Asian J Biol Life Sci., 2025; 14(3):747-757.
https://ajbls.com

Research Article

Privacy Preservation of Genomic and Medical Data

Sahiti Chamarthy[1], Nageswara Rao Moparthi[3], Kommaragiri Raghava Rao[4], Venkata Rakesh Chintala[5], Janaki Ramaiah Mekala[1,2,*], Charanya Gade[4], Reshitha Talluri[4]

[1]Department of Biotechnology, Koneru Lakshmaiah Education Foundation (KLEF), Guntur, Andhra Pradesh, INDIA.
[2]Department of Biosciences, School of Biosciences and Technology (SBST), Vellore Institute of Technology (VIT), Vellore, Tamil Nadu, INDIA.
[3]Amrita Vishwa Vidyapeetham, Amaravati, Andhra Pradesh, INDIA.
[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation (KLEF), Guntur, Andhra Pradesh, INDIA.
[5]Department of Metabolic Disorders, Sanjeevini Hospitals, Vijayawada, Krishna, Andhra Pradesh, INDIA.

ABSTRACT

Background: Cancer, characterized by abnormal and uncontrolled cell division, is a major global health concern and a prominent area of research. When a patient is diagnosed with a medical condition, their data is collected for further diagnosis, inference, and research. Given the genetic nature of DNA, the privacy of genomic data is critical, as it affects not only the research subjects but also their families and future generations. Errors in genetic data management can have long-term consequences, making secure data handling essential. **Materials and Methods:** Various methods have been employed to ensure the privacy and preservation of genomic data, including cloud computing, deep learning, and other advanced tools. These technologies aim to protect patient information while allowing controlled access for medical research, drug discovery, and clinical trials. Strict confidentiality must be maintained between patients and healthcare professionals, with appropriate legal and ethical considerations in place before sharing data. **Results:** The disclosure of medical and genomic data without consent can lead to harassment, discrimination, and emotional distress for patients. To address these concerns, innovative technologies, algorithms, and software have been developed to securely store and manage patient information. These advancements help mitigate risks associated with data breaches and unauthorized access. **Conclusion:** Enhancing privacy and security in medical and genomic data is crucial to prevent ethical and societal complications. When data is required for research purposes, obtaining informed patient consent through proper documentation and legal processes is necessary. This review highlights some of the latest algorithms and technologies that ensure the safe preservation of genomic data.

Keywords: Privacy, Cancer, Preservation, Genetic Data.

Correspondence:
Dr. Janaki Ramaiah Mekala
Professor, Department of Biotechnology, Koneru Lakshmaiah Education Foundation (KLEF), Guntur-522302, Andhra Pradesh, INDIA.
Email: janakiramaiah@kluniversity.in

Received: 21-08-2025;
Revised: 13-10-2025;
Accepted: 05-12-2025.

INTRODUCTION

2003's Human Genome Project[1] shows the value of genetic data. Next-generation sequencing technologies have made it possible to conduct whole genome sequence studies on many kinds of organisms. Preserving the genome involves effectively handling records of various physical damages.[2] Two Different means of conserving medical records like cleaning records, air-conditioners proper handling, fumigation for prevention of insect and rodent attacks, sun shades, energy-efficient lights for regulating temperature, fire extinguishers, microfilming of the records, maintaining proper security of the records so that no one has unauthorized access to records. Nevertheless, these conventional methods can be long and tiresome; hence, people now rely on software-based techniques such as cloud computing, deep learning, etc. Medical information is described by volume, velocity, variety, and value (4V).[3] Big data in medicine comes in many forms, such as gene and EMR data. EEG signals are now a vital source of data for medical information due to extensive research into deep learning and brain-computer interfaces.[4-6] Medical records may be used for public or private purposes. 1) Personal document-save in a few limited circumstances; this material should not be disclosed without the patient's consent as it is confidential. 2) Impersonal document-the record no longer qualifies as a personal record, and patient consent is no longer necessary. These documents could be helpful for study.[7] Application Programming Interfaces (APIs), Trusted Digital Repositories (TDRs), emulation, encapsulation, migration, normalization/conversion, cloud computing, refreshing, backup, and byte replication are some of the preservation techniques.[8] TDRs facilitate the long-term upkeep and conservation of digital records.[8-10] Refreshing makes it easier to copy data between media without sacrificing quality.[8,9,11] Byte replication enables

OPEN ACCESS

DOI: 10.5530/ajbls.20251484

Publishing Partner : ScienScript Digital. [www.scienscript.com.sg]

file storage and multiplication for a variety of systems.[8,9] Backup also includes creating multiple copies and keeping them in different distant places.[8] Emulation is a migration strategy that replicates an old, out-of-date system.[8,12,13] Metadata is the information that is recorded with digital records to characterize each one independently from other records. InterPARES 3 Project, 2013,[8] Encapsulation includes keeping digital records in their original format and including the necessary metadata.[8,14] Records are transferred from an outdated system to a recently established system for recordkeeping during migration.[8,13] The process of "converting digital records to standard format" that can be preserved is known as normalization.[8] Keeping records in (cloud) storage run by a third party is called cloud computing.[8] APIs provide a user interface that is easy to access recorded social media information (Digital Preservation Coalition, 2016a). Employing preservation file formats entails keeping in mind the format of the records prior to preservation (Figure 1).[8]

Cancer

Cancer is a growth formed by an organ as a result of organ defects and is also among the primary reasons for world fatalities. Cancer is not simple to identify at the initial stage. Additionally, if cancer is present beyond a specific stage, it will once more increase after healing. Medical image analysis is usually used to identify abnormalities in the body, including breast cancer, lung cancer, and brain tumor. Imaging scans are required for the detection of cancers such as breast, lung, and brain cancer. Developments in Artificial Intelligence (AI) and machine learning improve the quality of imaging such that it is faster and more accurate, increasing accuracy in diagnosis. Proper and prompt diagnosis greatly increases the chances of cure and survival.[15-21]

A New Chaos-Driven Privacy-Preserving Deep Learning Framework for Cancer Detection.

Cancer's primary signs include tiredness or excessive exhaustion, eating difficulties, swelling, inflammation, or bumps in a bodily part, discomfort, jaundice, coughing up blood, a fever, headache, and issues with eyesight or hearing.[22]

When doing genomics science, two of the fundamental values of science research must be weighed:

The necessity to safeguard research participants' confidentiality.

The need to share data extensively in order to maximize its usefulness for ongoing scientific research.

Acts

Confusion arises regarding who owns the genetic data when DNA/ tissue is removed from the body. Reports have suggested that it no longer belongs to the individual from whom it has been obtained but to the hospital/organization where it has been given reason being the signed consent forms by the individual. There is also a risk of the data being misused by the authorities leading to genetic discrimination and preferential treatment in cases of therapies. To curb this problem. Congress signed the Genetics Data Non-discrimination Act (GINA) into law in 2008. To safeguard patient medical records, the Health Insurance Portability and Accountability Act (HIPPA) was enacted.

