Related Works on Personalized Healthcare using Artificial Intelligence

Djelloul Daouadji Fadela

May 5, 2025

Abstract

This report summarizes recent research and developments in the field of AI-based personalized healthcare. It highlights the key contributions, methodologies, and challenges faced by existing systems.

1 Introduction

The use of Artificial Intelligence (AI) in healthcare has opened new avenues for personalized treatment. This report reviews existing studies and methods used in AI-driven healthcare solutions.

2 Related Works

3 Towards Realizing the Vision of Precision Medicine: AI-Based Prediction of Clinical Drug Response

Article Reference: [1]

Overview

This study uses machine learning to predict patient response to the epilepsy drug brivaracetam using integrated clinical and genomic data. The resulting model demonstrated strong performance (AUC: 0.76 training, 0.75 validation) and identified specific biomarkers associated with poor response. The research underscores the potential of ML models to support precision medicine and optimize clinical trials by targeting likely responders.

Dataset

- **Discovery dataset:** 235 adult patients from a phase III clinical trial (NCT01261325).
- **External validation dataset:** 47 patients from an independent trial (NCT00490035).

Processing

Clinical data included demographic and seizure-related information. Whole Genome Sequencing (WGS) data (20 million variants) was filtered down to 40 features through knowledge-driven extraction, focusing on epilepsy-related genes and drug mechanism (e.g., SV2A gene, eQTLs). Genetic features included mutational load scores, polygenic risk scores, and structural variant descriptors.

Model Building

Multiple ML models were evaluated: sparse multi-block PLS-DA, multimodal neural networks, elastic net, gradient-boosted decision trees (GBDT), and stacked classifiers. The best performance was achieved using a GBDT model integrating all data types.

Results

The GBDT classifier achieved: AUC (training): 0.76

Future Directions and Challenges

- Addressing high dimensionality and sparsity of genomic data.
- Integrating additional data types (e.g., EEG, imaging) to improve model performance.
- Generalizing models to other anti-epileptic drugs.
- Collaborating with regulatory bodies for clinical adoption.
- Increasing dataset size to enhance model performance (targeting 350 patients for AUC = 0.9).

4 Diabetes Prediction Using Machine Learning and Explainable AI Techniques

Article Reference: [2]

Overview

This study proposes an automated diabetes prediction system using ML and explainable AI. The system combines the public Pima Indian dataset with a private dataset collected from female workers in a Bangladeshi textile factory. The system addresses data imbalance, missing values, and is deployed for real-time prediction via web and mobile applications.

Dataset

- **Pima Indian Dataset:** 768 records, 268 diabetes-positive; includes 8 features.
- **RTML Private Dataset:** 203 female employees; features similar to Pima dataset but lacks insulin values.

Processing

- Zero values in the merged dataset were replaced with corresponding mean values and the dataset was separated into training and test sets using the holdout validation technique.
- Mutual information was used to measure the interdependence of variables and feature importance.
- A semi-supervised approach using the extreme gradient boosting technique (XGB regressor) was used to predict the missing insulin feature of the RTML dataset.

ML Approach

Various models were tested: decision trees, KNN, SVM, random forest, logistic regression, AdaBoost, XGBoost, bagging, and voting classifiers. Hyperparameters were tuned using GridSearchCV. The final model employed XGBoost with ADASYN for balancing.

Results

The best results were obtained using the XGBoost classifier with ADASYN:

- Accuracy: 81%
- F1 Score: 0.81
- AUC: 0.84

Challenges

- Missing insulin values required imputation via semi-supervised learning.
- Class imbalance necessitated oversampling (SMOTE, ADASYN).
- Limited private dataset size may hinder generalizability.

Future Directions

- Expanding dataset size for better robustness.
- Integrating fuzzy logic and optimization for improved prediction.

5 Integrating Machine Learning and Deep Learning Techniques for Advanced Alzheimer's Disease Detection through Gait Analysis

Article Reference: [?]

Overview

The paper aims to enhance early detection of Alzheimer's Disease (AD) by leveraging gait analysis combined with advanced machine learning (ML) and deep learning (DL) techniques. Gait abnormalities, such as reduced stride length and irregular cadence, are identified as early biomarkers for cognitive decline associated with AD. The study emphasizes the need for non-invasive, scalable diagnostic tools.

Dataset

Data were collected using wearable sensors and motion capture systems in both clinical and real-world environments, providing high-resolution temporal and spatial gait metrics. The dataset includes gait features like stride length, cadence, swing time, and gait variability, with some data sourced from publicly available repositories like the UCI Machine Learning Repository. Preprocessing steps involved normalization, handling missing data via median imputation, class balancing with SMOTE, and feature selection through Recursive Feature Elimination.

Processing

- Normalization: Features were scaled between 0 and 1 to standardize the data, ensuring that features with larger ranges (e.g., stride length) did not dominate the model training. - Handling Missing Data: Missing values were imputed using median substitution to maintain data integrity and reduce bias. - Class Imbalance: The Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic samples of the minority class (AD patients), addressing class imbalance issues. - Feature Selection: Recursive Feature Elimination (RFE) was used to identify the most significant gait features—such as stride length, gait variability, and cadence—to improve model performance. - Correlation Analysis: High correlations between key features (e.g., stride length and step length) validated their importance for prediction, informing feature selection.

ML Approach

The study employed a hybrid deep learning model comprising Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to classify individuals as healthy or at risk for AD. These models analyzed temporal-spatial gait features, capturing sequential patterns and irregularities. Traditional ML classifiers such as Random Forest and SVM were also evaluated for comparison.

Results

The hybrid CNN-RNN model achieved the highest accuracy of 93%, with other metrics like precision, recall, and F1-score also indicating strong performance. Traditional models like Random Forest and SVM performed well but with slightly lower accuracy (88% and 86%, respectively). These results demonstrate the potential of deep learning models in accurately detecting early AD.

Challenges

- The reliance on controlled datasets, which may not fully reflect real-world variability, impacting model robustness.
- The complexity and interpretability of deep learning models, posing a barrier for clinical acceptance.
- The need for large, diverse datasets to ensure generalizability.
- Integration into clinical workflows and validation through real-world testing.

Future Directions

- Incorporating multimodal data sources, such as MRI, PET scans, vocal, and cognitive measures, to improve diagnostic precision.
- Expanding datasets to include diverse populations and environmental conditions, enhancing model robustness.
- Developing explainable AI frameworks to improve interpretability and clinician trust.
- Extending studies to include longitudinal gait data for monitoring disease progression and enabling earlier detection.
- Conducting clinical pilot studies and developing affordable wearable technologies for widespread, low-resource application.

6 Comparison of the Solutions

The table below compares the reviewed studies based on disease domain, dataset, preprocessing methods, approach, and results.

Work	Disease/Domain	Dataset	Data Processing	Approach	Results
[1]	Epilepsy	Phase III (235) + Validation (47) patients	Clinical + WGS feature extraction (e.g., SV2A), mutational scores, PRS	Gradient-Boosted Decision Trees	AUC: 0.76 (train), 0.75 (validation)
[2]	Diabetes Prediction	Pima Indian (768) + RTML (203) records	Imputation, ADASYN, Mutual Info, Holdout Validation	XGBoost + Ensemble Methods (voting, bagging)	AUC: 0.84, Accuracy: 81%, F1 Score: 0.81
[?]	Alzheimer's Disease	300 instances from wearable sensors and motion capture	Normalization, median imputation, SMOTE, RFE, correlation analysis	Hybrid CNN-RNN (LSTM)	Accuracy: 93%, Precision: 92%, Recall: 91%, F1-Score: 91.5%, AUC-ROC: 95%

Table 1: Comparison of AI Approaches in Health Applications

7 Conclusion

Personalized healthcare using AI continues to evolve, offering significant potential to improve patient care. However, integration into real-world clinical settings remains an ongoing challenge.

References

[1] Johann de Jong, Ioana Cutcutache, Matthew Page, Sami Elmoufti, Cynthia Dilley, Holger Fröhlich, and Martin Armstrong. Towards realizing the vision of precision medicine: Ai based prediction of clinical drug response. *Brain*, pages 1738–1750, 03 2021.

[2] Isfafuzzaman Tasin, Tansin Ullah Nabil, Sanjida Islam, and Riasat Khan. Diabetes prediction using machine learning and explainable ai techniques. *Healthcare technology letters*, 2023.