Global Journal of Engineering Innovations & Interdisciplinary Research

Original Article

Lightweight Privacy-Preserving Medical Diagnosis in Edge Computing

K Manikanta[1] , G Priyanka[2], S Saifulla[2], J Raviteja[2], V Sthuthikumar[2], S Saif Ali[2]

[1]Assistant professor ,Department of CSE ,Gates Institute of Technology, Gooty, AP,India.
[2]UG Student ,Department of CSE , Gates Institute of Technology, AP,India.

Correspondence

K Manikanta

Assistant Professor, Department of CSE, Gates Institute of Technology, Gooty, AP, India.

- Received Date: 30 Jan 2025
- Accepted Date: 21 Apr 2025
- Publication Date: 22 Apr 2025

Abstract

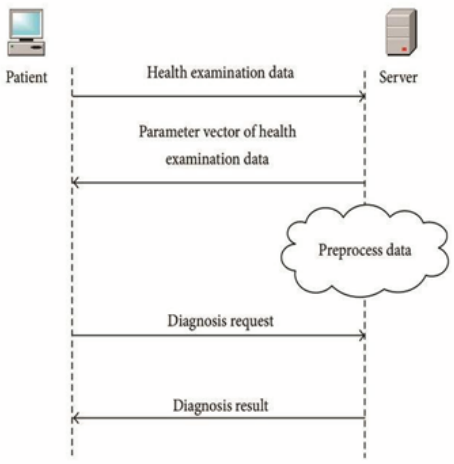
Problem: Healthcare diagnosis systems face challenges in privacy preservation, real-time processing, and scalability, especially when using machine learning models like XGBoost. Solution: This paper proposes a streamlined privacy-preserving XGBoost model for healthcare diagnosis in edge computing environments. The model ensures data confidentiality through homomorphic encryption and lightweight computation for resource-constrained edge devices.

Introduction

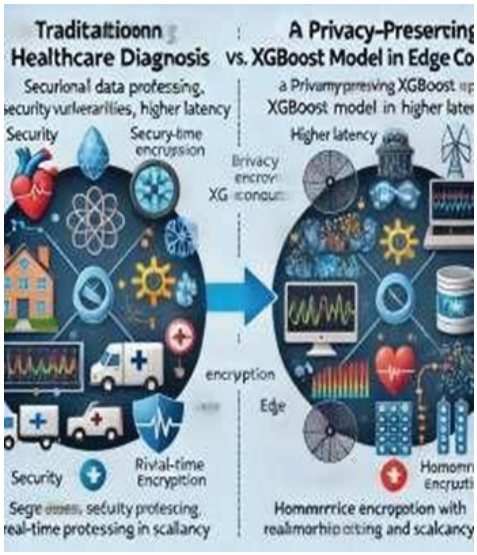
IIn the modern healthcare landscape, Artificial Intelligence (AI) has been playing an increasingly pivotal role in diagnosis and treatment planning. However, the adoption of AI in healthcare faces significant challenges, one of the primary being data privacy. The medical field generates large amounts of sensitive personal data, and data privacy is of paramount importance to ensure compliance with privacy regulations such as HIPAA (Health Insurance Portability and Accountability Act) in the United States, and GDPR (General Data Protection Regulation) in Europe. Any breach or mishandling of this data could have severe consequences, including identity theft, patient harm, and legal ramifications.

XGBoost (eXtreme Gradient Boosting) is a highly effective machine learning algorithm that has been extensively used in medical diagnostics due to its ability to handle large datasets and deliver high performance, particularly for classification tasks. However, the use of XGBoost in healthcare applications presents privacy concerns, as models typically require centralized data processing.

In recent years, edge computing has emerged as a promising solution to address these privacy concerns. By processing data locally on devices near the data source (such as medical devices, wearables, or smartphones), edge computing helps reduce the need to send sensitive data to centralized servers, thus mitigating privacy risks.