Laws and Regulations in the USA

Several rules and regulations protect participants in combined capitalized research. The United States Policy for the Protection of Human Subjects (often called the Common Rule), published in 1991, established the minimum ethical norm for government-funded human subjects research in America. The Common Rule was updated in 2017 to "modernize, simplify, and strengthen" supervision. The Common Rule mandates that each participant in all federally supported research initiatives under its definition of "human subjects" prior to participation. Additional federal rules and regulations protect the clinic, insurance, and employment sectors (Figure 2).

(A) Genetic notice law also known as informed consent will put constraints on data gathering, and analysis of DNA samples. (B) Genetic anti-discrimination law uses genetic information in making employment decisions such as hiring. (C) Genetic redisclosure law. Requires written consent of the individual every time a DNA sample or DNA sample analysis result is shared with a third party.

MATERIALS AND METHODS

Privacy Protection, Personalized Medicine, and Genetic Testing

There are numerous confidentiality choices for processing genetic information. Lauter *et al.* 2015 adapted some of the procedures applied in Genome-Wide Association Studies (GWAS) to process genomic information encrypted using homomorphic encryption.[23] These alternatives include.

Varlock method

This technique protects the beneficial, non-sensitive characteristics of structured DNA segments for scientific investigations while masking individual alleles inside genomic sequences. This technique can store and share raw aligned reads with an enhanced security layer. Individual alleles in the required gene sequences may be recovered and made available to patients, medical facilities, and academic researchers.[24]

Data masking strategy

It maintains the privacy of the patient and the confidentiality of data. It enables (direct care) direct use of medical records by approved healthcare professionals and privacy-respecting (non-direct care) secondary use by researchers. It facilitates the identification of the type of genetic data by augmenting the simple pseudonymization method with query-able encryption.

Novel Data Privacy-Preservation Protocol (NDPPP)

This outsources the files in the encrypted form to the cloud. The data consumers can download the encrypted files efficiently from the cloud service provider without any loss of data. An authentic protocol of authentication is formed between the trusted intermediary and the data recipient to mitigate various attacks. Moreover, the security was strengthened. Given the data leakage, data loss as well as data tampering can also be prevented and secure from different various forms of attacks such as impersonation attacks, eavesdropping, man-in-the-middle attacks, and a replay attack.

Cloud computing

Personalized medicine uses a patient's DNA data to identify problems and treat them. The future of healthcare will be completely transformed by the new model. With the aid of inexpensive personal genomics services, a doctor will use DNA sequence matching to diagnose and treat illnesses instead of standard clinical laboratory testing. Given that DNA data and computation are usually very large, cloud computing will be a suitable computing model. However, because DNA data is sensitive, it must be encrypted before being outsourced to the cloud. Cloud computing promises to provide a stronger privacy guarantee than current schemes by protecting DNA data privacy and search patterns. Our scheme needs just one communication round, which is required in real-world situations, in contrast to other interactive schemes currently in use. According to simulation results, communication costs are limited and computation overhead for real-world problem computation is acceptable. Furthermore, our scheme can be extended to general privacy-preserving pattern-matching problems that are commonly used in real-world applications, instead of just the genome matching problem (Figure 3).

The diagram shows a four-part system that uses personalized medicine. The doctor, a lab, or a personal genomics service provider sequences the patient's DNA from cell samples. After that, the DNA breakdown is encrypted and moved to the cloud. When a certified user, like a doctor, wants to perform genetic screening on a patient, they can order the test via the cloud and receive the results. To protect patient confidentiality, test sequencing-that is, a DNA pattern sequence-would also need to be encrypted. A trapdoor is the encrypted order. Genetic material is secured using Predicate Encryption (PE).

Under the clear subgroup decision assumption within a bilinear group, our method is provably secure. A proper authority such as a health care provider can send a request for genetic testing to the cloud based on a designated query sequence pattern. Search patterns and test outcomes are protected from the cloud server to provide strong privacy guarantee. Furthermore, the consent of the patient is necessary before the authorized entity can access anything. The genetic sequence was pre-processed with a suffix tree in a way that the computation of matching the sequences could be done within one round of communication.

Internet of Things-based Medical Big Data Privacy Protection Platform (MNSSp3)

It is a developed medical data-sharing platform that makes data analytics and data security tools available for a variety of data types. The system concentrates on big medical data transmission and sharing security and completes user and data segmentation to ensure medical data security in order to provide mining tools to users. It also gives users the option to independently add privacy algorithms (Figure 4).

Application Layer

The platform's application layer allows users to access data mining services, and its privacy policies can be broadly divided into two categories: This privacy methodology is applied by "predicting," and it primarily provides intelligent diagnostic research services; it can be applied to EMR or EEG data. SP1 type is "separating users from data," plus "protection with differential privacy the algorithm," and it primarily provides gene data query services and EMR data mining services. Transmission Layer: All of the medical data received from the last layer is kept in the database to identify different options, and the transmission layer can handle real-time data while verifying the accuracy of the information received.

Perception Layer

The perception layer is where data updates must be handled. This system database primarily stores information from open databases, associated hospitals, and research organizations. An IoT device can sense and stream private health records into the platform to help the user maintain his health. The platform provides a data update interface that allows medical information to be updated and access to the platform in real-time.

GWAS, or genome-wide association studies

GWAS model on united genetic records that protects privacy. We use secure Multi-Party Calculation (MPC) platforms to overlay the GWAS computations. In this manner, two participants in an online system can carry out safe GWAS calculations together without sharing their private information with third parties. For rapid prototyping, we use a state-of-the-art MPC framework called Compact Circuit Format (PCF).[25] Our experimental findings show that it is possible to obtain cross-institution GWAS computations that are both secure and efficient (Figure 5).

In the entire execution of the suggested layout, PCF has been used, beginning with Python data input analysis and moving on to automatically generated code via Bourne Once more Shell (Bash). The PCF/LCCYao compiler compiles the generated code, which

