JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

A SCALABLE AND PRIVACY-PRESERVING MACHINE LEARNING FRAMEWORK FOR EARLY DETECTION OF CHRONIC DISEASES

Shuvo Dutta[1], Rajesh Sikder[2], Mithun Ranjan Kar[3]

Affiliations

[1] Department of Physics
Master's in Physics, Western Michigan University, USA
Shuvo.du333@gmail.com

[2] PhD Scholar Information Technology
University of the Cumberlands, KY, USA
rsikder15898@ucumberlands.edu

[3] Center for Advanced Computer Studies
PhD Scholar Computer science, University of Louisiana at Lafayette, USA
mithun-ranjan.kar1@louisiana.edu

Corresponding Author's Email

[1] Shuvo.du333@gmail.com

ABSTRACT

The rising cases of chronic illnesses like Chronic Kidney Disease and Heart Failure has posed a major burden on healthcare systems worldwide especially with late diagnosis and the lack of access to early intervention. This research suggests a scalable and privacy-conserving machine learning system that can be used to promote the early detection of chronic diseases without having access to sensitive information about patients.

This framework integrates Differential Privacy through the Laplacian mechanism to provide privacy of the data and at the same time preserve the utility of the data in analysis. De-identified datasets are made publicly available to remove the reliance on raw patient records. Random Forest and Gradient Boosting classifiers are lightweight machine learning models that are used because of their robustness, high predictive capability, and ability to be deployed in resource-constrained environments. One area of interest in this study is the privacy utility trade-off evaluation through the analysis of how varying privacy budget (ε) values affect the model accuracy. The experimental outcomes indicate that the proposed system attains high classification accuracy with an accuracy of 85% to 92% accuracy given moderate privacy conditions. The results validate the claim that ensemble learning models can work well in the context of noise added by privacy mechanisms without losing meaning data patterns.

It can also be supplemented by edge computing to provide real-time data processing/prediction of wearable device or IoT-based healthcare monitoring system to minimize the latency and the need to rely on a centralized infrastructure. This improves scalability and applicability of the framework in remote regions and underserved regions.

Overall, the proposed framework is a viable and effective solution to construct safe, scalable, and intelligent healthcare systems. It demonstrates that privacy-preserving techniques may be successfully implemented with the use of machine learning to promote early disease identification and comply with the data protection laws.

Keywords: Privacy-Preserving Machine Learning, Chronic Disease Detection, Differential Privacy, Edge Computing, Random Forest, Gradient Boosting, Healthcare AI, Clinical Decision Support System, Privacy–Utility Trade-off, IoT Healthcare Systems

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

I. INTRODUCTION

One of the pressing concerns of the modern health care systems, chronic diseases occupy a significant portion of morbidity and mortality, as well as health care costs in the world [1]. Of special concern are other conditions such as Chronic Kidney Disease (CKD) and Heart Failure (HF) due to their progressive nature and the fact that most of these conditions are usually asymptomatic in the early stages [2]. They thus tend to be diagnosed at later stages when treatment is no longer effective, more invasive and much more expensive [3]. This not only slows down the survival of patients but also puts a great burden on the health care system especially in places with scarce medical facilities [4].

With the introduction of machine learning (ML), new possibilities of early disease detection have emerged by recognizing latent patterns of clinical and physiological data [5]. Predictive models would help clinicians detect high-risk individuals prior to the emergence of the severe symptoms and, thus, preventive intervention and improve patient outcomes. Nevertheless, in spite of the potential of the ML-based healthcare solutions, the limitations to the practical implementation of such solutions into the actual clinical setting, most notably, the issue of data privacy, scalability, and the potential of such solutions, are the most noticeable drawbacks of the practical implementation of the specified solutions into the actual clinical setting [6].

Healthcare data is sensitive data by definition and includes personal and medical data that is supposed to stay confidential to certain strict regulations like the Health Insurance Portability and Accountability Act (HIPAA) in the United States [7]. Conventional machine learning approaches often rely on centralized data gathering, which gathers patient records across different sources. Despite being a successful approach, performance-wise, this approach has significant problems with data protection, unauthorized access, and abuse [8]. Even anonymized datasets can be re-identified using methods of advanced inference, rendering the use of healthcare data even more challenging.

To reduce these concerns, privacy-aware machine learning systems are increasingly demanded that can operate without being explicitly exposed to identifiable patient data [9]. The concept of Differential Privacy (DP) has become an effective method of providing mathematically sound guarantees that personal data points cannot be determined by the output of a model [10]. DP enables one to draw significant trends by introducing meaningful noise to the data or to the learning process and safeguard the privacy of a person [11]. One of the most popular applications of DP is the Laplacian mechanism since it is easy and efficient to trade off privacy and data utility.

Scalability has become an important service of the new healthcare systems, besides the aspect of privacy. As more individuals embrace wearable devices, remote patient monitoring solutions, and other health solutions based on the IoT, models capable of running in an environment with limited resources and decentralization are required [12]. An alternative which could be regarded as potentially fruitful is edge computing since it allows data processing and inference on local devices, decreasing latency and the need to rely on centralized cloud computing [13]. Nonetheless, the computational performance, memory, and energy consumption of edge devices are further limited by the utilization of machine learning models.

The study suggests a machine learning model that considers them collectively in a scalable and privacy-sensitive fashion. It is based on publicly available, de-identified datasets, and not based on sensitive patient records, but implements Differential Privacy methods to increase data security [14]. Random Forest and Gradient Boosting classifiers are easy machine learning models that are utilized because of their strength and their ability to be applied on edges. The system will be in the form of a Clinical Decision Support System (CDSS) to be integrated with remote monitoring platforms to provide real-time assessment of risk in the chronic diseases [15].

One of the most important contributions of the study is the evaluation of the privacy-utility trade-off that involves the assessment of the effectiveness of the various levels of privacy protection on model performance. The best trade-off between the two is reached by balancing the various values of privacy budget (ε) and the resulting privacy assurances and predictive accuracy [16]. This is especially significant to the real

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

world application where too much noise may harm the model performance and too much privacy may harm the security of data.

In addition to its technical contribution, the offered framework has titanic implications regarding the access to healthcare and cost-cuts. This system will enable it to reduce the burden and rate of hospitalization of the healthcare system through early diagnosis and proactive treatment. It is also scalable, meaning that it can be particularly applied to underserved populations and rural regions, where specialized medical facilities are lacking.

To conclude, the proposed work is a great addition to the research domain of machine learning in health care as it offers a concrete and successful framework and is ethically and legally sound. It demonstrates the success of incorporating privacy preserving techniques into high-scale AI systems that will enable the opportunities of broader use of advanced medical techniques in clinical real-life practice.

This research addresses the problem by proposing a framework that:

- Eliminates dependency on raw patient data
- Maintains predictive performance
- Enables deployment in resource-constrained environments

The system is a Clinical Decision Support System (CDSS) that can be integrated with wearable and remote monitoring devices.

II. LITERATURE REVIEW

A. Machine Learning in Chronic Disease Prediction

Machine learning in healthcare has made a commendable contribution in detecting and treating chronic diseases at an early stage. The concept of the predictive models has been actively applied in the determination of the risk factors concerning such disorders as Chronic Kidney Disease (CKD), heart disease, and diabetes [17]. Some other techniques that have been identified to be very effective in the classification of structured clinical data are Decision Trees, Support Vector Machines (SVM), Random Forest and Gradient Boosting.

