

# HFR-MADM Based Healthcare Disease Prediction System

Automated dataset evaluation and intelligent disease risk prediction for clinical decision support

⚠ CHALLENGE

# The Dataset Quality Problem

## Critical Issue

Healthcare datasets vary dramatically in quality, completeness, and reliability

Selecting suboptimal datasets leads to poor prediction accuracy, potentially impacting patient care and clinical decisions. Manual dataset evaluation is time-consuming and subjective.

**The need:** An intelligent system that automatically evaluates, ranks, and selects the highest-quality datasets for disease prediction tasks.

# Project Objectives

---

## Compare Multiple Datasets

Systematically evaluate diverse healthcare datasets across multiple diseases

## Rank Using HFR-MADM

Apply Hesitant Fuzzy Rough Multi-Attribute Decision Making for objective scoring

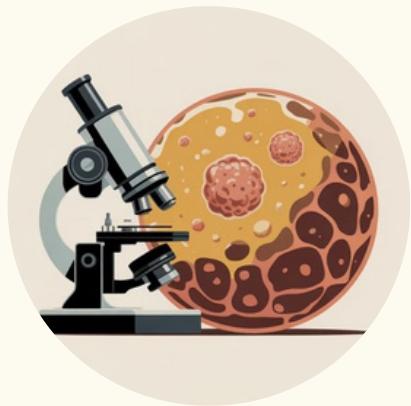
## Automatic Selection

Identify and select the optimal dataset based on weighted criteria

## Predict Disease Risk

Train machine learning models on the best-ranked dataset for accurate predictions