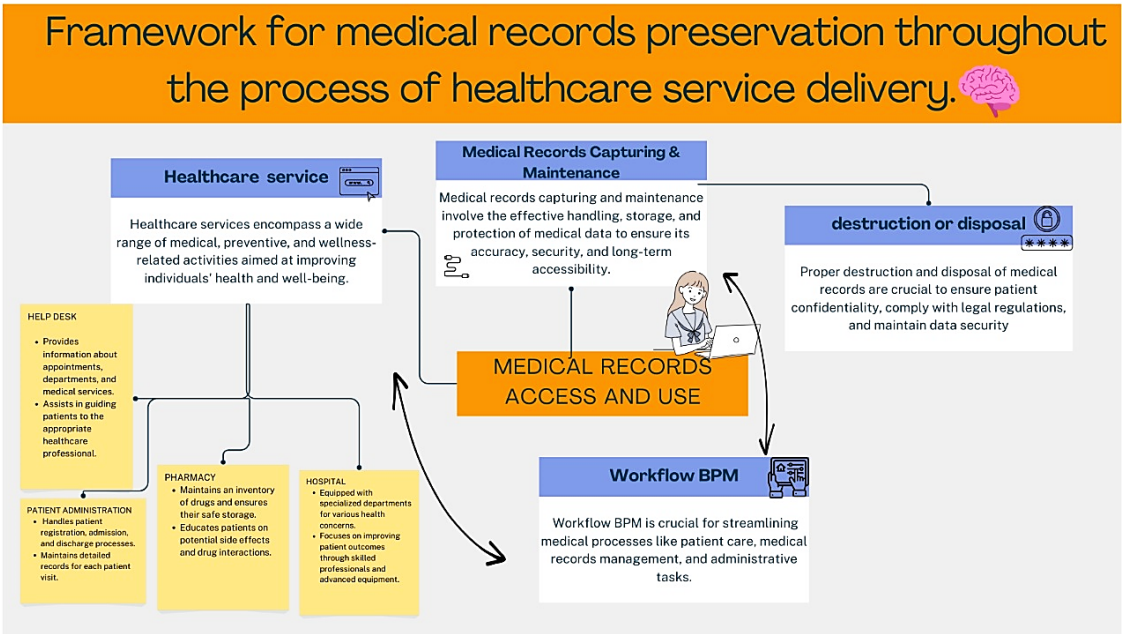

Figure 1: System of preserving medical records through the healthcare service delivery process.

is then run by the BetterYao runtime system and PCF simulator. Bash is used to construct various program components.

A distributed genetic data management system that uses close privacy settings (LDP) to maintain privacy

Offer a blockchain-based, within-differential privacy (LDP)-based, secure genetic information management system. Two storage types are used by the system that is recommended: semi-private storage for external users and private storage for internal employees. Only employees have access to the protected gene information because it is kept in private storage. Meanwhile, LDP permanently alters gene information that is kept in a semi-public format. Through LDP, noises get introduced into genomic data in all directions. In line with this, the owner's privacy is protected when common data is shared or utilized by a third party (Figure 6).

The concept makes use of a blockchain platform that has two storage options: private, semi-private storage and a Decentralized Application (DApp). In order to maintain privacy and security, even if the gene information is stored on both storage media, the sections that contain the owners' information have been hidden by both protecting them or including random genes. Semi-private storage is used to store gene data that contains noise. On the other hand, personal storage is the place where protected gene data are kept. With DApp, the gene data owner may monitor usage and communication information.

Deep Learning Method

To ensure secure transmission and prevent data theft, the collected data is encrypted before being sent across the channel. The effectiveness of the encryption method is evaluated using key security metrics such as correlation, entropy, contrast, structural content, and energy. For cancer diagnosis, a Convolutional Neural Network (CNN)-based model is applied to Magnetic Resonance Imaging (MRI) data. The model is enhanced using techniques like K-fold analysis, fine-tuning, and transfer learning to improve accuracy and performance (Figure 7).

The paradigm for cancer clinical diagnosis while protecting privacy is developed in two ways in this proposed research. First, the approaches are bit-plane extraction, DWT, and chaos to improve healthcare data security. Second, a CNN-based deep learning model is used to create the cancer diagnosis model. The clinical data is secured in the form of Magnetic Resonance Imaging images. After that, this information is sent to another testing facility without a non-invasive diagnostic device over a transmission link assisted by the Internet. Email can be used to communicate data; while safe, it is not sufficiently secure to guarantee data privacy. As a result, picture encryption is used to transmit MRI images securely. At the receiver end, these images are decrypted using the suggested encryption algorithms inverse.

k-PPD-ERT: A Distributed Privacy-Preserving Approach to Extremely Randomized Trees

Different data records were maintained across various sources, based on the assumption that the learning data was spatially partitioned. The focus of this work is on structured health data, such as those stored in spreadsheets, and the problem of classification. Our previous research introduced[26] a scalable framework for integrated machine learning that protects privacy and is based on the technique of the highly randomized tree. This approach has a linear overhead in relation to the number of cooperating parties, but it can effectively manage missing values. With k denoting the number of cooperating parties, we call our method k-PPD-ERT (Privacy-Preserving Distributed Extremely

Randomized Trees). We use the Wisconsin Diagnostic[28] and Heart Disease[27] databases, two well-known healthcare datasets, to assess performance.

Random Decision Tree (RDT)

The Bayes method of classification has a more complicated variant called the RDT. In contrast to other techniques, RDT is more effective at preserving current knowledge and avoiding information loss. A system dataset of diabetic patients was created to verify RDT's performance. The RDT technique efficiently retrieved and preserved the data with reduced calculation time and without information loss. A random design protects against discovering the full classification model or cases using a priori knowledge (Figure 8).

Deep Learning Natural Language Processing (DL NLP)

A technique for sharing a DL NLP model among cancer registries while protecting patient privacy by removing any information that could compromise patient privacy. This makes it possible for cancer registries to share information using NLP models without going against data privacy regulations.[29] Additionally, the suggested transfer learning with privacy-preserving models performs on par with the centralized model and the traditional transfer learning strategy without privacy-preserving.

Multiple (or m) iso-depth RDTs are constructed by the RDTs algorithm. Training and categorization are the two phases of the RDT algorithm. The fact that a random tree's structure is built entirely independently of the training data is one of the critical characteristics of RDTs.

Privacy of Cancer Data and Solutions

Cancer data should be highly classified as it can be extended to the patient's mental, emotional, and economic disadvantage. Legislative initiatives have been made at the state level to strengthen and safeguard the privacy of cancer data. Each cancer registry may have stringent guidelines for managing files and papers containing privileged information. The privacy of cancer

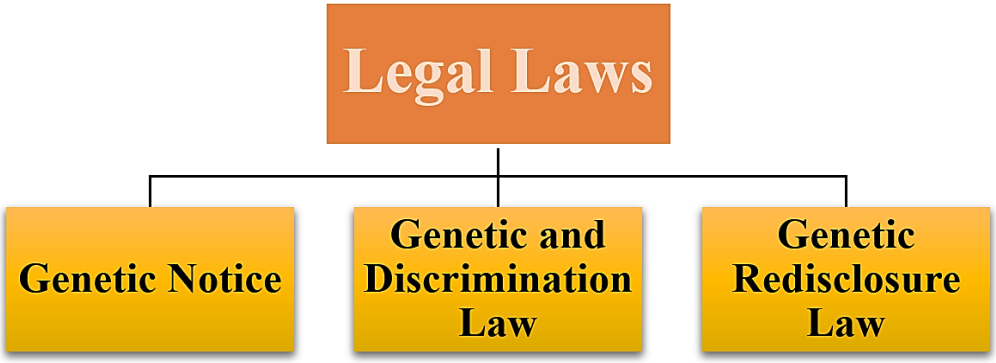

Figure 2: Legal approaches for genetic privacy.

