Grenze International Journal of Engineering and Technology, January Issue

Differential Privacy Enabled Healthcare AI for Thyroid Disease Prediction

Yashvitha PR[1], Srivarshini G[2], Sunitha S[3], Dr. D Sudha Devi[4] and Dr. V Radhamani[5]
[1-3]Student, Department of Computing (Data Science), Coimbatore Institute of Technology, Coimbatore, India.
Email: 71762132058@cit.edu.in, 71762132049@cit.edu.in, 71762132051@cit.edu.in
[4]Associate Professor, Department of Computing (Data Science), Coimbatore Institute of Technology, Coimbatore, India.
Email: sudhadevi@cit.edu.in
[5]Assistant Professor, Department of Computing (Decision and Computing Sciences), Coimbatore Institute of Technology, Coimbatore, India.
Email: vradhamani@cit.edu.in

***Abstract*— The increasing reliance on machine learning in healthcare, finance, and social media raises serious concerns about the protection of sensitive user data. Protecting patient data privacy while harnessing the power of artificial intelligence for disease prediction is a critical challenge in modern healthcare. Traditional machine learning models are vulnerable to privacy attacks such as membership inference and model inversion, which can expose individual records. To address this challenge, we propose a differential privacy enabled healthcare AI system for thyroid disease prediction, ensuring strong privacy guarantees without compromising diagnostic accuracy. The system employs Laplace noise for sensitivity-based perturbations, Gaussian noise for advanced composition scenarios, and exponential mechanisms for discrete optimization tasks. Experiments were conducted on thyroid datasets with privacy budgets (ε) ranging from 0.1 to 1.0 to systematically evaluate privacy-utility trade-offs and analyze their effects on performance metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that differential privacy can be effectively integrated into healthcare AI pipelines with controlled accuracy degradation, thereby fostering the development of secure, trustworthy, and privacy-preserving predictive systems for sensitive medical applications.**

***Keywords*— Differential Privacy, Privacy- Preserving Machine Learning, Laplace Mechanism, Gaussian Mechanism, Exponential Mechanism, Privacy-Utility Trade-off, Thyroid disease prediction.**

I. INTRODUCTION

Data-driven decision-making has been transformed by the widespread adoption of machine learning in social media, healthcare, and finance; however, this advancement has also raised serious privacy concerns. Conventional machine learning models are vulnerable to sophisticated privacy attacks such as membership inference and model inversion, which can unintentionally expose sensitive information from their training datasets. These vulnerabilities pose significant risks in critical domains where privacy protection is essential, such as financial fraud detection systems that process personal transaction data and medical diagnosis systems that handle patient health records.

Grenze ID: 01.GIJET.12.1.236

Finding the best possible balance between model utility and privacy protection is the main obstacle. Data anonymization and k-anonymity approaches are examples of traditional privacy protection strategies that have not been able to withstand contemporary adversarial attacks and do not offer formal mathematical privacy guarantees. Because of this constraint, more stringent privacy-preserving techniques that may provide demonstrable security while ensuring suitable model performance for real-world implementation are required.
Differential privacy (DP), which offers mathematically sound guarantees that restrict the information adversaries can glean about specific data points, has become the gold standard for privacy protection. By introducing precisely calibrated noise into the learning process, this approach makes sure that the presence or absence of any one data record has as little of an effect as possible on the model outputs. Using Support Vector Machines for healthcare data categorization, this research provides a thorough analysis of three differential privacy mechanisms: Laplace, Gaussian, and Exponential. We show that robust privacy protection and model utility across different privacy budget allocations are achievable through systematic experiments on thyroid illness prediction.

