SOUTHAMPTON SOLENT UNIVERSITY

DEPARTMENT OF SCIENCE AND ENGINEERING

MSc Applied AI and Data Science

Academic Year 2025-2026

Gudur Sahiti

Title of Your Research

Supervisor: **Zuhaib Khan**

January 2026

This report is submitted in partial fulfilment of the requirements of Southampton Solent University for the degree of MSc Applied AI and Data Science

Acknowledgment

I would like to express my sincere gratitude to my dissertation supervisor, **Mr Zuhaib Khan,** for their valuable guidance, constructive feedback, and continuous support throughout the development of this research. Their expertise and encouragement played a crucial role in shaping the methodology and overall direction of this study.

I would also like to thank the dissertation coordinator, **Mr Bacha Rehman** for their guidance, organisation, and support throughout the dissertation process, ensuring a clear structure, expectations, and timely progress.

I am grateful to the Southampton Solent University Library services for providing access to academic resources, research databases, and research support that contributed significantly to this work.

Additionally, I would like to acknowledge Kaggle for providing open-access datasets that enabled the experimental analysis conducted in this project.

Finally, I would like to thank my family and friends for their patience, encouragement, and continued support throughout my postgraduate studies.

Abstract

Stroke remains a major contributor to mortality and long-term functional impairment worldwide, placing sustained pressure on modern healthcare systems. Early identification of individuals at elevated risk is therefore essential to support timely intervention and reduce adverse clinical outcomes. However, conventional diagnostic approaches are largely reactive and often identify stroke only after neurological damage has occurred.

This dissertation proposes a machine learning–based decision-support framework for stroke risk prediction using structured clinical and lifestyle data. A hybrid modelling strategy is adopted, combining a CatBoost classifier with a multilayer perceptron (MLP) network to leverage both tree-based learning and non-linear representation capabilities. To address the pronounced imbalance between stroke and non-stroke cases, the Synthetic Minority Oversampling Technique (SMOTE) is applied exclusively to the training data.

Model performance is evaluated using multiple quantitative measures, with particular emphasis on recall and F1-score due to the clinical importance of minimising missed stroke cases. To enhance transparency and support clinical interpretability, Shapley Additive Explanations (SHAP) are employed to provide both global feature importance and local, instance-level explanations. Key contributing factors identified include age, hypertension status, average glucose level, and body mass index.

The proposed framework is implemented within an interactive Streamlit-based application, allowing users to input patient information and receive stroke risk predictions alongside interpretable explanations. This approach demonstrates the potential of explainable machine learning to support informed clinical decision-making in stroke risk assessment.

Table of Contents

List of figures

List of tables

Abbreviation's

Abbrevation	Full Form
AI	Artificial Intelligence
BMI	Body Mass Index
CatBoost	Categorical Boosting
CT	Computed Tomography
EDA	Exploratory Data Analysis
F1-Score	Harmonic Mean of Precision and Recall
GUI	Graphical User Interface
LightGBM	Light Gradient Boosting Machine
MLP	Multi-Layers Perceptron
MRI	Magnetic Resonance Imaging
ROC-AUC	Receiver Operating Characteristic – Area Under the Curve
ReLU	Rectified Linear Unit
SHAP	Shapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
XAI	Explainable Artificial Intelligence

1.Introduction

Stroke stands as one of the leading causes of mortality and long-term disability globally, presenting a significant challenge for modern healthcare systems. Survivors of stroke frequently endure enduring neurological impairments that impact mobility, cognition, speech, and overall quality of life. The toll of stroke extends beyond the individuals and families affected, imposing a considerable economic burden on healthcare infrastructures due to long-term treatment, rehabilitation, and lost productivity. Consequently, there is an increasing focus on preventive strategies aimed at the early identification of individuals at higher risk.

Clinically, a stroke occurs when the blood supply to the brain is interrupted, which can happen due to either vessel blockage **(ischaemic stroke)** or vessel rupture **(haemorrhagic stroke)**. Various clinical, and lifestyle-related factors contribute to the risk of stroke, including advanced age, hypertension, heart disease, abnormal glucose levels, obesity, smoking habits, and social determinants such as occupation and living environment. The coexistence of these risk factors complicates the prediction of strokes, making it a multi-faceted challenge.

Traditional stroke diagnosis heavily depends on clinical evaluations and medical imaging techniques, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). While these methods are effective for confirming a stroke after symptoms have appeared, they are primarily reactive and offer limited assistance for early risk assessment. Additionally, access to advanced diagnostic tools may be restricted in resource-limited healthcare environments. As a result, there is a pressing need for data-driven approaches that can help identify high-risk individuals prior to the onset of a stroke.

Recent developments in machine learning have enabled more data-driven approaches to healthcare risk assessment, particularly for conditions such as stroke where multiple interacting risk factors are involved. Unlike traditional statistical models, machine learning techniques can analyse large-scale clinical datasets and capture complex, non-linear relationships between demographic, clinical, and lifestyle variables. This capability is especially valuable in stroke prediction, where risk is influenced by a combination of age, comorbidities, and behavioural factors rather than a single determinant.

Despite these advantages, the application of machine learning in clinical decision-support systems remains challenging. Issues such as class imbalance in medical datasets, limited model transparency, and concerns surrounding clinical trust and interpretability continue to restrict widespread adoption. Addressing these challenges is therefore essential for developing reliable and responsible stroke risk prediction systems.

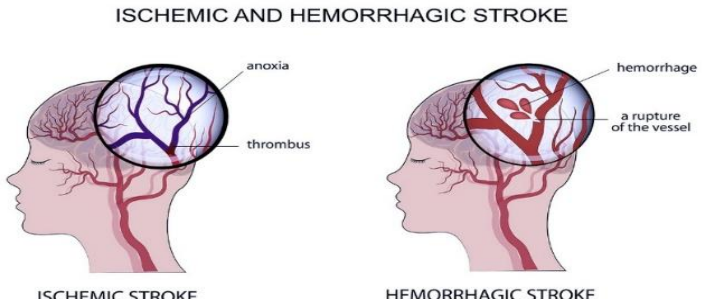


Figure 1: Illustration of ischemic and haemorrhagic stroke types affecting the brain. (source: Covenant Health, n.d.)

1.1 Problem Statement

Although many machine learning methods have been developed for predicting stroke risk, several limitations still impact their effectiveness and practical use. A key challenge is the inherent class imbalance in stroke datasets, where stroke cases are a small minority compared to non-stroke cases. Consequently, models trained on such data may achieve high overall accuracy but struggle to identify individuals who are truly at risk, resulting in poor recall for the stroke class. In a healthcare context, these false negatives are particularly concerning because overlooking high-risk cases can lead to serious clinical consequences.

Another significant limitation relates to model interpretability. Many high-performing machine learning and deep learning models function as black-box systems, offering little insight into how predictions are generated. In healthcare applications, the lack of transparency can undermine clinician confidence and raises ethical concerns regarding accountability and decision support. For machine learning models to be useful in practice, they must not only perform well but also provide interpretable explanations that align with clinical understanding.

These challenges highlight the need for a stroke prediction framework that addresses class imbalance, improves predictive robustness, and incorporates explainable artificial intelligence techniques, while remaining suitable for deployment as a decision-support system rather than a diagnostic tool.

1.2 Research Aim

- The primary aim of this research is to develop a stroke risk prediction decision-support system using a hybrid machine learning approach that combines gradient boosting and neural network models.

- The study seeks to improve predictive performance on an imbalanced healthcare dataset while maintaining transparency and interpretability through explainable artificial intelligence techniques.
- The system is designed to support analytical and educational use rather than clinical diagnosis.

1.3 Research Objectives

- To analyse and preprocess a publicly available stroke prediction dataset by handling missing values, encoding categorical variables, and performing feature engineering.
- To mitigate severe class imbalance in the dataset by applying the Synthetic Minority Oversampling Technique (SMOTE) to the training data.
- To develop and evaluate a baseline CatBoost classifier for stroke risk prediction using demographic and clinical features.
- To design and implement a Multi-Layer Perceptron (MLP) neural network with an appropriate preprocessing pipeline for numerical and categorical data.
- To construct a hybrid prediction model by combining CatBoost and MLP probability outputs through weighted probability fusion.
- To perform weight tuning and decision threshold optimisation to improve performance on the minority stroke class.
- To evaluate model performance using standard classification metrics, including accuracy, precision, recall, F1-score, and ROC-AUC.
- To use SHAP-based explanation techniques to examine how individual features influence model predictions and support transparent of results.
- To deploy the final hybrid model within a Streamlit-based decision-support system for user interaction and result visualisation.

1.4 Research Questions

- How effectively can a hybrid machine learning model combining CatBoost and a Multi-Layer Perceptron improve stroke risk prediction compared to individual models?
- How does probability-level fusion and threshold tuning influence the predictive performance of the hybrid model?
- Which demographic and clinical features contribute most significantly to stroke risk prediction?
- How can explainable artificial intelligence techniques, such as SHAP, improve the transparency and interpretability of stroke prediction models?
- How effectively can a Streamlit-based application present hybrid predictions and explainability outputs in a decision-support context?

2. Literature review

2.1 Stroke Risk Prediction and Clinical Context

Stroke is widely recognised as a leading cause of mortality and long-term disability, prompting extensive research into early risk assessment and prevention. Conventional stroke risk assessment approaches are largely based on clinical scoring systems and statistical models that utilise predefined risk factors such as age, blood pressure, diabetes status, and lifestyle indicators. While these tools are clinically interpretable, they often struggle to capture complex interactions among risk factors and may not generalise well across diverse populations.

A comparative evaluation of cardiovascular and stroke risk calculators by **Karmali et al. (2014)** and later reviewed in *Cardiovascular Stroke Risk Predictive Calculators: A Comparison* highlights limitations related to model rigidity, population bias, and reduced predictive accuracy when applied outside their original development cohorts. These findings have motivated the exploration of data-driven machine learning approaches capable of learning complex patterns from routinely collected healthcare data.

In addition to mortality, stroke is associated with significant long-term physical, cognitive, and socioeconomic consequences for patients and healthcare systems. Early identification of individuals at high risk is therefore essential not only for prevention but also for optimising healthcare resource allocation. Traditional clinical risk scores, while useful in practice, are often derived from linear assumptions and predefined thresholds, which may limit their ability to capture complex interactions among multiple risk factors. This limitation has contributed to growing interest in data-driven approaches that can adapt to diverse patient profiles and evolving healthcare data.

2.2 Machine Learning Techniques for Stroke Prediction

Machine learning approaches have gained increasing attention in stroke prediction research due to their ability to capture complex, non-linear relationships within clinical and demographic data. Unlike traditional statistical models, machine learning algorithms can learn intricate patterns from large-scale healthcare datasets. Zheng et al. (2025) evaluated several machine learning methods using population-level data from China and reported improved discriminative performance compared to conventional statistical techniques, particularly when assessed using recall and ROC-AUC. Their findings also highlighted the importance of addressing class imbalance to ensure reliable stroke risk prediction.

Related studies by Parvathi and Ellambotla (2023) and Soladoye, Akinwumi, and Zhang (2024) demonstrated that commonly used classifiers, including Logistic Regression, Support Vector Machines, and Random Forests, can achieve satisfactory performance on stroke datasets when supported by robust preprocessing pipelines. However, these studies consistently noted that severe class imbalance leads to biased predictions that favour the non-stroke class, thereby limiting clinical applicability.

To mitigate this issue, resampling strategies such as the Synthetic Minority Oversampling Technique (SMOTE) have been widely adopted. Chawla et al. (2002) introduced SMOTE to enhance minority class representation by generating synthetic samples. Subsequent stroke prediction studies, including Zheng et al. (2025), confirmed that applying SMOTE exclusively to training data improves recall and F1-score for stroke cases while maintaining acceptable precision.

Despite these advantages, the effectiveness of machine learning models in stroke prediction remains strongly influenced by data preprocessing choices and evaluation strategies. Several studies emphasise that performance outcomes vary substantially depending on feature selection, missing value handling, and metric selection. In imbalanced medical datasets, reliance on accuracy alone can be misleading, reinforcing the need to prioritise recall, F1-score, and ROC-AUC when evaluating stroke prediction models.

2.3 Gradient Boosting Models in Stroke Prediction

Gradient boosting models have emerged as highly effective for stroke prediction using structured tabular data. In *Machine Learning Based Brain Stroke Prediction Using Light Gradient Boosting Machine Algorithm*, **Wijaya, Rahman, and Lestari (2024)** demonstrated that LightGBM outperformed several traditional machine learning classifiers on the Kaggle stroke dataset. Their results highlighted the suitability of gradient boosting algorithms for capturing complex feature interactions in healthcare data.

While LightGBM requires explicit categorical encoding, **Dorogush, Ershov, and Gulin (2018)** introduced CatBoost as a gradient boosting framework designed to natively handle categorical variables and reduce target leakage through ordered boosting. These properties make CatBoost particularly suitable for healthcare datasets that combine demographic, clinical, and lifestyle-related attributes..

Despite their strong predictive performance, gradient boosting models are often criticised for limited interpretability. Feature importance scores alone may not adequately explain individual predictions, which is a significant concern for healthcare decision-support systems.

Gradient boosting approaches are well suited to datasets containing diverse feature types, as they can learn complex patterns without heavy reliance on manual feature construction. This capability is especially beneficial in healthcare applications, where input variables vary widely in measurement scale and clinical importance. Nevertheless, the layered structure of ensemble tree models can reduce transparency, requiring the use of additional explanation methods when applied in clinical decision-support settings.