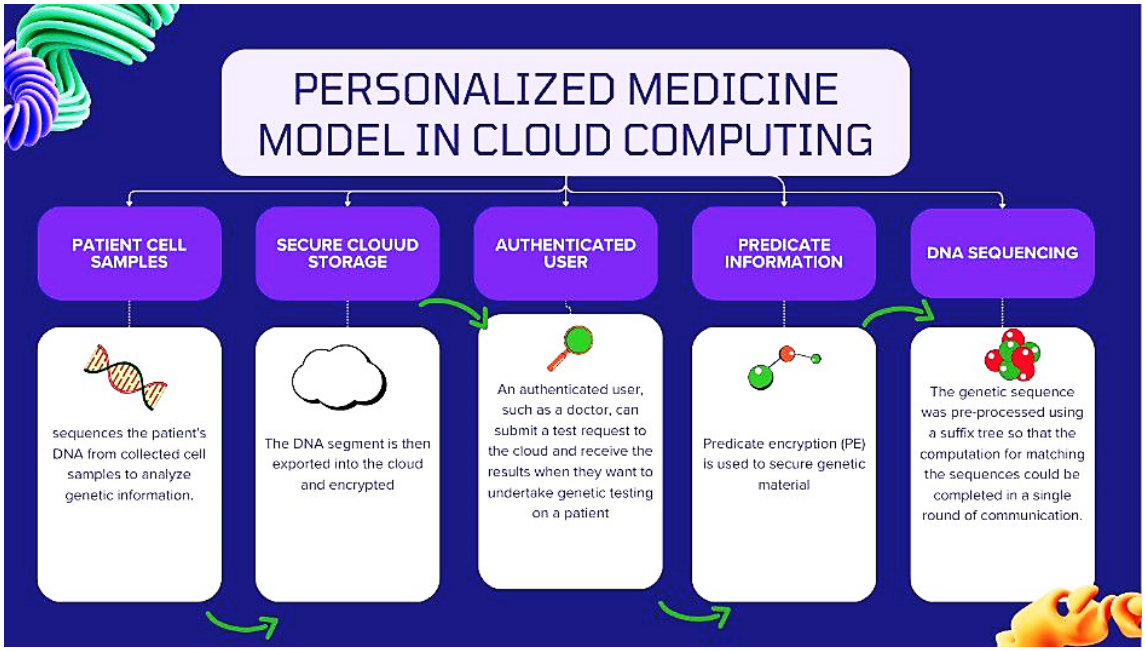

Figure 3: Cloud-based framework for personalized medicine.

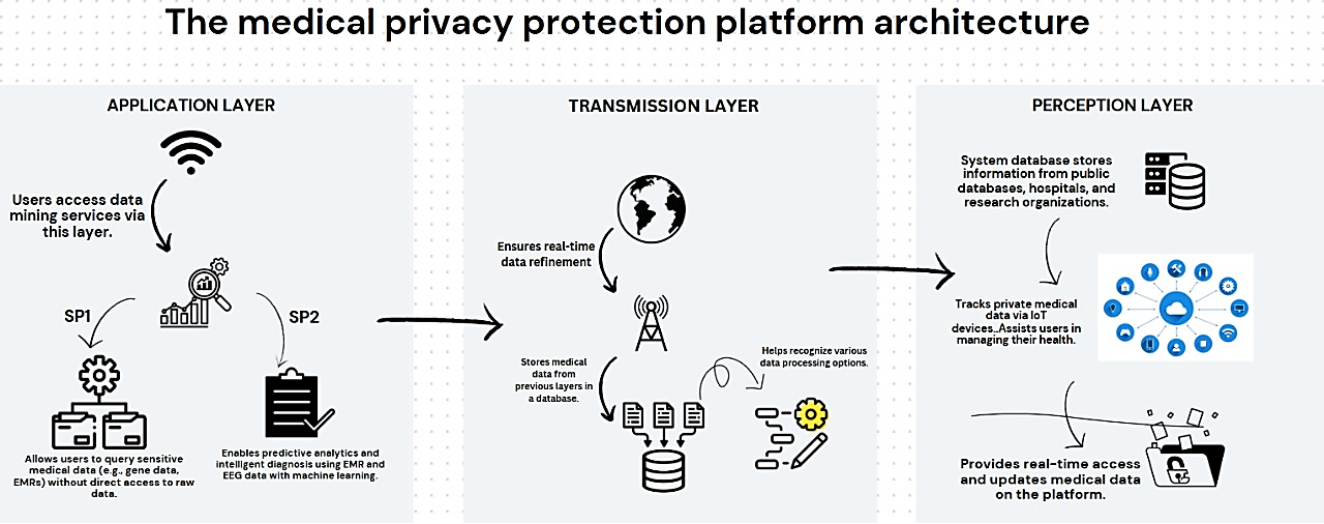

Figure 4: The architecture of the healthcare privacy protection platform.

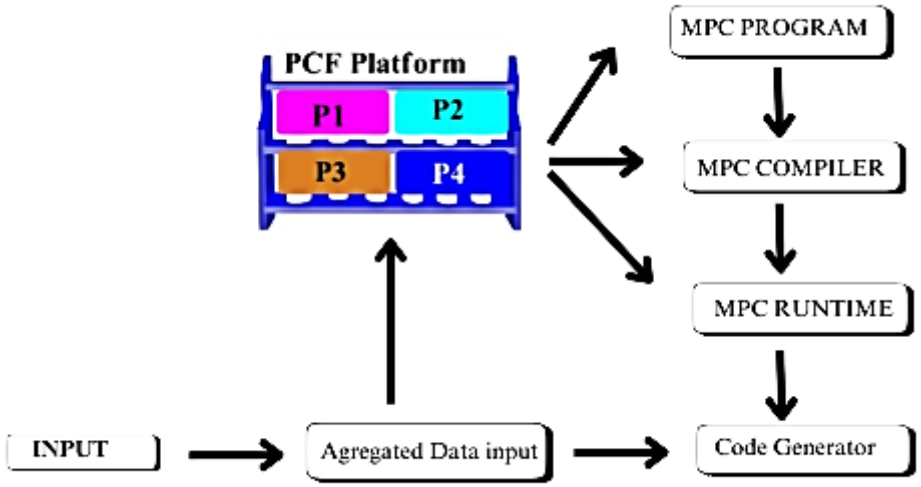

Figure 5: Federated genomic data collections facilitating privacy-enhanced GWAS analysis (System workflow).

patients is safeguarded because of such policies and processes. The importance of cancer registries comes from the fact that they gather precise and comprehensive cancer data that may be used for epidemiology and cancer control research, organizing public health programs, and enhancing patient care. In the end, all these initiatives lessen the impact of cancer. "A national system of cancer registries can help us understand the illness better and use our resources to the best effect in prevention and treatment," said Dr. Donna Shalala, a former Department of Health and Human Services secretary. Hospital-based registries and population-based registries are the two main categories of cancer registries. A specific cancer registry is a different kind of cancer registry. These are cancer registries created to gather and preserve information about a specific type of cancer. For instance, the Gilda Radner Familial Ovarian Cancer Registry is a unique cancer registry that collects data from households with two or more members who have the disease. Further specialized cancer registries compile information on lung, colorectal, or brain malignancies. These specialized cancer registries frequently offer learning opportunities for those interested in discovering more about a particular type of cancer as well as assistance for those who may be affected by it.