II. Literature Survey

In machine learning applications, differential privacy has emerged as a crucial security measure for sensitive health data. A thorough scoping assessment is given by Ficek et al. [1], who examine the advantages, disadvantages, and current approaches of DP in health research. With DP as a tool for safe public health data sharing, Dyda et al. [2] draw attention to the necessity of striking a balance between data usability and confidentiality. Practical DP approaches for the healthcare industry are shown by Subramanian [3], who shows methods that preserve model performance while reducing privacy risks. Ratra et al. In comparison to conventional anonymization techniques, differential privacy provides a higher level of protection when re-identification concerns are evaluated [4]. By offering a framework for differentially private model release, Sangeetha et al. [5] expand on this idea and guarantee model utility without jeopardizing patient anonymity. Dankar and El Emam's early pioneering work [6] established the framework for using DP in healthcare to stop patient disclosure, which has impacted contemporary privacy-preserving strategies.
The combination of differential privacy and federated learning (FL) has drawn a lot of interest in distributed and collaborative learning environments. Vunnam et al. [7] present DeepFed, a federated learning system that integrates DP to preserve model performance while securing healthcare data across several institutions. Mohammadi et al. [8] summarize the progress and difficulties in integrating DP with neural networks in their assessment of differentiated privacy in deep learning applications for healthcare. Zheng et al. [9] suggest sensitivity-aware differential privacy for federated medical imaging, which adjusts noise levels according to feature sensitivity. Ahmed et al. [10] create adaptive DP-enabled FL for COVID-19 detection, which maintains high diagnostic accuracy while protecting patient privacy. Recent research focuses on optimizing privacy- utility trade-offs.
Additional enhancements include system-level strategies that include federated learning, secure aggregation, and DP. Adnan [11] suggests a private and safe healthcare system that boosts confidence in sharing large amounts of data, and Haj Fares & Saad
[12] offer a framework that combines DP with secure aggregation in medical imaging, offering strong defense against data leakage in cooperative training settings. Collectively, these experiments show that differential privacy provides a potent mechanism for safeguarding sensitive healthcare data without materially affecting model performance, particularly when combined with federated learning and secure aggregation.
Differential privacy works well for both federated and centralized machine learning when it comes to safeguarding private medical information. The integration of DP with safe aggregation and federated learning improves privacy without appreciably impacting model performance. These methods offer a strong basis for developing safe, precise, and repeatable prediction models.

III. Methodology

A. Dataset Description and Selection

The experimental analysis uses the UCI Machine Learning Repository's thyroid illness dataset, which consists of 3,772 cases with 28 unique variables, including demographic data, medical history, and full laboratory test results. The binary classification of thyroid problems (positive/negative diagnosis) is covered by this dataset, which offers a realistic healthcare application scenario that highlights the usefulness of privacy-preserving machine learning in delicate medical fields. TSH, T3, TT4, T4U, FTI, and TBG levels are among the thyroid-

related laboratory tests included in the dataset, together with patient age, gender, medication status, and surgical history.

The selection of healthcare data is especially important for research that protects privacy because medical records include extremely sensitive personal data that needs to be protected strictly by laws like HIPAA. The evaluation is made simpler while retaining clinical relevance due to the problem's binary classification nature (thyroid illness present or absence). This makes it possible to clearly examine privacy-utility trade-offs without the complexity of multi-class scenarios.

B. Data Preprocessing and Feature Engineering

To ensure data quality and compatibility with differential privacy methods, extensive data preprocessing was put into place. A combination of numerical, boolean, and categorical variables in the original dataset needed to be systematically transformed. The feature space was expanded from 28 initial dimensions to 1,094 features following transformation by applying one-hot encoding to categorical data such as gender, medication status, and measurement indicators.

Periodic gaps in laboratory measurements were identified by missing value analysis and filled using domain-appropriate imputation techniques. Continuous variables, such hormone levels, were normalized to guarantee consistent sensitivity estimates across various measurement scales, whereas boolean features that represented medical disorders and treatments were maintained in their binary form. Throughout the preprocessing pipeline, class balance concerns were maintained by encoding the target variable as binary (0 for a negative thyroid state, 1 for a positive one).

A 70-30 train-test split was used in the dataset partitioning approach, yielding 2,640 training examples and 1,132 testing instances. With the majority class (positive thyroid condition) accounting for roughly 92% of cases and the minority class (negative condition) for 8% of the dataset, stratified sampling guaranteed proportionate representation of both classes in the training and testing sets.

C. Differential Privacy Mechanisms Implementation

Laplace Mechanism: The Laplace mechanism adds noise to function outputs from a Laplace distribution Lap(0, $\Delta f/\varepsilon$), where ε is the privacy budget and Δf is the global sensitivity. Sensitivity analysis for machine learning applications establishes the maximum alteration in model parameters brought about by the addition or deletion of a single data point. The implementation ensures that noise addition offers ε-differential privacy guarantees by calculating sensitivity based on the L1 norm of gradient changes. This technique offers unbiased estimates with a symmetric noise distribution and works especially well for continuous outputs.

