IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

# Secure Patient Data Anonymization and Readmission Risk Prediction

**Akilah Lavinia Falcao**
Forensic Science, Bangalore, India

**Abstract**: *The incorporation of machine learning into healthcare analytics promises to greatly enhance patient outcomes as well as minimize operational expenses. Hospital readmission prediction, especially within 30 days, is an essential activity intended to enhance quality of care as well as avoid financial penalties for healthcare facilities. Nonetheless, the utilization of sensitive patient information presents great risks to patient privacy, and thus the enforcement of strong privacy-preserving mechanisms is required. This project presents a comprehensive pipeline for developing a differentially private machine learning model to predict 30-day hospital readmission based on electronic health record (EHR) data. The approach includes combining patient demographic information, admission histories, and lab test data into a single large dataset. Data is thoroughly cleaned through addressing missing values, normalizing temporal fields, and validating chronology. A readmission label as a binary target is crafted in relation to the time elapsed between subsequent admissions. Laboratory information is converted through pivoting to build organized sets of features reflecting clinical measurements. A preprocessing pipeline is built using numerical scaling and one-hot encoding for categorical features, which yields a high-dimensional sparse feature matrix. A deep neural network with sparse input layers is trained using Differentially Private Stochastic Gradient Descent (DP-SGD) to guarantee that the model has strong privacy guarantees. The training procedure investigates the effect of different noise multiplier values and batch sizes, analyzing systematically the trade-offs between model utility and privacy loss. Smaller noise multipliers provide higher model utility but less robust privacy guarantees, whereas larger noise levels provide stronger privacy at the expense of performance. Likewise, batch size modifications are demonstrated to affect both learning dynamics and privacy budgets. Model accuracy is measured according to classification, and privacy loss is measured with the epsilon (ε) measure from differential privacy analysis. Results show that preserving decent model accuracy is possible alongside obtaining useful privacy guarantees. The research presents evidence of the use of machine learning and differential privacy together for health applications as an important step towards secure, ethical, and responsible data-driven health innovation.*

**Keywords**: Hospital Readmission Prediction, Differential Privacy, Electronic Health Records (EHR), Machine Learning in Healthcare, Privacy-Preserving Deep Learning



DOI: 10.48175/IJARSCT-26411