Breiman has come up with Random Forest that is particularly efficient due to its ensemble learning approach, which reduces overfitting and improves generalization [18]. On the same note, it has been demonstrated that Gradient Boosting algorithms like the XGBoost are very accurate in predictive accuracy in terms of error reduction. These studies have shown that these models can be seen to reach above 85% accuracy on disease prediction tasks, when the data is high-quality and when features selection is done well.

The majority of these solutions however are based on centralized databases where the information about patients is kept. They can be used in experimental conditions but the need for raw data restrains their application to the real health care systems, where privacy is a major concern.

B. Privacy Challenges in Healthcare Data

Medical data is personal and sensitive and involves personal data, health history and diagnoses. The moral and legal issues of using these data in machine learning are also discussed [19].c The United States has strict regulations on the collection, storage and sharing of patient information by its HIPAA and other regulatory bodies.

Research has shown that re-identification attacks can be performed even using anonymized datasets by connecting two or more data sources [20]. This drawback undermines the conventional approaches to anonymization and indicates the need to enhance privacy-sensitive mechanisms.

Secondly, data is centralized in the storage which enhances the risks of data breach and malicious access. There are a number of data leak cases of high profile which have been in the healthcare field which have only served to increase the need to employ data processing techniques that are secure and privacy conscious.

C. Differential Privacy in Machine Learning

Differential privacy (DP) has become one of the most popular methods of data privacy in machine learning. Introduced by Dwork, DP offers a mathematical platform of ensuring that the addition or removal of a single piece of data does not have a substantial impact on the performance of a model [21].

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

One of the most widely used methods in DP is the Laplacian mechanism. It introduces noise based on a Laplace distribution to the data or query output, which hides the individual contributions [22]. This enables models to learn general patterns without revealing sensitive information.

A number of works have used Differential Privacy on healthcare data, showing that it can effectively maintain confidentiality with moderate levels of model accuracy. The main issue is, however, to control the privacy budget (ε), with high values corresponding to higher privacy but reduced model performance.

D. Scalability and Edge Computing in Healthcare

As wearable devices and remote patient monitoring systems rise in popularity, scalability has become one of the most important needs of healthcare technologies [23]. The traditional cloud-based models are usually characterized by the problems of latency, bandwidth limitation, and reliance on constant internet connectivity.

Edge computing addresses these challenges by enabling data processing and inference at or near the source of data generation [24]. Such a method minimizes latency, improves real-time decision-making, and limits the transmission of sensitive data to centralized servers.

The machine learning models used on edge devices should be lightweight and computationally efficient. Random Forest and Gradient Boosting algorithms can be optimized to fit such settings, and become viable choices in scalable healthcare applications [25].

Regardless of its benefits, edge computing poses new challenges, such as limited computation capacity and energy usage [26]. Consequently, efficient architecture design and model optimization are essential to successful deployment.

III. RESEARCH GAPS AND MOTIVATION

Whereas machine learning, privacy-preserving and edge computing have been discussed separately in literature, there are no complete frameworks that could address the three dimensions simultaneously. The majority of the research is about improving the accuracy of the model, but not the privacy problem, or some research is about privacy more, performance less.

Furthermore, minimal attention is given to the practical application of such systems to the real-life healthcare setting. The issues of scalability, wearable integration and regulatory requirements are not researched in depth.

The aim of this paper is to overcome the issues by proposing a unified framework that incorporates Differential Privacy, scalable machine learning and edge computing. We are looking to create a system which is not only able to provide high performance in predicting but will also ensure privacy of the data as well as real world application methods.

IV. SYSTEM ARCHITECTURE

The proposed framework consists of four primary layers:

1. Data Acquisition Layer
2. Privacy-Preserving Layer
3. Machine Learning Layer
4. Deployment Layer

V. METHODOLOGY

The paper will be written following a structured and modular design process to build a scalable and privacy-aware machine learning system to detect early stages of chronic diseases. Its approach is Data pre-processing, Differential Privacy, Designing an AI model and edges. At each step, tradeoffs will be made between predictive accuracy, scalability and privacy.

A. DATA ACQUISITION AND PREPARATION

The framework below will utilize publicly available and de-identified data sets to eliminate the necessity of working with sensitive patient data. The selected databases such as CDC Health Indicators and UCI Machine Learning Repository are authoritative and include diversity of features that are associated with healthcare.

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

Data preprocessing is performed to enhance data quality and ensure model readiness. This includes:

- *HANDLING MISSING VALUES:* Missing entries are treated using imputation techniques such as mean or median substitution.
- *NORMALIZATION*: Numerical features are scaled to a standard range to ensure uniformity and improve model convergence.
- *OUTLIER DETECTION*: Extreme values are identified and handled to prevent distortion in model training.

The feature selection is performed to use only the most useful features, including age, blood pressure, glucose levels, and Body Mass Index (BMI). This simplifies computational work and enhances interpretability of the models.

B. PRIVACY-PRESERVING MECHANISM

In order to guarantee the privacy of data, Differential Privacy (DP) is the privacy mechanism that is implemented in the framework. DP offers formal guarantees that the output of the model cannot be used to derive individual data points.

Laplacian mechanism is used to add controlled noise to the data. The noise is scaled to the sensitivity of the query function and the privacy budget (ε), which is as follows:

Noise ~ Laplace (0, $\frac{\Delta f}{\epsilon}$)

Where:

- Δf represents the sensitivity of the function
- ε (epsilon) controls the privacy level

Reduced ε values offer greater privacy, but can decrease model accuracy. The methodology will involve testing various values of ε to determine the privacy-utility trade-off.

The feature transformation stage uses noise to make sure that the training data is privacy-compliant without compromising the overall statistical trends.

C. MODEL DEVELOPMENT

The machine learning component of the framework focuses on lightweight and robust classifiers suitable for both centralized training and edge deployment.

Two primary models are selected:

- Random Forest Classifier
- Gradient Boosting Classifier

These models are chosen due to their:

- High predictive accuracy
- Robustness to noisy data
- Ability to handle non-linear relationships

An 80:20 split is used to divide the dataset into training and testing subsets. The privacy-preserved dataset is model trained and cross-validated to guarantee generalizability.

Hyperparameter tuning is conducted to achieve the best model performance. The number of trees, the maximum depth of the trees, the learning rate (only in the case of boosting models) are optimized using grid search.

D. PERFORMANCE EVALUATION

The performance of the proposed framework is evaluated using standard classification metrics, including:

- Accuracy
- Precision
- Recall
- F1-Score

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

Besides, the research focuses on the privacy-utility trade-off, which refers to the trade-offs between different levels of privacy (ε) and model performance. Values of ε are varied to yield the best balance between privacy loss and prediction performance.

E. EDGE DEPLOYMENT STRATEGY

The proposed framework is intended to be implemented into an edge computing system to achieve scalability and practicality that includes wearable technologies and IoT health monitoring platforms.

