

Proposed Topic of Research

Analytics-Driven Big Data Framework for Predictive Healthcare: Early Detection and Risk Assessment of Chronic Diseases Using Electronic Health Records (EHRs).

Objective of the Proposed Research

To develop a predictive analytics system using big data techniques to identify early risk factors and progression patterns of chronic diseases from large-scale healthcare datasets.

Background

Healthcare systems are increasingly generating vast volumes of structured and unstructured data from EHRs, wearable devices, and clinical reports. Chronic diseases like diabetes and cardiovascular conditions contribute significantly to healthcare burdens. Traditional methods struggle to extract timely, actionable insights from this data. Big data analytics enables real-time processing and integration of heterogeneous data sources, uncovering hidden patterns and correlations. Predictive modeling can thus support early intervention and personalized care. Despite growing potential, challenges remain in data quality, integration, and interpretability, which this research aims to address through a robust, scalable big data analytics framework.

Methodology

The research will implement a big data analytics pipeline using platforms like Apache Hadoop and Spark for data ingestion, cleaning, and transformation from EHRs and sensor data. Advanced machine learning models—random forests, gradient boosting, and neural networks—will be trained to identify risk factors and predict disease onset. Feature engineering will incorporate demographic, clinical, and lifestyle variables. Model interpretability will be enhanced using SHAP or LIME. The system will be evaluated on real-world datasets (e.g., MIMIC-IV, eICU), using metrics such as AUC-ROC, precision, and recall. Cloud-based deployment will ensure scalability and real-time analysis capability.

Expected Outcome

The project will deliver a scalable, interpretable predictive analytics platform for chronic disease management, enabling earlier diagnosis, reducing hospitalizations, and supporting data-driven public health policies and personalized care strategies.