Contrarily, Rogith, *et al.* 2014 reported low levels of patient concern over the privacy of genomic data in their study.[30] Even if they are unlikely to directly benefit, a sizable majority of people with cancer are prepared to give their agreement to share their genomic data with research groups because they may understand the therapeutic and research benefits of genetic tests.

Homomorphic Encryption (HE)

One of the several finest options for cancer genomic data inference that protects privacy is the organization of the alteration data using biological intuition and bringing the dimensionality down to a manageable level. This genuine genomic dataset shows that

our method obtains a cancer prediction MAUC of 0.98 on the test dataset and can compute encoded genomic data at a rate of less than one second per patient (Figure 9).[31]

The training dataset for this threat mode comprises those who have consented to share their data and are available to the public. The training dataset will be used to create models for cancer prediction rather than being protected. Data must be secured when a new patient wishes to use their genome to test for cancer. So, in this threat model, the inference is private, but the training is not privacy-preserving.

Federated Learning

The infrastructure established and demonstrated here provides researchers with a systematic way to interact with cancer data regularly collected from multiple radiation oncology clinics. Consequently, this infrastructure has been evaluated through an application across a network of hospitals, assessing the performance of distributed/federated learning models and a data extraction pipeline for radiation oncology.[32]

The Cancer Biomedical Informatics Grid (caBIG)

This is the National Cancer Institute (NCI) Center for Bioinformatics is designing a data grid. Furthermore, it included NCI Specialized Programs of Research Excellence, NCI Community Cancer Centers, NCI-designated cancer centers, Clinical Trials Cooperative Groups, and other academic and business participants, adding up to more than 1,000 people at more than 80 institutions. This helps to connect researchers, doctors, and patients across the cancer community to create distributed computing systems that can speed up scientific breakthroughs and enhance patient outcomes. By integrating rigorous peer review into software development, connecting data models to a description-logic ontology underneath, facilitating robust data typing, and aiming to provide registered models for the information and records connected to an application.[33]

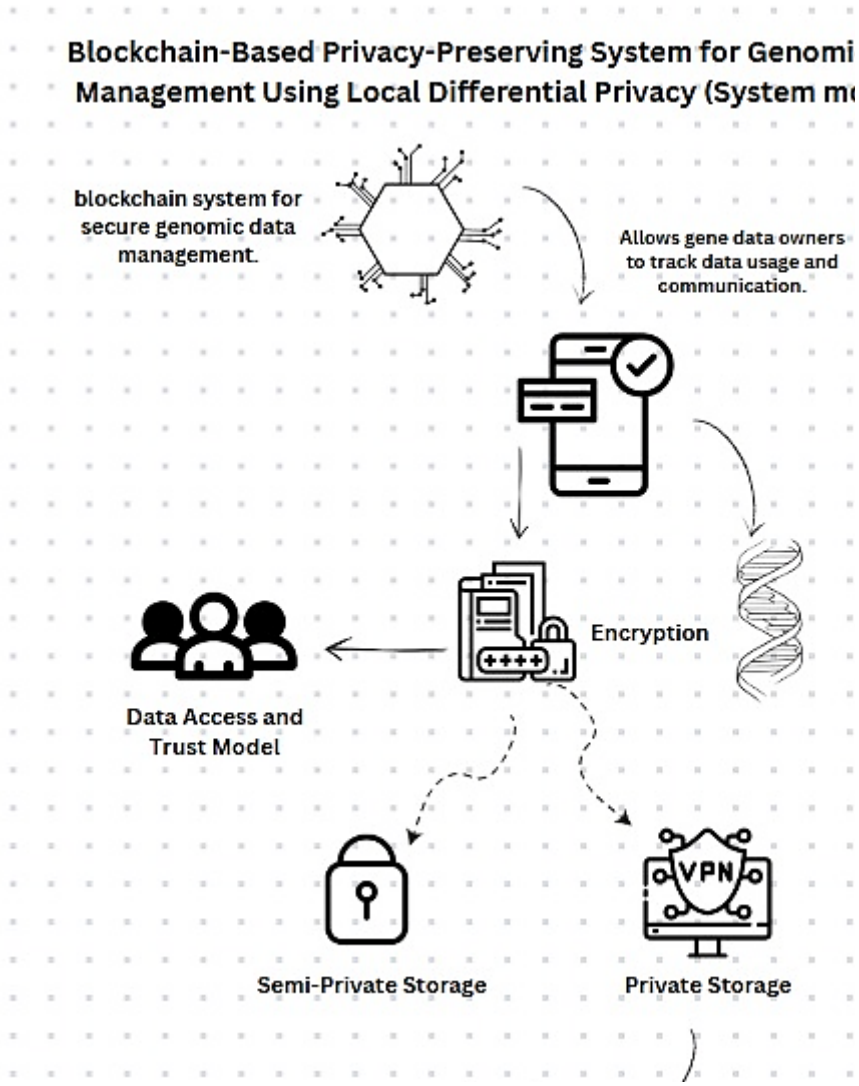

Figure 6: A Local Differential Security (LDP)-Based Distributed Secure Genetic Information Guidelines for Management.

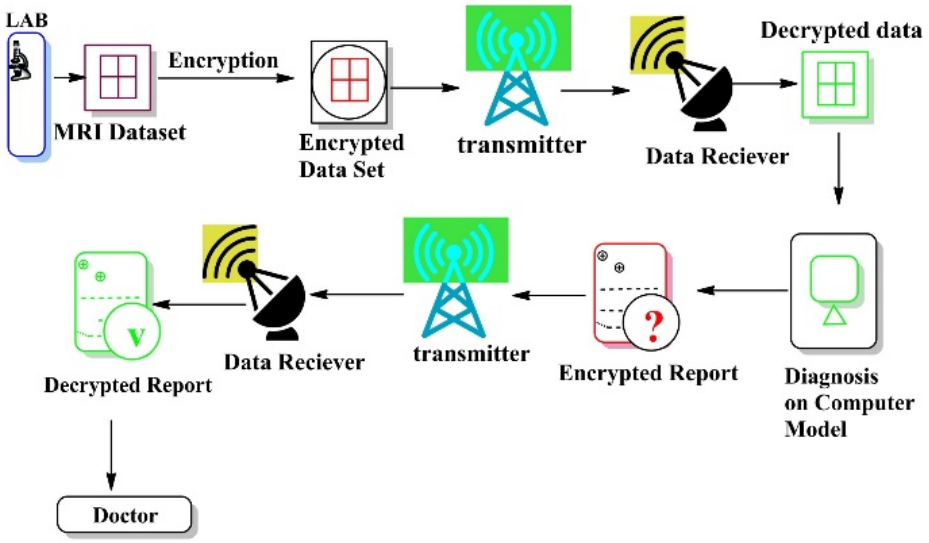

Figure 7: A Revolutionary Chaos-Based Deep Learning Model for Cancer Diagnosis that Preserves Privacy (Research flowchart).