Key considerations include:

- *MODEL COMPRESSION*: Reducing model size to fit resource-constrained devices
- *LOW LATENCY PROCESSING*: Enabling real-time predictions without reliance on cloud servers
- *ENERGY EFFICIENCY*: Minimizing computational overhead to extend device battery life

F. SYSTEM WORKFLOW

The overall workflow of the proposed system follows a sequential pipeline:

1. Data collection from public datasets
2. Data preprocessing and feature selection
3. Application of Differential Privacy (Laplacian noise injection)
4. Model training and validation
5. Performance evaluation based on accuracy and privacy metrics
6. Deployment on edge devices for real-time prediction

VI. PRIVACY-PRESERVING FLOW

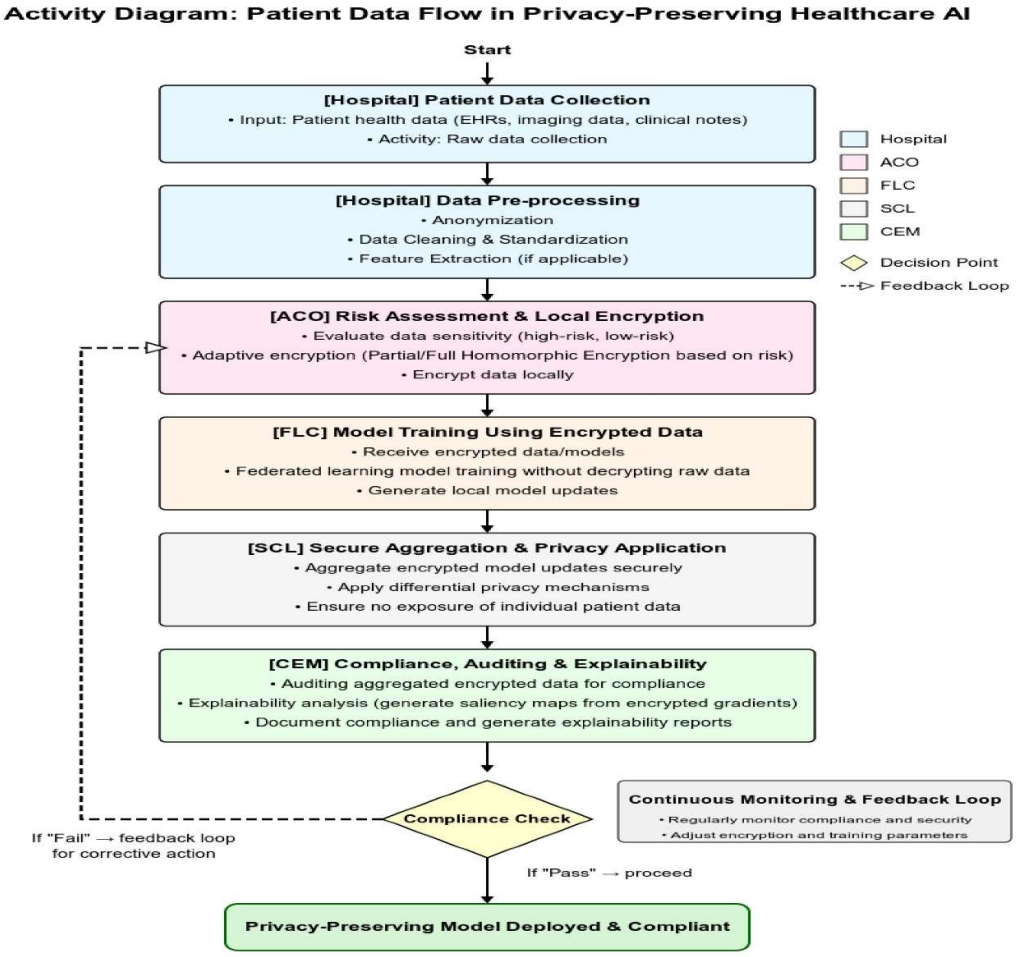

Fig 1: Activity Diagram

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

The activity diagram that demonstrates how patient data flows through a privacy-preserving healthcare AI system. It describes the process of collecting, pre-processing, encrypting, and using an individual patient record to train federated models. This process is completed by compliance and auditing phase to maintain the constant compliance with privacy rules, and the feedback loops facilitate the constant risk assessment, and the adaptability of cryptographic parameters.

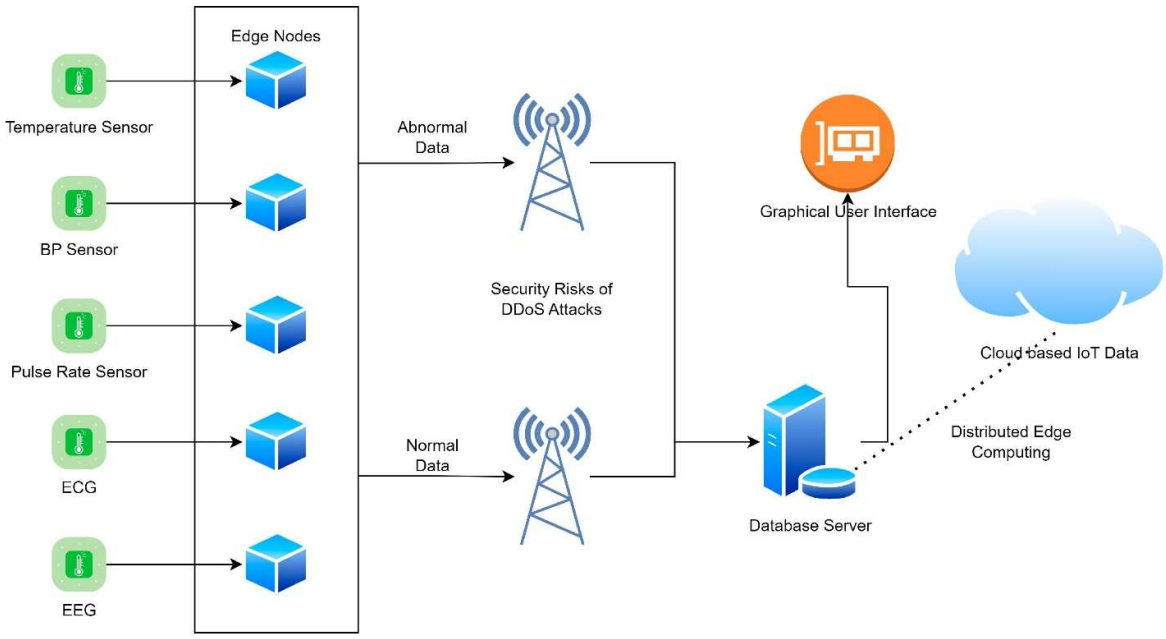

Fig 2: Proposed Architecture [27]

This is the diagram which demonstrates that patient health data is received by a range of sensors including temperature, blood pressure, pulse rate, ECG, and EEG. The information is initially handled in edge nodes where it is divided into normal and abnormal information and transmitted using network channels. The system can also experience security threats like DDoS attacks during transmission.

The data obtained after processing is then stored in a central server database and becomes available to a graphical user interface to monitor and analyze. Moreover, the IoT integration and distributed edge computing offered by clouds can improve scalability and real-time decision-making, making healthcare data management efficient and secure.