Gaussian Mechanism: The Gaussian mechanism uses noise from a normal distribution $N(0, \sigma^2)$, where (ε,δ)-differential privacy is ensured by carefully calibrating the variance σ^2. $\sigma \geq \Delta f\sqrt{(2\ln(1.25/\delta))/\varepsilon}$, where δ is the failure probability (usually set to 10^{-5}), yields the noise scale. This approach is perfect for iterative learning algorithms since it provides better composition qualities when several queries are executed. More effective use of the privacy budget over several training epochs is made possible by the Gaussian mechanism's compatibility with sophisticated privacy accounting methods.

Exponential Mechanism: The Exponential mechanism preserves differential privacy in situations when selection from discrete sets is necessary. For every conceivable output r, it allocates selection probabilities proportional to $\exp(\varepsilon u(D,r)/(2\Delta u))$, where $u(D,r)$ is a utility function and Δu is the utility sensitivity. This approach is useful for choosing model architectures, features, and hyperparameters in machine learning situations. In order to guarantee that higher-performing configurations obtain greater selection probability while upholding privacy restrictions, the implementation creates utility functions based on model performance criteria.

D. Privacy Budget Allocation and Experimental Design

According to recognized differential privacy standards, the experimental framework concentrates on epsilon values between 0.1 and 1.0, which indicate strong to moderate privacy protection levels. This range includes settings that balance privacy with practical utility requirements ($\varepsilon = 1.0$) and more relaxed settings ($\varepsilon = 0.1$), where individual data contributions are heavily concealed.

Direct comparison of privacy-utility trade-offs is made possible by the systematic evaluation technique, which applies each differential privacy mechanism independently across the epsilon spectrum. To guarantee reproducibility, each experimental setup is carried out using fixed random seeds, and performance measures are averaged over several runs to take stochastic noise fluctuations into consideration. In order to create utility standards for comparison analysis, the experimental design incorporates baseline model evaluation without differential privacy.

E. Model Architecture and Training Protocol

The fundamental learning technique is Support Vector Machines using Radial Basis Function (RBF) kernels, which were selected due to their proven efficacy in medical classification applications and their resilience to noise. To isolate the effects of differential privacy methods, the SVM implementation keeps the hyperparameters constant across all tests. Depending on the exact mechanism used, gradient perturbation and parameter noise injection are two ways that training involves differential privacy.

To identify suitable noise scales for every mechanism, the training methodology employs meticulous sensitivity analysis. With global sensitivity calculations, privacy requirements are maintained across all potential dataset alterations by taking into account the largest change in model parameters that could arise from changes to a single data point. Using common classification metrics like accuracy, precision, recall, and F1-score, performance evaluation pays close attention to class-imbalanced performance because of the features of the dataset.

IV. System Implementation

A. Architecture Overview and Design Principles:

- Using scikit-learn, numpy, and pandas, the privacy- preserving machine learning system is constructed on Python 3.8+ with a distinct separation of concerns and object-oriented design principles.
- The fundamental machine learning pipeline is unaffected when differential privacy methods are independently evaluated or altered.
- Modularity, scalability, and maintainability are prioritized in the design for long-term use and adaptability.
- The four main parts of the design are the Model Training Engine (SVM with privacy constraints), the Privacy Mechanisms Module (three differential privacy methods), the Data Handler (preprocessing), and the Evaluation Framework (performance evaluation).
- The modular approach preserves code quality and testing capabilities while guaranteeing simple extension to other privacy measures or machine learning algorithms.

