

Introduction

Traditional stroke prevention methods typically rely on generalized clinical guidelines and risk factor lists, which, while useful, do not capture the nuances of individual patient risk. In contrast, this study leverages advanced machine learning and deep learning models, sophisticated feature engineering, SHAP analysis for interpretability, and a personalized risk scoring system—all integrated into an API—to provide real-time, individualized stroke risk assessments. This data-driven approach offers a more precise, tailored foundation for public health strategies.

How This Study Modelled Stroke Risk

1. Feature Engineering:

- **Derived Features:** This study derived critical features such as quadratic age (Age^2 / *age_squared*) to capture non-linear risk acceleration and PCA components that distilled complex interactions among multiple clinical and sociodemographic variables. For example, PCA components like PCA_1 and PCA_3 abstracted the age-comorbidity spectrum and lifestyle-demographic influences, respectively.
- **Interaction Terms:** Features such as *age_hypertension* were engineered to quantify how the presence of hypertension interacts with age, revealing that younger patients with hypertension face a disproportionately high stroke risk.

2. SHAP Analysis:

- **Interpretability:** By applying SHAP analysis, this study identified and ranked the most impactful predictors (risk factors) of stroke risk. SHAP values provided clear, quantitative insights into which features (e.g., *age*, age^2 , *glucose levels*, and *cardiovascular conditions*) drive predictions.
- **Directional Insights:** The analysis showed that higher values of certain features (e.g., age and diabetic-range glucose) markedly increase stroke risk, while others (e.g., indicators of normal glucose levels or younger age) decreased risk.

3. Risk Scoring System:

- **Personalized Risk Scores:** We translated SHAP-derived feature importance into a 102-point risk scoring system that assigns points based on each patient's risk profile. This system categorizes patients into low, moderate, and high risk, providing a practical tool for clinicians and public health intervention.
- **Evidence-Based Stratification:** The scoring system, validated against our dataset, directly reflects the weighted contribution of each key predictor (risk factor)—ensuring that the risk stratification is both evidence-based and personalized.

4. API Development & Deployment:

- **Integration Across Platforms:** The entire stroke prediction system is encapsulated within a Flask API, enabling seamless integration with Electronic Health Records (EHRs), mobile applications, and web-based applications.
- **Real-Time Risk Assessment:** This API provides real-time risk scores and model explanations (via SHAP values), which facilitate immediate, data-driven decision-making in clinical and public health settings.

Data-Driven Public Health Strategies for Stroke Prevention

Based on the study's insights, the following strategies can be implemented:

1. Early Intervention:

- **Risk Identification:** Use the API to screen populations in real-time, identifying high-risk individuals early. For instance, patients with high scores driven by features such as advanced age, elevated diabetic-range glucose, and cardiovascular conditions can be prioritized.
- **Targeted Lifestyle Interventions:** Design intervention programs that focus on modifiable risk factors. For example, individuals with high SHAP contributions from obesity or hypertension should receive immediate lifestyle and medical interventions.

2. Personalized Health Recommendations:

- **Real-Time Feedback:** The risk scoring system provides personalized risk assessments that can be communicated directly to patients. If a patient exhibits high risk due to inactivity (as indicated by SHAP values), tailored exercise regimens can be recommended.
- **Individualized Care Plans:** Physicians can use these personalized risk profiles to develop customized treatment plans that address each patient's unique risk factors, ensuring that interventions are both efficient and effective.

3. Community-Level Insights:

- **Population Risk Mapping:** Aggregate risk scores from the API to generate geospatial heatmaps of stroke risk (provided we have geospatial data). These maps can identify high-risk communities, guiding public health officials in resource allocation and targeted outreach.
- **Resource Allocation:** Public health organizations like ADPHC can leverage these insights to design community-wide preventive programs, such as localized screening initiatives and educational campaigns in areas with elevated risk scores.

4. Integration with NLP & Chatbots for Patient Monitoring:

- **Continuous Monitoring:** Future integration of NLP-powered chatbots with the API can allow for continuous monitoring of patient-reported symptoms and lifestyle

factors. These chatbots can provide real-time alerts and recommendations based on evolving risk profiles.

- **Socioeconomic Assessments:** Chatbots can also assess socio-economic stressors (e.g., financial strain, access to care) that may contribute to stroke risk, ensuring that high-risk patients receive comprehensive, holistic support.

5. Policy Recommendations Based on Data:

- **Preventive Program Coverage:** Advocate for insurance and governmental support for preventive programs that use API-driven risk assessments to target interventions.
- **Public Health Campaigns:** Use data on key risk factors (e.g., the impact of early-onset hypertension or non-linear age effects) to inform public health messaging and campaigns, ensuring that they are targeted and evidence-based.

Conclusion

This study demonstrates that a data-driven approach—combining advanced feature engineering and modelling, SHAP interpretability, personalized risk scoring, and API integration—can significantly enhance stroke prevention strategies. By moving from generalized guidelines to precise, individualized risk assessments, public health initiatives can:

- Identify high-risk individuals early.
- Provide targeted, personalized interventions.
- Optimize resource allocation through real-time population monitoring.
- Support proactive public health policies.

Future Directions: Next steps include expanding the API to incorporate NLP-powered chatbots for continuous patient monitoring and integrating longitudinal data to refine dynamic risk models. Such innovations promise to further shift stroke prevention from a reactive to a proactive paradigm, ultimately reducing the burden of stroke.