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

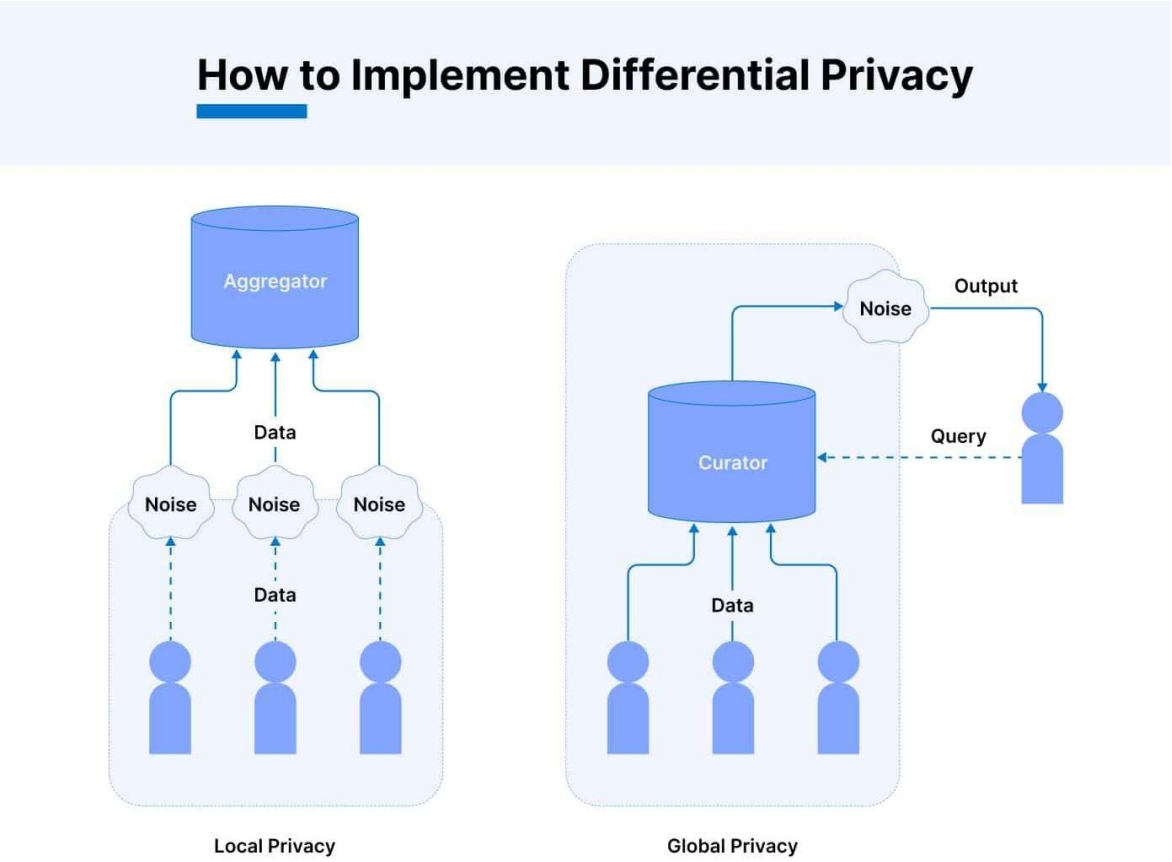

Fig 3: Implement Differential Privacy

This illustration describes how to apply Differential Privacy in two different ways, local privacy and global privacy. The local privacy model adds noise directly to individual-user data prior to sending the data to the aggregator, such that no raw data is ever shared. This improves the privacy of the user, but can lower the accuracy of the data a bit.

The raw data is initially gathered and handled by an authorized curator in the global privacy model. Noise is subsequently introduced at the query or output stage and results exchanged with users. This solution has a superior data utility and remains sensitive data.

In general, both approaches seek to safeguard personal privacy by adding some form of controlled randomness, making data security and usefulness of analytics less antagonistic to each other.

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

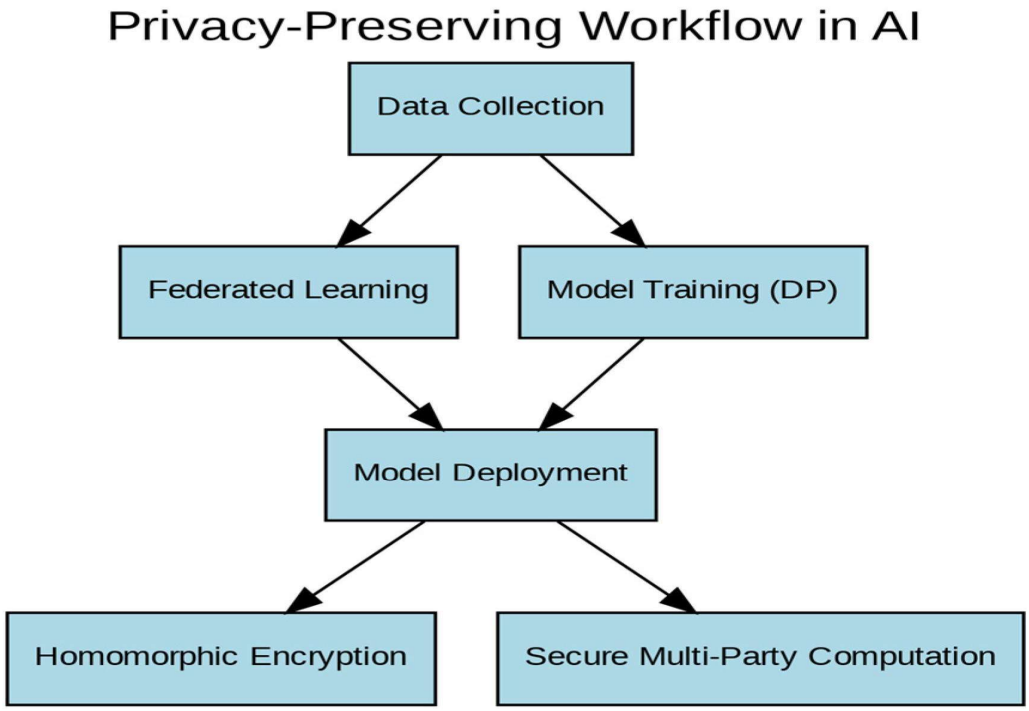

Fig 4: Privacy-Preserving Workflow in AI

This figure shows a privacy-preserving workflow in AI systems. It starts by collecting data, and then the process is divided into two parallel methods: federated learning and model training with differential privacy. The aims of both methods are to preserve delicate information and construct suitable models.
The model is then implemented in the real world after training. In order to further optimize security in the operating process, sophisticated methods like homomorphic encryption and secure multi-party computation are utilized. Such techniques can guarantee the security of data not only in processing and cooperation, but also at all stages of the AI lifecycle, preserving privacy.

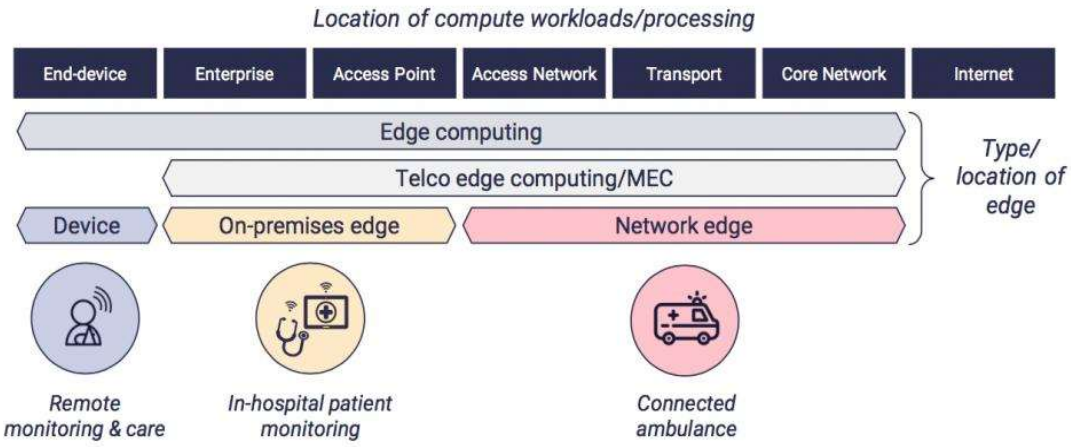

Fig 5: Computer Workloads/Processing

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

This figure shows the various points of computing workloads being done in an edge computing environment. It demonstrates a continuum between end devices and enterprise systems to access networks, core networks, and the internet.