B. Core System Components

- Data Handler Module: The data handling module incorporates error checking and validation into its robust preprocessing pipelines. The module manages feature scaling using StandardScaler, categorical encoding with scikit-learn's OneHotEncoder, missing value detection and imputation, and CSV data import. While memory- efficient processing manages huge datasets through batch processing capabilities, data integrity checks guarantee consistent feature dimensions across training and testing sets. For the sake of reproducibility and debugging, the module incorporates thorough logging of preprocessing stages and transformation parameters.
- Privacy Mechanisms Module: The three differential privacy mechanisms are implemented by this crucial component as separate classes that derive from a single DifferentialPrivacy abstract base class. The Laplace mechanism incorporates sensitivity analysis through gradient computation and L1 norm computations, and it makes use of numpy's laplace function with a scale parameter computed as global sensitivity/epsilon. With variance calculated using the formula $\sigma^2 = (\Delta f)^2 \cdot 2\ln(1.25/\delta)/\varepsilon^2$, where δ is set to 10^{-5} for practical purposes.
  The implementation of the exponential process involves probabilistic selection using numpy's random.choice function with calculated probability distributions and utility function optimization using Scipy's optimization methods. To guarantee that differential privacy guarantees are upheld during the training process, each method incorporates sensitivity bounds verification, noise production logging, and parameter validation.
- Model Training Engine: The training engine extends scikit-learn's SVC implementation by implementing a unique PrivacySVM class that combines Support Vector Machine learning with differential privacy requirements. In order to guarantee that noise injection takes place at the proper times during the learning method, the engine implements privacy protections during gradient computation and parameter updates. To preserve SVM convergence qualities under noise perturbation, the implementation involves dual optimization variables and careful handling of kernel computations.

In order to preserve consistency across all experiments, the training parameters are set with the RBF kernel (gamma='scale'), regularization parameter C=1.0, and default tolerance values. The engine incorporates features for model serialization and checkpoint saving to enable analysis continuation and experiment reproducibility.

C. Technical Implementation Details

- Noise Generation and Calibration: Sensitivity analysis conducted during system initialization necessitates meticulous calibration for precision noise injection. Through empirical estimate, the implementation determines global sensitivity by assessing the maximum parameter changes across perturbations in representative datasets. Conservative privacy assurances are established by caching sensitivity values and validating them against theoretical limitations.
  Explicitly defined seeds for reproducibility are used in the cryptographically safe random number generator provided by numpy. By carefully regulating the time of noise injection to take place after gradient computation but before parameter updates, theoretical privacy requirements are maintained while preventing privacy methods from interfering with SVM convergence qualities.
- Performance Monitoring and Logging: The system uses Python's logging module, which has adjustable verbosity settings, to implement thorough logging. At every training iteration, performance metrics such as accuracy, loss values, noise magnitudes, and privacy budget usage are gathered. During extensive tests, memory consumption monitoring guarantees effective resource utilization.
- Quality Assurance and Testing: Regression testing to verify privacy-utility trade-offs, integration testing for the entire pipeline, and comprehensive unit testing for every privacy mechanism are all part of the implementation. The validity of differential privacy guarantees under different epsilon configurations and dataset perturbations is confirmed using automated testing methods.

D. Deployment Configuration and Scalability

- The solution is compatible with both distributed deployment via Docker containerization and local execution. YAML files are used in configuration management to define evaluation metrics, privacy settings, and experimental parameters. For CPU- intensive calculations, the implementation uses Python's multiprocessing module to provide parallel processing capabilities for multiple epsilon evaluations and cross-validation processes.
- Options for database connectivity allow for direct interaction with healthcare data systems while upholding privacy regulations. In order to share experimental results without disclosing sensitive setup details, the system has built-in anonymization algorithms and data export options for results analysis and visualization.

V. RESULTS

A. Baseline Performance

On the thyroid disease dataset, the baseline Support Vector Machine model without differentiated privacy established a performance benchmark for evaluating the trade-off between privacy and usefulness with an accuracy of 95.36%. Without privacy restrictions, this baseline reflects the best possible utility (Figure 1).

```
--- New Sample Prediction ---
Input Sample Important Features:
  age                    : 45
  sex                    : F
  on thyroxine           : f
  TSH measured           : t
  TT4 measured           : t
  referral source        : SVHC

Predicted Class: N (Negative)
```

Figure 1. Sample prediction output

B. Differential Privacy Mechanism Comparison:

Gaussian Mechanism Performance: The Gaussian mechanism exhibited the best performance across all privacy levels. The prediction was performed after adding Gaussian noise to the data, simulating the privacy-preserving mechanism. The accuracy behaviour is illustrated in (Figure 2). At $\varepsilon = 0.1$, it achieved 91.92% accuracy with only a 3.44% accuracy loss compared to baseline (Figure 3). At $\varepsilon = 1.0$, accuracy improved to 95.63%, slightly exceeding the baseline due to regularization effects introduced by noise (Figure 4).

```
--- New Sample Prediction with Gaussian Privacy Mechanism---

Input Sample Important Features:
    age             : 45
    sex             : F
    on thyroxine    : f
    TSH measured    : t
    TT4 measured    : t
    referral source : SVHC

Predicted Class: N (Negative)
```

