AI-Driven Stroke Detection: A Data Science Approach for Early Diagnosis and Prediction

Team Members:

1- Youssef Haitham Mokhtar Mohammed

2- Hazem Ayman Hosny Abd-Elhamid

3- Ahmed Amr Afte Deghedy

4- Yomna Maged Ali Abdelwahaab

5- Hanaa Mahmoud Ahmed Morsi ElGabbas

6- Omar Gamal Zaki Ahmed

1. Introduction

Stroke is a leading cause of disability and death worldwide, with early detection playing a crucial role in improving patient outcomes.

Traditional diagnostic methods rely on clinical assessments, which can be time-consuming and subject to human error. Advances in data

science and artificial intelligence (AI) offer an opportunity to enhance stroke detection through machine learning (ML) models. This

proposal outlines an AI-driven approach to stroke detection, leveraging medical data to improve accuracy and speed in diagnosis.

2. Problem Statement

Current stroke detection methods face challenges, including delays in diagnosis and high false-negative rates. According to the World

Health Organization (WHO), strokes account for over 12% of deaths globally. Early and accurate detection can significantly reduce mortality

rates and improve treatment effectiveness. However, conventional methods often struggle with:

- Subjectivity in diagnosis
- Variability in clinical assessments
- High dependence on specialized medical professionals

3. Proposed Solution

To address these challenges, this project proposes developing a machine learning model capable of detecting strokes from patient health
records and demographic data. The proposed solution will:

- Utilize ML techniques for structured data analysis
- Leverage patient demographic and clinical data for predictive modeling
- Employ explainable AI (XAI) methods to ensure transparency and trust in the model's decisions

4. Methodology

4.1 Data Collection

- Data Sources: Public datasets (e.g., Kaggle Stroke Prediction Dataset), hospital medical records
- Data Preprocessing: Handling missing values, data normalization, feature selection

4.2 Model Development

- Machine Learning Techniques:
  - Classical models: Random Forest, XGBoost, Logistic Regression for structured data
- Evaluation Metrics:
  - Accuracy, Precision, Recall, F1-score, AUC-ROC for performance assessment

4.3 Model Validation & Deployment

- Validation: Cross-validation techniques to ensure model robustness
- Deployment: Cloud-based API integration for real-time stroke detection
- Interpretability: SHAP values and LIME for model explainability

5. Expected Outcomes

- Enhanced accuracy in stroke detection compared to conventional methods
- Faster diagnosis, leading to improved patient outcomes
- Integration potential with hospital systems for real-time analysis

6. Risk Assessment & Mitigation

- Data Bias & Imbalance: Addressed using oversampling, synthetic data augmentation
- Regulatory Compliance: Ensure adherence to HIPAA, GDPR for data privacy
- Model Interpretability: Utilize XAI techniques to build trust with medical professionals

7. Conclusion

This proposal presents an AI-driven approach to stroke detection that aims to enhance early diagnosis accuracy and speed. By leveraging machine learning techniques and explainable AI, this project seeks to provide a scalable and effective solution for real-world medical applications, ultimately improving stroke patient outcomes and reducing healthcare burdens.