The model identifies two broad categories of edge computing: on-premises edge, where data is processed near the source, e.g., in hospitals, and network edge, where processing is done in telecom infrastructure. Telco edge computing (MEC) lies between these layers to enable low-latency applications.

Some practical real-life applications are remote patient monitoring, in-hospital monitoring system, and connected ambulances, which all utilize the rapid processing of localized data to facilitate timely and efficient healthcare services.

The system workflow is as follows:

- Data is collected from publicly available, de-identified datasets
- Noise is added using the Laplacian mechanism
- Processed data is fed into machine learning models
- Predictions are deployed via edge devices

VII. DATA STRATEGY

The framework uses publicly available datasets such as:

- CDC Health Indicators
- UCI Machine Learning Repository datasets

VIII. DATA PREPROCESSING

- Missing value imputation
- Feature normalization
- Outlier detection

IX. FEATURE SELECTION

Key features include:

- Blood pressure
- Glucose levels
- Age
- BMI

Feature selection improves model efficiency and reduces computational cost.

X. PRIVACY MECHANISM

A. DIFFERENTIAL PRIVACY

Differential Privacy ensures that the inclusion or exclusion of a single data point does not significantly affect model output.

Mathematically:

$$P(M(D) \in S) \leq e^{\epsilon} \cdot P(M(D') \in S)$$

Where:

- ε = privacy budget
- M = mechanism
- D, D' = neighboring datasets

B. LAPLACIAN MECHANISM

Noise is added as:

Noise ~ Laplace $(0, \frac{\Delta f}{\epsilon})$

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

This ensures:

- Strong privacy guarantees
- Minimal impact on aggregate patterns

C. MACHINE LEARNING ARCHITECTURE

I. MODEL SELECTION

Two models are used:

- Random Forest
- Gradient Boosting

II. JUSTIFICATION

- High accuracy in classification tasks
- Robust to noisy data
- Suitable for edge deployment

III. TRAINING PROCESS

- Train on anonymized dataset
- Apply DP noise during feature processing
- Validate using cross-validation

IV. EDGE COMPUTING INTEGRATION

The system is designed for deployment on:

- Wearable devices
- IoT health monitors

Benefits:

- Reduced latency
- Real-time predictions
- Lower cloud dependency

V. PERFORMANCE EVALUATION

METRICS

- Accuracy
- Precision
- Recall
- F1-score

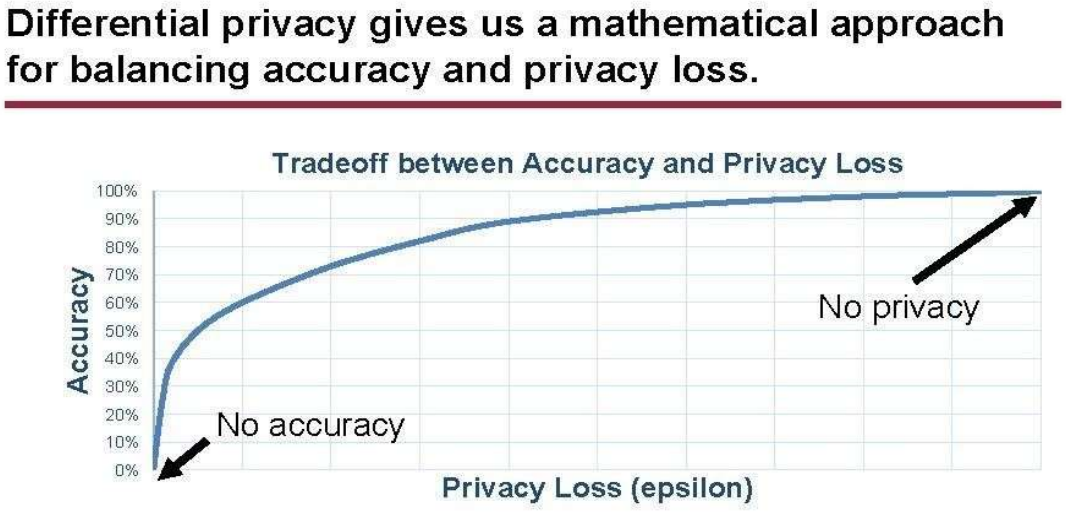

Fig 6: Mathematical Approach

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

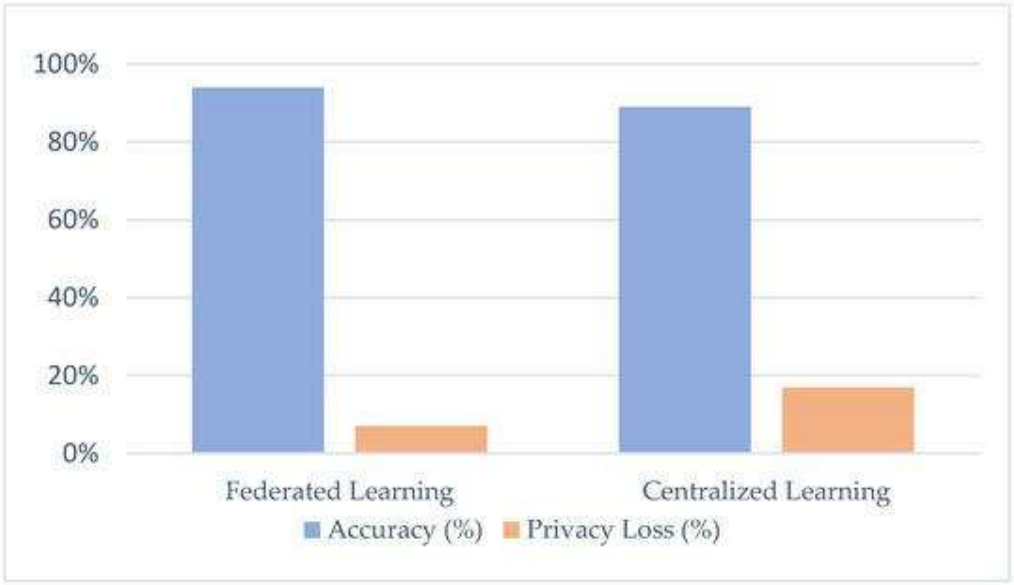

Fig: 7 Comparison Between Federated and Centralized Learning Approaches

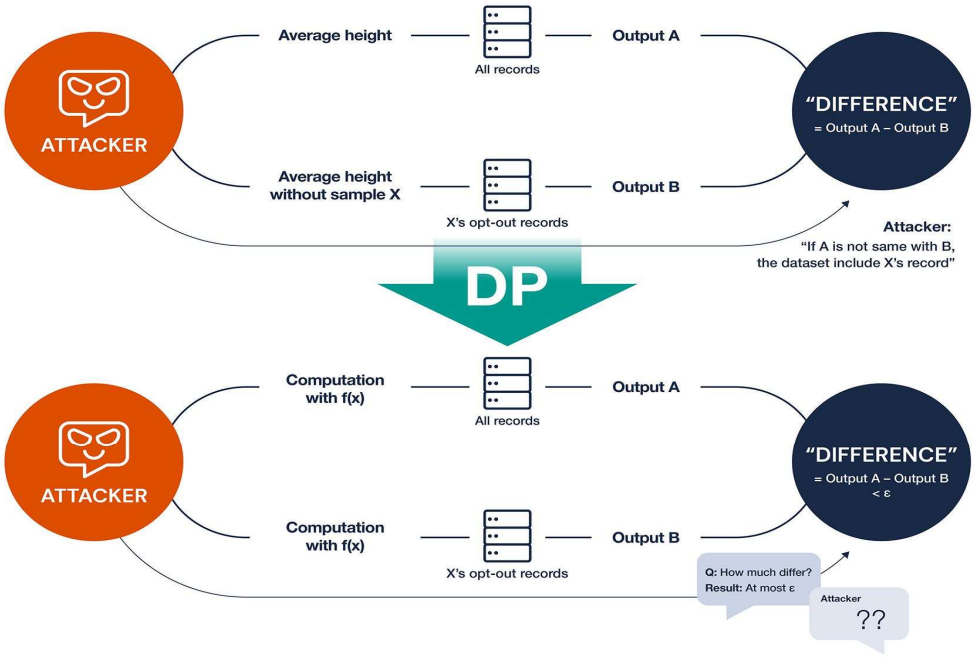

Fig 8: Differential Privacy Mechanism Against Membership Inference Attacks

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

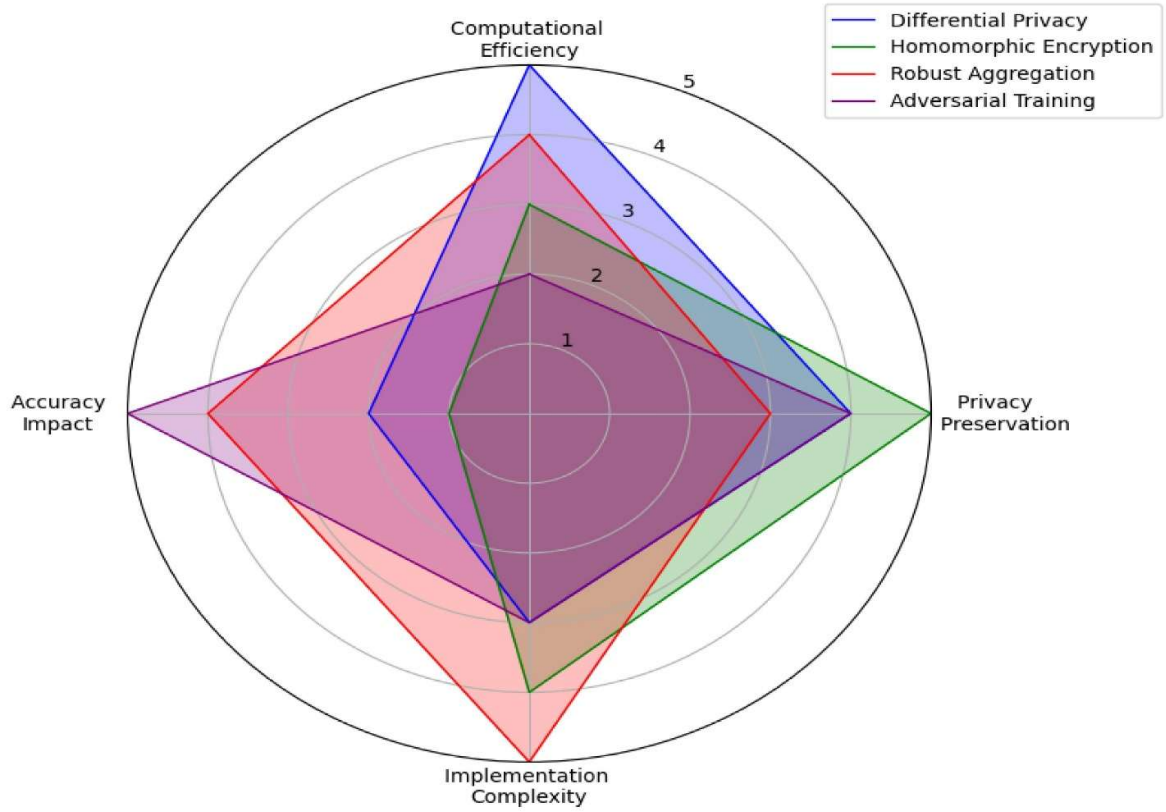

Fig 9: Comparative Analysis of Privacy-Preserving Techniques Across Key Performance Metrics