Figure 2. Sample prediction output with Gaussian Privacy Mechanism

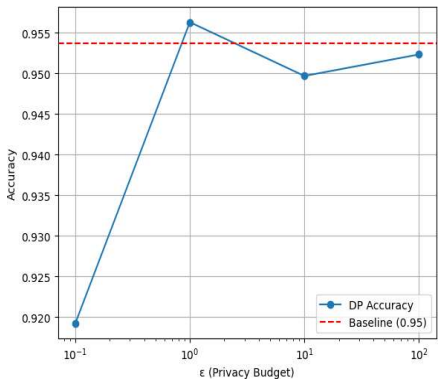


Figure 3. Accuracy vs Privacy Budget

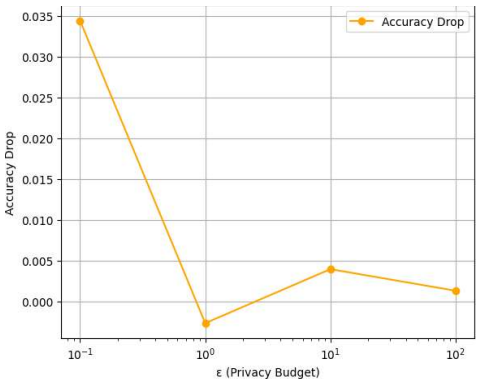


Figure 4. Accuracy Drop vs Privacy Budget

```
=== Baseline Model (No DP) ===
Accuracy: 0.9536

=== Differential Privacy Results ===

--- ε = 0.1 ---
Accuracy       : 0.9192
Accuracy Drop  : 0.0344
Classification Report:
              precision    recall  f1-score   support

           0       0.33      0.05      0.09        58
           1       0.93      0.99      0.96       697

    accuracy                           0.92       755
   macro avg       0.63      0.52      0.52       755
weighted avg       0.88      0.92      0.89       755


--- ε = 1.0 ---
Accuracy       : 0.9563
Accuracy Drop  : -0.0026
Classification Report:
              precision    recall  f1-score   support
...
    accuracy                           0.95       755
   macro avg       0.83      0.84      0.83       755
weighted avg       0.95      0.95      0.95       755
```

Figure 5. Model performance comparison

Laplace Mechanism Performance: The Laplace mechanism demonstrated better performance. The prediction was performed after adding Laplace noise to the data, enforcing privacy guarantees. As shown in (Figure 6), the mechanism experienced a moderate utility loss as privacy increased. The Laplace mechanism showed a performance with 94.70% accuracy at ε = 1.0 (0.66% drop) and 89.14% accuracy at ε = 0.1 (6.23% reduction) (Figure 7). It exhibited higher utility loss at stricter privacy settings (Figure 8).

```
--- New Sample Prediction with Laplace Privacy Mechanism---

Input Sample Important Features:
    age              : 45
    sex              : F
    on thyroxine     : f
    TSH measured     : t
    TT4 measured     : t
    referral source  : SVHC

Predicted Class: N (Negative)
```

