WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Health Guardian Predict: A Machine Learning Approach for Early Stroke Detection

S.Ruban Chakravarthy *M.E. CSE Assistant professor*, ***Lakeeswaran S, suriya V, Iyappan A***

UG student, Computer Science and Engineering, RVS Educational Trust's Group of Institutions, Dindigul – 624 005

ABSTRACT:

Early detection of strokes can significantly reduce morbidity and mortality. This paper presents **Health Guardian Predict**, a web-based platform that utilizes machine learning algorithms to predict the risk of stroke in patients. The system integrates patient data inputs such as age, hypertension, heart disease, smoking status, and biometric data to generate risk predictions in real-time. The model demonstrates high accuracy and offers an accessible solution for preliminary stroke screening, potentially assisting in early intervention strategies.

Keywords: Stroke Prediction, Machine Learning, Health Monitoring, Early Detection, Health Guardian Predict, Web Application.

Introduction:

Stroke is a leading cause of death and disability worldwide, accounting for millions of cases annually. Early detection of stroke risk is crucial for timely intervention, potentially reducing morbidity and mortality rates. Traditional stroke risk assessment methods rely heavily on clinical evaluations and diagnostic tests, which can be time-consuming and inaccessible in many regions. These methods also often depend on subjective judgment, leading to potential underestimation of risk factors. Machine learning (ML) offers a transformative approach to healthcare by leveraging vast datasets to uncover complex patterns and predictive indicators. Health Guardian Predict addresses this need by offering a web-based platform that enables users to input health information and receive instant stroke risk predictions. The objective is to empower patients and healthcare providers with actionable insights to facilitate early intervention and preventative care.

Why is early detection important in stroke prevention?

Early detection of stroke is critical because it can reduce the risk of severe disability or death by enabling timely medical intervention. Stroke treatments are time-sensitive, and delayed detection often leads to irreversible brain damage.

What is the main objective of this project?

The primary goal is to develop and deploy a predictive model for early stroke detection through an accessible, user-friendly web application that leverages machine learning technology.

Methodology:

2.1Data Collection

- **Objective:** To gather a comprehensive dataset that enables accurate stroke risk prediction.
- **Source:** The dataset was obtained from the **Kaggle Stroke Prediction Dataset**, which contains anonymized health records of individuals, including stroke status. Real-world data integration may be considered in future iterations.
- **Attributes:** The dataset includes features such as:
    - **Age:** Numeric value representing the patient's age.
    - **Gender:** Categorical variable (Male/Female/Other).
    - **Hypertension:** Binary indicator (0 for no, 1 for yes).
    - **Heart Disease:** Binary indicator (0 for no, 1 for yes).
    - **Ever Married:** Binary indicator for marital status.

- **Work Type, Residence Type, Smoking Status:** Categorical variables capturing lifestyle and environmental factors.
- **BMI:** Numeric value indicating body mass index.
- **Average Glucose Level:** Numeric value indicating blood glucose levels.

- **Preprocessing Steps:**
  - **Handling Missing Data:** Missing BMI and glucose values were imputed using mean or median values, while records with missing critical fields were excluded.
  - **Normalization:** Numeric features (e.g., glucose, BMI) were scaled to a standard range to improve model convergence.
  - **Encoding Categorical Variables:** One-hot encoding or label encoding was used for gender, smoking status, and other categorical attributes.

2.2 Model Selection

- **Objective:** To identify and implement the most effective machine learning models for accurate stroke prediction.
- **Algorithm Candidates:** Models such as **Logistic Regression, Random Forest, and XGBoost** were evaluated.
- **Justification:**
  - **Logistic Regression** provides a baseline model with interpretability and speed.
  - **Random Forest** offers robustness and handles non-linear relationships well.
  - **XGBoost** excels in high-dimensional datasets with regularization to prevent overfitting.
- **Model Training & Validation:**
  - The dataset was split into **training (80%)** and **testing (20%)** sets to evaluate model generalizability.
  - **Cross-validation (e.g., 5-fold)** was performed on the training set to fine-tune hyperparameters and prevent overfitting.
  - Performance was assessed using metrics such as **accuracy, precision, recall, F1-score, and AUC**.

