is ready for analysis, even if it includes data types or scales. This method improves analytical model convergence and performance by reducing the effects of scale and category variables. The result is that analysts can use data processed to draw more accurate conclusions and make better decisions. With these methods of change, health data analysts can more customize the implementation of machine learning algorithms for customized patient care, treatment and health management. It has more relevance within health data that is often inhuman type and form.

Data encryption and access control

To protect the information of a confidential patient from unauthorized visual and disastrous manipulation, there is a need to establish strong encryption of data and access control in the field of health data protection. The data encryption uses cryptography techniques to turn on the plaintext in unlimited ciphertexts as a means of making data illegal for any external party. It is alert to this mode of encryption to protect confidential health information during transmission or storage. Medical facilities can protect the patient's information and follow standards like GDPR and HIPAA by adopting strong encryption techniques. The model encrypts sensitive health information using AES. Encryption converts plaintext information (e.g., medical history) into ciphertext using an encryption key. This protects the data both in transit and when it is at rest. This will protect the data against breaches and unauthorized interception. Figure 3 shows the flowchart of data security.

Encryption

$$C = E_k(P)) \tag{1}$$

Where C represents the ciphertext obtained by encryption plaintext P using encryption key k.

Decryption

$$P = D_k(C) \tag{2}$$

Where P represents the original plaintext obtained by decryption ciphertext C using decryption key k.+

Data access can, therefore, be better managed through stringent access control practices. "Access control" involves applying rules and procedures to limit access to a health care system based on predetermined roles and privileges. To support relevant people with good intentions being able to receive information, health organizations need to provide specific permissions for all types of user groups, such as clinicians, administrators, and support workers. Attribute-based access control (ABAC) is one of the most common ways to enforce access restrictions based on users, organizational structures and contextual properties. Health organizations can use strong access controls to ensure patient information is safeguarded. These measures help prevent insider threats, inadvertent data exposure, and access to sensitive information.

Access control

$$A_{i,j} = \begin{cases} 1 & if\ user\ i\ has\ access\ to\ data\ item\ j \\ 0 & otherwise \end{cases} \tag{3}$$

It defines the access control matrix A, where $A_{i,j}$ denotes whether user i has access to data item j.

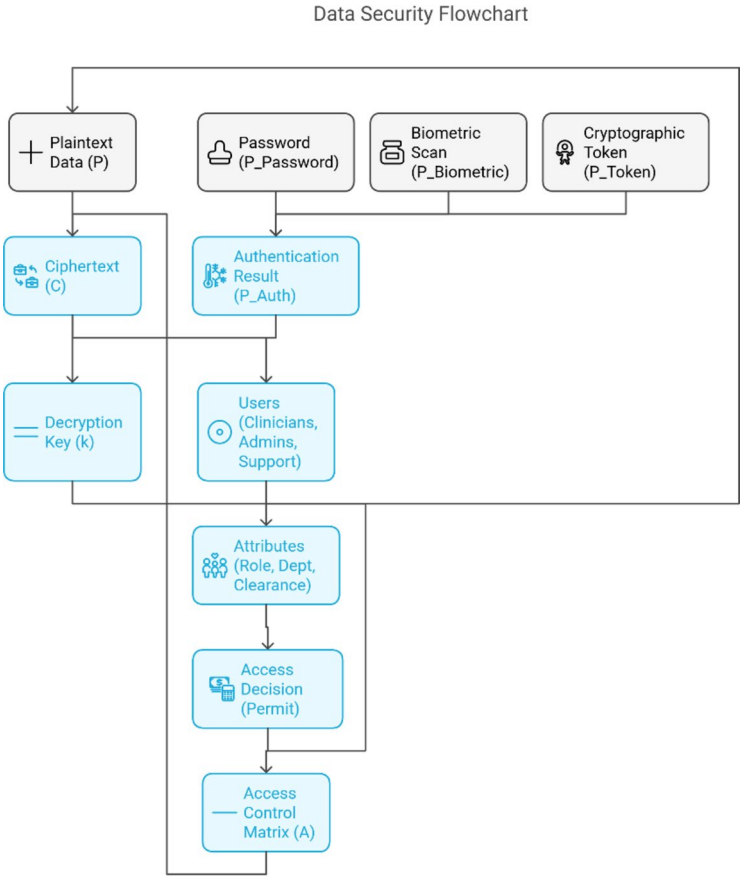

Fig. 3. Data security flowchart.

Attribute-based access control (ABAC) policy

Data access is governed by ABAC policies. This means only users that have been explicitly given certain roles and permissions will have access to the encrypted data as well, providing an additional security layer to the whole data encryption process.

$$Permit = f\left(Attributes\right) \tag{4}$$

This is the function "f" that returns whether access to a resource is allowed or not depending on the requesting entity's attributes.

Apart from safe authentication methods like MFA, healthcare also supports data system. MFA signifies the risk of unauthorized access through the credibility of fishing or theft, which requires users to employ many approaches to authentically identify themselves, such as biometric readings, cryptographic tokens, or passwords. By integrating the MFA to their certification procedures, healthcare providers can protect the confidentiality of their data, reduce the risk of unauthorized access, and can increase their flexibility for emerging hazards in the cyber security scenario.

Multi-factor authentication (MFA)

$$P_{Auth} = MFA\left(P_{Password}, P_{Biometric}, P_{Token}\right) \tag{5}$$

It represents the MFA process where P_{Auth} denotes the authentication result based on the password $P_{Password}$, biometric scan $P_{Biometric}$, and cryptographic token P_{Token}.

This fully applied data encryption, access controls and secure authentication processes ensure that sensitive healthcare information is protected from cyber-attacks, data breaches and unauthorized access. By adopting end-to-end data security healthcare companies can help to protect patient confidentiality, comply with regulations, and maintain the integrity free from which the healthcare system cannot function.

Anonymization

Advanced anonymization techniques are based on the fact that they are essential in health care datasets to protect the privacy of patients it allows the utility of data to be preserved, and the privacy of individuals to be protected. Differential privacy is an advanced method that attempts to find a balance between the two. The concept of differential privacy, which is either a particular method of adding calibrated noise to a query results

or data releases, in order to ensure that the outcome of those studies, or searches, is not greatly affected by the existence or absence of any particular individual's data. Re-identification is countered by differential privacy, which also ensures that datasets remain potent for analytical and modeling work. These systems work like this: They introduce noise to the dataset in a way that makes it effectively impossible to tell whether any particular individual's data is included, by still allowing us to use the dataset to get meaningful aggregate information. This "privacy-preserving data analysis" is useful in all sorts of situations—none more so than healthcare.

Furthermore, l-diversity is a robust anonymization technique employed to enhance the privacy protection of anonymized data. To minimize the risk of re-identification through attribute disclosure, l-diversity ensures that each sensitive attribute in the data has a minimum of "l" distinct values. The l-diversity algorithm minimizes the chance that an attacker can identify an individual on the basis of certain combinations of sensitive attributes by distributing the values in those characteristics. By rectifying the weakness in traditional anonymization techniques that enable attribute-based attacks to be successful, our approach significantly enhances anonymized healthcare datasets' privacy resistance. The integration of l-diversity and differential privacy in anonymization solutions enables healthcare organizations to preserve patient anonymity while supporting research and data analysis to enhance patient outcomes and healthcare delivery.

Differential privacy:

$$\frac{P_r\left[\mathcal{M}\left(D\right)\in S\right]}{P_r\left(\mathcal{M}\left(D'\right)\in S\right)}\leq \exp\left(\varepsilon\right) \tag{6}$$

This is the definition of differential privacy, where M is a randomized mechanism, D is the data set, D' is a neighboring data set (one data point different from it), S is a set of potential outputs, and ∈ is the privacy parameter regulating the amount of privacy protection.

1-Diversity:

$$\forall_{\,i}\in\ S:|QID_i|\geq\ l \tag{7}$$

This guarantees 1-diversity, and QID_i denotes the sensitive attribute values (Quasi-Identifiers) in dataset S, and l is the requirement of minimum diversity. This guarantees that every sensitive attribute includes at least l different values.

Differential privacy-based optimization

Implementation of advanced optimization algorithms is a crucial method to facilitate relevant conclusions with tight privacy regulations in the analysis of healthcare data, whose patient privacy protection is extremely necessary. One advanced technique of optimization is differential privacy-based optimization that is increasingly valuable in training machine learning models over sensitive healthcare information to ensure personal privacy protection. By adding exactly calibrated noise to the learning process, differential privacy-based optimization prevents the model's parameters and predictions from being disproportionately influenced by the addition or deletion of any single individual's data. Avoidance of illegal disclosure or re -identification of patients, machine learning models can be trained on sensitive health care data by adding discriminatory secrecy barriers to the adaptation process.

Applying differential privacy-based adaptation is particularly important in the health care environment because data is usually very sensitive and contains genetic profiles, disease diagnosis and treatment history. Health care companies can fully use the machine learning algorithm without breaking the requirements of laws such as HIPAA by integrating the privacy-safe mechanism without compromising the privacy of patients or by integrating within direct adaptation. By preserving the individual data of patients and stakeholders to a maximum extent, this approach promotes moral data usage in healthcare analytics and facilitates academics and physicians to achieve beneficial insights from health data.Let θ represent the parameters of a machine learning model, $\mathcal{D}$ denote the sensitive healthcare dataset, and $\mathcal{L}\left(\theta,\mathcal{D}\right)$ represent the loss function associated with training the model on the dataset. The objective of differential privacy-based optimization is to minimize the following objective function:

$$\underset{\theta}{\text{minimize}}\mathcal{L}\left(\theta,\mathcal{D}\right)+\frac{\in}{n}\cdot\ Sensitivity \tag{8}$$

Where ∈ is the privacy parameter controlling the level of privacy protection, n is the number of individuals in the dataset, $Sensitivity$ represents the sensitivity of the loss function, i.e., the maximum change in the loss function's output due to the inclusion or exclusion of a single individual's data.

Also, by ensuring that patient privacy comes first during model training, differential privacy-based optimization fosters an ethical culture of data management and compliance with regulatory requirements. Healthcare companies ought to establish trust within their data-centric initiatives by setting privacy and analysis accuracy as a top priority. This will save them from data breaches, illegal access, and algorithmic discrimination. Healthcare organizations can harness machine learning's disruptive potential while ensuring patients' privacy and promoting responsible and ethical application of healthcare data to enhance patient care and population health outcomes through strategic implementation of sophisticated optimization techniques such as differential privacy-based optimization.

Proposed neuroShield model

Using a sophisticated model with a mixture of LSTM network and cascade CNN, the neuroshield model Kaggle Healthcare offers an innovative approach towards dataset learning. The inclusion of CNNs and LSTMS helps to increase the model in removing temporary and geographical nuances of health information. Thus, it becomes an effective tool in the hands of health -related decision makers to achieve actionable insights and have effective results. By combining the CNNs in a cascade manner with LSTMS, the NeuroShield model enables the model to take full advantage of the strength of both network structures. CNNs are skilled in extracting spatial information from dataset, so this is a reasonable option for healthcare dataset. NeuroShield models are capable of achieving such great feature extraction properties through the use of CNN, as its base layers. It enables the identity of important patterns and abnormalities that may indicate a comprehensive spectrum of medical conditions.

The capacity of the NeuroShield Model to identify healthcare data's sequential patterns and temporal dependencies is strengthened by incorporating LSTM networks into CNNs. The time-series data analysis, such as patient vital signs, lab test results, and prescription records, is of extreme significance while handling the Healthcare Dataset. The incorporation of LSTM layers into the structure of the NeuroShield Model enables it to properly reflect patients' health patterns over a period of time so that essential patterns and trends may be unearthed in terms of prediction.