Figure 6. Sample prediction output with Laplace Privacy Mechanism

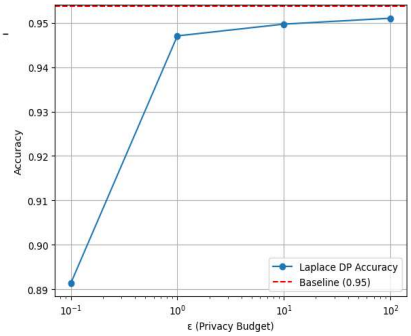


Figure 7. Accuracy vs Privacy Budget with Laplace Noise

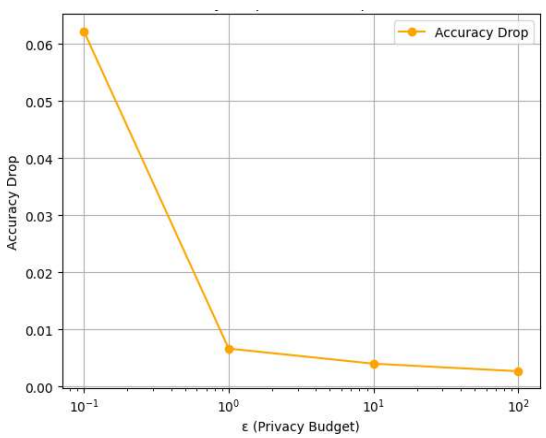


Figure 8. Accuracy Drop vs Privacy Budget with Laplace Noise

```
=== Baseline Model (No DP) ===
Accuracy: 0.9536

=== Differential Privacy Results with Laplace N

--- ε = 0.1 ---
Accuracy       : 0.8914
Accuracy Drop  : 0.0623
Classification Report:
              precision    recall  f1-score   s

           0       0.12      0.07      0.09
           1       0.93      0.96      0.94

    accuracy                           0.89
   macro avg       0.53      0.51      0.52
weighted avg       0.86      0.89      0.88


--- ε = 1.0 ---
Accuracy       : 0.9470
Accuracy Drop  : 0.0066
Classification Report:
              precision    recall  f1-score   s
...
    accuracy                           0.95
   macro avg       0.83      0.83      0.83
weighted avg       0.95      0.95      0.95
```

Figure 9. Model performance comparison

Exponential Mechanism Performance: The prediction was performed after adding noise to the data, simulating the privacy-preserving effect of the exponential mechanism (Figure 10). The exponential method was most sensitive to privacy constraints, achieving 87.15% accuracy at ε = 0.1 (8.21% drop) (Figure 11). However, it recovered to baseline performance with 95.36% accuracy at ε = 1.0 (Figure 12), showing sharp threshold behavior (Figure 13).

```
--- New Sample Prediction with Exponential Privacy Mechanism---

Input Sample Important Features:
    age              : 45
    sex              : F
    on thyroxine     : f
    TSH measured     : t
    TT4 measured     : t
    referral source  : SVHC

Predicted Class: N (Negative)
```

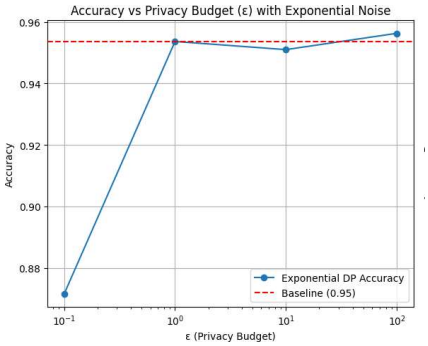
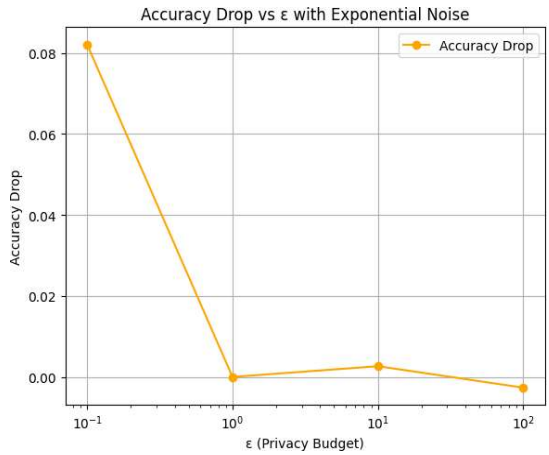


Figure 10. Sample prediction output with Exponential Privacy Mechanism

Figure 11. Accuracy vs Privacy Budget with Exponential Noise




```
=== Baseline Model (No DP) ===
Accuracy: 0.9536

=== Differential Privacy Results with Exponential

--- ε = 0.1 ---
Accuracy       : 0.8715
Accuracy Drop  : 0.0821
Classification Report:
              precision    recall  f1-score   sup

           0       0.07      0.05      0.06
           1       0.92      0.94      0.93

    accuracy                           0.87
   macro avg       0.49      0.50      0.49
weighted avg       0.86      0.87      0.86


--- ε = 1.0 ---
Accuracy       : 0.9536
Accuracy Drop  : 0.0000
Classification Report:
              precision    recall  f1-score   sup
...
    accuracy                           0.96
   macro avg       0.84      0.86      0.85
weighted avg       0.96      0.96      0.96
```

Figure 12. Accuracy Drop vs Privacy Budget with Exponential Noise

Figure 13. Model performance comparison

C. Privacy-Utility Trade-off Analysis

The experimental results show that each mechanism exhibits unique privacy-utility trade-off patterns. Even at high privacy levels, the Gaussian technique maintains the best trade-off with the least amount of accuracy loss. There is a logarithmic relationship between accuracy and epsilon values, with diminishing returns when privacy budgets rise above ε = 1.0 (Figure 14).