The Electronic Health Record (EHR) Standards for India were released by the Ministry of Health and Family Welfare (MoH &FW) in September 2013. A collection of different medical records created during any clinical interaction or event is called an Electronic Health Record (EHR).

Biological Importance of Privacy

Improved personalized treatment and the development of targeted medications for certain diseases can be facilitated by the growing DNA collection. If researchers have wide access to the data, this can be accomplished. Although genomic data is publicly accessible, the untrustworthy third party may raise privacy issues. Although they take time, the unresolved privacy issues continue to consume a lot of resources. In this case, personal genomic data can be correctly identified as an individual's identity since, as demonstrated by the actual usage of DNA sequence in forensics, even a small fraction of the sequence is sufficient to identify a person or a relative. The genetic sequences are therefore linked to a significant security concern.

The outcomes of the data analysis provide the final piece of the missing component. Researchers' input strategy and data characteristics will be revealed through query processing and query output, even though privacy-preserving mechanisms are utilized for data computation and sharing. Researchers are less likely to find these issues, though, and assaults against the Beacon service's aggregated data highlight how important privacy.[34] The purpose of the Beacon service project is to assess the data owner's readiness for safe, straightforward procedural sharing of genetic data.[35] Identifying those who have the disease for treatment and cure, stopping its spread, and eventually eliminating its origin have all been part of public health practice since the late 19th century. Whether it is through the establishment of a hospital or clinic's medical chart, illness reporting platforms, or research into causes and cures, methods of accomplishing each of these aims have nearly always involved the sharing of some private information. Making sure that these remarkable advancements in public health are sustained without infringing on people's right to privacy has proven to be an ongoing problem. HIPAA (Health Insurance Portability and Accountability Act of 1996) has raised a lot of awareness about these problems in the last several years. The "Administrative Simplification" modification to this complicated federal law brought in previously unheard-of rules governing the gathering, storing, utilizing, and disclosing of medical data in a variety of contexts. There was soon widespread bewilderment and perhaps fear. In 1997 the American Civil Liberties Union condemned the legislation as “dangerous” and “yet another of the Clinton Administration privacy dodges.” The group continued to demand that “patients must be given the option of a ‘paper only’ record-thereby keeping confidential information out of computer networks and databases” and claimed that the American public wants their consent to be obtained before using their medical information. The privacy of medical records is protected by a number of federal and state regulations, including HIPAA. It is crucial to carefully weigh the benefits and drawbacks of finding a balance between societal public health goals and individual privacy concerns, even if some persons may use hyperbole or strong language.

Cancer surveillance and privacy and confidentiality concerns (United States)

Risks of breach of privacy

"People are adopting 'privacy-protective' practices to protect themselves from what they perceive to be unwanted and damaging uses of their health information." Patients may choose to pay for medical care out of pocket, "doctor-hope to prevent having access to all of their medical records committed to a single practitioner, conceal information, lie, or even forego care entirely in order to protect their privacy.

There are serious repercussions from such "privacy-protective" behavior:

- The patient can receive subpar care, increasing the chance of illnesses going undiagnosed and poorly treated.
- When a patient fails to provide accurate and complete information, the doctor's ability to diagnose and treat them is compromised.
- If a doctor falsifies diagnosis or treatment codes on claim forms, keeps separate records for internal use only, or sends insufficient information for claims processing in an effort to get a patient to communicate more fully, the integrity of the data leaving the doctor's office may be compromised. The diagnosis and treatment that are derived from the patient's information may be inaccurate, deficient, and not totally representative of the state of the patient or care.[36-38]

Current solutions

Federal law and regulations

The purpose of the Privacy Rule in HIPAA is to safeguard patient privacy, especially regarding medical records. The rule governs the use and disclosure of data, known as Protected Health Information (PHI), by healthcare providers, health plans, and other covered entities, which frequently has a direct impact on public health research and surveillance, even though it does not directly regulate these activities. Certain public health surveillance investigations and actions are designated as high-priority exceptional circumstances under the Privacy Rule, which waives the need for authorization to release PHI and does not need informed consent. To prevent interfering with public health research and surveillance, specific measures are provided.

State laws and regulations

State laws that specifically govern cancer surveillance are a more complex matter. Although these laws exist in every state, there are differences in the procedures and clauses pertaining to privacy protection, access, release, and usage. In the future, these differences might grow more significant. For instance, several states enter "case-sharing" arrangements with other states where their inhabitants may likely seek medical attention in an effort to guarantee full registration of their insured population. It can be challenging to gather this kind of information while guaranteeing adherence to reporting regulations in other states. As multi-state research groups attempt to adhere to numerous and potentially conflicting policies and procedures, differences in policies regarding data release and patient interaction for research purposes can present administrative and logistical obstacles.

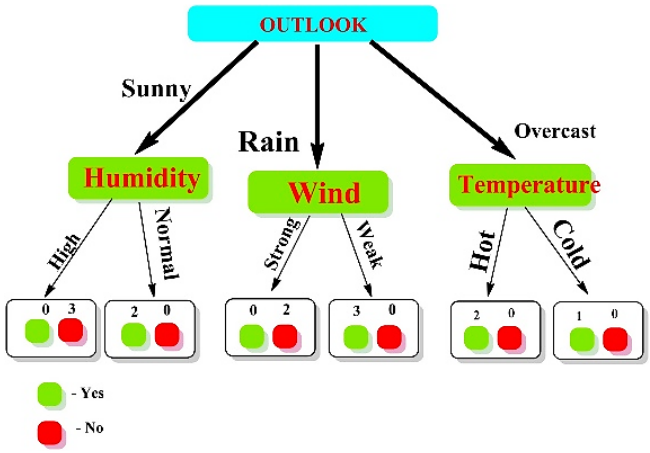

Figure 8: Sample Random Decision Tree (RDT).