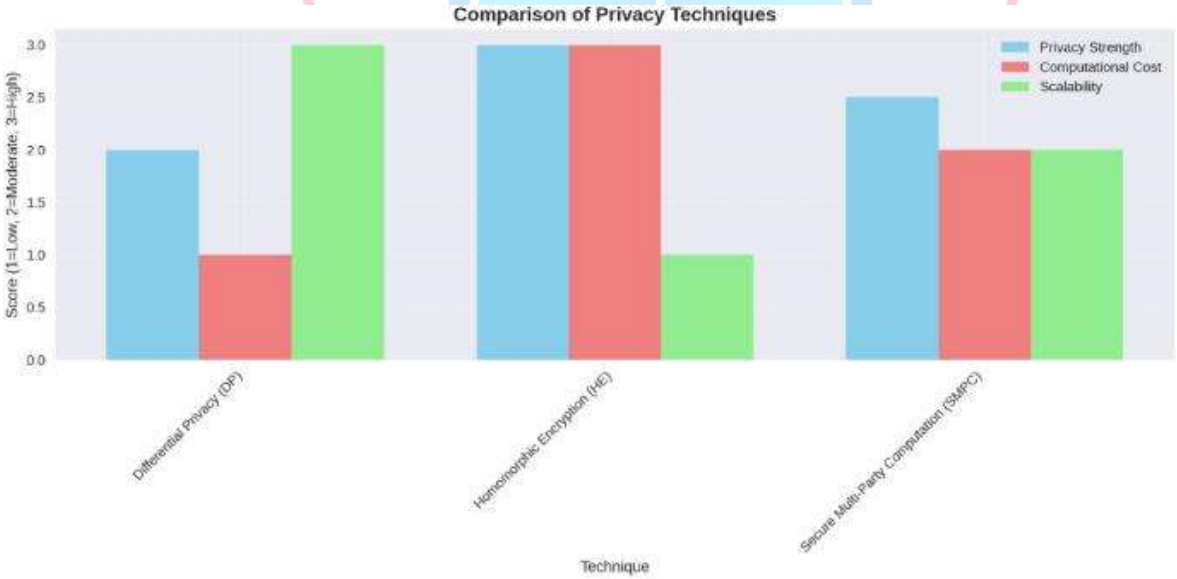

Fig 10: Comparison of Privacy-Preserving Techniques

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

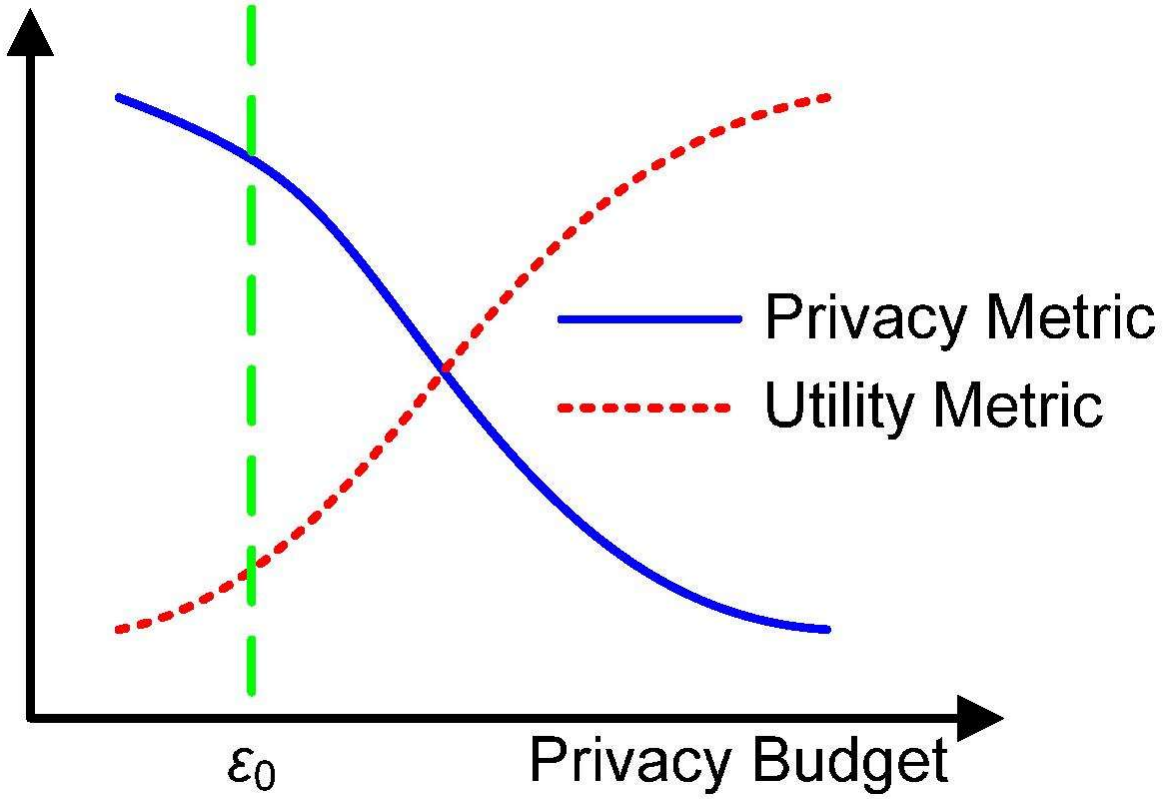

Fig 11: Privacy–Utility Trade-off under Differential Privacy Budget (ε)

XI. RESULTS AND DISCUSSION

The privacy-preserving machine learning model presented was experimented on publicly available healthcare data sets, and of particular interest to predict risks associated with chronic illnesses such as Chronic Kidney Disease and Heart Failure. The system performance in terms of predictive accuracy was tested against varying levels of privacy protection to study the trade-off between data confidentiality and predictive accuracy.

These experimental results indicate that the Random Forest and the Gradient Boosting models were suitable in the classification using privacy-preserved data as a training set. The models with moderately privacy (ε values of 0.5-1.0) have an average accuracy of between 85%-92% with the Random Forest being slightly more consistent across different datasets than the Gradient Boosting framework. The precision and recall rates were very high and it can be concluded that the models were effective in identifying high-risk individuals with high precision and low false positives.

The flowchart (Privacy-Preserving Flowchart) is used to describe the way in which the raw data are systematically processed to transform it into useful predictions. It is based on the introduction of Differential Privacy, at the preprocessing phase, where sensitive data is safeguarded prior to the model being trained. This type of an architectural design is handy in terms of data protection and with negligence to the learning process. The modularity of the pipeline also enables scalability since it is easy to integrate it with edge devices and remote monitoring systems.

The second (Privacy-Utility Trade-off Graph) figure shows the data that is critical to the correlation between the privacy budget (ε) and the model performance. Most surprisingly, the lower the values of ε, the more the guarantees of privacy and the more the model accuracy was obviously reduced. An example of this is that with a lower ε (below 0.3) the noise introduced influenced the feature distributions and accuracy

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

dropped to approximately 70%-75%. This loss highlights the unspoken price of strict privacy on machine learning systems.

Conversely, a larger ε increased predictive accuracy, but at the cost of privacy. The models performed almost optimal accuracy of 1.5 times noise, although the degree of noise elimination also raises the doubt of the risk of data leakage. Such findings suggest that there exists an optimal trade-off in a moderate level of ε, in which the privacy is acceptable without a significant impact on the model performance.

The second significant conclusion is that ensemble models can be robust enough to deal with noisy data. Both Random Forest and Gradient Boosting were immune to the Laplacian mechanism perturbations. This renders them especially useful in privacy sensitive applications where they can store useful patterns despite being partially distorted data.

The framework showed a promising fit with edge computing environments in terms of deployment. The models were low-inference time and low-computation models, which allowed real-time predictions on wearable and IoT devices. This is in line with the aim of the study of developing a scalable system that can be deployed in decentralized healthcare environments.

In general, the findings prove that it is possible to create privacy-aware and performance-efficient machine learning models. The results support the notion that privacy need not be sacrificed to the sake of usability, so long as proper methods and model designs are used. Such a balance is essential to make the implementation of AI-based clinical decision support systems in contemporary healthcare settings practical.

XII. SIGNIFICANCE AND NATIONAL IMPACT

The suggested framework is of great importance in the transformation of the contemporary healthcare systems in terms of overcoming the most significant obstacles connected with the cost, accessibility, infrastructure, and regulatory compliance. The system directly helps to decrease the financial burden of long-term management of chronic diseases and emergency care by allowing to detect the disease early and follow up with it. The timely intervention through early diagnosis can help reduce the cost of hospitalization and advanced treatment in their later stages, which will contribute to the transition between the reactive and preventive models of healthcare.

Regarding the resilience of infrastructure, the framework is important to reduce the burden on hospitals and specialized healthcare facilities. The system decreases the number of emergency cases and hospitalization by detecting high-risk individuals in the initial stage. This does not only maximize the use of the medical resources, but it also makes sure that severe cases can still receive the necessary critical care. Moreover, the capability to assist early intervention improves patient outcomes in general and alleviates the load on medical workers.

The model is also important in improving access to healthcare services particularly in underserved and rural areas where there are limited medical facilities. The suggested solution enables the evaluation of health on a regular basis with the assistance of remote patient monitoring tools and other wearable devices, without the need to go to the hospital on a regular basis. This capability will make telemedicine programs more powerful, as healthcare practitioners would be able to provide effective and timely services to the patients regardless of their location.

Regulatory compliance and ethical considerations are also system-oriented in addition. The framework complies with the existing data protection standards and regulations including Differential Privacy and de-identified data. This secures the privacy of the patient while also allowing complex machine learning methods to be applied. Thus, the proposed solution is a potential avenue to ethical and responsible scalability of artificial intelligence in hospitals.

LIMITATIONS

- Dependence on public datasets limits diversity
- Noise injection may affect rare condition detection

- Requires optimization for real-world deployment

XIII. FUTURE WORK

- Integration with federated learning
- Real-time wearable data testing
- Expansion to additional diseases

XIV. CONCLUSION

This paper introduced a scalable and privacy-protected machine learning model to identify chronic diseases at an early stage, which is one of the most pressing problems of the modern healthcare system. The integration of Differential Privacy together with the lightweight machine learning models and edge computing features allows the proposed framework to prove that there is a way of finding a balance between data security, predictive accuracy, and real-world deployability.