Classification performance investigation shows that at strict privacy levels (ε = 0.1), minority class performance degrades somewhat, but all algorithms maintain great precision and recall for the majority class (Figure 15). Across all mechanisms and privacy levels, the weighted average F1-scores stay above 0.86, indicating strong classification ability.

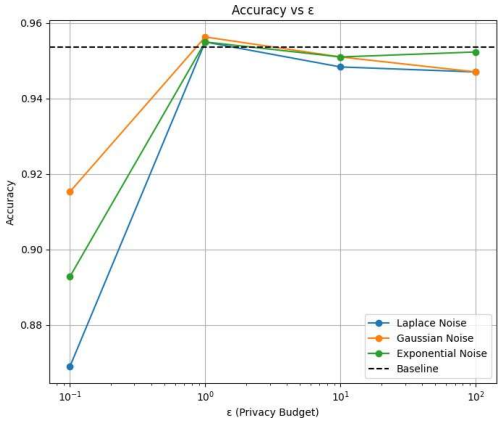


Figure 14. Accuracy comparison of noise mechanisms vs Privacy Budget (ε)

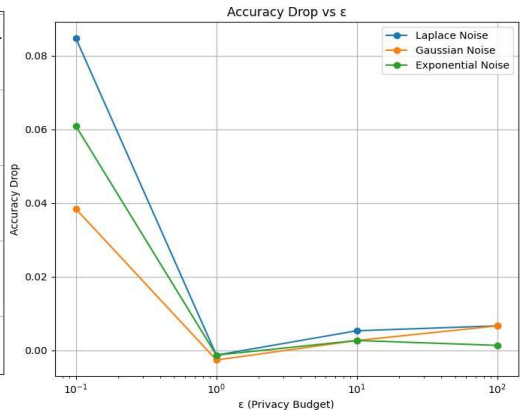


Figure 15. Accuracy Drop Comparison of Noise Mechanisms vs Privacy Budget (ε)

VI. Discussion and Future Work

A. Key Findings

The experimental evaluation shows that machine learning pipelines may successfully incorporate differential privacy with reasonable utility costs. In healthcare applications, the Gaussian technique is the most practical option because it provides the best privacy-utility balance. According to the findings, privacy budgets that are moderate (ε > 1.0) offer robust protection while preserving performance that is close to baseline.

Some methods show minor performance benefits when ε = 1.0, indicating that the regularization effect of differential privacy noise can sometimes improve generalization, according to the observed performance patterns. There are significant ramifications for real-world deployment scenarios from this discovery.

B. Limitations

The current study restricts generalizability across various problem domains and data characteristics by concentrating on binary classification using a single dataset. Only epsilon- differential privacy is taken into account in the privacy analysis; more complex variations like Rényi differential privacy and concentrated differential privacy are not examined. The implementation is predicated on a trusted curator model in which a single entity applies the privacy mechanism. Different privacy rules and techniques are needed for distributed or federated learning scenarios with numerous data holders.

C. Future Directions

Future research should investigate adaptive privacy budget allocation schemes that maximize value according to dataset properties and query significance. For distributed privacy- preserving machine learning, integration with federated learning frameworks is a viable avenue.

In order to manage the privacy budget across several inquiries and model updates, more complex composition theorems and privacy accounting techniques may be necessary. Further research into domain-specific utility functions for the Exponential mechanism may enhance its functionality in specific applications.

Important areas for improving practical deployment include the creation of automated privacy budget allocation methods and privacy-preserving hyperparameter optimization approaches.

VII. CONCLUSION

In order to classify healthcare data, this work offers a thorough framework for privacy-preserving machine learning that makes use of differential privacy protections. Across various privacy budget allocations, the experimental evaluation shows that adequate model utility may be preserved while formal privacy obligations are upheld.