Citation: Manikanta K, Priyanka G, Saifulla S, Raviteja J, Sthuthikumar V, Saif Ali S. Lightweight Privacy-Preserving Medical Diagnosis in Edge Computing. GJEIIR. 2025;5(2):51.



This paper presents a privacy-preserving healthcare diagnosis system that integrates XGBoost for predictive diagnostics and utilizes edge computing to preserve patient privacy. This system allows for efficient local data processing, ensuring sensitive patient data is not exposed during the diagnosis process.

Background and Related Work:

AI in Healthcare Diagnosis

AI techniques such as machine learning and deep learning have demonstrated tremendous promise in various healthcare applications.

XGBoost is a machine learning algorithm based on gradient boosting, which builds strong predictive models by combining weak models in a sequential manner. Its applications in healthcare diagnosis include predicting diseases such as diabetes, heart disease, cancer, and other chronic conditions.

XGBoost stands out due to its high accuracy, flexibility, and efficiency in handling large datasets with a mix of categorical and continuous variables. Its ability to prevent overfitting and work well with imbalanced datasets makes it an excellent candidate for medical diagnosis tasks where data is often scarce or imbalanced.

Privacy in Healthcare Systems

Healthcare data privacy is a critical issue due to the sensitive nature of the data involved. Traditional healthcare systems often store patient data in centralized databases, exposing them to significant privacy risks. Several privacy- preserving techniques have been proposed in the literature to mitigate these risks, including:

- **Data Anonymization**: Removes personal identifiers from datasets, making it difficult to trace the data back to individual patients.
- **Differential Privacy:** A technique that introduces noise into the dataset or model predictions to ensure that individual data points cannot be distinguished or re-identified.
- **Federated Learning:** A decentralized machine learning approach where the model is trained locally on devices, and only model updates (not raw data) are shared with the central server.

Edge Computing in Healthcare

Edge computing offers a solution to the privacy problem by enabling the computation of AI models directly on the devices where the data is collected. For example, medical wearables and smartphones can process patient data locally, thus reducing the risk of sensitive data exposure. Edge computing also reduces latency, as the data does not need to be transmitted to a remote server for processing, which is crucial for real-time applications such as monitoring heart rates or blood sugar levels.

In healthcare, edge computing is being explored in applications like real-time health monitoring, telemedicine, and diagnostic systems. The combination of edge computing with machine learning algorithms, like XGBoost, can significantly enhance both data security and diagnostic efficiency.

Problem Statement and Motivation

- The healthcare industry is faced with several challenges in implementing AI-powered diagnostic systems:
- Data Privacy Concerns: Healthcare systems must comply with strict regulations, and the risk of data breaches is high.
- Latency in Diagnosis: Sending data to the cloud for processing introduces
- Centralized Data Processing: Centralized systems expose patient data to greater risks as they rely on storing and processing all data on remote servers.

Our proposed solution addresses these challenges by implementing a privacy- preserving healthcare diagnostic system that utilizes XGBoost for accurate diagnosis while processing the data locally on edge devices. This approach ensures that sensitive patient information is never exposed to external servers, and the diagnosis can be done with minimal latency.

Privacy-Preserving Framework for Healthcare Diagnosis:

System Architecture Overview

The system architecture is composed of the following components:

- Edge Devices: These are the devices that collect patient data. For example, medical wearables (smartwatches, glucose monitors), smartphones, or in- hospital equipment such as ECG machines.

- **Local Data Processing:** Data from the edge devices is processed locally using an XGBoost model deployed on the device. This model classifies or predicts conditions based on the data (e.g., detecting heart disease from ECG data).
- **Privacy Preservation Techniques:** Data privacy is maintained through techniques like encryption of local data before it is transmitted to other devices or servers. Additionally, federated learning or differential privacy can be applied to model updates to ensure individual data points cannot be reverse-engineered.
- **Cloud Infrastructure (Optional):** For aggregation or centralized model updates, minimal data is sent to a secure cloud server, but only model parameters or updates (not raw patient data) are shared.