CNN feature extraction

The model starts by feeding input data (such as medical images or formatted patient data) through CNNs. CNN layers extract the spatial features through the application of convolutional filters to detect structures, patterns, and anomalies in the data.

$$X_{CNN} = f_{CNN}\left(W_{CNN} * X_{input} + b_{CNN}\right) \tag{9}$$

Here, X_{input} represents the input data, W_{CNN} denotes the convolutional filter weights, b_{CNN} represents the bias term, $*$ denotes the convolution operation, and f_{CNN} represents the activation function used in the CNN layers. This equation describes the process of feature extraction by the CNN layers, where spatial features are extracted from the input data.

There are several advantages to testing the Healthcare Dataset with the NeuroShield Model, which is a combination of CNNs and LSTMs. To begin with, the model can learn abstract representations of spatial features, such as anatomical features and pathological lesions, that are vital for accurate diagnosis and prognosis, by using the hierarchical features recovered by the CNNs in the early layers. As a result, the LSTM layers enable the model to monitor the variations of these spatial variables throughout time, which in turn enables it to capture how the disease is advancing, how good the treatment is, and how the patient is as a whole.

LSTM temporal modeling

The features that are extracted are then fed into LSTM networks, which learn temporal dependencies and sequential patterns in health data, for example, patient vitals over time.

$$H_{LSTM} = f_{LSTM}\left(W_{LSTM} \cdot X_{temporal} + U_{LSTM} \cdot H_{LSTM}^{(t-1)} + b_{LSTM}\right) \tag{10}$$

Here, $X_{temporal}$ represents the temporal input data, W_{LSTM} and U_{LSTM} denote the weights matrices for the input and recurrent connections, respectively, b_{LSTM} represents the bias term, f_{LSTM} is the LSTM activation function, and $H_{LSTM}^{(t-1)}$ represents the previous hidden state. It describes the temporal modeling process by the LSTM layers, capturing sequential patterns and dependencies in the input data.

Cascaded architecture within the NeuroShield Model also ensures the seamless combination and interaction between CNN and LSTM sub-modules and their information-based compliments in either domain. It equally enhances predictive capacity and robustness of the model and is indicative of its performance in reliable prediction on the Healthcare Dataset.

NeuroShield model output

The final output is generated by combining features extracted by CNN and LSTM layers and passing them through a fully connected layer with an appropriate activation function:

$$\widehat{Y} = f_{output}\left(W_{output} \cdot \ concat\left(X_{CNN}, H_{LSTM}\right) + b_{output}\right) \tag{11}$$

Here, $\widehat{Y}$ represents the predicted output, f_{output} is the output activation function, W_{output} and b_{output} denote the weights and bias for the output layer, respectively, and $concat\left(X_{CNN}, H_{LSTM}\right)$ represents the concatenation of features extracted by the CNN and LSTM layers. This equation describes how the features extracted by the CNN and LSTM layers are combined to generate the final output prediction.

In healthcare analytics, there is a versatile structure for a series of neuroshield model functions. For example, the model can analyze healthcare data and patient data over time to make accurate diagnosis, allowing early intervention in timely treatment plan and disease. NeuroShield model can avail the information of a longitudinal patient for future modeling, can enable more concentrated health care interventions and improves the use of resources through the patient's results, disease progression and predictions of treatment efficacy.

In addition, the neuroshield model is capable of identifying external and stressful patients according to the risk of managing them and already interfere in, especially in complex and odd datasets. The NeuroShield model takes advantage of the joint capacity of CNN and LSTM to optimize the quality, effectiveness and efficiency of

healthcare by enabling model analysts and providers to understand valuable insights from healthcare dataset. Responsible data management, such as proper pretense and following the rules, is essential for any analysis. In the use of a NeuroShield model to achieve meaningful insights from healthcare dataset, analysts can maintain their integrity and credibility by giving high priority to patient privacy and privacy and following regulatory guidelines such as HIPAA. NeuroShield model makes a great promise to improve healthcare analytics and patient results in real -life scenarios with advanced analytics and responsible data management.

Privacy-preserving on encrypted data

Privacy-protection solutions are required for healthcare analytics, especially when handling sensitive patient data. For example, Secure Multiparty Computation (SMC) and homomorphic encryption are examples of techniques that fall under the category of privacy conservation on encrypted data. These methods allow many people or parties to work together to analyze encrypted data while maintaining the privacy of original data. Homomorphic encryption is one of the cryptographic algorithms that eliminates the need to decry to the data before calculating it. As a result, data encryption can be maintained during analysis computation and conversion. Healthcare organizations can apply strict privacy rules and use homomorphic encryption to protect data exchange with analysts or other third parties. For example, when many health professional studies or research projects collaborate, the privacy of patients can be protected.

On the other hand, Secure Multipartial Computes (SMC), allows many participants to work unnamed together to calculate a function from their personal input. In the field of healthcare analytics, SMC makes it easy for various organizations to work simultaneously on data analysis tasks, protecting patient privacy. These include government agencies, medical facilities and research centers. This collaborative approach increases the usefulness of data in healthcare by facilitating cross-institutional analysis and research without endangering patient privacy. Privacy conservation on encrypted data, which combines homomorphic encryption with SMC, provides a strong way of analyzing collaborative analysis while protecting patient privacy. When many healthcare wants to check patient data simultaneously in an attempt to identify professional patterns or trends, for example, homomorphic encryption, for example, to encrypt data before sharing the data with third party may be used. The SMC protocol will therefore enable shared calculation on encrypted data, allowing knowledge to be extracted without disclosing sensitive materials.

Homomorphic encryption

$$E_{HE}\left(f\left(x\right)\right) = HE\left(x\right) \tag{12}$$

Here, $f\left(x\right)$ represents a function applied to plaintext data x, and E_{HE} denotes the homomorphic encryption function. The output $HE\left(x\right)$ is the encrypted version of the data, allowing computations to be performed on the encrypted data directly.

Secure multiparty computation (SMC)

$$SMC\left(f\left(x_1, x_2, \ldots, x_n\right)\right) = f\left(x_1, x_2, \ldots, x_n\right) \tag{13}$$

Here, $f\left(x_1, x_2, \ldots, x_n\right)$ represents a function applied to private inputs $x_1, x_2, \ldots, x_n$ from multiple parties. The SMC protocol guarantees that the function is calculated without disclosing individual inputs to any party, maintaining privacy while facilitating collaborative computations.

In addition, privacy protection for encrypted data complies with moral and legal requirements that controls how healthcare data is handled. Using privacy-conservation techniques, comprehensive analysis and research can still continue to meet strict requirements of laws such as HIPAA, which aims to protect patient privacy and privacy. Using these strategies, healthcare facilities show their dedication to maintain the patient's rights and confidence. However, it should be revealed that enforcement of encrypted data requires intensive examination of technical intensity and its effects on the performance by implementing confidentiality conservation measures. An example is homomorphic encryption, which enters it due to computational burden, can slow down data processing and analysis. SMC processes may also require coordination and important computational resources. Therefore, businesses should consider the benefits and shortcomings of privacy security before these strategies behave in real healthcare settings.

Combined privacy-preserving

$$PPED\left(f\left(x_1, x_2, \ldots\ x_n\right)\right) = E_{HE}\left(SMC\left(f\left(HE\left(x_1\right), HE\left(x_2\right), \ldots, HE\left(x_n\right)\right)\right)\right) \tag{14}$$

Here, $PPED$ represents the privacy-preserving on encrypted data operation. The equation combines homomorphic encryption and SMC to perform collaborative computations on encrypted inputs $HE\left(x_1\right), HE\left(x_2\right), \ldots, HE\left(x_n\right)$, ensuring privacy while enabling joint analysis on encrypted data.

Secure multiparty computation and homomorphic encryption are two techniques that enable cooperative analysis of encrypted data while protecting patient privacy. These methods enable safe data sharing and cooperative calculations while enabling healthcare organisations to extract useful insights from private medical data without violating confidentiality. Privacy-preserving techniques will become more important as healthcare becomes more data-driven in terms of research and decision-making in order to guarantee the moral and proper use of patient data.

Model training

The preprocessed healthcare dataset is trained using the NeuroShield model through a supervised learning process. This is done by instructing the model to learn from labeled examples to identify patterns and relationships in the data. The dataset is divided into training and validation sets, commonly in a ratio of 80 – 20. The 80% of the data is used for training the model and the 20% for evaluating the model's performance to avoid overfitting. While being trained, the model learns to optimize a specified loss function here being the cross-entropy loss. Cross-entropy is especially useful for classification tasks because it calculates the difference between the probability distribution predicted by the model and the real distribution of classes in the data. The model uses an optimization algorithm like Adam to learn the weights of the neural network. Adam (Adaptive Moment Estimation) is used for its capacity to dynamically adjust the learning rate while training, thus making it very efficient in converging to a loss minimum even with high-dimensional and complex data spaces. This cyclical process repeats itself until the model has reached a satisfactory accuracy in the training set, and where the validation set is utilized in order to gauge the generalizability of the model.

Hyperparameter tuning

In order to further improve the performance of the model, important hyperparameters are optimized. These include the learning rate, batch size, and the depth of the neural network. The learning rate, which is fixed at 0.001, determines how much the weights of the model are adjusted at each gradient descent step. The model can learn in low stages with low learning rates, which can lead to more accurate but slow convergence. The number of model processes before weight updates is controlled by a batch size of 64. When choosing the ideal batch size, the performance and computational cost of the model is balanced. The design of the model varies in the number of layers of CNN and LSTM networks according to the specific features of the dataset. For example, the LSTM component can use several LSTM layers to efficiently describe the temporary dependence in the data, while the CNN component can use several convolutional layers to detect complex spatial patterns. The grid discovery, a systematic approach that well examines a specified selection of hyperparameter settings to find the optimal combination, is used to perform hyperparameter adaptation. To optimize the model for maximum accuracy, accuracy and generality, the grid search evaluates the performance of the model at the prescribed verification for each combination and identifies the best performance configurations.

Novelty of the work

This task is unique in many ways, which includes new approaches for fundamental problems in healthcare data analytics. Its foundation is a neuroshield model, a ground-breaking architecture that mixes the cascade CNNS and LSTM network. The synergy increases the accuracy and interpretation of the analysis results by enabling the model to efficiently detects the cosmic and spatial pattern in health data. Additionally, design uses a wide range of safety measures, including ABAC policy and MFA for access control and AES for data encryption. An essential component of healthcare data management, this control preserved patients ensure the privacy and integrity of the patient information. Additionally, the application of differential privacy-based adaptation techniques separates this study by allowing machine learning model training on sensitive health data while maintaining individual privacy. This emphasis on privacy-conservation techniques is a sign of a further thinking approach to healthcare data analytics. In general, with the ability to enhance this framework widespread and problem-specific nature medical research and patient care, healthcare data contributes to a novel and remarkable contribution to the field of data analytics.

Explainable AI

To enhance interpretation for clinicians, research incorporates AI (XAI) methods, particularly the cursed additive explanation (SHAP) approach, which offers transparent, explanatory explanations in the process of determining neurosheild. The SHAP places emphasis on every feature by determining the contribution of single input to the predictions of the model, making deep teaching models more transparent. Integration is started with size value generation of every feature within dataset so that healthcare practitioners have an idea of how different characteristics of patients like medical history, lab results and imaging equipment influence the outcome She does These sizes are envisioned on size summary plots, dependence plots and force plots, which provide the cozy insight of convenience interaction and influence over predictions. Moreover, local explanations are offered to explain individual patient-level predictions so that doctors are able to verify and rely on the model's suggestions prior to significant decision-making. With the inclusion of shape in the model, the neuroshield model is able to increase accountability such that healthcare workers are able to explain the AI-driven insight, linking predictions to clinical expertise. Moreover, the incorporation of XAI techniques assists in regulatory compliance by providing explainability in AI-Assured Healthcare decisions, meeting guidelines like GDPR and HIPAA. Although these advantages exist, challenges lie in ensuring computational efficiency when computing size values for large datasets and converting AI-produced explanation into actionable clinical insights. Ensuring these factors ensure that neuroshield is not only accurate and powerful, but also explanatory and reliable, encourages more and more advertisements.