2.4 Neural Networks for Stroke Prediction

Neural networks, particularly Multi-Layer Perceptron's (MLPs), have been explored as an alternative approach for stroke risk prediction. **Reddy and Kumar (2023)** reported that neural networks can model complex nonlinear relationships between clinical variables and achieve competitive predictive performance when combined with appropriate preprocessing techniques such as feature scaling and categorical encoding.

However, several studies, including **Zheng et al. (2025)**, noted that neural networks are sensitive to class imbalance and dataset size, often requiring careful hyperparameter

tuning to avoid overfitting. Furthermore, neural networks are typically regarded as black-box models, providing limited transparency regarding how predictions are generated. This lack of interpretability restricts their standalone applicability in healthcare contexts where explainability is essential.

Furthermore, neural network performance in healthcare applications is often constrained by dataset size and quality. Stroke datasets are typically limited in size and highly imbalanced, which can reduce the effectiveness of deep learning models that require large amounts of data to generalise well. As a result, while neural networks offer strong representational capacity, their standalone use in stroke prediction may be less reliable without complementary modelling approaches or additional regularisation strategies.

2.5 Ensemble and Hybrid Learning Approaches

To overcome the limitations of single-model approaches, ensemble and hybrid learning techniques have gained increasing attention. **Chakraborty et al. (2024)** proposed a stacked ensemble framework combining Random Forest, Decision Tree, and K-Nearest Neighbour classifiers for stroke prediction using the Kaggle dataset. Their results demonstrated that ensemble learning can significantly improve predictive performance androbustness on imbalanced healthcare data.

Similarly, **Burman et al. (2024)** and **Alsubaie et al. (2024)** showed that stacking-based ensembles outperform individual classifiers by leveraging complementary learning patterns. However, these stacking architectures introduce additional complexity and may reduce interpretability due to the involvement of meta-learners.

An ensemble machine learning and data mining approach presented by **Aish, Dhingra, and Wani (2024)** further demonstrated that combining diverse classifiers improves stroke prediction accuracy. Nevertheless, most ensemble approaches focus on stacking rather than simpler probability-level fusion strategies, which can offer improved transparency and controllability through weight and threshold tuning.

In addition to stacking-based ensembles, simpler hybrid approaches that operate at the probability level have gained attention due to their interpretability and ease of implementation. Probability-level fusion allows individual models to contribute proportionally to the final prediction, enabling explicit control over model influence through weight and threshold tuning. Such approaches can offer a practical compromise between performance improvement and system transparency, particularly in healthcare decision-support contexts.

2.6 Explainable Artificial Intelligence in Healthcare

Explainability is a critical requirement for the adoption of machine learning models in healthcare. Black-box models that provide predictions without justification are often met with resistance from clinicians due to concerns regarding trust, accountability, and ethical responsibility. As a result, explainable artificial intelligence (XAI) techniques have become an essential component of modern healthcare analytics.

SHapley Additive exPlanations (SHAP) is one of the most widely used XAI methods for interpreting machine learning models. SHAP assigns contribution values to individual features based on cooperative game theory, enabling both global and local interpretability. In healthcare applications, SHAP has been used to identify influential risk factors and explain individual patient predictions, supporting transparent and informed decision-making.

Integrating SHAP with gradient boosting models has been shown to be particularly effective, as it provides consistent explanations for complex models while maintaining high predictive performance. This makes SHAP well suited for stroke prediction systems that aim to balance accuracy and interpretability.

In recent healthcare studies, explainable AI techniques have been increasingly adopted to support clinician trust and model validation, particularly in high-risk prediction tasks such as cardiovascular and stroke risk assessment. By providing feature-level explanations, XAI methods enable clinicians and researchers to verify whether model predictions align with established medical knowledge. This alignment is especially important in decision-support systems, where transparency can influence user acceptance and responsible deployment.

Despite its advantages, SHAP-based explainability also has limitations. SHAP explanations describe how features contribute to a model's prediction but do not imply causal relationships between risk factors and clinical outcomes. As a result, SHAP should be interpreted as a tool for model transparency rather than clinical causation. Nevertheless, when used alongside robust predictive models, SHAP offers a practical and widely accepted approach for enhancing interpretability in healthcare machine learning applications.

Beyond model interpretation, explainable AI also supports regulatory compliance and ethical accountability in healthcare applications. Transparent models enable auditing of predictions and facilitate communication between technical developers and domain experts. By integrating explainability into both model analysis and system deployment, healthcare decision-support tools can better align with responsible AI principles and promote informed usage.

3. Methodology

3.1 Data Cleaning

3.1.1 Dataset Description

The dataset utilised in this research is a publicly accessible healthcare dataset developed for stroke risk prediction tasks. It supports a binary classification objective, where the outcome variable denotes the occurrence of stroke, coded as 1 for stroke cases and 0 for non-stroke cases. The dataset comprises a combination of demographic, clinical, and lifestyle-related variables that are commonly recognised as influential factors in stroke risk assessment.

The dataset includes the following key features:

- Demographic attributes: age, gender, residence type
- Clinical attributes: hypertension status, heart disease history, average glucose level, body mass index
- Lifestyle and social attributes: smoking status, employment type, marital status

Table 1: Dataset Description

Feature Name	Data Type	Description
Age	Numerical	Age of Individual
Gender	Categotical	Biological sex
Hypertension	Binary	Presence of hypertension
Heart_Disease	Binary	Presence of heart disease
Ever_married	Categorical	Marital status
Work_type	Categorical	Employment category
Residence_type	Categorical	Urban or rural residence
Avg_glucose_level	Numerical	Average blood glucose level
Bmi	Numerical	Body mass index
Smoking_status	Categorical	Smoking habits
Age_group	Categorical	Engineered age category
stroke	Binary	Target variable

3.1.2 Data Loading

The dataset was imported into the Python environment utilizing the Pandas library. An initial examination was carried out to understand the structure of the dataset, encompassing the number of instances, types of features, and distribution of target classes. This step was crucial for confirming the successful import of the data and for identifying any potential data quality issues early on.

Descriptive statistics were generated for numerical features to examine their distributions, ranges, and central tendencies. Categorical variables were analyzed to identify unique values and determine the need for encoding strategies during preprocessing.

3.1.3 Handling Missing values

Exploratory analysis identified missing values in the BMI feature. Since BMI is a crucial clinical indicator for stroke risk, removing instances with missing values could result in significant data loss and potential bias.

To solve this issue, median imputation was applied to the BMI feature. Median imputation was selected over mean imputation due to its robustness to outliers and skewed distributions, which are common in healthcare data. This approach ensured that all records retained valid BMI values while preserving the overall distribution of the feature.

3.1.4 Feature Engineering

To improve the representation of stroke risk related to age, a new categorical feature named **"age_group"** was created from the continuous age variable. The age groups were defined to reflect significant life stages, enabling models to capture the non-linear risk patterns associated with ageing.

Feature engineering was performed prior to model training to ensure consistency across all modelling approaches. The engineered feature was included in both the CatBoost and MLP pipelines where applicable.

```
#step 3 - create an age group feature
df["age_group"] = pd.cut(
    df["age"],
    bins=[0, 18, 40, 65, 120],
    labels=["Child", "Young Adult", "Adult", "Senior"]
)
sns.countplot(data=df, x="age_group")
plt.title("Age Group Distribution")
plt.show()
```

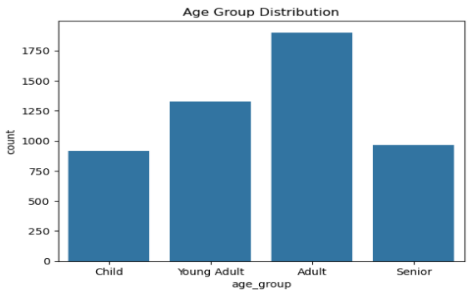


Figure 2: Age group distribution after feature engineering.

3.1.5 Class Imbalance Analysis

Exploratory examination of the target variable indicated a pronounced imbalance between classes, with stroke cases forming only a small proportion of the dataset. This imbalance can negatively influence model training by encouraging predictions that favour the majority non-stroke class, thereby reducing the model's ability to correctly detect stroke events.

To address this challenge, the Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training data during model development. By generating synthetic minority-class instances, SMOTE enhances class representation and supports

improved learning of stroke-related decision boundaries, while preserving the integrity of the unseen test set and avoiding data leakage.

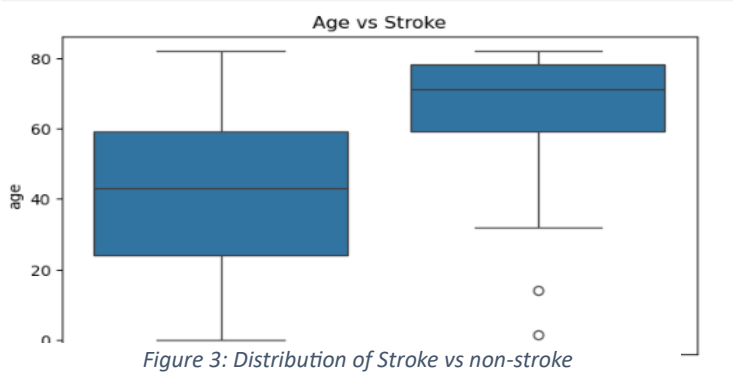


Figure 3: Distribution of Stroke vs non-stroke

3.1.6 Exploratory Data Analysis (EDA)

Exploratory data analysis was conducted to examine relationships between features and the target variable. Visualisations and summary statistics were used to identify trends and patterns associated with stroke risk.

Key observations from the EDA include:

- Stroke incidence increases significantly with age.
- Individuals with hypertension or heart disease exhibit higher stroke prevalence.
- Elevated average glucose levels are associated with increased stroke risk.
- BMI and smoking status show varying influence across different subgroups.

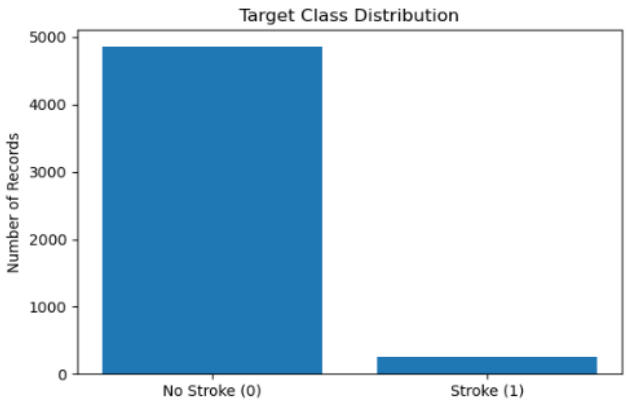


Figure 4: Comparison of age distributions btw stroke and non-stroke

3.1.7 Summary of data preparation

The data preparation stage established a reliable foundation for model development by ensuring that the dataset was well-structured and suitable for machine learning analysis. Core preprocessing activities included treating missing values, constructing engineered features, transforming categorical attributes into numerical representations, and managing class imbalance. Collectively, these steps improved data quality and consistency, supporting the development of stable, interpretable, and effective stroke prediction models in subsequent stages of the study.

3.2 CatBoost Baseline Model

3.2.1 Rationale for Selecting CatBoost

CatBoost was selected as the baseline machine learning model in this study due to its suitability for structured healthcare datasets containing both numerical and categorical features. Stroke prediction datasets commonly include categorical attributes such as gender, work type, residence type, marital status, and smoking status, which often require extensive preprocessing when using traditional machine learning algorithms.

Unlike many gradient boosting frameworks, CatBoost natively processes categorical variables without requiring explicit one-hot encoding. This reduces dimensionality expansion, minimises information loss, and improves learning efficiency on tabular medical data. Additionally, CatBoost employs an ordered boosting strategy that mitigates target leakage, thereby enhancing model generalisation, which is particularly important when working with imbalanced healthcare datasets.

The CatBoost model was implemented as a baseline to establish a strong reference point for evaluating the effectiveness of more complex modelling approaches developed later in this research.

3.2.2 Data Input and Dataset Preparation

The CatBoost model was trained on a cleaned and pre-processed version of the stroke dataset (stroke_clean.csv). This dataset was prepared from the raw stroke data after conducting exploratory analysis and performing necessary preprocessing steps, such as handling missing values and feature engineering. Utilizing the cleaned dataset ensured that all features were complete and consistently formatted before the model training process.

The dataset was loaded into the Python environment using the Pandas library, which facilitated efficient data manipulation and integration with machine learning pipelines. At this stage, the dataset maintained the original class imbalance between stroke and non-stroke cases, allowing for the application of imbalance handling techniques during model training.

```
#step 2 Load Cleaned Dataset


CLEAN_FILE = "../data/stroke_clean.csv"

df = pd.read_csv(CLEAN_FILE)
print("Shape:", df.shape)
df.head()
```

```
Shape: (5110, 13)
```

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke	age_group
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1	Senior
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	28.1	never smoked	1	Adult
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1	Senior
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1	Adult
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1	Senior

Figure 5: Data Input Values

3.2.3 Feature and Target Variable Definition

For supervised model development, the stroke dataset was structured by separating explanatory variables from the prediction outcome. The input features represented patient-level characteristics spanning demographic, clinical, and lifestyle domains, including age, biological sex, cardiovascular conditions, metabolic indicators, behavioural factors, employment status, and residential context.