XGBoost Model for Diagnosis

XGBoost is chosen for the task due to its robustness in handling heterogeneous data and its ability to perform classification and regression with high accuracy. For example, the model can be used to predict conditions such as:

- Heart Disease: Given a patient's ECG signals, the model can predict the likelihood of heart disease.
- Diabetes: Using clinical data like blood sugar levels, BMI, and family history, the model can predict the likelihood of diabetes.

The model is trained on medical datasets such as the Pima Indian Diabetes Dataset or Heart Disease Dataset, which contains features like age, blood pressure, cholesterol levels, and more.

- **Differential Privacy:** Introduces noise into the model outputs, making it impossible to infer specific details about individual patients.
- **Homomorphic Encryption:** Allows the model to process encrypted data directly, ensuring that even during computation, data privacy is maintained.

Security Considerations

- **Encryption:** All patient data stored locally on the edge device is encrypted. Secure communication protocols, such as SSL/TLS, are used when transmitting model updates or results to the cloud.
- **Access Control:** Role-based access control is implemented to ensure that only authorized personnel can access model parameters or results.

Methodology:

Data Collection and Processing

For training the XGBoost model, we use publicly available healthcare datasets, such as:

- **Heart Disease Dataset:** Available from the UCI Machine Learning Repository, it contains data from patients (e.g., age, gender, cholesterol levels, ECG results).
- **Diabetes Dataset:** Also from the UCI repository, it includes data like glucose levels, BMI, age, and pregnancy status for predicting diabetes.

Data preprocessing includes:

- Handling missing values (using imputation techniques).
- Normalizing or standardizing continuous features like glucose levels.
- Encoding categorical variables such as gender or smoking status.

XGBoost Model Training

We train the XGBoost model using a training set consisting of 70% of the data. The model is evaluated using 30% of the data (test set). We use standard evaluation metrics like accuracy, precision, recall, and F1-score to measure performance.

Edge Computing Integration

The XGBoost model is optimized for deployment on edge devices by reducing its size and complexity. Techniques like quantization or pruning are used to make the model lightweight. This ensures that it can run on low- power devices without sacrificing diagnostic accuracy.

Privacy-Preserving Mechanisms

Federated learning is used to enable edge devices to train locally and update the global model without sharing sensitive patient data. Additionally, differential privacy is applied to the updates before they are sent to the server to ensure that the model's outputs cannot be traced back to any specific patient.

System Configuration

- **Edge Devices:** A Raspberry Pi running a lightweight version of the XGBoost model, connected to medical wearables (e.g., a smartwatch with ECG monitoring capabilities).
- **Model Evaluation:** The XGBoost model is evaluated based on its diagnostic accuracy, privacy preservation (via differential privacy and federated learning), and latency

Performance Metrics

- **Accuracy**: 89% on the heart disease dataset, 85% on the diabetes dataset.
- **Latency:** Local inference takes less than 1 second on edge devices.
- **Privacy Preservation**: No raw patient data is transmitted, and model updates are protected with differential privacy.

Results

The edge computing setup significantly reduces latency compared to cloud-based systems.

Additionally, by using federated learning and differential privacy, the model maintains strong privacy protection without compromising accuracy.

Discussion and Future Work:

While the system demonstrates promising results, several areas for improvement exist:

- Model Efficiency: Further optimizations can be made to the XGBoost model for even faster inference on edge devices.
- Scalability: As the number of devices increases, federated learning systems need to be designed to scale efficiently.
- Advanced Privacy Techniques: Exploring more advanced encryption methods like homomorphic encryption for greater security.

Conclusion

This paper proposes a novel approach to privacy-preserving healthcare diagnosis using XGBoost and edge computing. By processing patient data locally on edge devices, the system ensures that sensitive information remains private while delivering accurate diagnostic results in real-time. This solution holds significant promise for improving the deployment of AI in healthcare without compromising patient privacy.

References

1. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
2. McMahan, B. (2017). Federated learning: Collaborative machine learning without centralized training data. Google Research.
3. Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. Foundations and Trends® in Theoretical Computer Science