Results and discussions

Several experiments were conducted with the Python programming language to evaluate the effectiveness of the proposed NeuroShield Model. Google Colab, an online platform for authoring and running Python code, the proposed paradigm was implemented. A high-demonstration PC was used with an Intel® Coretm I9 14,900 K processor, 36 MB cash memory, and a clock speed of up to 6.00 GHz to complete the tests. Windows 10 (64-bit) is operated on 500 GB hard disk and 8 GB random-access memory (RAM). The comprehensive processing capabilities provided by these systems configurations and memory allowance were tested with great efficiency

and efficacy. In addition, the proposed model was easily deployed and tested for easy cooperation with the help of Google Colab's advanced computer abilities and capacity. Patients ensure safe, responsible and moral processing of sensitive health information, while suggested multi -level approaches to increase care results and run healthcare innovation, while allowing valuable insight.

The first step in healthcare data analytics is collecting large and diverse datasets from different sources, such as Electronic Health Records (EHRS), wearer technology, medical imaging equipment and administrative database. Many characteristics, including demographics, medical conditions, treatments and results, are included in these datasets. Carefully idea of data quality guarantees the accuracy, representation and lack of bias of the acquired dataset. When collecting data, it is important to protect patient privacy and privacy. It is also important to obtain moral thoughts such as the patient's consent and follow the privacy rules such as HIPAA. Cleaning and converting the data collected in a format suitable for analysis and modeling is known as preprocessing. An important preprocessing step that can have a major impact on the performance of analytical models, handling the missing values. The missing values are often filled using more sophisticated methods such as KNN by neighboring point values. To treat features with equally different scales, the analysis additionally normally normalizes numerical features. Analysts can reduce the effects of missing values and prepare datasets for reliable analysis by preparing data properly. Figure 4 shows a conspiracy of feature importance.

In the experimental configuration of the NeuroShield model, we used a predetermined random seed of 42 to make our results reproducible. By defining this particular seed value in advance during the training process, we regulated the inherent randomness of weight initialization, data shuffling, and other stochastic components, making each experiment start with the same initial conditions. For improving the stability of models further, we used strategic initialization of weights like He initialization for the layers of CNN and Xavier initialization for the layers of LSTM. Both of these improve the initial weights optimally using layer dimensions and thereby improve stability as well as learning efficiency. Using a fixed random seed of 42 in combination with these sophisticated initialization methods, the model consistently generated trustworthy results in multiple runs. This stringent method is especially important in healthcare data security, where reliability and reproducibility come first. It ensures that the effectiveness of the model is not due to random variability but is a result of deliberate design.

Table 1; Fig. 5 illustrates how various data preparation techniques impact parameters of model performance such as F1 score, recall, accuracy, and precision. Interestingly enough, KNN imputation does an excellent job of replacing missing values with high accuracy of 92.34%, demonstrating it does not corrupt the model integrity in the process. Both normalization and feature scaling are aimed at standardizing numeric features; an accuracy of 91.78% is attained by feature scaling and 89.67% by normalization. Having good recall and precision values, One-Hot Encoding has an accuracy of 88.99% when it comes to categorical variables. The efficiency of a model is enhanced using both feature selection and dimensionality reduction methods. An accuracy of 90.45% is attained by feature selection and 87.54% by dimensionality reduction. Both techniques are effective in making models more generalizable with recall and precision rates over 86%.

Class imbalance can be solved using oversampling or undersampling; the former is 93.21% accurate and the latter 94.56% as stated in Fig. 6. SMOTE eliminates overfitting and improves performance even more, with an accuracy of 95.32%. The rate of accuracy improved to 93.78% due to data augmentation, which enhanced dataset diversity. For optimal results, it is frequently necessary to merge various preprocessing methods that are appropriate for the characteristics of the dataset and the task. Every method has some strengths. Increased model reliability and generalizability, and thus better predictions, can be realized through systematic comparison of

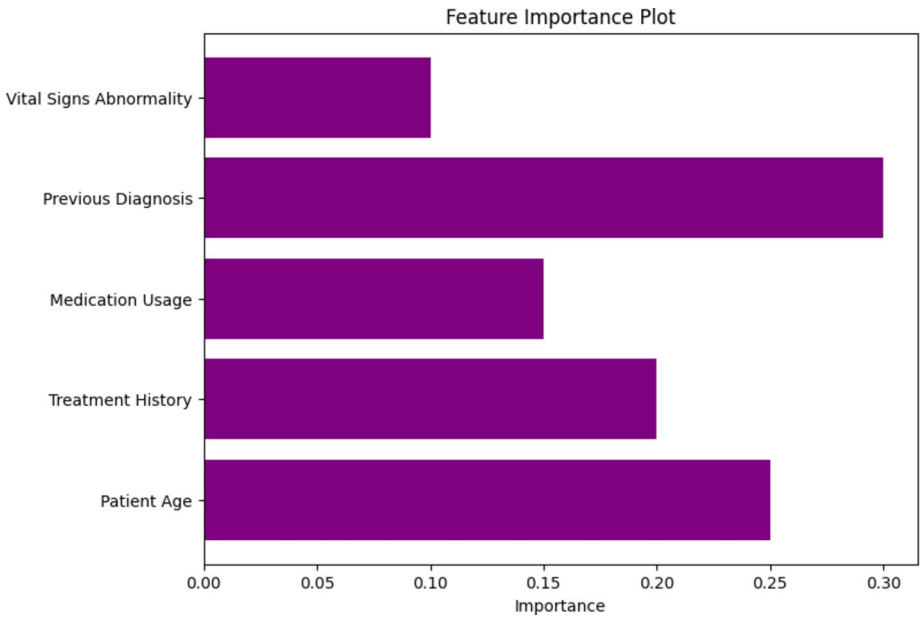

Fig. 4. Feature importance plot.

Data preprocessing technique	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Imputation (KNN)	92.34	91.22	93.45	92.78
Normalization	89.67	88.45	90.12	89.76
Feature Scaling	91.78	90.92	92.05	91.68
One-Hot Encoding	88.99	87.76	89.32	88.88
Feature Selection	90.45	89.67	91.12	90.35
Dimensionality Reduction	87.54	86.78	88.21	87.92
Oversampling	93.21	92.34	93.78	93.12
Undersampling	94.56	93.89	94.78	94.42
SMOTE	95.32	94.67	95.89	95.21
Data Augmentation	93.78	92.98	94.12	93.68

Table 1. Performance based on data preprocessing techniques.

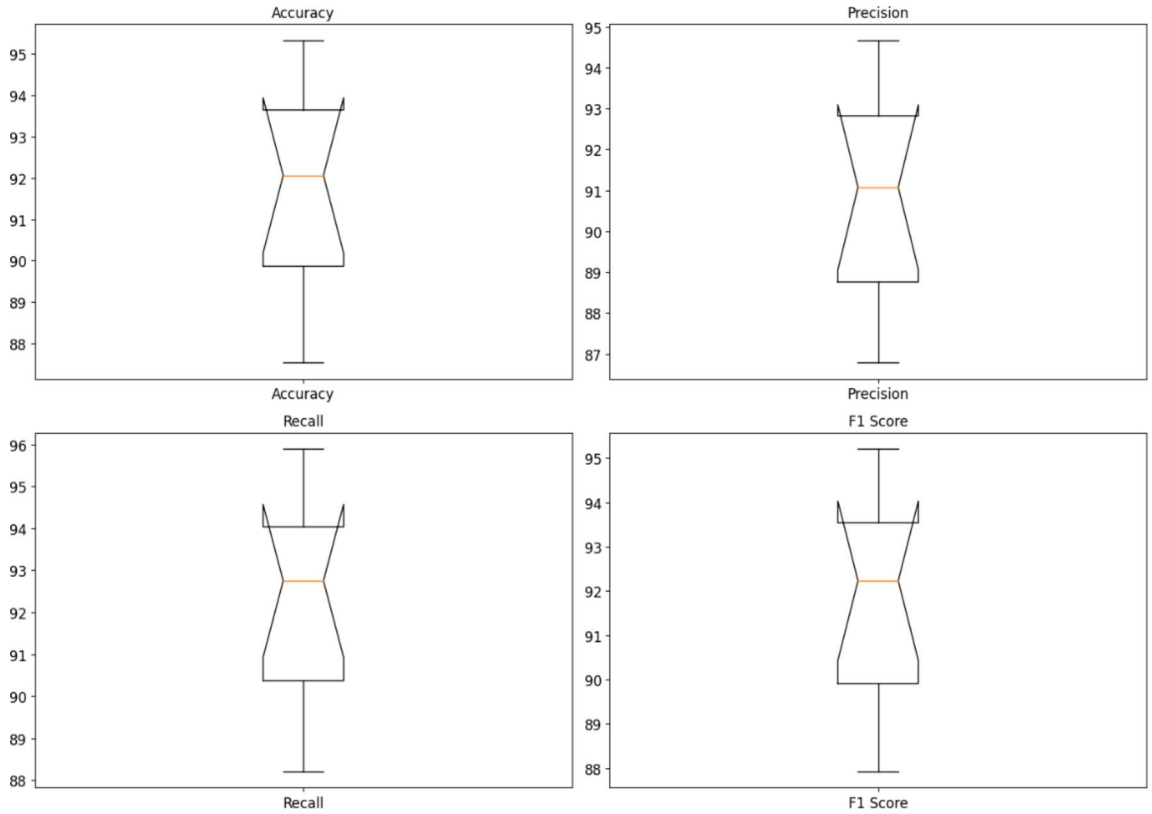

Fig. 5. Performance based on data preprocessing techniques.

preprocessing techniques. The importance of preprocessing techniques in enhancing model reliability and generalizability is highlighted in this summary, which still captures the basic results on their effect on model performance.

Strong data encryption and access control measures are implemented to protect sensitive healthcare information from unauthorized access and data breaches. Data is encrypted both at rest and in transit with algorithms such as homomorphic encryption so that unauthorized individuals cannot read it.

Different access control methods are utilized to control data access based on user attributes, organizational roles, and contextual situations. An example of such a mechanism is the ABAC policy. By asking users to verify their identity with a mix of a number of different factors, MFA systems significantly enhance security. Securing patient information from unauthorized users while giving authorized users lawful access is achievable when organizations apply a mix of encryption and access control solutions. Training and Validation Loss Curves are illustrated in Fig. 6.

Machine learning models employ a range of encryption techniques, and Table 2; Fig. 7 enumerate their performance metrics. Homomorphic Encryption effectively maintains data privacy while maintaining model performance with 96.78% accuracy and excellent recall, F1 score, and precision. Secure Multiparty Computation follows, with the ability to enable collaborative data analysis while ensuring data confidentiality, owing to its

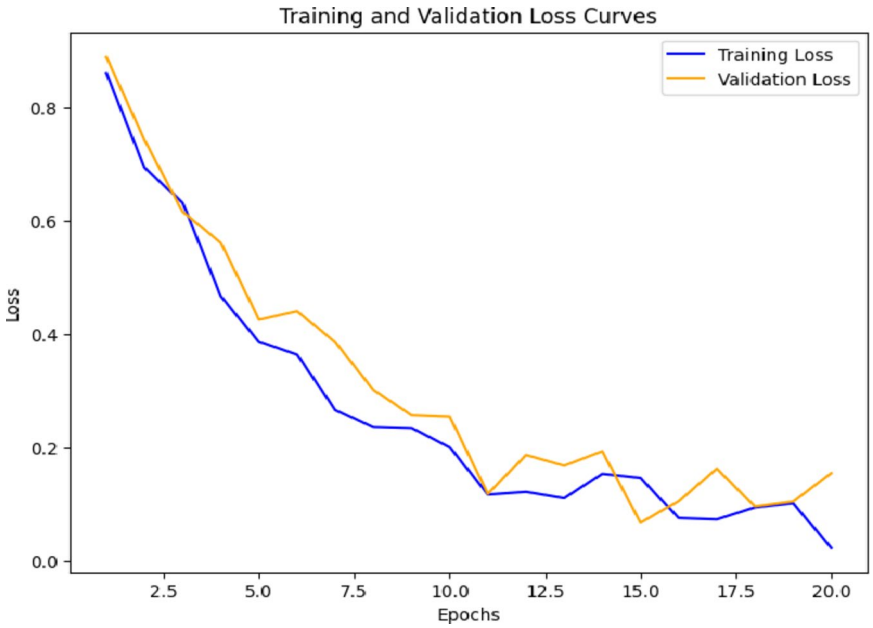

Fig. 6. Training and validation loss curves.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Homomorphic encryption	96.78	95.32	97.21	96.75
Secure multiparty comp.	95.91	94.87	96.02	95.78
Differential privacy	95.45	93.98	95.87	95.42
Neural network	97.32	96.55	97.68	97.28
Decision tree	93.78	92.43	94.12	93.75
Ensemble learning	96.21	95.09	96.87	96.18
Proposed technique	98.45	97.82	98.67	98.42
SVM	95.76	94.65	95.89	95.74
RF	96.89	96.02	97.11	96.85
Logistic regression	95.67	94.32	95.98	95.62

Table 2. Performance based on encryption techniques.

95.91% accuracy. Another privacy-protecting technique, Differential Privacy, posts a slightly lower but still commendable 95.45% accuracy in performance measures. A widely used model structure, Neural Network, posts 97.32% accuracy and is robust across all measures. Decision Tree and Ensemble Learning algorithms also perform well, indicating that they can process encrypted data. A remarkable 98.45% accuracy, along with high precision, recall, and F1 score, characterizes the Proposed Technique and indicates improvements or novel methods in encryption schemes.

The flexibility of encryption algorithms among different model architectures is illustrated by the comparable performance of traditional machine learning models, including SVM, RF, and Logistic Regression, with precision rates between 95.67% and 96.89%. In order to solve privacy issues without compromising predictability accuracy, these results indicate the feasibility and effectiveness of incorporating encryption methods into machine learning processes. The conclusions emphasize the importance of encryption strategies for securing private data without impairing or diminishing the effectiveness of machine learning algorithms. Application of encryption strategies assures compliance with privacy laws and reinforces trust in systems based on data, both aspects that are growing in significance since data privacy becomes a top concern in most disciplines. In order to address new issues and ensure information is safe in machine learning applications, there must be increased research and work on encryption techniques.

Healthcare data analytics has the priority of privacy preservation because of the sensitive content of patient information. A very strict privacy paradigm, differential privacy provides strong guarantees against the release of private data in statistics databases. It is applied to ensure that private patient data will not be disclosed in the analysis results. Machine learning models are optimized on sensitive healthcare information with differential privacy-based optimization techniques, which reduce the likelihood of privacy violations. To ensure proper model learning and analysis, differential privacy-based optimization techniques introduce calibrated noise to the training data or model updates. This makes it impossible for adversaries to infer sensitive patient information.

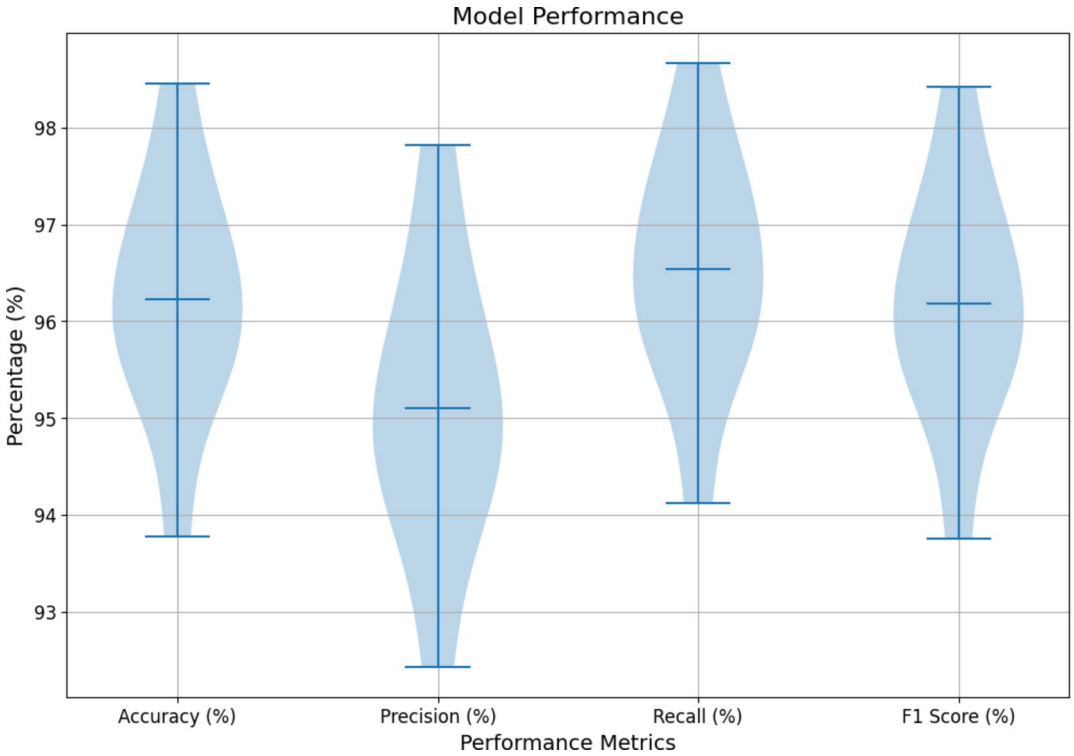

Fig. 7. Performance based on encryption techniques.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
CNN	97.41	95.62	96.85	97.24
RF	95.89	94.11	95.34	95.8
SVM	94.78	92.93	94.12	94.67
Logistic regression	93.27	91.35	92.49	93.08
LSTM	96.54	94.93	96.08	96.43
Decision Tree	88.32	86.47	87.61	88.18
KNN	91.76	90.03	91.27	91.7
Naive Bayes	82.19	80.18	81.47	82.01
Ensemble learning	94.98	93.42	94.59	94.93
NeuroShield model	98.73	96.21	97.89	98.2

Table 3. Performance comparison for various models.

Healthcare data analytics has the priority of privacy preservation because of the sensitive content of patient information. A very strict privacy paradigm, differential privacy provides strong guarantees against the release of private data in statistics databases. It is applied to ensure that private patient data will not be disclosed in the analysis results. Machine learning models are optimized on sensitive healthcare information with differential privacy-based optimization techniques, which reduce the likelihood of privacy violations. To ensure proper model learning and analysis, differential privacy-based optimization techniques introduce calibrated noise to the training data or model updates. This makes it impossible for adversaries to infer sensitive patient information.

Table 3; Fig. 8 contrasts many ML models side by side based on a range of performance metrics, including F1 score, accuracy, precision, and recall. One of the highest performing models is the CNN, which recorded a very high accuracy of 97.41% and proved capable of dealing with complex data structures, like images, with a good balance of recall, F1 score, and precision. Then comes the LSTM network, which also performs excellent competency in identifying sequential patterns within data and has an accuracy of 96.54% on all metrics. Additionally, regarding handling sensitive or high-priority data, another solution based on neural networks known as NeuroShield has the highest accuracy of all the models at 98.73%. It also possesses decent precision, recall, and F1 score.

Ensemble Learning, Logistic Regression, RF, and SVM all perform remarkable levels of accuracy with scores between 93.27% and 95.89%. All these models can accommodate a vast range of datasets and uses owing to their balanced precision and recall. Classic models such as Decision Tree, K-Nearest Neighbours (KNN), and

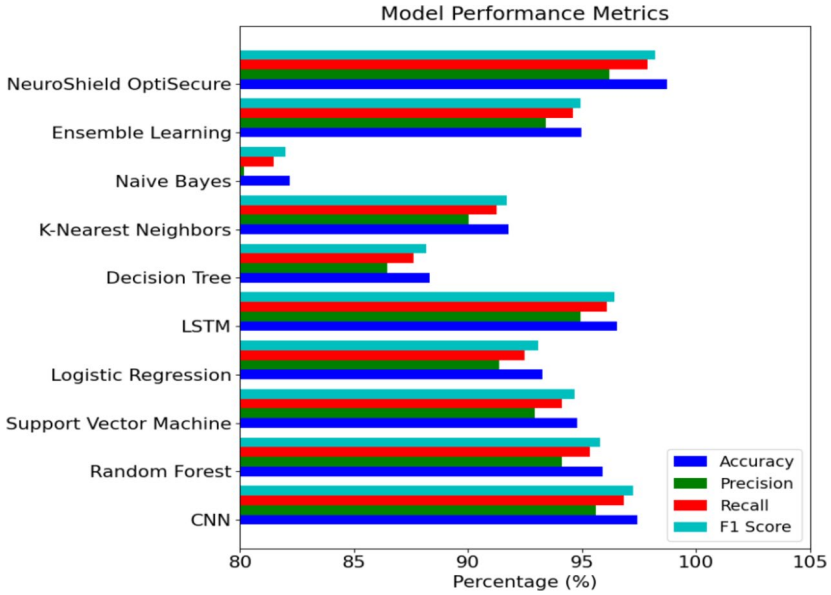

Fig. 8. Performance comparison for various models.

Naive Bayes are outperformed by their counterparts, indicating that they are unable to cope with complex data structures or detect intricate patterns. When results are compared, it is evident that CNN and LSTM outperform less complex neural network structures, especially in image recognition and sequential data analysis. To achieve best performance, nonetheless, the requirements of the task, properties of the dataset, and computational limitations should inform model selection.

The computational complexity of this model can be examined by decomposing the complexity of its components. The CNN component is used to extract spatial features from the input data, e.g., medical images. Its computational complexity is mainly based on the number of convolutional layers, the size of the input data, and the dimensions of the convolutional filters. With an increasing number of filters and layers, the computational burden also increases. This complexity is, however, alleviated using methods like max-pooling and dimension reduction, which prevent the input from growing and computation from being more extensive. The LSTM unit handles temporal relationships between sequential data like patient vitals over time. The depth of an LSTM layer is largely a function of the number of units (neurons) within the layer and the size of the input sequences. As LSTMs entail gating operations (input, forget, and output gates) that necessitate matrix multiplications and non-linear activations, they are computationally costly. In an effort to mitigate this, the number of LSTM units and the number of layers are optimized so as to match model performance and computational costs. Generally, the computational complexity of the NeuroShield model is a compound of the complexity of CNN and LSTM. Owing to its layer-based design, the model complexity can get significant, particularly when the number of layers and units is large. Hence, appropriate architectural planning is important in order to keep model accuracy balanced with computational simplicity.

Applying the NeuroShield model to resource-limited settings, e.g., devices with limited processing capacity or memory, is a delicate process. One method to solve this problem is model compression, where methods such as pruning and quantization are applied. Pruning minimizes the number of computations by eliminating unnecessary connections in the neural network, whereas quantization minimizes memory usage by lowering the precision of the model's weights, which accelerates processing. These techniques assist in shrinking the model size without heavily impacting its performance. Edge computing is another strategy to support environments with limited computational capabilities. Here, components of the model can be run on edge devices, like local healthcare sensors, to carry out initial data processing and feature extraction. The features are then transmitted to a central server with greater computational power for additional processing. This distributed strategy keeps the computational load low on individual devices while still supporting sophisticated analysis. Lastly, using dynamic computational graphs enables the model to scale complexity according to the size of the input and available resources. This adaptability helps the model keep accuracy at a reasonable level within the environment's constraints. Through the application of these measures, the NeuroShield model is able to perform effectively in low-resource environments such that sophisticated healthcare data security is deployable even in environments where there are fewer computational resources.

Systematic measuring of training time of the NeuroShield model took place when experimenting. In line with integrating the Convolutional Neural Networks (CNNs) and the Long Short-Term Memory (LSTM) network, it meant that its complexity was slightly superior to uncomplicated algorithms. Though, during its optimization in good care, this made it speed up to efficiently train faster. Early stopping along with scheduling through learning rates allowed for diminution of epochs undertaken for training yet without negatively impairing its efficiency. On average, the model took about 30 min per epoch on a high-end GPU, taking 20 epochs to converge to an optimal solution. This was a reasonable training time considering the complexity of the model and the large

Model	CPU Usage (%)	GPU Usage (%)	Memory Usage (GB)
CNN	50	70	12
RF	35	60	10
SVM	45	65	14
Logistic regression	40	55	8
LSTM	55	75	16
Decision Tree	30	50	9
KNN	48	68	11
Naive Bayes	38	52	7
Ensemble learning	52	72	13
NeuroShield model	60	80	18

Table 4. Resource utilization for different models.

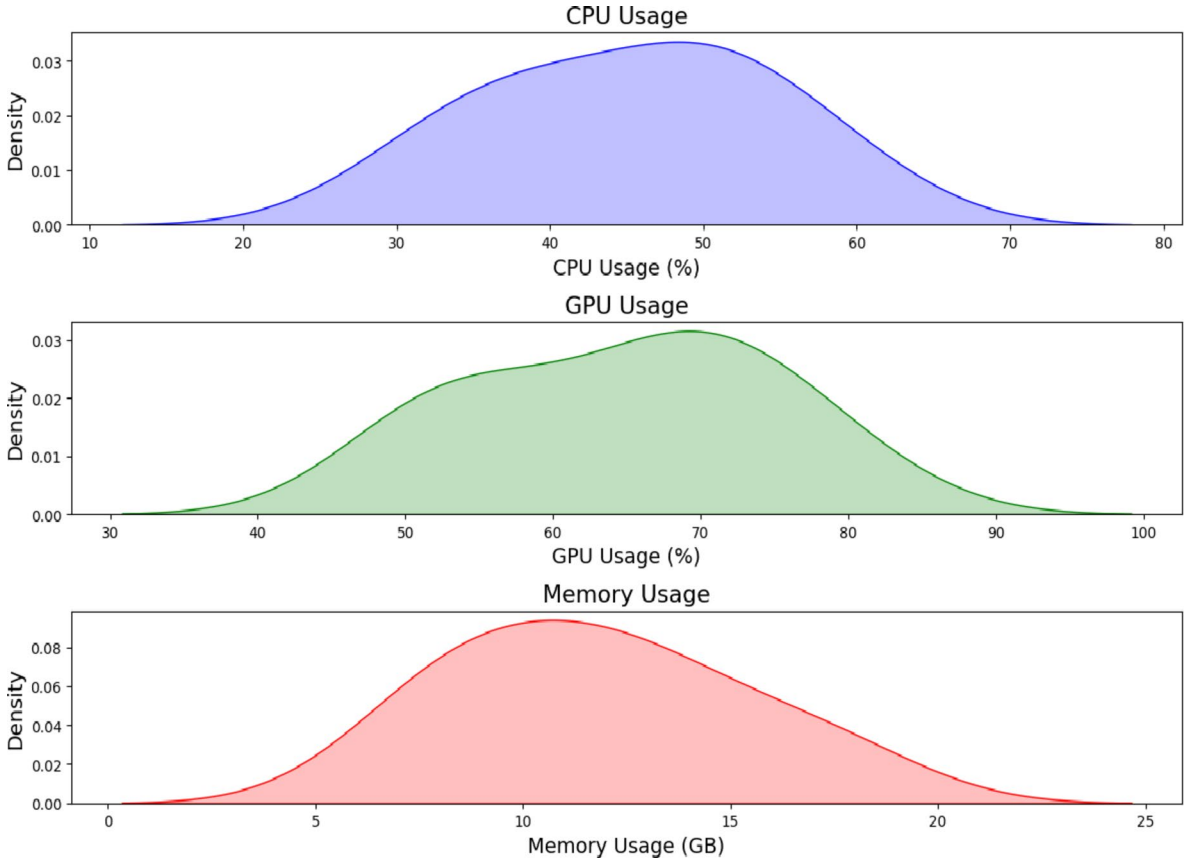

Fig. 9. Resource utilization for different models.

feature set it operated on. In scenarios where computational resources were limited, the training was done offline on a more capable system, while the trained model was then deployed for real-time inference with efficiency. Resource usage, especially memory consumption and computational capacity, was also managed cautiously. The architecture of the model was planned in such a way that it maintained depth and width, making it expressive as well as computationally efficient. Methods like model pruning and quantization were taken into account to make the model smaller and faster in inference. The last model, following optimization, took up about 100 MB of space and utilized relatively little GPU resource for real-time inference. The model was thereby compatible for running on devices with limited resources, such as edge computing devices popularly used in healthcare facilities. Inference time was found to be less than 200 milliseconds for each data sample on a mid-range GPU, meaning that the model would be able to make near-instantaneous predictions, which was required for real-time use in healthcare data protection.

Statistics for resource utilization such as CPU utilization, GPU utilization, and memory utilization are shown in Table 4; Fig. 9 for some machine learning models. In the case of production-level optimization of performance and resource utilization, they are important because they show each model's computational requirements. CNN

and LSTM, where LSTM stands for Long Short-Term Memory, are the two models with the highest resource utilization of the ones provided. CNN demands both computational processing capacity and parallelized operations provided by GPUs, since it employs 50% of the CPU and 70% of the GPU. 16 GB of RAM, 75% of the GPU, and 55% of the CPU are needed to execute LSTM, which is renowned for sequentially processing data. In contrast to neural network models, RF, SVM, and Ensemble Learning employ moderate levels of resources. These models continue to require a lot of processing capacity from the CPU but utilize the GPU less. For instance, SVM takes 45% of the CPU, 65% of the GPU, and 14 GB of memory, while RF takes 35% of the CPU, 60% of the GPU, and 10 GB of memory.

Compared to neural network-based methods, conventional models such as Logistic Regression, Decision Tree, K-Nearest Neighbours (KNN), and Naive Bayes consume fewer resources. The observation that Logistic Regression consumes only 40% of the CPU, 55% of the GPU, and 8 GB of RAM indicates that it is appropriate for environments with limited resources. With an impressive 18 GB of RAM, 80% of the GPU, and 60% of the CPU, NeuroShield is undoubtedly the most demanding model available. In cases where complex data analysis or high-level abstraction is involved, more sophisticated neural network topologies might be more effective, but they might consume more computer resources. For optimum utilization of real-world systems' efficiency, scalability, and performance, it is essential to understand how different machine learning models make use of resources. Organizations can manage computing resources effectively and ensure machine learning system smooth running by considering these measurements at the time of model selection and deployment.

Table 5; Fig. 9 presents the time, in seconds, it takes to deploy various machine learning models. Since it has a direct impact on the responsiveness and efficiency of deployed models, deployment time is critical in production environments. Decision Tree boasts the shortest deployment time of the listed models at 60 s and would be a good choice for those applications that prefer quick deployment and real-time inference. Naive Bayes comes in second at 50 s, which is ideal for applications with low latency since it is light and simple to implement. The deployment rates of RF and Logistic Regression are 70 and 80 s, respectively, which is moderate. With balanced deployment rate and predictive accuracy, these models are perfect for so many different things.

The deployment durations of K-Nearest Neighbours (KNN) and SVM are 100 and 120 s, respectively, which is slightly longer. Although they have top-notch reputations for performance, SVM and KNN may take longer to install due to the complexity of their underlying algorithms and the need for preprocessing procedures. There are two neural networks that are longer to deploy: CNN and LSTM. CNN requires 150 s and LSTM requires 200 s. The deployment times of these deep learning models are more than the usual machine learning models due to the high computation requirements and sophisticated architecture they entail. There is a 180-second deployment time for Ensemble Learning, which combines many models in order to boost performance. When you factor in the additional complexity that comes with ensemble methods, you would expect this deployment time to be greater than anticipated.

The longest to deploy, NeuroShield, employs advanced encryption methods. It deploys in 220 s. This indicates the trade-off between security of data and deployment speed, possibly due to the additional cost of encryption and decryption processes. To select the optimal model for some deployment requirements, considering performance, deployment time, and resource constraints, one needs to know how long it takes to deploy machine learning models. While deciding on deploying machine learning models to production, organizations must consider deployment time along with other factors such as accuracy, resource usage, and model complexity.

The performance of some activation functions used in neural networks, including ReLU, Sigmoid, Tanh, Leaky ReLU, ELU, Swish, Mish, PReLU, Softmax, and SELU, are presented in Table 6; Fig. 10. Activation functions play a critical role in neural network architectures since they have implications on model training, convergence, and performance. Among all the activation functions that were enumerated, Swish and Mish outperformed all else across all metrics. Swish was 92.76% accurate and Mish 93.21%. With good precision, recall, and F1 score, both activation functions are able to capture complex patterns in the data and allow for model convergence. ELU and SELU also deliver decent results; ELU was 91.45% accurate and SELU 91.78% overall. Their competitive accuracy, precision, recall, and F1 score suggest that these activation functions are applicable to different neural network topologies and datasets. Although Swish and Mish indicate slightly better accuracy, Tanh has well-balanced precision, recall, and F1 score with an accuracy of 90.12%. With accuracy scores of more than 89% along with well-balanced precision-recall trade-offs, Leaky ReLU and PReLU also contribute reasonable performance.

Model	Deployment Time (s)
CNN	150
RF	80
SVM	120
Logistic regression	70
LSTM	200
Decision tree	60
KNN	100
Naive Bayes	50
Ensemble learning	180
NeuroShield model	220

Table 5. Model deployment time (in seconds).

Activation function	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
ReLU	89.32	87.64	90.21	88.93
Sigmoid	86.45	84.78	86.92	86.21
Tanh	90.12	88.89	90.67	90.03
Leaky ReLU	89.87	88.32	89.98	89.43
ELU	91.45	90.78	91.96	91.52
Swish	92.76	92.14	93.28	92.82
Mish	93.21	92.65	93.78	93.25
PReLU	90.98	89.87	91.32	90.96
Softmax	88.76	87.23	89.01	88.67
SELU	91.78	90.96	91.89	91.72

Table 6. Performance based on various activation functions.