The prediction target was defined as stroke occurrence and encoded as a binary variable, where one indicated a recorded stroke event and zero denoted its absence. This representation enabled the learning algorithms to model relationships between patient attributes and stroke risk in a clear and systematic manner.

3.2.4 Test – Train Data Split

Model evaluation was conducted by dividing the dataset into separate training and testing partitions using an 80:20 allocation.This approach allows the model to learn underlying patterns from the training data while reserving a separate portion for unbiased performance validation.

To maintain consistency in class representation between the subsets, stratified sampling was applied with respect to the target variable. This ensured that both the training and test sets preserved the original distribution of stroke and non-stroke cases — a crucial consideration given the dataset's imbalance, where stroke occurrences were relatively rare. Maintaining proportional representation across the splits improves the reliability of evaluation metrics and reduces bias during model assessment.

3.2.5 Identification of Categorical Features

Categorical features were programmatically identified based on their data types and explicitly passed to the CatBoost model using the cat_features parameter. This enabled CatBoost to apply its internal encoding mechanisms optimised for categorical data.

Explicit identification of categorical features ensures that CatBoost handles these variables appropriately, without requiring manual encoding or transformation, and preserves the inherent relationships within categorical attributes.

```
#step - 5 Detect categorical columns

cat_cols = X_train.select_dtypes(include=["object", "category"]).columns.tolist()
cat_idx = [X_train.columns.get_loc(c) for c in cat_cols]

print("Categorical cols:", cat_cols)
print("Categorical idx:", cat_idx)
```

```
Categorical cols: ['gender', 'ever_married', 'work_type', 'Residence_type', 'smoking_status', 'age_group']
Categorical idx: [0, 4, 5, 6, 9, 10]
```

3.2.6 Handling Class Imbalance Using Class Weighting

The stroke dataset exhibits severe class imbalance, with stroke cases representing a small minority of the total observations. Training a model on such imbalanced data without correction may result in biased predictions favouring the majority non-stroke class.

To address this issue, class imbalance was handled using class weighting within the CatBoost framework. The ratio of non-stroke to stroke instances in the training set was calculated and supplied to the model through the **scale_pos_weight** parameter. The scale_pos_weight parameter was calculated as the ratio of non-stroke to stroke instances in the training data and supplied to the CatBoost classifier. This approach increases the penalty associated with misclassifying stroke cases, encouraging the model to improve sensitivity toward the minority class while preserving the original data distribution in the test set.

Class weighting was chosen over resampling techniques at this stage to preserve the original data distribution in the test set and avoid potential data leakage.

3.2.7 Model Training and Hyperparameter Configuration

The CatBoost classifier was trained using a fixed set of hyperparameters selected to balance model complexity and generalisation. Key hyperparameters included the number of boosting iterations, learning rate, tree depth, and evaluation metric. The Area Under the Receiver Operating Characteristic Curve (ROC-AUC) was used as the primary evaluation metric during training due to its suitability for imbalanced classification tasks.

An evaluation set was provided during training to monitor performance and enable early stopping, allowing the model to retain the best-performing iteration. A fixed random seed was used to ensure reproducibility of results.

3.2.8 Model Evaluation Metrics

Following model training, the CatBoost classifier was assessed on an independent test dataset to examine its predictive effectiveness. A range of evaluation measures was used to capture different aspects of performance, including overall accuracy, class-specific detection capability, and the model's ability to distinguish between stroke and non-stroke cases.

To support a more detailed error analysis, a confusion matrix was constructed to summarise correct and incorrect predictions across both classes. This allowed closer inspection of misclassification patterns, particularly the model's effectiveness in detecting stroke instances.

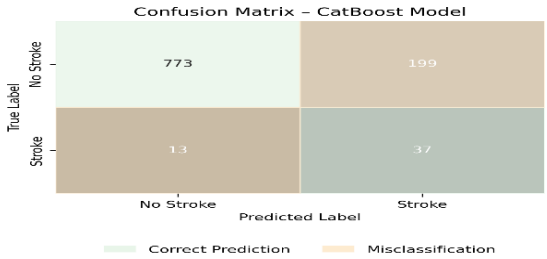


Figure 7: CatBoost Accuracy

3.2.9 Feature Importance and Model Persistence

Feature importance scores were extracted from the trained CatBoost model to identify variables contributing most strongly to stroke risk prediction. These importance values provided preliminary insights into influential risk factors and supported subsequent explainability analysis using SHAP.

Finally, the trained CatBoost model and associated metadata, including feature names and categorical feature indices, were saved to disk. These saved artefacts were later reused in the hybrid modelling stage and deployed within the Streamlit-based decision-support system.

3.2.10 Model Selection – Algorithm Approach

CatBoost is a gradient boosting algorithm based on decision trees and was selected as the primary tree-based model in this study. It is particularly well-suited to tabular healthcare data due to its ability to capture non-linear feature interactions and handle mixed data types effectively.

A key motivation for selecting CatBoost is its robustness when dealing with imbalanced datasets. The algorithm supports class weighting, allowing the minority stroke class to be emphasised during training without altering the original data distribution. This is especially important in clinical risk prediction tasks, where false negatives (missed stroke cases) can have severe consequences.

In this study, CatBoost was trained using probability outputs rather than hard class labels. This design choice enables downstream threshold tuning and hybrid probability fusion, rather than relying on a fixed default classification thresho

3.2.11 Summary

In summary, CatBoost was selected as an effective baseline model due to its strong suitability for structured clinical data containing mixed feature types. Its ability to natively process categorical variables, combined with class weighting to address data imbalance, makes it particularly well suited for stroke risk prediction. The model provides stable performance without extensive preprocessing and offers a reliable probabilistic output that supports further optimisation through threshold tuning and hybrid integration.

3.3 Multi-Layer Perceptron (MLP) Model

3.3.1 Working of Neural Network Model

In addition to tree-based learning, a neural network model was implemented to explore the ability of deep learning techniques to capture complex nonlinear relationships within the stroke prediction dataset. While gradient boosting models such as CatBoost are highly effective for structured tabular data, neural networks provide a fundamentally different learning paradigm that can complement tree-based approaches.

A Multi-Layer Perceptron (MLP) was selected due to its suitability for supervised classification problems involving heterogeneous clinical and demographic features. The inclusion of an MLP model allows for comparative analysis between traditional machine learning and neural network approaches and forms a critical component of the hybrid modelling strategy developed later in this research.

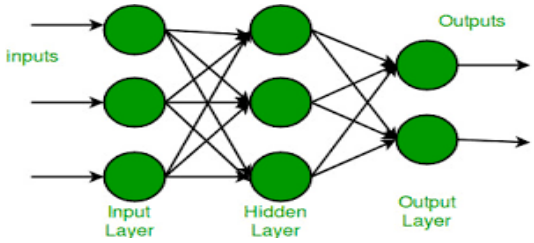


Figure 8: Multi Layered values

3.3.2 Input Data and Pre-Model Processing

The MLP model was trained using the cleaned stroke dataset produced during the data preparation phase. This dataset incorporated missing value handling and feature engineering steps performed during exploratory data analysis. Training the neural network on a cleaned dataset ensured that all input features were complete, consistent, and suitable for numerical computation.

As with the CatBoost model, the unique identifier attribute was excluded from the modelling process to prevent the network from learning patterns unrelated to stroke risk. This step ensured that model predictions were based solely on clinically relevant information.

3.3.3 Feature Definition and Dataset Partitioning

For supervised learning, the dataset was organised by clearly distinguishing input variables from the prediction target. Stroke occurrence was modelled as a binary outcome, indicating whether an individual had experienced a stroke or not. The input variables comprised a combination of demographic characteristics, clinical indicators, and lifestyle-related factors relevant to stroke risk.

To evaluate model generalization on unseen data, the dataset was divided into 80% training and 20% testing sets, using stratification based on the target variable. This approach ensured both subsets maintained a representative proportion of stroke and non-stroke cases, which is crucial in healthcare predictions to properly evaluate performance.

3.3.4 Feature Scaling and Categorical Encoding Strategy

Neural networks are sensitive to the scale of input features and cannot directly process categorical variables represented as text. To address these requirements, a structured preprocessing pipeline was constructed using a column-wise transformation strategy.

Numerical features were standardised using z-score normalisation, ensuring that all continuous variables contributed proportionally during optimisation. This step improves convergence stability and prevents features with large numeric ranges from dominating the learning process.

Categorical variables were transformed using one-hot encoding, converting each category into a binary indicator. To enhance robustness and deployment compatibility, the encoding strategy was configured to ignore previously unseen categories rather than raising errors during prediction. This design choice supports practical deployment within an interactive decision-support system.

All preprocessing steps were learned exclusively from the training data and then applied unchanged to the test set, ensuring that evaluation results were not influenced by information leakage.

```
preprocessor = ColumnTransformer(
    transformers=[
        ("num", StandardScaler(), num_cols),
```

Figure 9: Feature scaling and categorical encoding pipeline

3.3.5 Handling Class Imbalance Using SMOTE

The stroke dataset is characterised by a strong imbalance between classes, with stroke cases forming only a small proportion of the total records. When models are trained on such skewed data, they tend to favour the majority non-stroke class, which can result in poor detection of stroke events.

To reduce this bias, the Synthetic Minority Oversampling Technique (SMOTE) was applied after preprocessing and restricted to the training data only. Rather than duplicating existing samples, SMOTE generates new minority-class instances by interpolating between similar stroke cases, improving class representation during learning.

Resampling was applied exclusively to the training set, while the test set was kept unchanged. This approach ensures unbiased model evaluation and allows performance metrics to reflect true generalisation rather than the effects of artificial data balancing.

```
X_train_p = preprocessor.fit_transform(X_train)
X_test_p = preprocessor.transform(X_test)
```

Figure 10: SMOTE resampling applied during MLP training

3.3.6 Neural Network Architecture Design

The MLP model was implemented with a feed-forward architecture comprising two hidden layers. This configuration balances expressive capacity and generalisation, allowing the network to learn non-linear feature interactions without excessive complexity.

Rectified Linear Unit (ReLU) activation functions were employed due to their computational efficiency and ability to mitigate vanishing gradient issues. Model optimisation was performed using the Adam optimiser, which adaptively adjusts learning rates and is widely adopted for neural network training due to its robustness.

To reduce overfitting risk, L2 regularisation was applied, and early stopping was enabled. Early stopping monitors validation performance during training and halts optimisation when improvements stagnate, thereby improving generalisation on unseen data.

```
mlp = MLPClassifier(
    hidden_layer_sizes=(64, 32),
    activation="relu",
    solver="adam",
    alpha=0.0005,
    max_iter=400,
    random_state=42,
    early_stopping=True
)
```

Figure 11: MLP architecture configuration used in this study

3.3.7 Model Training Procedure

The MLP model was trained using the SMOTE-balanced training dataset. Training was conducted over a fixed maximum number of iterations, with early stopping determining the effective stopping point. A consistent random seed was used to ensure reproducibility across experimental runs.

The training process optimised a cross-entropy-based loss function, aligning with the probabilistic nature of stroke risk prediction. Model convergence was monitored internally through validation performance metrics.

3.3.8 Prediction and Decision Threshold

After model training, the MLP classifier produced probability scores for each instance in the test dataset. These scores represent the estimated likelihood that an individual belongs to the stroke class rather than a direct categorical decision.

For initial evaluation, a probability cut-off value of 0.5 was used to convert these continuous outputs into binary predictions. This threshold was selected to provide a consistent baseline for assessing model behaviour before exploring alternative threshold values during the hybrid optimisation stage.

3.3.9 Model Evaluation Metrics

Model performance was assessed using multiple quantitative metrics, including accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC). Due to the substantial class imbalance present in the stroke dataset, greater emphasis was placed on recall and F1-score, as these measures better reflect the model's ability to correctly identify stroke cases.

Additionally, confusion matrices were generated to provide a detailed breakdown of prediction outcomes. This enabled closer analysis of false positive and false negative errors, which is particularly important in clinical decision-support settings where incorrect classifications can have serious implications.

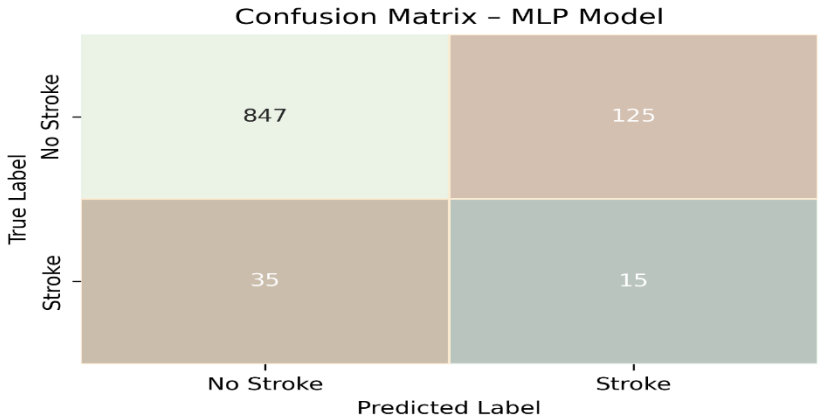


Figure 12: Confusion Matrix for MLP

3.3.10 Reason for Choosing Multilayer Perceptron (MLP)

In addition to CatBoost, a Multilayer Perceptron (MLP) neural network was implemented to model complex non-linear relationships within the data. Neural networks can learn representations of features in high dimensions and may uncover subtle patterns that tree-based models might overlook.