With 91.92% accuracy even under strict privacy constraints ($\varepsilon = 0.1$) and near-baseline performance at moderate privacy levels, the Gaussian method works best for real-world applications. Comparing Laplace, Gaussian, and exponential methods methodically offers important information for choosing the best privacy-preserving strategies depending on the needs of the application.

This research creates benchmark performance measurements for privacy-utility trade-offs in healthcare machine learning and offers useful implementation guidance for differential privacy deployment in sensitive data applications. A wider adoption of privacy-preserving machine learning methods is supported by the framework's modular architecture, which allows adaptability across different domains and datasets.

In order to address important privacy concerns while preserving operational efficacy, the results validate the viability of implementing differentially private machine learning systems in production settings. The state-of-the-art in privacy-preserving AI is advanced by this work, which also lays the groundwork for further studies in safe and reliable machine learning applications.

REFERENCES

[1] Joseph Ficek, Wei Wang, Henian Chen, Getachew Dagne, Ellen Daley, "Differential privacy in health research: A scoping review," Journal of the American Medical Informatics Association, 2021, DOI: https://doi.org/10.1093/jamia/ocab135.

[2] Amalie Dyda, Michael Purcell, Stephanie Curtis, Emma Field, Priyanka Pillai, Kieran Ricardo, Haotian Weng, Jessica C. Moore, Michael Hewett, Graham Williams, Colleen L. Lau, "Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality," Patterns, vol. 2, no. 12, 2021, DOI: https://doi.org/10.1016/j.patter.2021.100366. for Healthcare Data," 2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA), San Antonio, TX, USA, pp. 95–100, 2022, DOI: https://doi.org/10.1109/IDSTA55301.2022.9923037.

[3] Ritu Ratra, Preeti Gulia, Nasib Singh Gill, "Evaluation of Re-identification Risk using Anonymization and Differential Privacy in Healthcare," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 2, 2022, DOI: https://doi.org/10.14569/IJACSA.2022.0130266.

[4] S. Sangeetha, G. Sudha Sadasivam, A. Srikanth, "Differentially private model release for healthcare applications," International Journal of Computers and Applications, vol. 44, no. 10, pp. 953–958, 2022, DOI: https://doi.org/10.1080/1206212X.2021.2024958.

[5] Fida Kamal Dankar, Khaled El Emam, "The application of differential privacy to health data," Proceedings of the 2012 Joint EDBT/ICDT Workshops, pp. 158–166, 2012, DOI: https://doi.org/10.1145/2320765.2320816.

[6] Nithin Vunnam, Deepak Venkatachalam, Bhaskar Yakkanti, "DeepFed: Federated Learning with Differential Privacy for Secure Healthcare Data Analysis," American Journal of Data Science and Artificial Intelligence Innovations, vol. 1, 2021, DOI: https://www.ajdsai.org/index.php/publication/article/view/45.

[7] Mohammadi, M., et al., "Differential Privacy for Deep Learning in Medicine: A Scoping Review," arXiv preprint arXiv:2506.00660, 2025, DOI: https://doi.org/10.48550/arXiv.2506.00660.

[8] Zheng, H., Yang, Y., Zhao, H., et al., "Sensitivity-Aware Differential Privacy for Federated Medical Imaging," Sensors, vol. 25, no. 9, 2025, DOI: https://doi.org/10.3390/s25092847.

[9] Ahmed, K., et al., "Adaptive Differential Privacy–Enabled Federated Learning for COVID-19 Detection," Frontiers in Medicine, 2024, DOI: https://doi.org/10.3389/fmed.2024.1409314.

[10] Adnan, M., "A Secure and Privacy-Preserving Approach to Healthcare Systems Using Federated Learning," Symmetry, vol. 17, no. 7, 2025, DOI: https://doi.org/10.3390/sym17071139.

[11] Haj Fares, A., Saad, W., "Towards Privacy-Preserving Medical Imaging: Federated Learning with Differential Privacy and Secure Aggregation," arXiv preprint arXiv:2412.00687, 2024, DOI: https://doi.org/10.48550/arXiv.2412.00687.

[12] R. Subramanian, "Differential Privacy Techniques