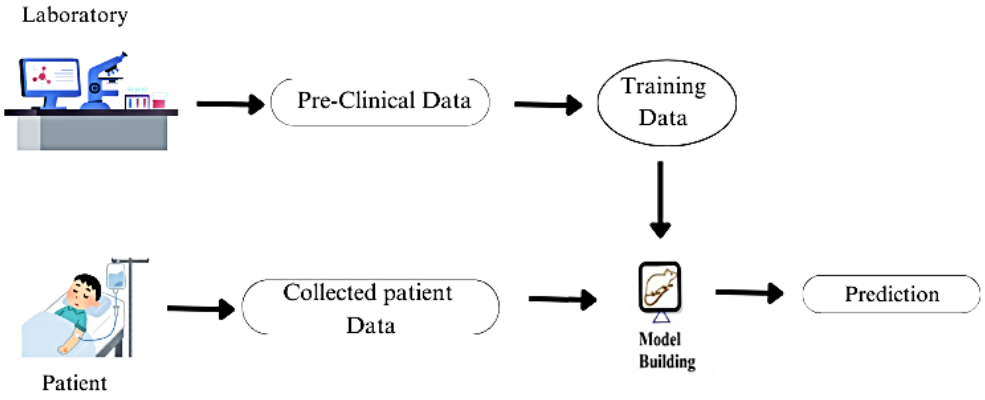

Figure 9: Homomorphic Encryption (HE) Threat Model for privacy inference.

CONCLUSION

Curbing the various types of cancer has become the interest of many scientists and researchers globally. To aid the same, many patients agree to provide their data to find a cure for this deadliest disease because they appreciate the therapeutic and research benefits of genetic tests, even though they are unlikely to reap benefits. But the data collected should be confidential to avoid biasedness and effects on future generations. People, as well as organizations, have the privilege of deciding how, when, and to what degree personal data about them is shared with others. This includes the right to remain anonymous. It is possible to utilize a variety of confidentiality technologies to handle sensitive data in medicine properly. As described in the text, various approaches, algorithms, technologies, models, etc., have been designed to achieve privacy and preserve the collected data. Some of these models were also useful for diagnosing cancer in patients. Patenting for a successful design prevents people from plagiarizing the model. It will require an amalgamation of scientific and social solutions which take into consideration the data's context is employed to provide genetic data with the necessary degrees of privacy. With the development of security-enhancing technology and regulatory frameworks for the use of genetic data, pressure is mounting to protect genomic privacy.

ACKNOWLEDGEMENT

The concept was created, written, organized, edited, and directed by Janaki Ramaiah Mekala. The manuscript was written by Sahiti R. Chamarthy. Reshitha Talluri and Charanya Gade drew Figures. Nageswara Rao Moparthi and Raghava Rao helped in manuscript preparation and network and cloud computing aspects. Rakesh has helped in medical genomics.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ABBREVIATIONS

AI: Artificial Intelligence; **APIs:** Application Programming Interfaces; **ADME:** Absorption, Distribution, Metabolism, and Excretion; **caBIG:** Cancer Biomedical Informatics Grid; **CNN:** Convolutional Neural Network; **Dapp:** Decentralized Application; **DEGs:** Differentially Expressed Genes; **DL NLP:** Deep Learning Natural Language Processing; **DWT:** Discrete Wavelet Transform; **EEG:** Electroencephalogram; **EHR:** Electronic Health Record; **EMR:** Electronic Medical Records; **GINA:** Genetics Data Non-discrimination Act; **GWAS:** Genome-Wide Association Studies; **HE:** Homomorphic Encryption; **HIPAA:** Health Insurance Portability and Accountability Act; **IoT:** Internet of Things; **LDP:** Local Differential Privacy; **MoH & FW:** Ministry of Health and Family Welfare; **MPC:** Multi-Party Calculation; **MRI:** Magnetic Resonance Imaging; **NCI:** National Cancer Institute; **NDPPP:** Novel Data Privacy-Preservation Protocol; **PCF:** Portable Circuit Format; **PE:** Predicate Encryption; **PHI:** Protected Health Information; **RDT:** Random Decision Tree; **TDRs:** Trusted Digital Repositories; **WQI:** Water Quality Index; **4V:** Volume, Velocity, Variety, and Value.

FUNDING

SERB supported this work, Government of INDIA, with grant number EMR/2017/001201, and DBT, Govt. of INDIA, for funding with grant number BT/PR20836/MED/30/1727/2016. The grants were received by the corresponding Dr. M. Janaki Ramaiah. All the authors thank KLEF management for its infrastructural support and encouragement.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ETHICAL APPROVAL

No author-performed studies involving humans or animals are included in this article.

REFERENCES

1. Guttmacher AE, Collins FS. Welcome to the Genomic Era. New Engl. J. Med. 2003; 349 (10): . https://doi.org/10.1056/nejme038132
2. Anyira IE, Onoriode OK, Nwabueze AU. The Role of Libraries in the Preservation and Accessibility of Indigenous Knowledge in the Niger Delta Region of Nigeria. Library Philosophy and Practice 2010; 1.
3. Wu X, Zhang Y, Wang A, *et al.* MNSSp3: Medical Big Data Privacy Protection Platform Based on Internet of Things.*Neural Comput. Appl*. 2020; 34(14): 11491-505. https://doi.org/10.1007/s00521-020-04873-z.
4. Chen Y, Ding S, Xu Z, Zheng H, Yang S. Blockchain-Based Medical Records Secure Storage and Medical Service Framework. J. Med. Syst. 2018; 43(1). https://doi.org/10.1007/s10916-018-1121-4.
5. Vayena E. Biomedical Big Data: New Models of Control over Access, Use and Governance. J. Bioethical Inquiry 2017; 14(4): 501-13. https://doi.org/10.1007/s11673-017-9809-6.
6. Bibault J-E, Giraud P, Burgun A. Big Data and Machine Learning in Radiation Oncology: State of the Art and Future Prospects.*Cancer Lett*. 2016; 382(1): 110-7. https://doi.org/10.1016/j.canlet.2016.05.033.
7. Thomas, J. Medical Records and Issues in Negligence. Indian Journal of Urology. 2009; 25(3): 384. https://doi.org/10.4103/0970-1591.56208.
8. Magama, B. Strategies for Preservation of Digital Records in Masvingo Province of Zimbabwe. ESARBICA J. 2017; 37: 18-38.
9. Adu KO. Framework for Digital Preservation of Electronic Government in Ghana, 2015. https://uir.unisa.ac.za/bitstream/10500/20118/6/thesis_adu_kk.pdf.
10. Ngulube P. "Ghosts in our Machines": Preserving Public Digital Information for the Sustenance of Electronic Government in Sub-Saharan Africa.*Mousaion* 2012; 30(2): 128-36.
11. Rinehart A, Prud'homme, P.-A, Huot AR. Overwhelmed to Action: Digital Preservation Challenges at the under-Resourced Institution. OCLC Systems & Services. 2014; 30 (1): . https://doi.org/10.1108/oclc-06-2013-0019
12. Ngoepe M, Van Der Walt T. Strategies for the Preservation of Electronic Records in South Africa: Implications on Access to Information. Innovation (Pietermaritzburg) 2009; 38(1). https://doi.org/10.4314/innovation.v38i1.46971.
13. Lowry J, and Nduna V. Digital records management and preservation. In XXIII ESARBICA General Conference, Victoria Falls. 2015: 8-12.
14. Thomas S. Selecting the right preservation strategy. 2006.
15. Tellez D, Litjens G, Van Der Laak J, Ciompi, F. Neural Image Compression for Gigapixel Histopathology Image Analysis. IEEE Trans. Pattern Anal. Mach. Intell. 2021; 43(2): 567-78. https://doi.org/10.1109/tpami.2019.2936841.
16. Botterill T, Lotz T, Kashif AS, Chase JG. Reconstructing 3-D Skin Surface Motion for the DIET Breast Cancer Screening System. IEEE Trans. Med. Imag. 2014; 33(5): 1109-18. https://doi.org/10.1109/tmi.2014.2304959