The MLP architecture used in this study consists of two hidden layers with Rectified Linear Unit (ReLU) activation functions. This configuration provides sufficient representational capacity while reducing the risk of overfitting. Model optimisation was performed using the Adam optimiser, which adaptively adjusts learning rates and is well suited to neural network training.

Unlike CatBoost, the MLP model is sensitive to feature scaling and categorical encoding. Therefore, a structured preprocessing pipeline involving standardisation of numerical

features and one-hot encoding of categorical variables was applied. To reduce the impact of class imbalance on neural network learning, the Synthetic Minority Oversampling Technique (SMOTE) was applied entirely to the training data.

3.3.11 Summary

In summary, the MLP model was employed to capture complex non-linear relationships within the stroke dataset that may not be fully represented by tree-based methods alone. When supported by appropriate preprocessing, feature scaling, and SMOTE-based imbalance handling, the neural network demonstrated improved sensitivity to stroke cases. The inclusion of the MLP model contributes complementary learning behaviour to the overall framework and provides valuable diversity for hybrid modelling.

3.4 Hybrid CatBoost-MLP Model

3.4.1 Main work of the Hybrid Modelling Approach

Single-model approaches can be effective; however, they frequently represent only a limited subset of the patterns found within complex healthcare datasets. In the fcators of stroke prediction, clinical and demographic variables interact in both linear and non-linear manners, and there is no single modelling technique that can be relied upon to capture all pertinent relationships. To overcome this limitation, this research employed a hybrid modeling strategy.

The hybrid model integrates a gradient boosting algorithm (CatBoost) with a neural network model (MLP) at the probability level. CatBoost excels in learning structured relationships and managing categorical variables, while the MLP is adept at modeling complex non-linear interactions following feature scaling and encoding. By combining these models, the hybrid approach seeks to harness their complementary strengths and enhance predictive robustness, particularly for minority stroke cases.

3.4.2 Inputs to the Hybrid Model

The hybrid framework does not train a new classifier from scratch. Instead, it operates on the probabilistic outputs generated by two independently trained models:

- The CatBoost classifier trained using class weighting
- The MLP classifier trained using a preprocessing pipeline and SMOTE

Both models were trained and evaluated on the same train-test split to ensure consistency. The hybrid model uses the predicted stroke probabilities produced by each model on the test set as its inputs.

This design ensures that the hybrid model reflects learned patterns from both modelling paradigms without introducing additional feature transforms or training instability.

```
#loading catboost and mlp

model_cb = CatBoostClassifier()
model_cb.load_model("../models/catboost_baseline.cbm")
print("CatBoost loaded")

# CatBoost schema
cb_feature_names = joblib.load("../models/catboost_feature_names.pkl")
cb_cat_cols = joblib.load("../models/catboost_categorical_cols.pkl")

# MLP bundle
mlp_bundle = joblib.load("../models/mlp_bundle.pkl")
pre = mlp_bundle["preprocessor"]
mlp = mlp_bundle["model"]
mlp_feature_names = mlp_bundle["feature_names"]
```

Figure 13: Loading and preparation of model outputs for hybrid prediction.

3.4.3 Probability- Level Fusion Strategy

The hybrid model combines the predicted probabilities of CatBoost and MLP using **weighted probability fusion**. For each instance in the test set, a hybrid probability score is computed as a linear combination of the individual model probabilities.

The hybrid probability is defined as:

$$P_{hybrid} = w \cdot P_{CatBoost} + (1 - w) \cdot P_{MLP}$$

where:

- $P_{CatBoost}$ is the predicted probability from the CatBoost model,
- P_{MLP} is the predicted probability from the MLP model,
- w *is a weighting parameter between* 0 *and* 1.

This formulation allows explicit control over the contribution of each model. A higher value of $w$increases reliance on CatBoost predictions, while a lower value increases the influence of the MLP.

```
weights = [0.3, 0.5, 0.7]
results = []

for w in weights:
    p_hybrid = w*p_cb + (1-w)*p_mlp
    m = evaluate_probs(y_test, p_hybrid, threshold=0.5)
```

Figure 14: Hybrid probability fusion formulation

3.4.4 Weight Selection and Exploration

To identify an effective balance between the two models, multiple weight values were evaluated. The hybrid notebook explores different weight configurations (for example, values closer to CatBoost dominance, equal contribution, and MLP dominance).

Rather than assuming equal importance, weight tuning enables empirical assessment of how each model contributes to stroke risk prediction. This is particularly important because CatBoost and MLP may differ in sensitivity to minority stroke cases. The exploration of multiple weight values ensures that the hybrid model is not arbitrarily biased toward one model.

```
weights = [0.3, 0.5, 0.7]
results = []

for w in weights:
    p_hybrid = w*p_cb + (1-w)*p_mlp
```

Figure 15: Weight selection process for the hybrid model

3.4.5 Decision Threshold Optimisation

In addition to weight tuning, decision threshold optimisation was performed as part of the hybrid modelling process. The default classification threshold of 0.5 is often suboptimal in imbalanced healthcare datasets, where the cost of false negatives is high.

To address this, the hybrid probabilities were evaluated using multiple threshold values (including values lower than 0.5). Lowering the threshold increases sensitivity to stroke cases by classifying instances with moderate predicted risk as positive.

Threshold tuning was conducted systematically by evaluating model performance across several thresholds. This allowed identification of a threshold that achieved a more favourable balance between recall and precision, with particular emphasis on improving F1-score and ROC-AUC performance.

3.4.6 Hybrid Model Evaluation Metrics

The hybrid CatBoost–MLP framework was assessed using accuracy, precision, recall, F1-score, and ROC-AUC to allow structured comparison with the individual classifiers. Due to the imbalanced class distribution, recall and F1-score were prioritised as key indicators of stroke detection capability. The evaluation focused on determining whether combining probabilistic outputs resulted in more balanced performance than either model when applied independently.

3.4.7 Confusion Matrix Analysis

A confusion matrix was created for the optimised hybrid model to examine prediction outcomes in greater detail. This analysis provided insight into the distribution of correct classifications and misclassifications, particularly false negatives and false positives. The results show that threshold optimisation reduced missed stroke cases compared to the individual models, while introducing a limited increase in false positives. From a clinical decision-support perspective, this trade-off is acceptable, as prioritising stroke detection is critical for reducing potential adverse outcomes.

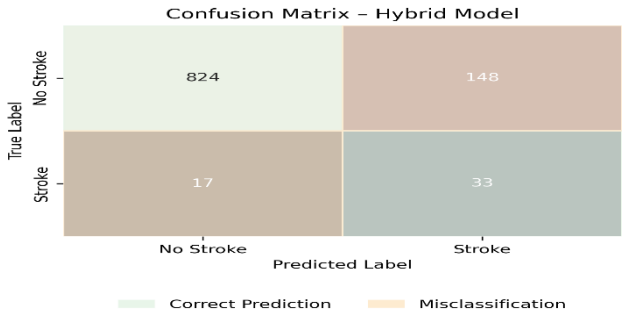


Figure 16: Confusion Matrix for Hybrid Model

3.4.8 Reason for Choosing Multilayer Perceptron (MLP) Neural Network

Rather than selecting a single model, this study adopts a hybrid modelling strategy that combines the probabilistic outputs of CatBoost and MLP. The hybrid framework does not train a new classifier; instead, it integrates the predicted stroke probabilities generated by each model.

The hybrid probability is calculated as a weighted linear combination of the individual model outputs:

$$\boldsymbol{P_{\text{hybrid}} = w \cdot P_{\text{CatBoost}} + (1 - w) \cdot P_{\text{MLP}}}$$

where w is a weighting parameter between 0 and 1.

This approach allows flexible control over the relative influence of each model. Higher weight values increase reliance on CatBoost predictions, while lower values emphasise the MLP output. Multiple weight configurations were evaluated empirically to identify the optimal balance between sensitivity and overall predictive performance.

3.4.9 Summary

In summary, the hybrid modelling approach was designed to integrate the complementary strengths of CatBoost and MLP classifiers through probability-level fusion. By combining robust tree-based predictions with neural network sensitivity and optimising both model weights and decision thresholds, the hybrid model obtains a more balanced adjustment between recall and precision. This approach is particularly well suited for imbalanced healthcare datasets, where reducing false negatives is a critical priority, and forms the foundation of the deployed decision-support system.

3.5 Explainability and Model Interpretation Using SHAP

3.5.1 Motivation for Model Explainability in Stroke Prediction

In healthcare-related machine learning applications, predictive performance alone is insufficient for practical adoption. Models must also provide transparent explanations that allow clinicians and analysts to understand how predictions are generated. In stroke risk prediction, interpretability is particularly important because incorrect or unexplained predictions may lead to loss of trust and ethical concerns.

Many advanced machine learning approaches, including gradient boosting methods and neural networks, are commonly described as black-box models due to their complex internal structures. To improve transparency, explainable artificial intelligence (XAI) techniques were integrated into this study. In particular, Shapley Additive Explanations (SHAP) were applied to the CatBoost component of the hybrid model to quantify feature contributions and support clinically meaningful interpretation.

3.5.2 Selection of SHAP for Model Interpretation

SHAP was selected as the explainability technique due to its strong theoretical foundation and suitability for tree-based models. SHAP is based on cooperative game theory and assigns contribution values (SHAP values) to individual features, indicating how much each feature contributes to a particular prediction.

Tree-based SHAP implementations are computationally efficient and provide consistent explanations at both the global and local levels. Given that CatBoost is a tree-based gradient boosting model, SHAP is particularly well-suited for interpreting its predictions without requiring model approximation.

3.5.3 Input Model and Data for SHAP analysis

SHAP analysis was applied to the **trained CatBoost model**, which had been previously selected as a strong baseline and as a core component of the hybrid framework. The same cleaned dataset and feature schema used during CatBoost training were reused to ensure consistency between prediction and explanation.

The SHAP explainer was initialised using the trained CatBoost model and a representative subset of the dataset. This allowed SHAP values to be computed for each feature with respect to stroke risk prediction.

3.5.4 Global Model Interpretation – Summary Plot

Global model interpretability was examined using a SHAP summary visualisation, which aggregates feature contributions across the entire dataset. This representation illustrates both the relative influence of each feature and the direction in which feature values affect predicted stroke risk.

The summary plot enables identification of the most influential predictors by showing how high and low feature values shift model outputs. Consistent with clinical understanding,

variables such as age, average glucose level, hypertension, and body mass index were observed to have substantial impact on prediction outcomes.

The SHAP summary plot was generated and saved for inclusion in the dissertation and the deployed decision-support system.

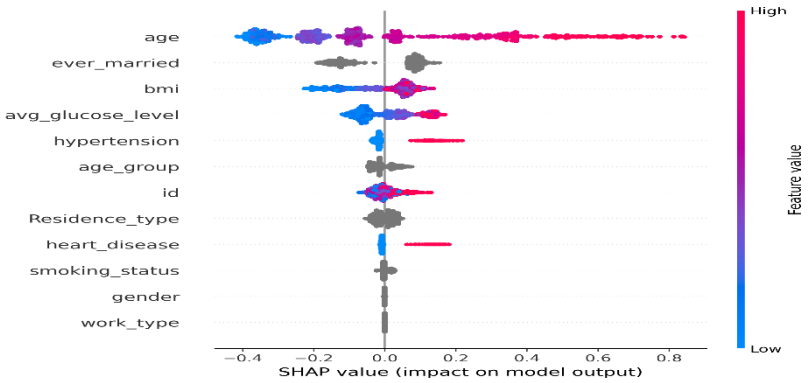


Figure 17: SHAP summary plot

3.5.5 Global Feature Importance Bar Plot

Alongside the SHAP summary visualisation, a SHAP-based feature importance bar chart was produced to rank predictors according to their mean absolute contribution to the model output. This representation offers a concise overview of global feature influence, allowing key stroke risk factors to be identified more clearly.

The bar plot complements the summary plot by offering a ranked interpretation of features without showing individual sample distributions. This is particularly useful for high-level reporting and comparison with clinical expectations.

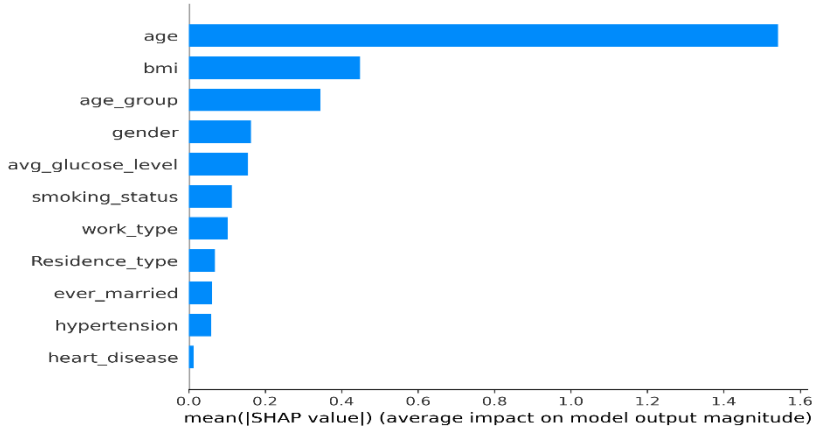


Figure 18: SHAP feature importance bar plot

3.5.6 SHAP Waterfall Plot for Individual Predictions

While global explanations describe overall model behaviour, local explanations are essential for understanding individual predictions. To achieve this, SHAP waterfall plots were generated for selected instances.

The waterfall plot illustrates how each feature contributes to pushing the prediction away from the baseline expectation toward a higher or lower stroke risk. Features with positive SHAP values increase predicted risk, while features with negative values reduce it.

This form of explanation is particularly valuable in a decision-support context, as it enables users to understand why a specific individual is classified as high or low risk.

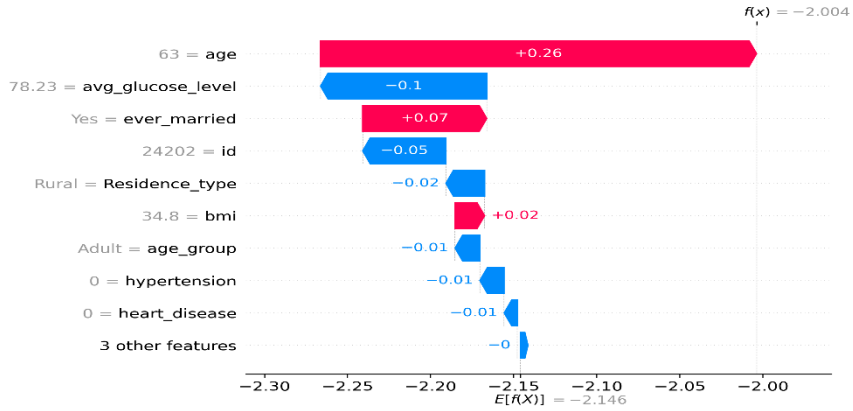


Figure 19: SHAP waterfall plot for an individual case

3.5.7 Role of SHAP in the Hybrid Decision-Support System

Although SHAP explanations were generated using the CatBoost model, they play a crucial role in enhancing the interpretability of the overall hybrid framework. By explaining the CatBoost component, SHAP provides insight into the most influential factors driving stroke risk predictions, even when the final decision is based on hybrid probability fusion.

This design balances predictive performance with transparency and supports ethical use of machine learning in healthcare decision-support systems.

3.5.8 Summary

This chapter presented the application of SHapley Additive exPlanations (SHAP) to interpret the CatBoost component of the hybrid stroke prediction framework. Both global and local explanations were generated to provide insight into feature contributions and individual predictions. The integration of SHAP enhances transparency, supports trust in model outputs, and enables meaningful interpretation within the deployed decision-support system.

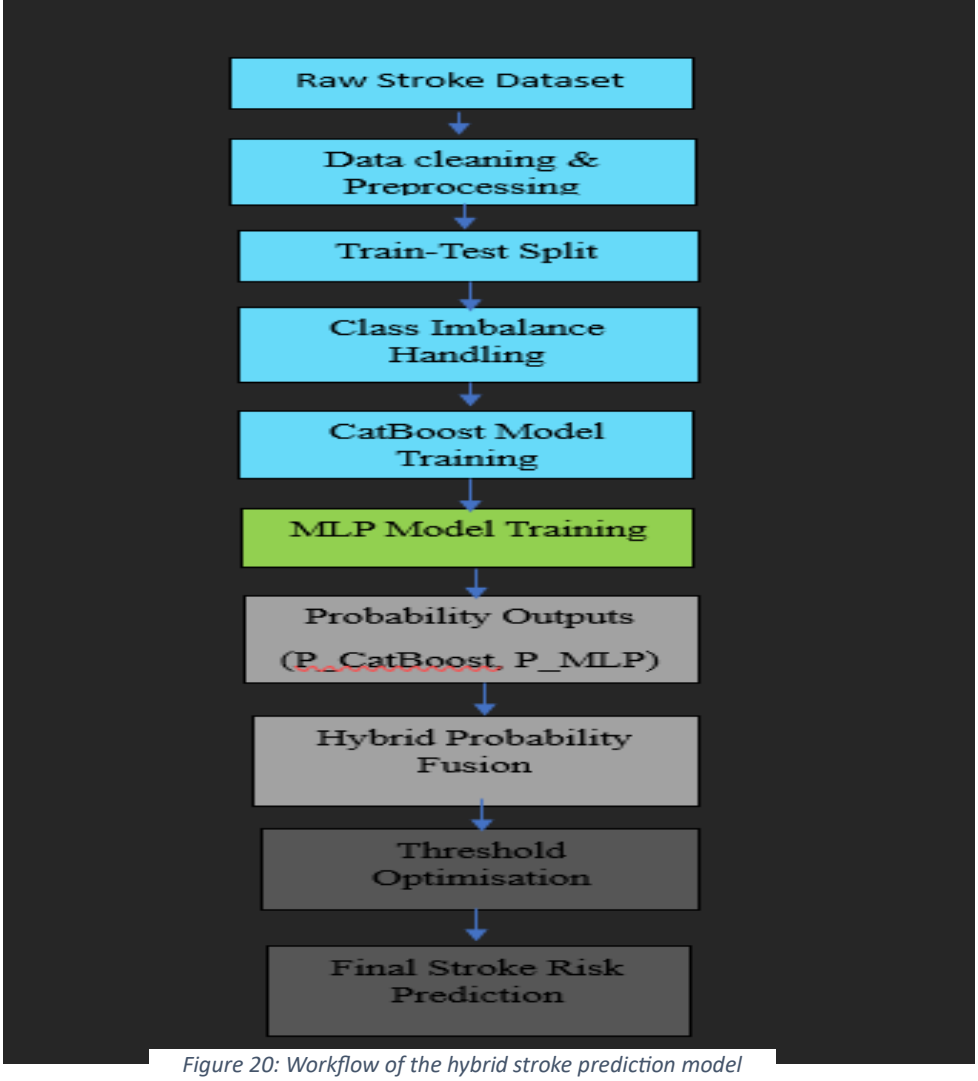


Figure 20: Workflow of the hybrid stroke prediction model

Fig 20 illustrates the end-to-end workflow of the proposed hybrid stroke prediction framework, showing how raw clinical inputs are progressively processed to produce a final stroke risk decision. Each stage of the pipeline is structured to maintain data reliability, methodological consistency, and clinical relevance.

The process begins with data cleaning and preprocessing to handle missing values and formatting inconsistencies. The dataset is subsequently partitioned into training and testing subsets using stratified sampling, ensuring that the original class proportions are maintained. This step is essential due to the low prevalence of stroke cases within the dataset.

To mitigate class imbalance, dedicated handling strategies are applied prior to model training. Two models are trained in parallel: CatBoost, which effectively captures non-linear relationships in structured tabular data, and a Multi-Layer Perceptron (MLP), which learns complex feature interactions following feature scaling and encoding. Each model is trained independently to exploit complementary learning strengths.

Rather than producing direct class labels, both models generate probability outputs, which are combined using a weighted hybrid probability fusion approach. This design enables flexible control over each model's contribution without introducing additional model complexity.

Finally, threshold optimisation is applied to convert hybrid probabilities into binary stroke risk predictions. Adjusting the decision threshold allows the model to prioritise recall, which is particularly important in clinical screening scenarios where missed stroke cases may have serious consequences.

4 System Implementation and Development

4.1 Purpose of the Decision- Support System

Although predictive performance is central to machine learning research, its practical impact is realised only when models are delivered through an accessible and interpretable interface. To demonstrate the real-world applicability of the proposed hybrid stroke prediction approach, a web-based decision-support system was developed using the Streamlit framework.

The purpose of this system is to provide an interactive interface through which users can input patient-related information and receive stroke risk predictions generated by the trained machine learning models. The application is designed for **educational and analytical decision-support purposes**, rather than clinical diagnosis. By translating model outputs into an intuitive interface, the system bridges the gap between technical model development and user-facing applications, highlighting how machine learning can support informed decision-making in healthcare contexts.

4.2 System Architecture Overview

The decision-support system follows a modular architecture in which each component performs a clearly defined role. The overall architecture is designed to ensure consistency between experimental results and deployed predictions, while maintaining transparency and reproducibility.

At a high level, the system workflow consists of the following stages:

- User input collection through the graphical interface
- Input validation and formatting
- Loading of pre-trained machine learning models and preprocessing pipelines
- Generation of stroke risk probabilities using individual models
- Hybrid probability fusion and decision threshold application
- Presentation of prediction results and explainability outputs

This architecture ensures that no model retraining occurs within the application itself. Instead, all predictions are generated using pre-trained models that were previously evaluated during experimentation, thereby guaranteeing alignment between research findings and system behaviour.

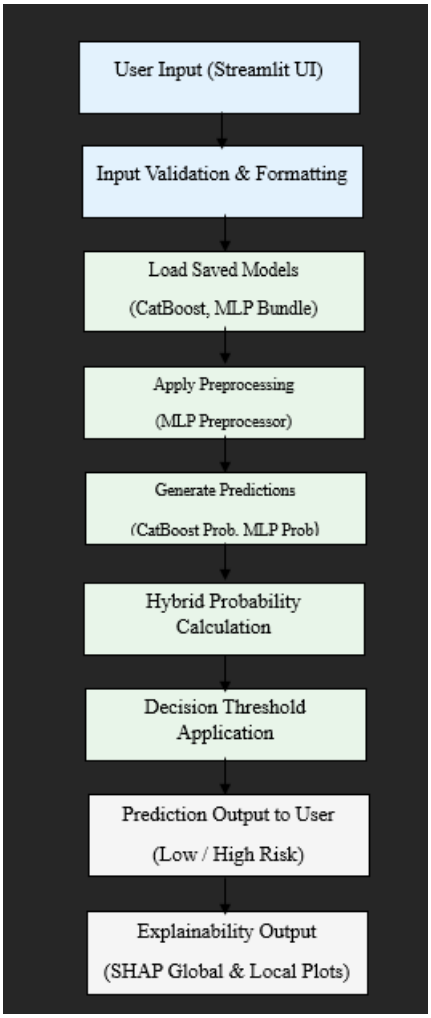


Figure 21: Workflow of the Streamlit decision-support system

Figure 21 illustrates the end-to-end workflow of the Streamlit-based stroke risk decision-support system. The process begins with user-provided patient data, which undergoes validation and formatting to ensure input reliability. Pre-trained CatBoost and MLP models, along with their corresponding preprocessing components, are then loaded. The validated input data is transformed consistently with the training phase before probabilistic predictions are generated by both models. These probabilities are combined using a hybrid fusion strategy, followed by the application of a decision threshold to classify stroke risk. Finally, SHAP-based explainability outputs are generated to support transparent and interpretable clinical decision-making.

4.3 Integration of Trained Models

The Streamlit application integrates multiple trained machine learning components to generate stroke risk predictions. Specifically, the system loads:

- A trained CatBoost Classifier
- A bundled MLP model, including its preprocessing pipeline
- Configuration files defining feature order and categorical handling

These components are loaded from persistent storage at runtime, ensuring that the same models evaluated during the experimental phase are reused during deployment. This approach avoids inconsistencies that could arise from retraining or redefining models within the application.

By enforcing a consistent feature schema and preprocessing strategy, the system ensures that user inputs are transformed in the same manner as the training data. This design choice is critical for maintaining prediction reliability and avoiding subtle discrepancies between experimental evaluation and deployed behaviour.

4.4 Graphical User Interface (GUI) Design

The graphical user interface was developed using Streamlit due to its suitability for rapid deployment of data-driven applications and its strong integration with Python-based machine learning workflows. Streamlit enables the creation of interactive web applications with minimal overhead, making it particularly appropriate for demonstrating machine learning decision-support systems in an academic context.

The interface is organised into multiple tabs, each corresponding to a distinct functional component of the system. This tab-based design improves usability by clearly separating prediction, explainability, and informational content.

The Prediction tab allows users to enter relevant demographic and clinical attributes, such as age, hypertension status, glucose level, body mass index, and lifestyle factors. Upon submission, the system computes and displays stroke risk probabilities generated by each model as well as the final hybrid probability.

4.5 Explainability Integration within the system

Interpretability is a key requirement for healthcare-related decision-support systems. To address this, the application integrates explainability outputs generated using SHapley Additive exPlanations (SHAP). Rather than computing SHAP values dynamically, which can be computationally expensive, pre-generated SHAP visualisations are loaded and displayed within the application.

The Explainability tab presents global SHAP summary plots and feature importance visualisations, allowing users to understand which variables most strongly influence stroke risk predictions. These explanations enhance transparency and support user trust by demonstrating that predictions are grounded in clinically relevant factors.

By incorporating explainability directly into the interface, the system ensures that predictive outputs are accompanied by meaningful contextual information rather than being presented as opaque numerical scores.

4.6 Summary

This chapter described the implementation and deployment of a Streamlit-based decision-support system designed to operationalise the proposed hybrid stroke prediction framework. By integrating trained machine learning models, probability-level fusion, and explainability outputs within an interactive interface, the system demonstrates how predictive models can be translated into accessible and transparent decision-support tools.

The deployment component complements the experimental work presented earlier in this dissertation and highlights the practical applicability of the proposed approach, reinforcing its relevance within applied artificial intelligence and data science contexts.

5 Results and Discussion

5.1 Train – Test Split Strategy

To assess the performance of the stroke prediction models, the cleaned dataset was split into training and testing sets using an 80:20 ratio. Stratified sampling was applied to ensure that the proportion of stroke and non-stroke cases remained similar in both subsets. This is important for imbalanced datasets, as it helps avoid biased evaluation results. By preserving the original class distribution, the models are evaluated more fairly and the reported performance metrics better reflect how the models would behave when applied to real-world stroke risk prediction tasks.

5.2 Comparative Model Performance

This section reports the predictive performance of the CatBoost, MLP, and hybrid models on the held-out test dataset. Model effectiveness was evaluated using multiple performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, in order to provide a balanced assessment under imbalanced class conditions.

The CatBoost model achieved strong baseline performance, benefiting from native categorical feature handling and class weighting. It demonstrated high accuracy and precision, indicating reliable identification of non-stroke cases. However, recall for stroke cases was comparatively lower, suggesting that some high-risk individuals were not identified under default decision thresholds.

The MLP model, trained using standardised inputs and SMOTE-based oversampling, showed improved recall relative to CatBoost. This indicates greater sensitivity to stroke cases, although this improvement was accompanied by an increase in false positives, reflecting a trade-off between sensitivity and specificity.

The hybrid CatBoost–MLP model outperformed both individual models in terms of balanced performance. By combining probabilistic outputs and optimising model weights and decision thresholds, the hybrid approach achieved higher recall and F1-score, demonstrating improved identification of stroke cases while maintaining acceptable precision. These results support the hybrid model as the most suitable candidate for deployment in a clinical decision-support context.

5.3 Model Evaluation Results

Model performance was assessed using multiple evaluation metrics to capture different aspects of predictive behaviour. Accuracy provided a general measure of correctness, while precision reflected the reliability of predicted stroke cases. Recall was considered the most critical metric, as failing to identify a true stroke case carries a higher clinical risk than generating false alarms. The F1-score was used to balance precision and recall, and ROC-AUC evaluated the model's ability to discriminate between stroke and non-stroke cases across varying thresholds.

Model performance was assessed using multiple evaluation metrics to provide a comprehensive assessment:

- **Accuracy**: Overall correctness of predictions
- **Precision**: Proportion of predicted stroke cases that are correct
- **Recall (Sensitivity)**: Ability to correctly identify stroke cases
- **F1-score**: Harmonic mean of precision and recall
- **ROC-AUC**: Ability to discriminate between stroke and non-stroke cases across thresholds

Table 2: Evaluation metrics used for stroke prediction models

Metric	Description	Importance
Accuracy	Overall correctness	Baseline
Precision	Positive prediction quality	False alarms
Recall	Sensitivity to stroke	Critical
F1-score	Balance metric	Imbalanced data
ROC-AUC	Discrimination ability	Threshold- independent

5.4 Confusion Matrix Analysis

Confusion matrix analysis was conducted to gain deeper insight into classification errors beyond aggregate metrics. The CatBoost model exhibited a lower number of false positives but a higher number of false negatives, indicating conservative prediction behaviour. In contrast, the MLP model reduced false negatives but increased false positives, reflecting a more aggressive stroke detection strategy

The hybrid model demonstrated the most favourable error distribution, achieving a noticeable reduction in false negatives compared to both individual models. Although this resulted in a modest increase in false positives, the trade-off was considered acceptable given the clinical priority of identifying at-risk patients. In stroke risk assessment, early identification is preferable even at the cost of additional follow-up screening, reinforcing the suitability of the hybrid model.

5.5 Class Imbalance Handling Strategy

Given the severe class imbalance, different mitigation strategies were applied across models. CatBoost addressed imbalance through class weighting, increasing the cost of misclassifying stroke cases. The MLP model relied on SMOTE to synthetically balance the training data, improving minority class learning.

Given the severe class imbalance, different mitigation strategies were applied across models. CatBoost addressed imbalance through class weighting, increasing the cost of misclassifying stroke cases. The MLP model relied on SMOTE to synthetically balance the training data, improving minority class learning.

5.6 Reproducibility and Environment

The hybrid model benefited from these complementary strategies and further improved performance through probability fusion and decision threshold optimisation. Lowering the classification threshold increased recall while maintaining stable ROC-AUC, demonstrating that threshold tuning provides an effective mechanism for aligning model behaviour with clinical risk priorities.

5.7 Summary

The experimental results demonstrate that a hybrid modelling strategy combining CatBoost and MLP classifiers provides superior and more clinically meaningful performance compared to individual models. By leveraging complementary learning characteristics and optimising probability fusion and decision thresholds, the hybrid model achieved improved sensitivity to stroke risk while maintaining acceptable precision. These findings justify the selection of the hybrid model as the final predictive engine for the deployed decision-support system.

6. Limitations

6.1 Dataset Size and Representativeness

The stroke prediction dataset used in this study is publicly available and relatively limited in size. While it contains relevant demographic and clinical features, it may not fully represent the diversity of real-world populations. Factors such as geographic variation, ethnicity, and access to healthcare services are not explicitly captured. As a result, the generalisability of the proposed models to broader populations may be limited.

6.2 Class Imbalance in Stroke Data

Stroke cases represent a small minority of the dataset, resulting in a highly imbalanced classification problem. Although class imbalance was addressed using class weighting for CatBoost and SMOTE for the MLP model, these techniques do not fully eliminate imbalance-related challenges. In particular, oversampling methods may introduce synthetic patterns that do not perfectly reflect real patient data.

6.3 Limited Feature scope

The dataset includes a restricted set of features, primarily focusing on basic demographic and clinical indicators. Important factors such as genetic information, detailed medical history, medication usage, and lifestyle behaviours are not available. The absence of these variables may limit the model's ability to capture more complex risk patterns associated with stroke.

6.4 Lack of Temporal and Longitudinal Data

The dataset used in this study represents a snapshot of patient information rather than longitudinal data collected over time. Stroke risk is influenced by changes in health status and behaviour across time, which cannot be captured using static data. The models developed therefore, cannot account for temporal trends or evolving risk factors.

6.5 Absence of External Validation

The models developed in this study were evaluated using a publicly available dataset and were not validated on external or clinical datasets. While internal evaluation provides useful insight into predictive performance, results may vary when applied to real-world healthcare environments due to differences in patient populations, data collection procedures, and clinical practices.

Without external validation, it is not possible to determine how well the proposed models would generalise to unseen populations. For this reason, the developed system is intended strictly for educational and analytical decision-support purposes rather than direct clinical deployment. External validation using hospital or population-level datasets would be a necessary step before considering real-world adoption.

7. Conclusion

This dissertation presented the development of a stroke risk prediction decision-support system using machine learning techniques, with a particular focus on addressing class imbalance, model interpretability, and practical deployment. Using a publicly available stroke dataset, comprehensive data preprocessing and exploratory analysis were conducted to ensure data quality and suitability for modelling.

Three predictive approaches were implemented and evaluated: a CatBoost classifier, a Multi-Layer Perceptron (MLP) neural network, and a hybrid model that combines both approaches through probability-level fusion. The CatBoost model demonstrated strong baseline performance due to its native handling of categorical features, while the MLP model captured non-linear relationships following appropriate preprocessing and SMOTE-based imbalance handling. Building on these models, the hybrid CatBoost–MLP framework achieved more balanced predictive performance through weight tuning and decision threshold optimisation, leading to improved sensitivity to stroke cases.

To enhance transparency, SHapley Additive exPlanations (SHAP) were applied to interpret model predictions at both global and local levels. The integration of explainability supports user understanding of feature contributions and aligns the system with responsible and transparent AI principles. Furthermore, the trained models and explainability outputs were deployed within a Streamlit-based graphical user interface, demonstrating how machine learning models can be translated into an accessible decision-support tool.

Overall, this study demonstrates that hybrid modelling combined with explainability and deployment can provide a robust and interpretable approach to stroke risk prediction, suitable for educational and analytical decision-support purposes.

8 Future Work

While the proposed system demonstrates promising performance, several avenues for future work remain. Firstly, the use of larger and more diverse datasets could improve model generalisability and robustness across different populations. Incorporating data from multiple sources or healthcare settings would help reduce potential bias introduced by a single public dataset.

Secondly, future research could explore the use of temporal or longitudinal data to capture changes in patient health over time. Time-series modelling approaches may enable more accurate assessment of evolving stroke risk and provide earlier warnings for high-risk individuals.

Additional work could also investigate alternative ensemble strategies, such as stacking or boosting-based ensembles, to further enhance predictive performance. Moreover, fairness and bias analysis could be incorporated to evaluate whether model predictions differ across demographic groups.

Finally, clinical validation remains an important step for future development. Collaborating with healthcare professionals to evaluate the system's outputs in real-world scenarios would provide valuable feedback and support further refinement. With these extensions, the proposed framework could evolve into a more comprehensive and clinically relevant decision-support system.

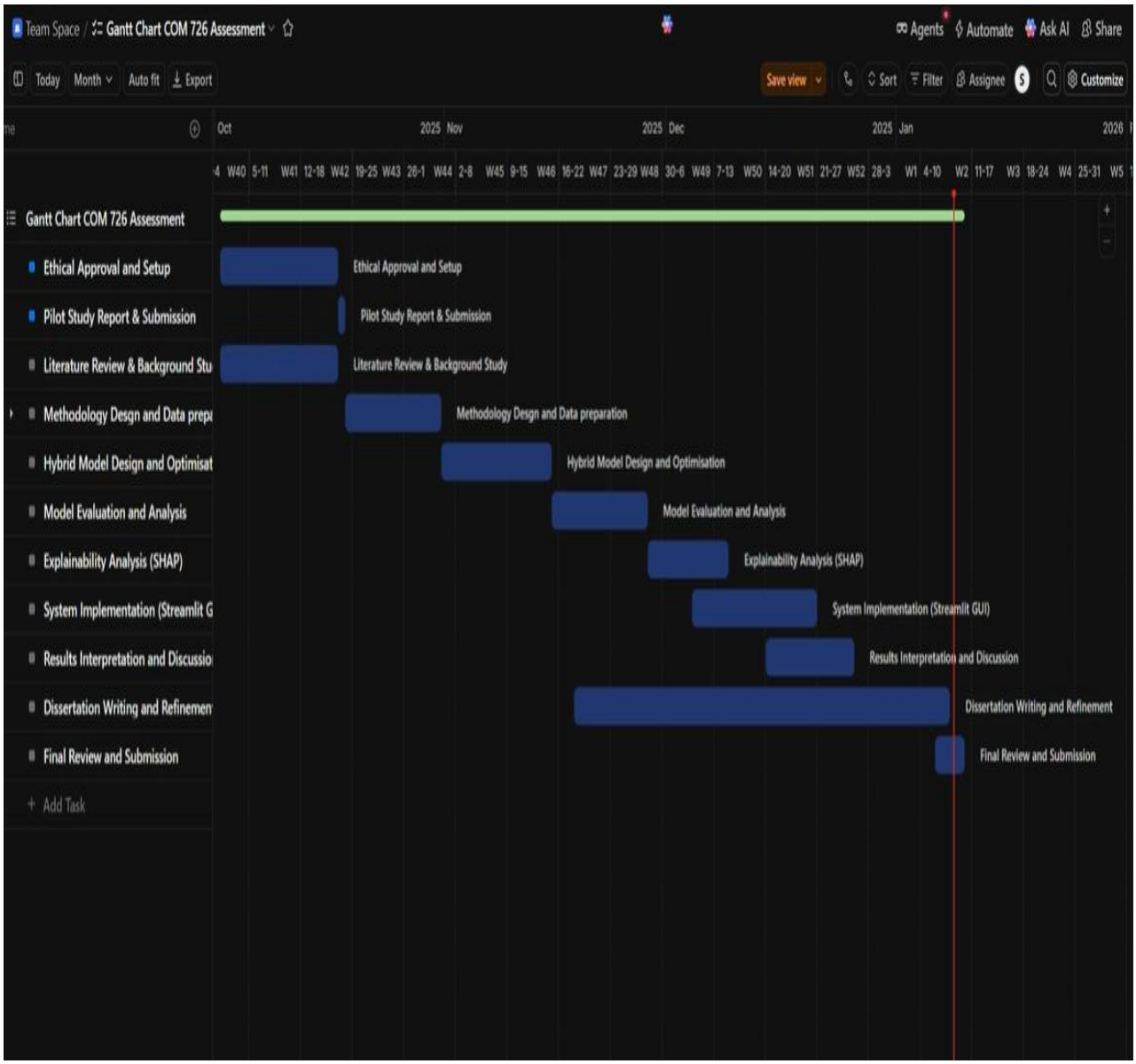
9. References

1. Adeel, A., Gogate, M., Hussain, A. and Whitmer, W.M. (2022) *An interpretable approach with explainable AI for heart stroke prediction*. **Diagnostics**, 14(2), 128.
2. Akinwumi, O., Al-Khalidi, H.R. and Yusuf, S. (2025) *Cardiovascular stroke risk predictive calculators: a comparison*. **Egyptian Heart Journal**, 77, 31
3. Alam, M., Rahman, M.M. and Hossain, M.S. (2024) *Predicting stroke occurrences: a stacked machine learning approach with feature selection and data preprocessing*. **BMC Bioinformatics**, 25, 138.
4. Ali, S., Khan, M.A. and Khan, S. (2023) *Improving stroke prediction accuracy through machine learning and synthetic minority over-sampling*. **Journal of Computing & Biomedical Informatics**, 7(2), 566.
5. Chakraborty, S., Banerjee, A. and Ghosh, R. (2024) *An ensemble machine learning and data mining approach to enhance stroke prediction*. **Bioengineering**, 11(6), 672.
6. **Chaudhary, R., Gupta, A., Jatana, N., Malik, S., Gepreel, K.A., Asmat, F. and Mohanty, S.N. (2025)**
   Predicting stroke risk: An effective stroke prediction model based on neural networks. *Journal of Neurorestoratology*, 13(1), 100156.
7. Ding, Y., Zhou, X. and Li, Y. (2024) *Predicting stroke recurrence using segmented neural network aggregation*. **Frontiers in Neurology**, 15, 1187423.
8. Dritsas, E. and Trigka, M. (2024) *Evaluating machine learning models for stroke prediction based on clinical variables*. **Applied Computing and Informatics**, 20(1), pp. 15–32.
9. Islam, M.M., Poly, T.N. and Li, Y.C. (2022) *Application of explainable artificial intelligence in EEG-based stroke prediction*. **IEEE Access**, 10, pp. 108923–108935.
10. **Kleindorfer, D.O., Towfighi, A., Chaturvedi, S., Cockroft, K.M., Gutierrez, J., Lombardi-Hill, D., Kamel, H., Kernan, W.N., Kittner, S.J., Leira, E.C., Lennon, O., Meschia, J.F., Nguyen, T.N., Pollak, P.M., Santangeli, P., Sharrief, A.Z., Smith, S.C., Turan, T.N. and Williams, L.S. (2021)**
    2021 guideline for the prevention of stroke in patients with stroke and transient ischemic attack. *Stroke*, 52(7), pp. e364–e467.
11. **Liu, Y., Zhang, Y., Fang, Y., Chen, X. and Wang, J. (2017)**
    A machine learning-based stroke prediction model using demographic and clinical features. *PLoS ONE*, 12(8), e0185402.

12. Mia, M.S., Rahman, M.A. and Hasan, M. (2024) *Hybrid machine learning approaches for stroke prediction using SMOTE*. **Journal of Biomedical Informatics**, 147, 104326.
13. Park, J., Kim, J. and Lee, H. (2020) *Real-time stroke risk prediction using SMOTE-based machine learning*. **Sensors**, 20(12), 3462.
14. Rahman, M.M. and Hasan, M. (2023) *Machine learning-based stroke prediction using imbalanced clinical datasets*. **Healthcare Analytics**, 3, 100143.
15. Reddy, K. and Kumar, S. (2023) *Stroke risk prediction with a hybrid deep transfer learning framework*. **IEEE Journal of Biomedical and Health Informatics**, 27(8), pp. 3921–3932.
16. Santos, J., Ramos, L. and Ferreira, P. (2025) *Automatic prediction of stroke treatment outcomes: latest advances and perspectives*. **Biomedical Engineering Letters**, 15, pp. 467–488.
17. **Smith, M., Brown, A., Lee, J. and Patel, R. (2024)**
Machine learning algorithms for predicting hemorrhagic transformation in ischemic stroke: A systematic review. *Journal of Neurology*, 271, pp. 4567–4582.
18. **Wang, W., Kiik, M., Peek, N., Curcin, V., Marshall, I.J., Rudd, A.G., Wang, Y., Douiri, A., Wolfe, C.D.A. and Bray, B.D. (2020)**
A systematic review of machine learning models for predicting outcomes of stroke using structured data. *PLoS ONE*, 15(6), e0234722.
19. Wijaya, D.R., Nugroho, L.E. and Pratama, A. (2024) *An ensemble data mining approach for stroke prediction*. **Discover Data**, 2, 70.
20. **Yadav, P., Sharma, R., Singh, A. and Verma, S. (2025)**
Automatic prediction of stroke treatment outcomes: Latest advances and perspectives. *Biomedical Engineering Letters*, 15(1), pp. 1–15.
21. **Zhang, H., Li, Q., Sun, X. and Chen, L. (2025)**
Stroke risk prediction with deep learning techniques: A comparative study. *Discover Data*, 3(1), 70.
22. **Zhao, Y., Lin, Z., Wang, J. and Xu, K. (2025)**
Unlocking the potential of deep learning in brain stroke prognosis: A systematic review. *Artificial Intelligence Review*, 58, pp. 1–32.
23. **Zhou, Y., Chen, H., Li, S. and Wang, X. (2023)**
Explainable artificial intelligence for stroke outcome prediction: A clinical perspective. *Methods of Information in Medicine*, 62(4), pp. 205–214.

24. Covenant Health (n.d.) Types of stroke. Available at: https://www.covenanthealth.com/cumberland/services/stroke-care/types-of-stroke/ (Accessed: 8 January 2026).

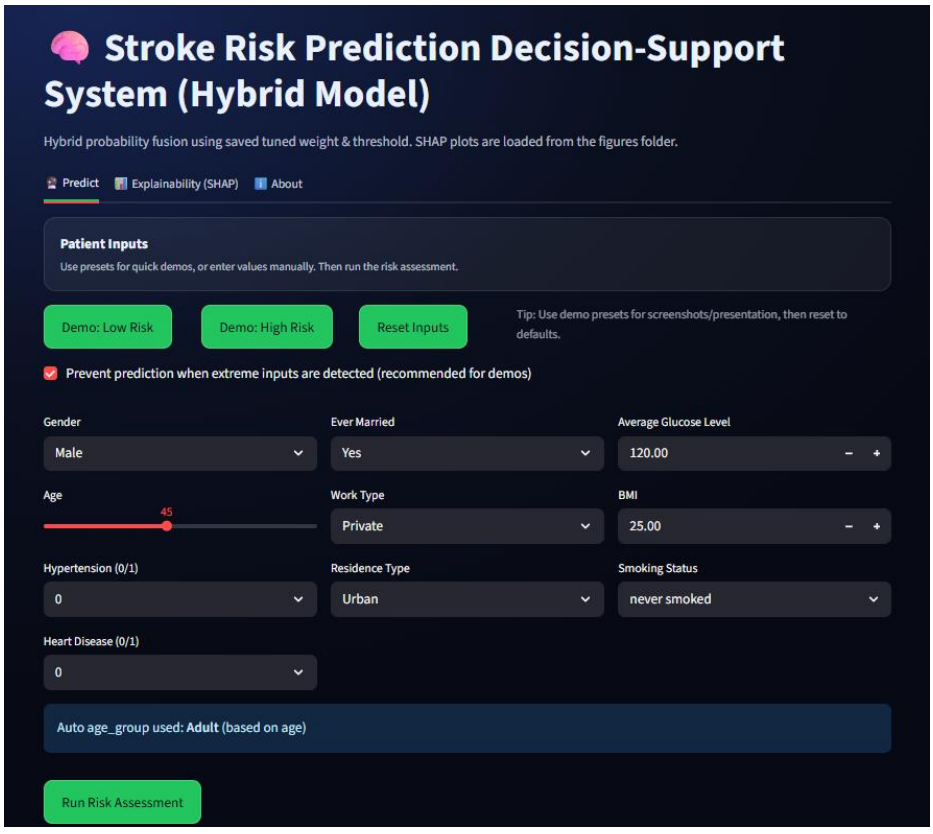
Appendix - Figures



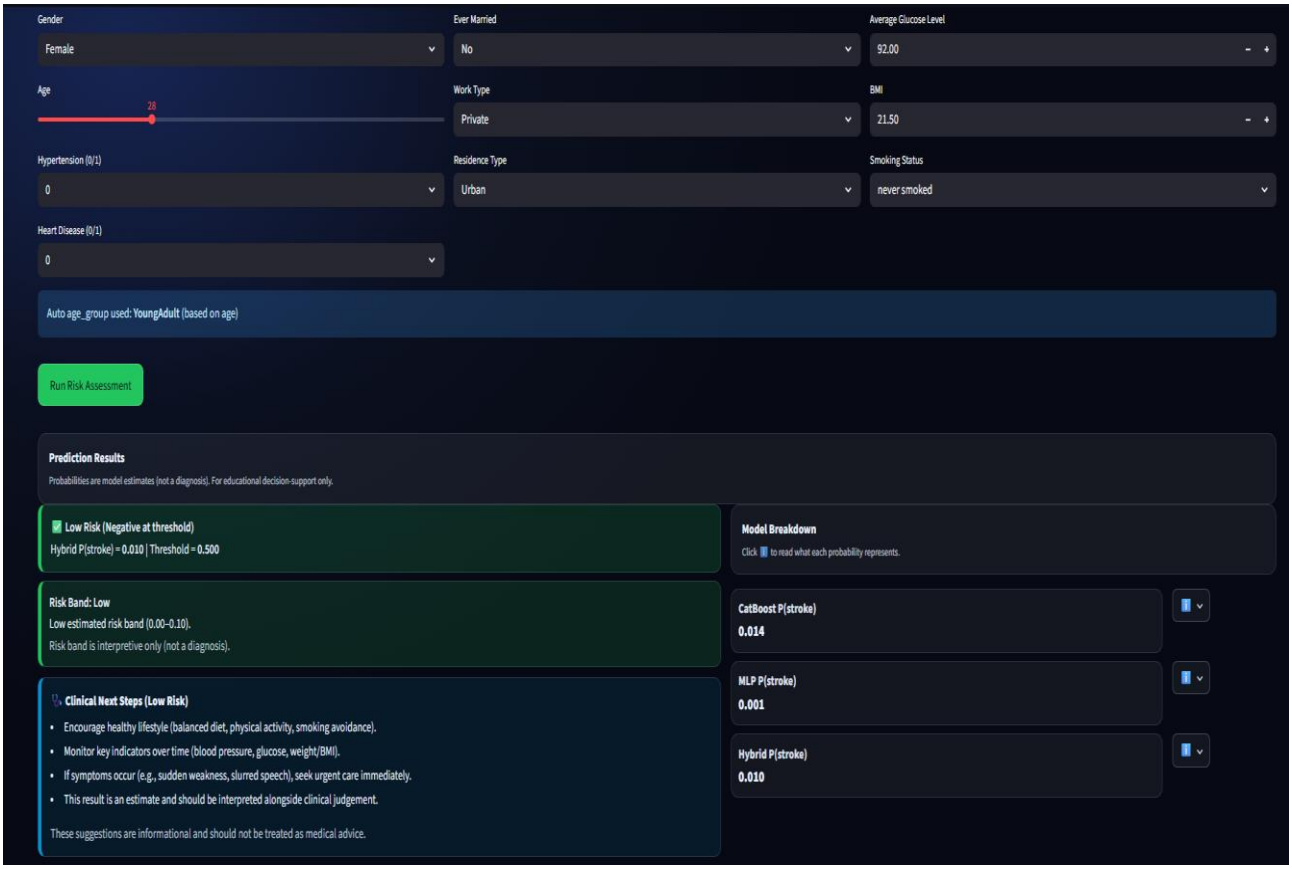
Appendix 1 – Gantt Chart

```python
# TAB 1 — Prediction

with tab1:
    st.markdown(
        '<div class="card"><div class="card-title">Patient Inputs</div>'
        '<div class="card-sub">Use presets for quick demos, or enter values manually. Then run the risk assessment.</div></div>',
        unsafe_allow_html=True
    )
    st.write("")

    b1, b2, b3, b4 = st.columns([1.2, 1.2, 1.2, 3])
    with b1:
        if st.button("Demo: Low Risk", key="preset_low"):
            apply_preset(DEMO_LOW); st.rerun()
    with b2:
        if st.button("Demo: High Risk", key="preset_high"):
            apply_preset(DEMO_HIGH); st.rerun()
    with b3:
        if st.button("Reset Inputs", key="reset_btn"):
            reset_inputs(); st.rerun()
    with b4:
        st.caption("Tip: Use demo presets for screenshots/presentation, then reset to defaults.")

    # ✅ Toggle to block extreme inputs
    st.checkbox(
        "Prevent prediction when extreme inputs are detected (recommended for demos)",
        key="block_extreme_inputs"
    )

    st.write("")
    c1, c2, c3 = st.columns(3)

    with c1:
        gender = st.selectbox("Gender", ["Male", "Female", "Other"],
                              index=["Male", "Female", "Other"].index(st.session_state["gender"]),
                              key="gender")
        age = st.slider("Age", 0, 100, int(st.session_state["age"]), key="age")
        hypertension = st.selectbox("Hypertension (0/1)", [0, 1],
                                    index=[0, 1].index(int(st.session_state["hypertension"])),
                                    key="hypertension")
        heart_disease = st.selectbox("Heart Disease (0/1)", [0, 1],
                                     index=[0, 1].index(int(st.session_state["heart_disease"])),
                                     key="heart_disease")
```

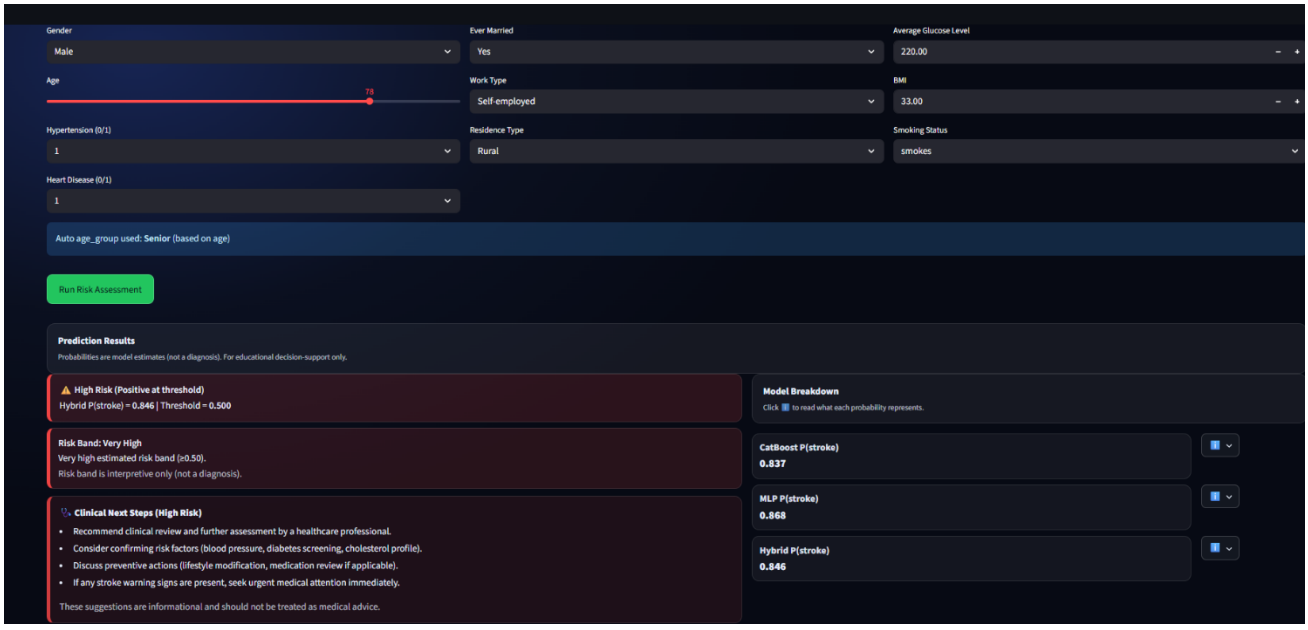
Appendix 2 – Code Snippet of GUI Prediction



Appendix 3 – Stroke Risk Prediction GUI Prediction



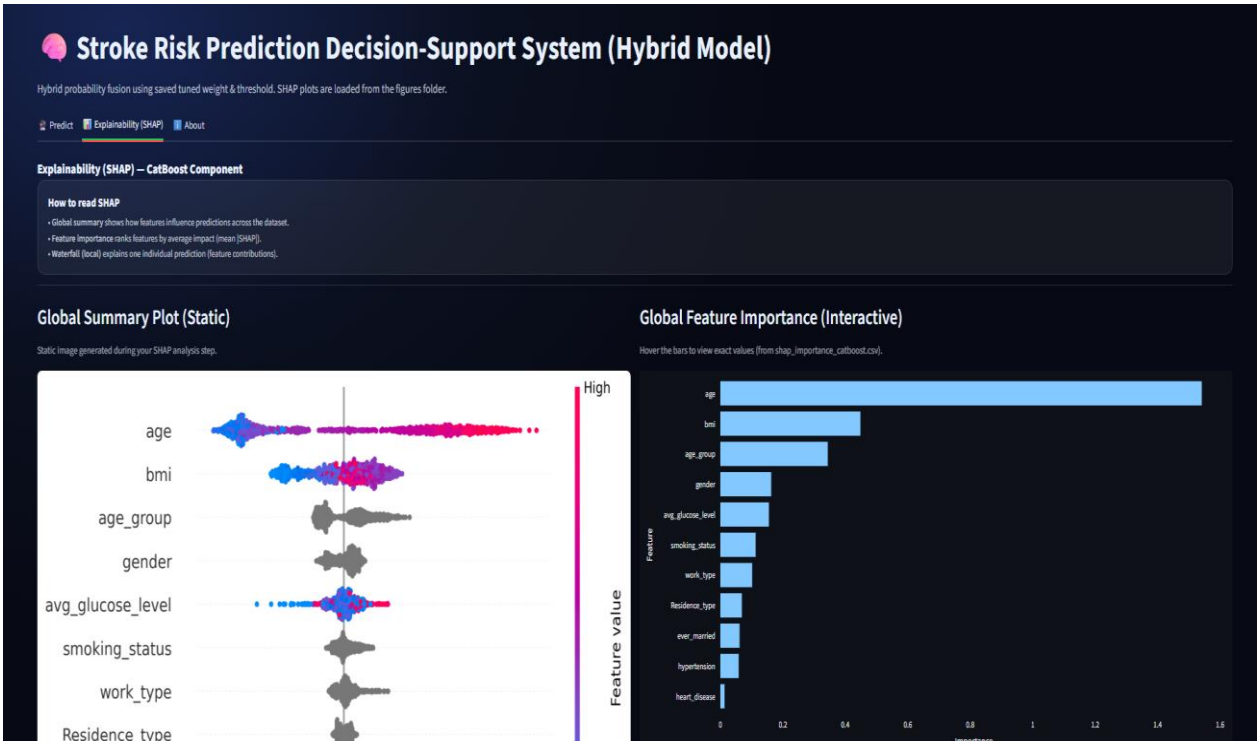
Appendix 3 – GUI Prediction for Low Risk



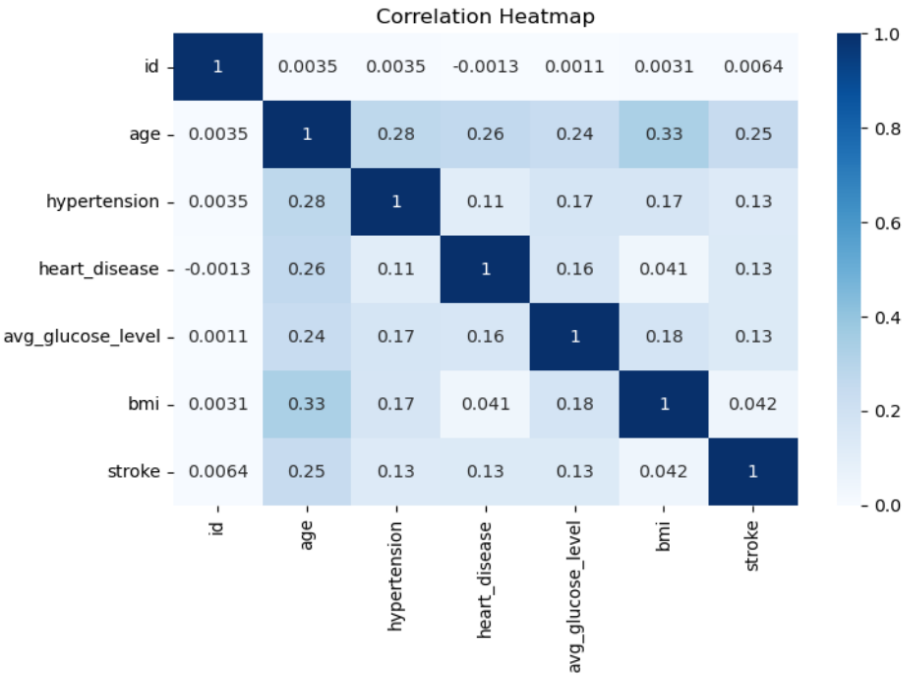
Appendix 4 – GUI prediction for High Risk

```
# TAB 2 — SHAP

with tab2:
    st.markdown('<div class="section-title">Explainability (SHAP) — CatBoost Component</div>', unsafe_allow_html=True)
    st.markdown(
        """
<div class="card">
  <div class="card-title">How to read SHAP</div>
  <div class="card-sub">
    • <b>Global summary</b> shows how features influence predictions across the dataset.<br>
    • <b>Feature importance</b> ranks features by average impact (mean |SHAP|).<br>
    • <b>Waterfall (local)</b> explains one individual prediction (feature contributions).
  </div>
</div>
        """,
        unsafe_allow_html=True
    )
    st.markdown('<div class="hr"></div>', unsafe_allow_html=True)

    summary_img = os.path.join(FIG_DIR, "shap_summary_catboost.png")
    shap_csv = os.path.join(FIG_DIR, "shap_importance_catboost.csv")

    colA, colB = st.columns(2)

    with colA:
        st.markdown("### Global Summary Plot (Static)")
        st.caption("Static image generated during your SHAP analysis step.")
        if os.path.exists(summary_img):
            st.image(summary_img, use_container_width=True)
        else:
            st.warning("shap_summary_catboost.png not found in figures/")

    with colB:
        st.markdown("### Global Feature Importance (Interactive)")
        st.caption("Hover the bars to view exact values (from shap_importance_catboost.csv).")
        if os.path.exists(shap_csv):
            fi = pd.read_csv(shap_csv).copy()
            cols_lower = {c.lower(): c for c in fi.columns}
            feat_col = cols_lower.get("feature", fi.columns[0])
            imp_col = fi.columns[1] if len(fi.columns) >= 2 else fi.columns[0]
```

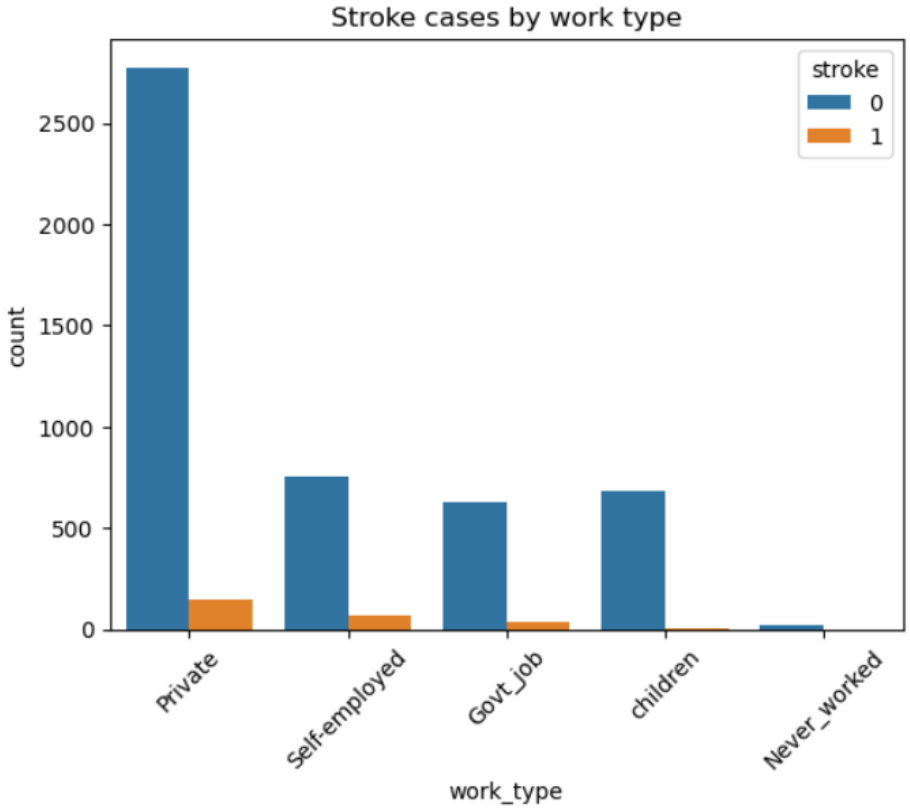
Appendix 5 – Code snippet of SHAP



Appendix 6: Stroke Risk Prediction Decision- (Explanability SHAP)



Appendix 7: Correlation Heatmap



Appendix 8: Stroke cases by work where (0= Non stroke, 1=Stroke)

Appendix – Ethical Form

Downloaded: 08/01/2026
Approved: 02/12/2025

Sahiti Gudur
Science and Engineering

Dear Sahiti

PROJECT TITLE: Enhancing Stroke Risk Prediction through a Hybrid CatBoost–Neural Network Framework
APPLICATION: Reference Number 003185

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 02/12/2025 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 003185 (form submission date: 16/10/2025);

If during the course of the project you need to deviate significantly from the above-approved documentation please email the ethics team.

Yours sincerely

Zuhaib Khan
Deputy Chair
Science and Engineering

Approval Letter

Application 003185

Section A: Researcher Details

Date application started:
Wed 15 October 2025 at 21:41

First name:
Sahiti

Last name:
Gudur

Email:
0gudus19@solent.ac.uk

Research project title:
Enhancing Stroke Risk Prediction through a Hybrid CatBoost–Neural Network Framework

Last updated:
02/12/2025

Department:
Science and Engineering

Course:
Artificial Intelligence and Data Science

Applying as:
Student

Proposed project start date
16/10/2025

Supervisor

Name	Email
Zuhaib Khan	zuhaib.khan@solent.ac.uk

Section B: General Questions

Are you dealing with human participants?
No

Is your research going to involve vulnerable groups in any way?
No

Will the study require the cooperation of a gatekeeper?
No

Are you dealing with Personal or Sensitive topics?
No

Will there be audio, video or photographic recording of participants?
No

Will the project involve any risk to the participants' health?
No

Could the study induce psychological stress or anxiety, or cause harm or negative consequences beyond the risks encountered in normal life to participants or researchers?
No

Does your research involve the storage of records on a computer, electronic transmissions, or visits to websites, which are associated with terrorist or extreme groups or other security sensitive material?
No

Will the project involve the development for export of 'controlled' goods regulated by the Export Control Organisation (ECO)?
No

Will the project involve financial inducement offered to participants other than reasonable expenses and compensation for time?
No

Will this study be submitted for ethical review to an external organisation?
No

Will any aspect of your research take place overseas?
No

Will the study involve use of deception?
No

Does your project involve using children as your participant population? If 'yes', for children under the age of 18, their own consent (where possible) and parental / guardian consent is required this must be written consent).
No

Are there any other ethical issues or risks of harm raised by the study that have not been covered by previous questions?
No

Appendix – Link

Dataset Link:

(https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset)

Streamlit GUI Link:

(http://localhost:8501/)