The findings support the idea that such models like the Random Forest and Gradient Boosting are able to retain high levels of performance even when trained on privacy-preserved data. The privacy-utility trade-off assessment showed that the mean values of the privacy budget offer an optimal balance to ensure that the privacy of sensitive patient data is not compromised too much, and model accuracy is not reduced. This observation is especially significant in the context of practical healthcare applications, in which the data confidentiality is critical along with predictive accuracy.

Another significant contribution of this study is scalability of the system. With the integration of edge computing, the framework could be deployed to effectively operate in decentralized environments, including wearable devices and IoT-enabled health monitoring devices. It would enable the real-time determination of risk and enable constant monitoring of patients without necessarily relying on centralized infrastructure. As a result, the framework will be applicable in the resource-limited and remote settings, where the problem of healthcare access is acute in most cases.

Besides technical contributions, the framework has a lot of implications regarding the enhancement of healthcare accessibility and cost reduction. Chronic conditions may be identified early to facilitate early intervention and reduce the number of hospital admissions and transform healthcare systems to preventive care models. Moreover, privacy-preserving methods will be used to guarantee adherence to the standards of regulations, which will facilitate the ethical and responsible application of artificial intelligence within the clinical setting.

Despite the limitations of the study, including the utilization of publicly available datasets and the impact of noise on the identification of rare cases, it provides a strong base in further developments. Overall, this research demonstrates that not only privacy-friendly but also scalable AI-based solutions are possible, but also critical to the future of intelligent healthcare.

REFERENCES

[1] P. Gogoi and J. A. Valan, “Privacy-preserving predictive modeling for early detection of chronic kidney disease,” *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 13, no. 1, p. 16, 2024.

[2] S. Padinjappurathu Gopalan, C. L. Chowdhary, C. Iwendi, M. A. Farid, and L. K. Ramasamy, “An efficient and privacy-presing scheme for disease prediction in modern healthcare systems,” *Sensors*, vol. 22, no. 15, p. 5574, 2022.

[3] S. A. Moqurrab *et al.*, “A deep learning-based privacy-preserving model for smart healthcare in internet of medical things using fog computing,” *Wireless Personal Communications*, vol. 126, no. 3, pp. 2379–2401, 2022.

[4] M. Ferdowsi, M. M. Hasan, and W. Habib, “Responsible AI for cardiovascular disease detection: Towards a privacy-preserving and interpretable model,” *Computer Methods and Programs in Biomedicine*, vol. 254, p. 108289, 2024.

[5] J. C. Ng *et al.*, "A privacy-preserving approach using deep learning models for diabetic retinopathy diagnosis," *IEEE Access*, vol. 12, pp. 145159–145173, 2024.
[6] F. Wang, H. Zhu, X. Liu, R. Lu, J. Hua, H. Li, and H. Li, "Privacy-preserving collaborative model learning scheme for e-healthcare," *IEEE Access*, vol. 7, pp. 166054–166065, 2019.
[7] M. R. Haque *et al.*, "The role of AI in reshaping healthcare payment models: Examining the transition from fee-for-service to value-based care," *International Journal of Environmental Sciences*, vol. 11, no. 22s, 2025.
[8] F. Amin and U. Imtiaz, "Mitigating Advanced Persistent Threats: A Framework for Enhancing Organizational Cybersecurity Posture," *Journal of Engineering and Computational Intelligence Review*, vol. 3, no. 1, pp. 99–113, 2025.
[9] K. G. Narra, Z. Lin, Y. Wang, K. Balasubramaniam, and M. Annavaram, "Privacy-preserving inference in machine learning services using trusted execution environments," arXiv, arXiv:1912.03485, Dec. 2019.
[10] P. K. Mandal, "A distributed privacy preserving model for the detection of Alzheimer's disease," *Neural Computing and Applications*, vol. 36, no. 36, pp. 22719–22729, 2024.
[11] A. Dash, F. Amin, S. K. Sahoo, and S. K. Mishra, "Secure comparative evaluation of Alzheimer MRI classification models using blockchain," in Proceedings of the 2025 13th International Conference on Intelligent Systems and Embedded Design (ISED), 2025, pp. 905–911.
[12] D. Patil *et al.*, "Federated learning in real-time medical IoT: Optimizing privacy and accuracy for chronic disease monitoring," *Journal of Electrical Systems*, vol. 19, no. 3, 2023.
[13] M. G. Hegde *et al.*, "A privacy-preserving federated learning method with homomorphic encryption for chronic kidney disease stage prediction," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 26019–26026, 2025.
[14] Imtiaz, U. (2025). Investigating the impact of phishing attacks on organizational cybersecurity posture. *Spectrum of Engineering Sciences*, 1758-1780.
[15] S. Yuan *et al.*, "A privacy-preserving platform oriented medical healthcare and its application in identifying patients with candidemia," *Scientific Reports*, vol. 14, no. 1, p. 15589, 2024.
[16] X. Yang and J. Li, "A clustering-based federated deep learning approach for enhancing diabetes management," *Healthcare Analytics*, vol. 7, p. 100392, 2025.
[17] R. S. Mondal, M. N. A. Bhuiyan, and L. Akter, "Machine learning for chronic disease predictive analysis," *Applied IT & Engineering*, vol. 2, no. 1, pp. 1–11, 2024.
[18] E. Badidi, "Edge AI for early detection of chronic diseases," *Future Internet*, vol. 15, no. 11, p. 370, 2023.
[19] R. Torkzadehmahani *et al.*, "Privacy-preserving artificial intelligence techniques in biomedicine," *Methods of Information in Medicine*, vol. 61, no. S1, pp. e12–e27, 2022.
[20] S. Iqbal *et al.*, "Privacy-preserving collaborative AI for distributed deep learning," *Multimedia Tools and Applications*, vol. 83, no. 33, pp. 80051–80073, 2024.
[21] A. H. K. Choain *et al.*, "Integrating blockchain-enhanced enterprise systems," *Journal of Engineering and Computational Intelligence Review*, vol. 1, no. 1, pp. 51–63, 2023.
[22] M. Rowshon *et al.*, "Impact of generative AI and virtual reality on mental health," *International Journal of Mental Health Research and Global Education*, vol. 3, no. 1, pp. 784–796, 2025.
[23] T. A. Shiva *et al.*, "Optimizing early intervention strategies for neurodiverse children," *Apex Journal of Social Sciences*, vol. 3, no. 1, pp. 30–52, 2024.
[24] M. Shahinuzzaman *et al.*, "Mental health of women breast cancer survivor," *Jagannath Univ. J. Earth Life Sci.*, vol. 5, no. 1, pp. 1–12, 2019.
[25] L. Čepová *et al.*, "Improving privacy-preserving LSTM for encrypted MRI images," *Scientific Reports*, vol. 14, no. 1, p. 20218, 2024.

JOURNAL OF ENGINEERING AND COMPUTATIONAL INTELLIGENCE REVIEW (JECIR)
ISSN (Online): 3105-4811 ISSN (Print): 3106-969X
Volume 4, Issue 1, 2026
https://jecir.com/

[26] H. Kasyap and S. Tripathy, "Privacy-preserving decentralized learning framework," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 17, no. 2s, pp. 1–24, 2021.
[27] A. K. Alnaim and A. M. Alwakeel, "Machine-learning-based IoT–edge computing healthcare solutions," *Electronics*, vol. 12, no. 4, p. 1027, 2023.