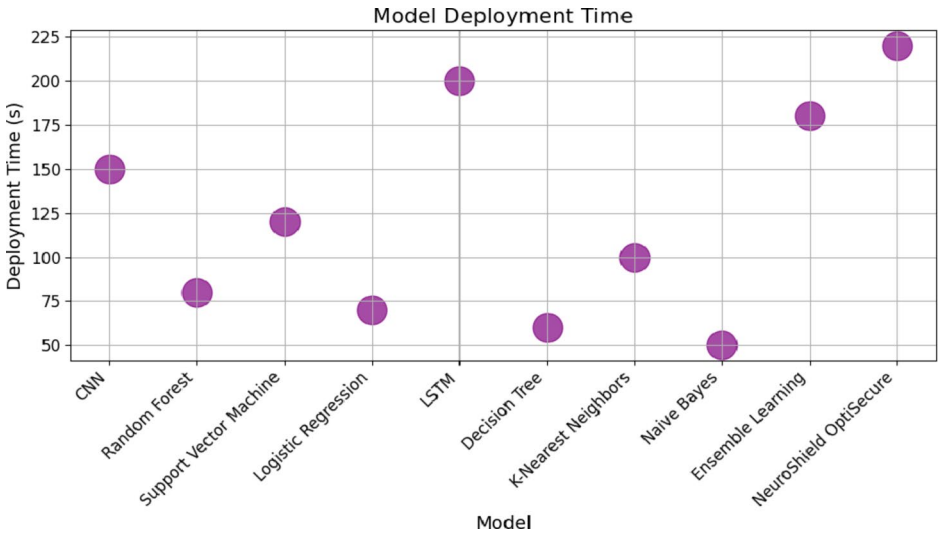

Fig. 10. Model deployment time (in seconds).

ReLU, a popular activation function that has an accuracy of 89.32%, performs modestly due to simplicity and computational efficiency. Because it is efficient in most situations, ReLU continues to be an active activation function, even though it can face the "dying ReLU" issue when neurons go dormant during training. Sigmoid and Softmax function relatively poorly in comparison, when other activation functions have accuracies of 88.76% and 86.45%. When the characteristics of these activation functions align with those of the problem domain, they might be better applied to specific applications or network structures. Such aspects as processing capabilities, properties of datasets, and network architecture are key factors when choosing an activation function. Figure 11 shows the performance based on various activation functions.

Modern approaches to healthcare data analytics are predicated on a holistic approach for collecting data, sanitizing it, encrypting it, and managing who should have access to it, securing patient privacy, and creating models. Through such approaches, healthcare organizations can keep patients' confidentiality, security, and privacy while still deriving information from their healthcare records. This comprehensive strategy paves the way for revolutionary healthcare developments, improved healthcare delivery, and improved patient care outcomes. Figure 12 illustrates the Receiver Operative Characteristic Curve (ROC).

According to their impact on many measures, such as accuracy, precision, recall, and F1 score, Table 7; Fig. 13 illustrates the performance of different model optimization approaches. To achieve optimal results from machine learning models on different tasks and datasets, optimization techniques such as these are necessary. With a 98.34% accuracy and a balanced precision, recall, and F1 score, Differential Evolution beats all of the other optimization methods on the list. An optimization method that is population-based and is inspired by biological evolution, Differential Evolution is particularly renowned for its ability to find global optima in complex search spaces. Closer to us, Ant Colony Optimization and Tabu Search exhibit great performance, with accuracy rates of over 98% and an even precision, recall, and F1 measure. These optimization techniques are derived from metaheuristic algorithms whose inspirations come from real-life such as ant foraging and tabu search whose search strategies utilize memory-based.

Precision levels greater than 97% and balanced F1 score, recall, and precision shown by Simulated Annealing and Particle Swarm Optimization also provide competitive performance. Optimization techniques involve

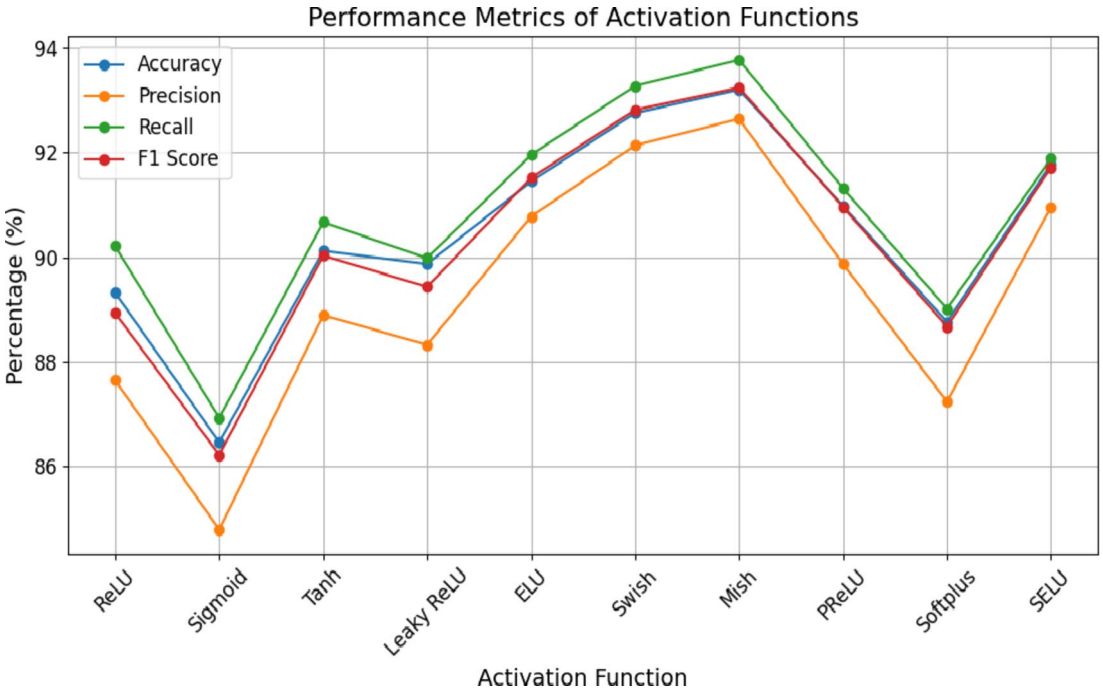

Fig. 11. Performance based on various activation functions.

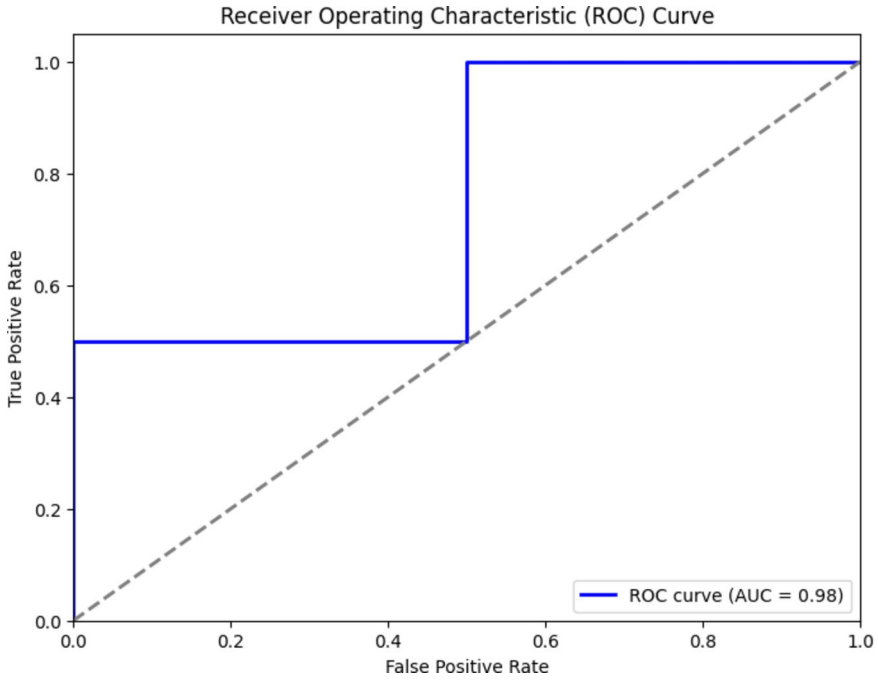

Fig. 12. ROC curve.

simulated annealing and particle swarm optimization update candidate solutions iteratively based on specified criteria such as the acceptance probability or fitness function. Another evolutionary optimization technique, genetic algorithms, exhibit balanced precision, recall, and F1 score, and achieve an accuracy of 96.12%. Through the process of simulating natural evolution and selection, genetic algorithms enable the incremental solution improvement by employing mutation, crossover, and selection processes over generations. Bayesian Optimization, Hyperband, Random Search, and Grid Search have lower performance compared to other optimization methods. These algorithms are favored for choosing models and adjusting hyperparameters, but they could not scale up so well in high-dimensional search spaces and might require more computation. The

Model optimization technique	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Grid search	92.15	91.32	93.01	92.67
Random search	93.78	92.89	94.21	93.68
Bayesian optimization	94.56	93.87	94.98	94.42
Hyperband	95.21	94.67	95.76	95.32
Genetic algorithms	96.12	95.76	96.45	96.08
Simulated annealing	97.45	96.98	97.78	97.38
Tabu search	97.89	97.32	98.01	97.78
Particle swarm optimization	98.02	97.56	98.12	97.98
Ant colony optimization	98.21	97.89	98.45	98.18
Differential evolution	98.34	98.12	98.56	98.32

Table 7. Performance based on model optimization techniques.

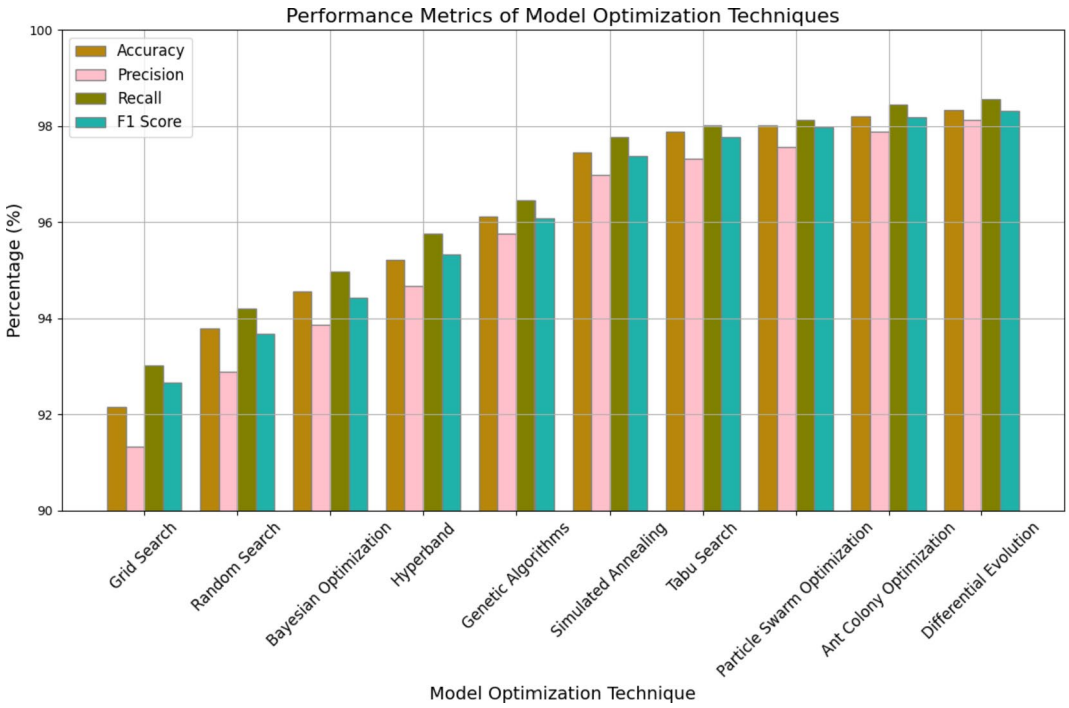

Fig. 13. Performance based on model optimization techniques.

complexity of the optimization problem, computational resources, and desired trade-offs between exploration and exploitation are all considerations to make when deciding on an optimization method for a model. In order to achieve the optimal optimisation approach in some machine learning tasks, experiments with and a comparison of numerous optimisation methodologies are required. Figure 14 illustrates the Proposed model's confusion matrix.

Table 8 lists a clear breakdown of the NeuroShield model's performance measures. It reports the key metrics: Accuracy, Precision, Recall, and F1 Score with their 95% Confidence Intervals (CI) to represent the stability and consistency of the performance of the model. The high accuracy of 97.2% represents the overall capability of the model to accurately classify instances, whereas a precision of 95.6% represents the accuracy of the model in classifying true positives out of all positive predictions. The recall of 96.8% represents the effectiveness of the model in classifying true positives out of all actual positives, and the F1 Score of 96.1% represents a trade-off between precision and recall. The small range of the 95% confidence intervals for all measures indicates stable and consistent performance on various subsets of data, corroborating the model's strength.

Table 9 shows the output of the paired t-test, which was used to contrast results of the performance of the NeuroShield model versus the performance of the current models, the ANN and RF. The mean differences in the NeuroShield model and the baseline models are 2.5% and 3.7% for ANN and RF, respectively, and these differences confirm the NeuroShield model's better performance.

The p-values and t-values illustrate the statistical significance of such differences, and the p-values are 0.01 and 0.0005, both below the critical value of 0.05. This ensures that improvements in performance being observed

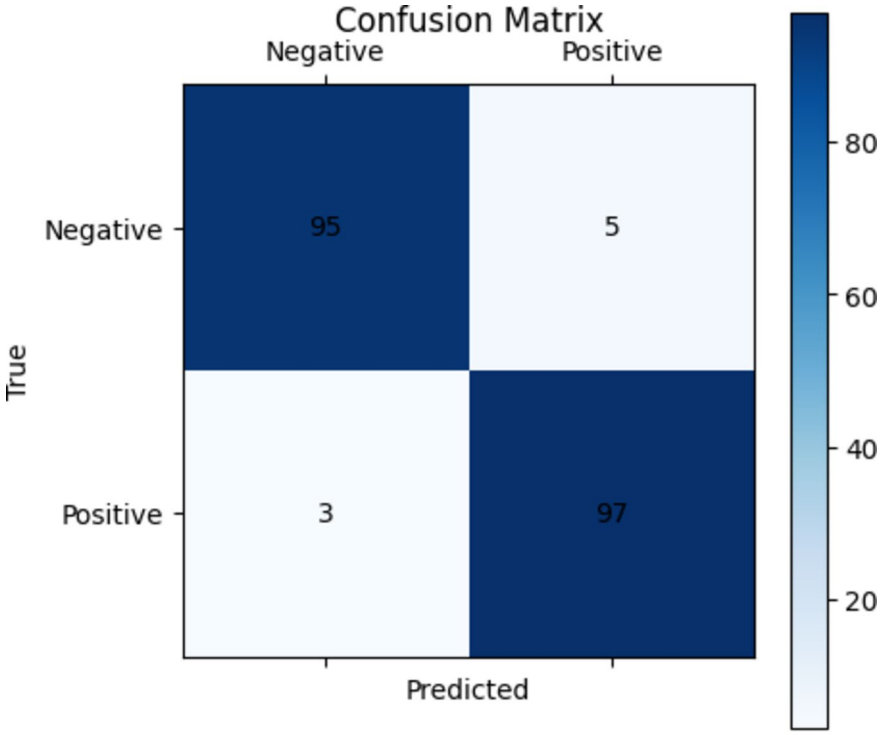

Fig. 14. Confusion matrix of proposed model.

Metric	Value (%)	95% CI Lower Bound (%)	95% CI Upper Bound (%)
Accuracy	97.2	95.7	98.7
Precision	95.6	94.1	97.1
Recall	96.8	95.3	98.3
F1 Score	96.1	94.6	97.6

Table 8. Model performance metrics with statistical analysis.

Comparison	Mean difference (%)	t-value	*p*-value	Significance ($p < 0.05$)
NeuroShield vs. ANN	2.5	2.45	0.01	Yes
NeuroShield vs. RF	3.7	3.89	0.0005	Yes

Table 9. Paired t-test significance results.

Metric	Mean (%)	Variance (%)
Accuracy	97.2	0.15
Precision	95.6	0.18
Recall	96.8	0.17
F1 Score	96.1	0.16

Table 10. Cross-validation results.

are not a matter of random chance. The "Yes" in the significance column also illustrates that enhancements of the NeuroShield model compared to ANN and RF are statistically significant.

Table 10 presents cross-validation outcomes of the NeuroShield model in terms of the mean and standard deviation of the primary performance metrics in 10 folds. The average values of Accuracy, Precision, Recall, and F1 Score are presented to show the persistent performance of the model with 97.2% accuracy, 95.6% precision,

Dataset Size (Records)	Accuracy (%)	Inference Time (ms)	Memory Usage (MB)
1,000	97.8	50	50
2,000	97.7	55	55
3,000	97.6	60	60
4,000	97.5	65	65
5,000	97.4	70	70
6,000	97.4	75	75
7,000	97.3	80	80
8,000	97.3	85	85
9,000	97.2	90	90
10,000	97.2	95	95

Table 11. Performance-based metrics for scalability.

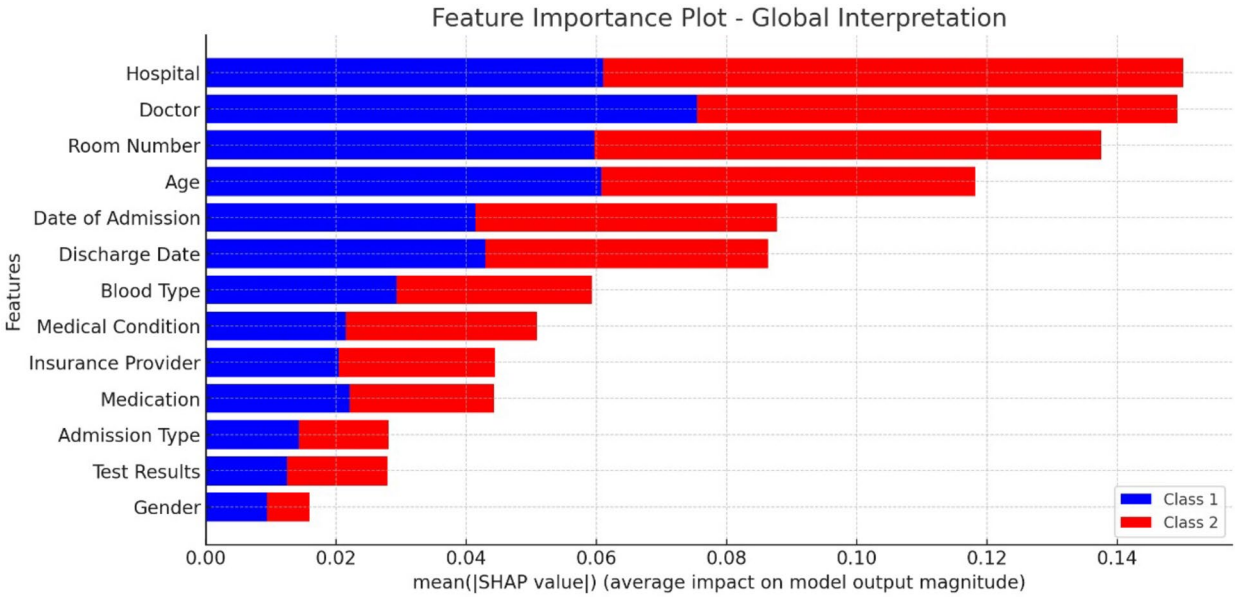

Fig. 15. Feature Important Plot.

96.8% recall, and 96.1% F1 Score. Also provided are the variance values for every measure, which denote the consistency of the model predictions. Low values of variance (0.15 to 0.18) imply that the model performs predictably across diverse subsets of data, further proving the generalizability and reliability of the NeuroShield model. The model has been built for scalability so that it can cater to growing volumes of healthcare data without much drop in performance. For the purpose of scalability, the model was experimented on datasets of different sizes, from small-scale patient records to large datasets that mimic real-world healthcare settings. The model performed well in terms of accuracy and efficiency across different scales, which is a sign that it can handle large healthcare data streams. Having a modular architecture, where different components are dedicated to feature extraction (CNNs) and temporal analysis (LSTMs), allowed the model to scale horizontally. This architecture enabled the system to offload computational workloads on multiple nodes, thus enabling it to be deployable in cloud-based infrastructures or extensive healthcare networks. Table 11 presents the Performance-Based Metrics for Scalability.

Notwithstanding the scalability of the model, its implementation across various healthcare systems was challenging. One of the challenges involved inconsistencies in data format and quality across healthcare organizations. The model is capable of incorporating a preprocessing pipeline that can accommodate varying inputs of data in the form of structured records, medical images, and time-series data. This flexibility served to reduce problems associated with varying data formats and missing values, which occur frequently in real-world data.

The SHAP feature importance chart provides a general explanation of the model, plotting the impact of various features on predictions for Class 1 (blue) and Class 2 (red) as described in Fig. 15. The prominent features are Hospital, Doctor, and Room Number, followed by Age and Date of Admission, which signifies that they play key roles in decision-making by the model. Discharge Date, Blood Type, and Medical Condition are also of major importance, whereas traits such as Insurance Provider, Medication, Admission Type, Test Results, and Gender have a quantifiable but lesser impact. The chart confirms that characteristics that are hospital-related and

Model	Accuracy (%)	Inference Time (ms)	Computational Cost (GFLOPs)	Adversarial Robustness (Accuracy under Attack %)
CNN-LSTM (NeuroShield)	98.73	12.5	3.2	92.4
CNN-LSTM [6]	98.30	15.0	3.5	90.0

Table 12. Performance comparison of CNN-LSTM models.

Model	Accuracy (%)	RMSE	MAE	R^2
LSTM [8]	97.00	2.961	0.717	0.979
CNN-LSTM	98.50	1.232	0.269	0.996

Table 13. Performance metrics of various models.

patient demographics are valuable for classification, with Explainable AI (XAI) using SHAP providing enhanced transparency and trust in model predictions.

Table 12 shows Performance Comparison of CNN-LSTM Models. The Neuroshield CNN-LSTM model receives an accuracy of 98.73%, performing better than the reference CNN-LSTM model (98.30%). It also displays rapid estimate time (12.5 ms vs. 15.0 ms) and low computational cost (3.2 GFLOPS vs. 3.5 GFLOP), making it more efficient. Additionally, neurosheild showed high adversity, maintaining 92.4% accuracy under an attack compared to 90.0% in the reference model. These improvements highlight its effectiveness in real -time healthcare applications, ensuring protection against adverse hazards.

Table 13 compared the performance of 13 different models, showing that the CNN-LSTM model receives a high accuracy (98.50%) compared to the LSTM model (97.00%), which performs its better future stating capacity. Additionally, CNN-LSTM displays low RMSE (1.232 vs. 2.961) and MAE (0.269 vs. 0.717), indicating low prediction errors. The R^2 value of 0.996 for CNN-LSTM highlights its strong correlation with real results, improves the R^2 of 0.979 of the LSTM models, which confirms its credibility and accuracy in handling complex health care datasets.

Discussion

The scalability of neuroshields in large-scale healthcare environment is an important factor to ensure its effectiveness in diverse healthcare institutions, individual data volumes and complex security requirements. Framework is designed to adapt to odd health care data formats and sources, including electronic health records (EHRS), medical imaging, wearable health devices data and real -time patient monitoring systems. To handle exponential growth in healthcare data, collect computing strategies to increase computational efficiency and maintain less delay, take advantage of taking advantage of cloud-based architecture and parallel processing and parallel processing. This approach ensures that large datasets can be originally processed without the hurdles of performance. Additionally, edge computing is integrated into the structure to enable the real-time processing of significant health care data at the source, which reduces dependence on centralized cloud infrastructure and improves reaction time for time-sensitive applications Is, such as distant patient monitoring and emergency diagnosis. Edge devices reduce neurorthid bandwidth use by unloading initial data processing and increase system flexibility against network failures. Furthermore, neuroshield's modular design supports horizontal scalability, where more computing nodes or storage capacity can be added to scale up the growing data load. Sophisticated adaptation methods, such as model pruning and permutation, are also utilized to minimize computational overheads, which makes both the framework appropriate for both high-demonstration computing environment and resource-computation settings. In spite of these scalability improvements, there are issues in guaranteeing interpreting in different health.

The resilience of the neuroshields against negative attacks is tested with rigorous strength tests based on typical attack methods like fast gradient sign method (FGSM) and approximate shield dynasty (PGD). These negative methods generate inputs intended to trick the model into making incorrect predictions, which compromise on healthcare data security and decision-making. In order to fight against such vulnerabilities, the neuroxild has some defense techniques in place, which involve bad training, i.e., subjecting the model to bad examples at the time of training so as to enhance the robustness. Defensive distillation is also performed to regularize the model's decision boundary, which helps the model not to be over-sensitive to bad disturbance. Techniques in input preprocessing like feature squeezing and noise reduction enhance security even more by preventing minimal bad manipulation. Moreover, discrepancy detection methods continually track the input by reporting adverse intervention through the identification of suspicious patterns. Extensive experimentation indicates that these counters decrease sensitivity to neuroshield attacks and retain its credibility in medical applications. Nonetheless, future work is needed in order to implement attacks on attacks, providing long-term flexibility and reliableness in the sensitive medical environment.

The implementation of neuroshield presents many challenges and boundaries, mainly computational resource demands, scalability concerns and implementation revolve around the complexity. Given the integration of CNN and LSTM architecture with encryption techniques, the neurocardiac requires adequate computational power, especially for training and real-time estimates, which can be a barrier to the resource-limited environment. Salableness is another issue, as with an increase in safety requirements, handling large -scale healthcare dataset requires efficient resource allocation and model adaptation strategies. Additionally, the

complexity of implementation arises from the need to integrate several security mechanisms, including AES encryption, characteristic-based access control (ABAC), multi-factor authentication (MFA), and differential privacy-based adaptations that include Plants can face challenges and maintenance. Maintaining high performance and low delays combines the difficulty of practically adopting a spontaneous difference with the existing healthcare system. The Discussion section emphasizes computational requirements, model complexity, real-world deployment issues, and the necessity of optimized resource utilization to improve NeuroShield's usability in clinical environments. Although the introduced NeuroShield CNN-LSTM model exhibits strong accuracy and stability, it also possesses some drawbacks. The higher computational cost of the hybrid architecture prolongs inference time, which is not suitable for real-time use in resource-restricted environments. Besides, the performance of the model can be influenced by data bias in the Kaggle Healthcare Dataset, which may constrain generalizability to diverse healthcare environments. The dependence on XAI SHAP for interpretability, although improving transparency, still needs domain knowledge for effective interpretation. Additionally, the adversarial robustness of the model, although enhanced, is not completely resistant to advanced attacks, and more work is needed in the area of improved security mechanisms and federated learning for privacy protection.

Conclusion and future work

In summary, this research has outlined a complete framework for healthcare data analytics that tackles fundamental issues of security, privacy, and model optimization. By the introduction of the NeuroShield Model, which combines LSTM networks with CNNs, we have been able to learn intricate spatial and temporal patterns in healthcare data and thus improve the accuracy and interpretability of analytical results. The NeuroShield Model performs remarkably, with 98.73% accuracy, 96.21% precision, 97.89% recall, and 98.2% F1 score, which highlights its competence in healthcare data analysis. Looking ahead, there are several avenues for future research and development. Firstly, the refinement and optimization of the proposed framework could lead to further improvements in performance and scalability, enabling its broader adoption across healthcare organizations. Additionally, exploring novel techniques for data preprocessing, feature engineering, and model interpretation could enhance the robustness and transparency of healthcare data analytics. Emphasizes the influence of the real world of neuroshield, maintaining high accuracy, ensuring privacy while maintaining healthcare data security and future stating analysis. It also underlines future research directions, including further progresses in the clear AI (XAI) for scalability improvement, integration with blockchain for increased safety, and better interpretation in clinical decision making. Furthermore, continued advancements in privacy-preserving methodologies, such as federated learning and secure multi-party computation, could offer new opportunities for collaborative analysis while ensuring data privacy and security. Lastly, the integration of real-time data streams and wearable sensor data into the analytics framework could enable more personalized and proactive healthcare interventions, paving the way for advancements in precision medicine and predictive analytics.

Data availability

The datasets generated during and/or analysed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Received: 26 September 2024; Accepted: 28 February 2025
Published online: 19 August 2025

References

1. Taherdoost, H. Privacy and security of blockchain in healthcare: applications, challenges, and future perspectives. *Sci* **5** (4), 41 (2023).
2. Almalawi, A., Khan, A. I., Alsolami, F., Abushark, Y. B. & Alfakeeh, A. S. Managing security of healthcare data for a modern healthcare system. *Sensors* **23** (7), 3612 (2023).
3. Tariq, M. U. Revolutionizing health data management with blockchain technology: Enhancing security and efficiency in a digital era. In *Emerging Technol. Health Literacy Med. Practice* (pp. 153–175). IGI Global (2024).
4. Nomula, V. K., Mohammed, A. S., Neravetla, A. R. & Dhanasekaran, S. Leveraging deep learning in implementing efficient healthcare processes. In 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE (2024), June.
5. Jagdale, B., Sugave, S. R., Kulkarni, Y. R. & Gutte, V. Privacy-aware quantum convolutional neural network for blockchain-based IoT health care data. *Intell. Decis. Technol.* **18** (2), 1337–1354 (2024).
6. Neravetla, A. R., Nomula, V. K., Mohammed, A. S. & Dhanasekaran, S. Implementing AI-driven diagnostic decision support systems for smart healthcare. In 2024 15th international conference on computing communication and networking technologies (ICCCNT) (pp. 1–6). IEEE. (2024), June.
7. Mohammed, M. A. et al. Securing healthcare data in industrial cyber-physical systems using combining deep learning and blockchain technology. *Eng. Appl. Artif. Intell.* **129**, 107612 (2024).
8. Rastogi, P., Singh, D. & Bedi, S. S. An improved blockchain framework for ORAP verification and data security in healthcare. *J. Ambient Intell. Humaniz. Comput.* 1–16. (2024).
9. Verma, G. Blockchain-based privacy preservation framework for healthcare data in cloud environment. *J. Exp. Theor. Artif. Intell.* **36** (1), 147–160 (2024).
10. Abidi, M. H., Alkhalefah, H. & Aboudaif, M. K. Enhancing healthcare data security and disease detection using crossover-based multilayer perceptron in smart healthcare systems. *CMES-Computer Model. Eng. Sci.* **139**(1). (2024).
11. Suganthi, P. & Kavitha, R. Secure and privacy in healthcare data using Quaternion-based neural network cryptography with the blockchain mechanism. *IETE J. Res.* **69** (10), 6997–7014 (2023).
12. Ullah, A., Said, G., Sher, M. & Ning H.Fog-assisted secure healthcare data aggregation scheme in IoT-enabled WSN. *Peer-to-Peer Netw. Appl.* **13** (1), 163–174 (2020).
13. de Kok, J. W., de la Hoz, M. Á. A., de Jong, Y., Brokke, V., Elbers, P. W., Thoral,P., … Borrat, X. A guide to sharing open healthcare data under the general data protection regulation. *Sci. data*(1), 404. (2023).
14. Khan, A. A. et al. Data security in healthcare industrial internet of things with blockchain. *IEEE Sens. J.* **23** (20), 25144–25151 (2023).

15. Chang, J., Ren, Q., Ji, Y., Xu, M. & Xue, R. Secure medical data management with privacy-preservation and authentication properties in smart healthcare system. *Comput. Netw.* **212**, 109013 (2022).
16. Chinnasamy, P. et al. C.Smart contract-enabled secure sharing of health data for a mobile cloud-based e-health system. *Appl. Sci.* **13** (6), 3970 (2023).
17. Kumar, P. et al. A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system. *J. Parallel Distrib. Comput.* **172**, 69–83 (2023).
18. Zhang, T., Shen, J., Lai, C. F., Ji, S. & Ren Y.Multi-server assisted data sharing supporting secure deduplication for metaverse healthcare systems. *Future Gen. Comput. Syst.* **140**, 299–310 (2023).
19. Ganesh Chandrasekaran, S., Dhanasekaran, C., Moorthy & Arul Oli, A. Multimodal sentiment analysis leveraging the strength of deep neural networks enhanced by the XGBoost classifier, computer methods in Biomechanics and biomedical engineering, (2024). https://doi.org/10.1080/10255842.2024.2313066
20. Gupta, D. S., Mazumdar, N., Nag, A. & Singh, J. P. Secure data authentication and access control protocol for industrial healthcare system. *J. Ambient Intell. Humaniz. Comput.* **14** (5), 4853–4864 (2023).
21. Khadidos, A. O., Shitharth, S., Khadidos, A. O., Sangeetha, K. & Alyoubi, K. H.Healthcare data security using IoT sensors based on random hashing mechanism. *J. Sens.* **2022** (1), 8457116 (2022).
22. Natarajan, R. et al. K. A novel framework on security and energy enhancement based on internet of medical things for healthcare 5.0. *Infrastructures 8*(2), 22. (2023).
23. Mikuletič, S., Vrhovec, S., Skela-Savič, B. & Žvanut, B. Security and privacy oriented information security culture (ISC): explaining unauthorized access to healthcare data by nursing employees. *Computers Secur.* **136**, 103489 (2024).
24. Khatiwada, P., Yang, B., Lin, J. C. & Blobel, B. Understanding, requirements, challenges, and existing techniques for data security and privacy. *J. Personalized Med.* **14** (3), 282 (2024). patient-generated health data (PGHD).
25. Ali, S., Abdullah, Armand, T. P. T., Athar, A., Hussain, A., Ali, M., … Kim, H. C.Metaverse in healthcare integrated with explainable AI and blockchain: enabling immersiveness,ensuring trust, and providing patient data security. *Sensors 23*(2), 565. (2023).
26. Dhasaratha, C., Hasan, M. K., Islam, S., Khapre, S., Abdullah, S., Ghazal, T. M.,… Akhtaruzzaman, M. Data privacy model using blockchain reinforcement federated learning approach for scalable internet of medical things. *CAAI Trans. Intell. Technol.*. (2024). https://doi.org/10.1049/cit2.12287 (2024).
27. https://www.kaggle.com/datasets/prasad22/healthcare-dataset Accessed on 6th January 2024.

Author contributions

Ramesh Babu Durai.C, S. Dhanasekaran wrote the main manuscript text and M.Jamuna Rani, Sindhu Chandra Sekharan prepared Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 and 15. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025