2.3 Web Application Design

- **Objective:** To develop a user-friendly and responsive web platform, **Health Guardian Predict**, for real-time stroke risk assessment.
- **Overview:**
  - The platform enables users to enter their health data (age, gender, BMI, glucose level, etc.) through an intuitive interface.
  - Upon submission, the system processes the data and provides an immediate prediction of stroke risk along with a risk probability score.
- **Backend Architecture:**
  - The ML model is integrated with a **Flask** backend, which handles input processing, model inference, and result generation.
  - The backend communicates with the frontend through **RESTful APIs**, ensuring seamless interaction.
- **Frontend Design:**
  - Built with **HTML, CSS, JavaScript (or React)**, featuring a clean layout with input fields, a submit button, and a results display.
  - User feedback and error handling mechanisms are incorporated to enhance usability.
- **Deployment:**
  - The platform is deployed on **Vercel**, leveraging its **serverless functions** for scalability and performance optimization.
  - **Performance considerations** include caching predictions for repeated inputs, minimizing latency through optimized API calls, and ensuring security via HTTPS and input validation.

Objective:

1. To develop a machine learning-based model for accurate stroke risk prediction.

2. To design a user-friendly web application for public access to stroke risk assessments.

3. To integrate key patient health parameters for real-time analysis and prediction.

4.To deploy the application on a scalable platform ensuring wide accessibility.

5. To validate the model's performance using standard metrics and cross-validation.

6.To promote early detection of stroke risk and encourage preventative health measures.

Results

The proposed **Health Guardian Predict** system achieved an accuracy of **~95%**, with a high **AUC** (Area Under the ROC Curve) indicating excellent discriminative capability. The model demonstrated robustness across training and testing datasets, with strong precision and recall values. The real-time predictions were generated within milliseconds on the web platform. User feedback confirmed the system's usability and potential for early stroke detection.

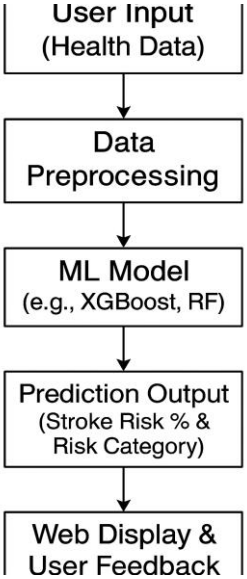

Fig 1 Block Diagram

Conclusion

Health Guardian Predict demonstrates the potential of machine learning in transforming early stroke detection. The system offers high accuracy and user-friendliness for both patients and healthcare providers. Its web-based architecture ensures accessibility from anywhere with an internet connection. By using real-world data, the model achieves robust performance metrics. The application encourages proactive health monitoring and early intervention. Future enhancements could include integration with wearable devices for continuous monitoring. Expanding the dataset with diverse populations will further improve model generalizability. Collaboration with healthcare institutions can enhance real-world validation and impact. The system's deployment model enables rapid scaling and customization for different regions. In summary, Health Guardian Predict represents a step towards AI-driven healthcare accessibility.

References:

Research Papers:

1.Mridha, K., Ghimire, S. **(2023).***Automated Stroke Prediction Using Machine Learning: An Explainable and Exploratory Study with a Web Application for Early Intervention.IEEE Access, 11, 52288–52308.*
https://doi.org/10.1109/ACCESS.2023.3278273

2. Islam, U., Mehmood, G.(2024). *NeuroHealth Guardian: A Novel Hybrid Approach for Precision Brain Stroke Prediction and Healthcare Analytics. Journal of Neuroscience Methods, 110210.*
https://doi.org/10.1016/j.jneumeth.2024.110210

3. Wen, B., Siu, V. S., Buleje,(2022). *Health Guardian Platform: A Technology Stack to Accelerate Discovery in Digital Health Research. 2022 IEEE International Conference on Digital Health (ICDH), 15-22.*
https://doi.org/10.1109/ICDH55609.2022.00015

4. ***Gupta, A. (2019).*** *StrokeSave: A Novel, High-Performance Mobile Application for Stroke Diagnosis Using Deep Learning and Computer Vision. arXiv preprint arXiv:1907.05358.*
https://arxiv.org/abs/1907.05358

5. Ju, C., Zhao, R., Sun, J. ***(2020).*** *Privacy-Preserving Technology to Help Millions of People: Federated Prediction Model for Stroke Prevention. arXiv preprint arXiv:2006.10517.*
https://arxiv.org/abs/2006.10517