17. Litjens G, Debats OA, Barentsz JO, *et al.* Computer-Aided Detection of Prostate Cancer in MRI. IEEE Trans. Med. Imag. 2014; 33(5): 1083-92. https://doi.org/10.1109/tmi.2014.2303821.
18. Mohamed S, Salama MMA. Prostate Cancer Spectral Multifeature Analysis Using TRUS Images. IEEE Trans. Med. Imag. 2008; 27(4): 548-56. https://doi.org/10.1109/tmi.2007.911547.
19. Islam A, Reza SMS, Iftekharuddin KM. Multifractal Texture Estimation for Detection and Segmentation of Brain Tumors. IEEE Trans. *Biomed. Eng*. 2013; 60(11): 3204-15. https://doi.org/10.1109/tbme.2013.2271383.
20. Wang X, Chen H, Gan C *et al.* Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis. IEEE Trans. Cybern. 2020; 50(9): 3950-62. https://doi.org/10.1109/tcyb.2019.2935141.
21. Feng X, Song L, Wang, S, *et al.* Accurate Prediction of Neoadjuvant Chemotherapy Pathological Complete Remission (PCR) for the Four Sub-Types of Breast Cancer. IEEE Access. 2019; 7: 134697-706. https://doi.org/10.1109/access.2019.2941543.
22. Rizzo A, Santoni M, Mollica V, *et al.* Peripheral Neuropathy and Headache in Cancer Patients Treated with Immunotherapy and Immuno-Oncology Combinations: The MOUSEION-02 Study. Expert Opin. Drug Metab. Toxicol. 2021; 17(12): 1455-66. https://doi.org/10.1080/17425255.2021.2029405.
23. Lauter K, López-Alt A, Naehrig M. Private Computation on Encrypted Genomic Data. In Springer eBooks; 2015: 3-27. https://doi.org/10.1007/978-3-319-16295-9_1.
24. Hekel R, Budis J, Kucharik M, *et al.* Privacy-Preserving Storage of Sequenced Genomic Data. BMC Genomics 2021; 22(1). https://doi.org/10.1186/s12864-021-07996-2
25. Kreuter B, Mood B, Shelat A, Butler K. PCF: A Portable Circuit Format for Scalable Two-Party Secure Computation. 22nd USENIX Security Symposium 2013, 321-336.
26. Aminifar A, Rabbi F, Lamo Y. Scalable Privacy-Preserving Distributed Extremely Randomized Trees for Structured Data With Multiple Colluding Parties. ICASSP 2021-2021 IEEE Int. Conf. Acoust., Speech and Signal Processing (ICASSP) 2021. https://doi.org/10.1109/icassp39728.2021.9413632.
27. Detrano R, Jánosi A, Steinbrunn W, *et al.* International Application of a New Probability Algorithm for the Diagnosis of Coronary Artery Disease. Amer. J. Cardiol. 1989; 64(5): 304-10. https://doi.org/10.1016/0002-9149(89)90524-9.
28. Mangasarian OL, Street WN, Wolberg WH. Breast Cancer Diagnosis and Prognosis Via Linear Programming. Oper. Res. 1995; 43(4): 570-7. https://doi.org/10.1287/opre.43.4.570.
29. Alawad M, Yoon H-J, Gao S, *et al.* Privacy-Preserving Deep Learning NLP Models for Cancer Registries. IEEE Trans. Emerging Topics in Comput. 2021; 9(3): 1219-30. https://doi.org/10.1109/tetc.2020.2983404
30. Rogith D, Yusuf RA, Hovick SR, *et al.* Attitudes Regarding Privacy of Genomic Information in Personalized Cancer Therapy. J. Am. Med. Informatic. Assoc. 2014; 21(e2): e320-e5. https://doi.org/10.1136/amiajnl-2013-002579.
31. Sarkar E, Chielle EO, Gürsoy G, *et al.* Scalable Privacy-Preserving Cancer Type Prediction with Homomorphic Encryption. *arXiv* (Cornell University) 2022. https://doi.org/10.48550/arxiv.2204.05496.
32. Field M, Thwaites DI, Carolan M, Delaney GP, Lehmann J, Sykes J, *et al.* Infrastructure platform for privacy-preserving distributed machine learning development of computer-assisted theragnostics in cancer. J Biomed Informatics [Internet] 2022; 134: 104181. Available from: https://doi.org/10.1016/j.jbi.2022.104181.
33. Manion FJ, Robbins RJ, Weems WA, Crowley RS. Security, and Privacy Requirements for a Multi-Institutional Cancer Research Data Grid: An Interview-Based Study. BMC Med Inform Decis Mak. 2009; 9 1). https://doi.org/10.1186/1472-6947-9-31
34. Wang S, Jiang X, Tang, H.; *et al.* Community Effort to Protect Genomic Data Sharing, Collaboration and Outsourcing. Npj Genomic Med. 2017; 2(1). https://doi.org/10.1038/s41525-017-0036-1.
35. Bu D, Wang X, Tang, H. Real-Time Protection of Genomic Data Sharing in Beacon Services. PubMed. 2018; 2017: 45-54. https://pubmed.ncbi.nlm.nih.gov/29888039
36. Goldman J. Protecting Privacy To Improve Health Care. Health Affairs. 1998; 17 (6): . https://doi.org/10.1377/hlthaff.17.6.47
37. Protecting personal health information in research: understanding the HIPAA privacy rule (2004) US Department of Health and Human Services, NIH publication number 03-5388
38. Clinton medical privacy recommendations undercut state protections, ACLU says [press release] (10/28/1997) American Civil Liberties Union. Available from: http://www.aclu.org/news/ n102897a.html.

Cite this article: Chamarthy S, Moparthi NR, Raghava KR, Chintala VR, Mekala JR, Gade C, *et al*. Privacy Preservation of Genomic and Medical Data. Asian J Biol Life Sci. 2025;14(3):747-57.