# Healthcare Datasets Evaluated



## Breast Cancer

Tumour characteristics and diagnostic features



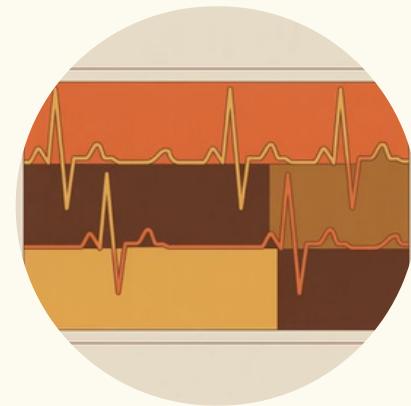
## Brain Stroke

Neurological indicators and risk factors



## Diabetes

Metabolic markers and glucose levels



## Heart Disease

Cardiovascular health parameters



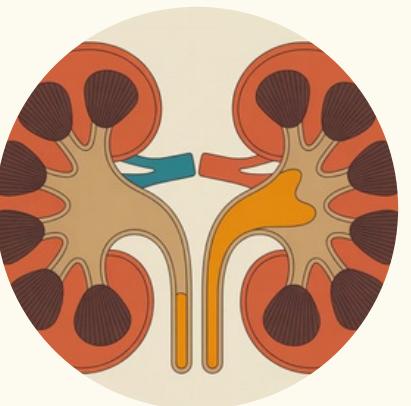
## Hypertension

Blood pressure and circulatory data



## Liver Disease

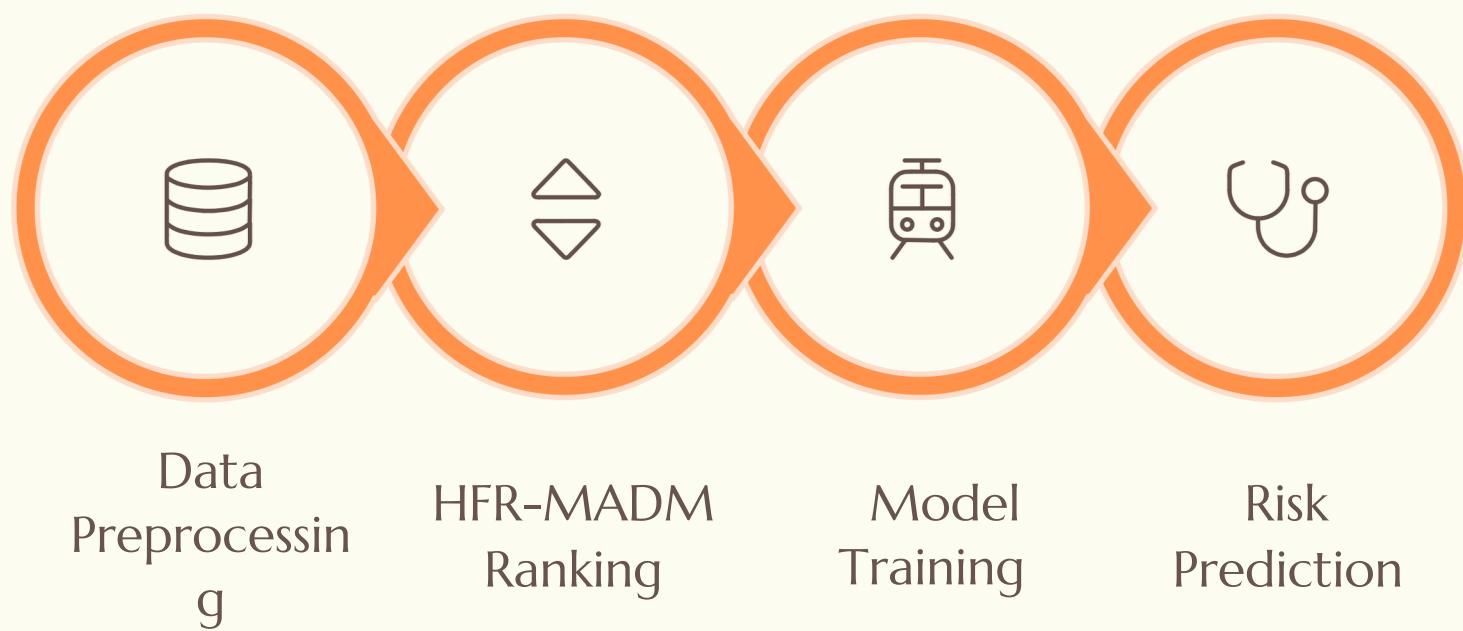
Hepatic function and enzyme markers



## Kidney Disease

Renal function indicators

# System Methodology



- 1 **Preprocessing**  
Clean and prepare raw healthcare data
- 2 **Ranking**  
Apply HFR-MADM evaluation criteria
- 3 **Training**  
Build logistic regression models
- 4 **Prediction**  
Generate patient risk assessments

# Dataset Preprocessing Pipeline

## Missing Value Removal

Eliminate incomplete records to ensure data integrity

## Duplicate Elimination

Remove redundant entries for accurate analysis

## Categorical Encoding

Transform categorical variables into numerical format

## Feature Separation

Isolate predictive features from target variables



# Model Architecture and Evaluation

## Logistic Regression

Chosen for interpretability and clinical explainability

## 80-20 Split

Training and testing data partition for robust validation

## StandardScaler

Feature normalisation for improved model convergence

**Evaluation Metrics:** Model performance assessed using accuracy scores, F1-scores for class balance, and confusion matrices for detailed classification analysis.

# Interactive System Features



## Interactive Dashboard

Intuitive web interface for clinicians and data scientists to explore results



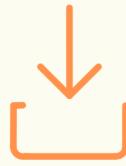
## Dataset Ranking Visualisation

Clear presentation of HFR-MADM scores and selection rationale



## Performance Metrics

Comprehensive display of accuracy, precision, recall, and F1-scores



## Feature Importance

Visual analysis of which clinical variables drive predictions



## Risk Prediction

Individual assessment form for real-time disease risk evaluation



# Deployed Application

## Try It Now

Streamlit-based web application deployed on Streamlit Cloud

[Launch Demo](#)

### Key Results Achieved:

- Automatic selection of optimal datasets with transparent scoring
- High prediction accuracy across multiple disease categories
- Clear visualisation of medical risk factors and model confidence
- Real-time patient risk assessment capability

**Technologies:** Python, Streamlit, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn