STROKE PREDICTION SYSTEM

Be ready to know more about our system

TABLE OF CONTENTS

01 OVERVIEW

02 PROBLEM

03 SOLUTION

04 DATASETS

05 CHALLENGES

06 MODELS

OVERVIEW

- A stroke happens when the blood supply to part of the brain is interrupted or reduced, preventing brain tissue from getting oxygen and nutrients. Within minutes, brain cells begin to die.

- Types of Stroke:
  - Ischemic Stroke: Caused by a blocked artery (about 85% of cases).
  - Hemorrhagic Stroke: Caused by a ruptured blood vessel (about 15% of cases).
- Stroke is the 2nd leading cause of death and a major cause of disability worldwide.

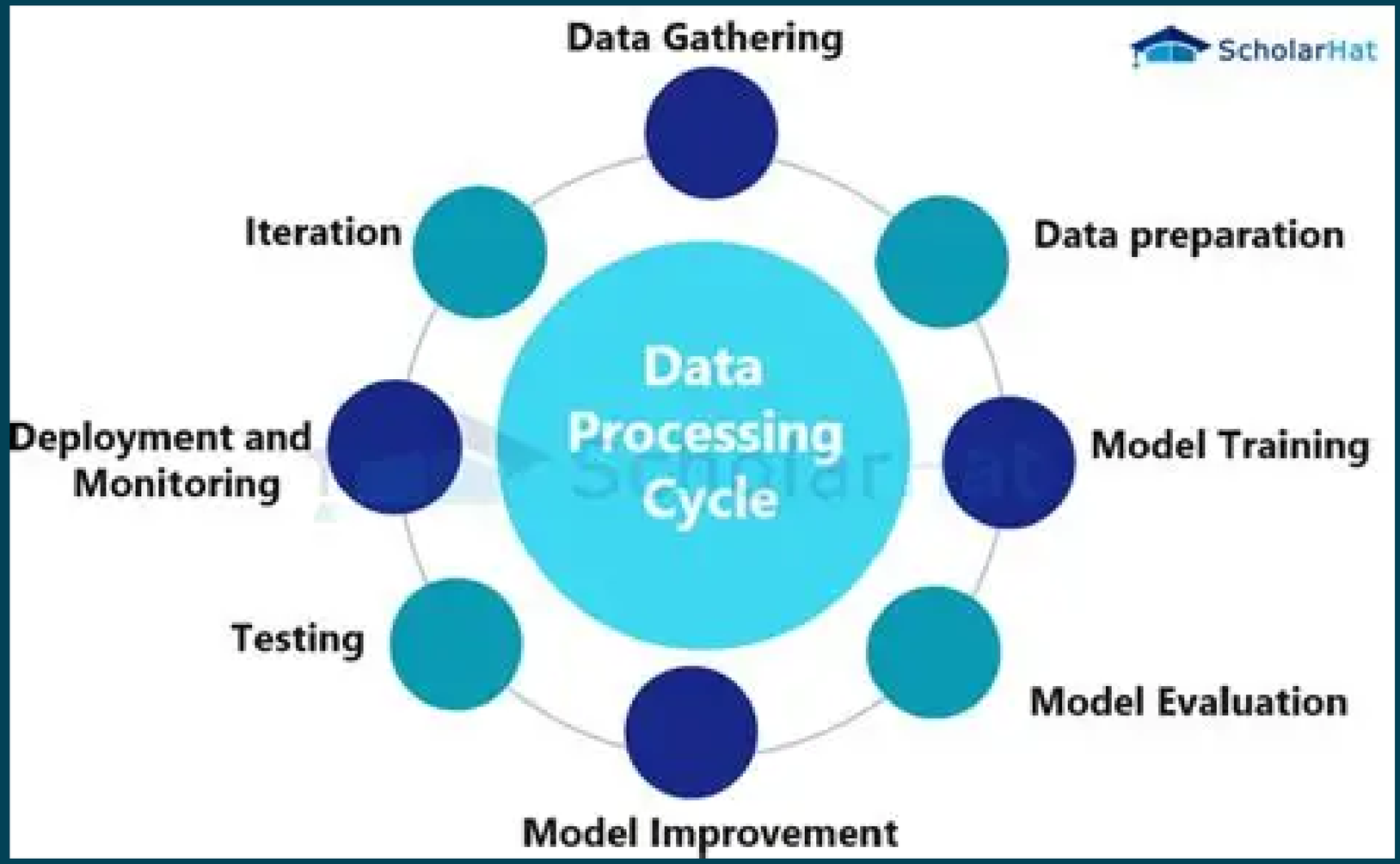
PROBLEM

Early detection of stroke type is difficult but critical for treatment.
Misdiagnosis or late diagnosis leads to higher mortality and long-term disabilities.

SOLUTION

- Develop a Machine Learning & Deep Learning-based Stroke Prediction System.
- Uses patient records + brain MRI images for better accuracy.
- Helps doctors with early decision support.

CYCLE FOR MACHINE LEARNING



DATASETS

TABULAR DATASET (MACHINE LEARNING)

- Contains patient info: age, gender, hypertension, heart disease, smoking status, etc.
- Goal: Predict if a patient is at risk of stroke.

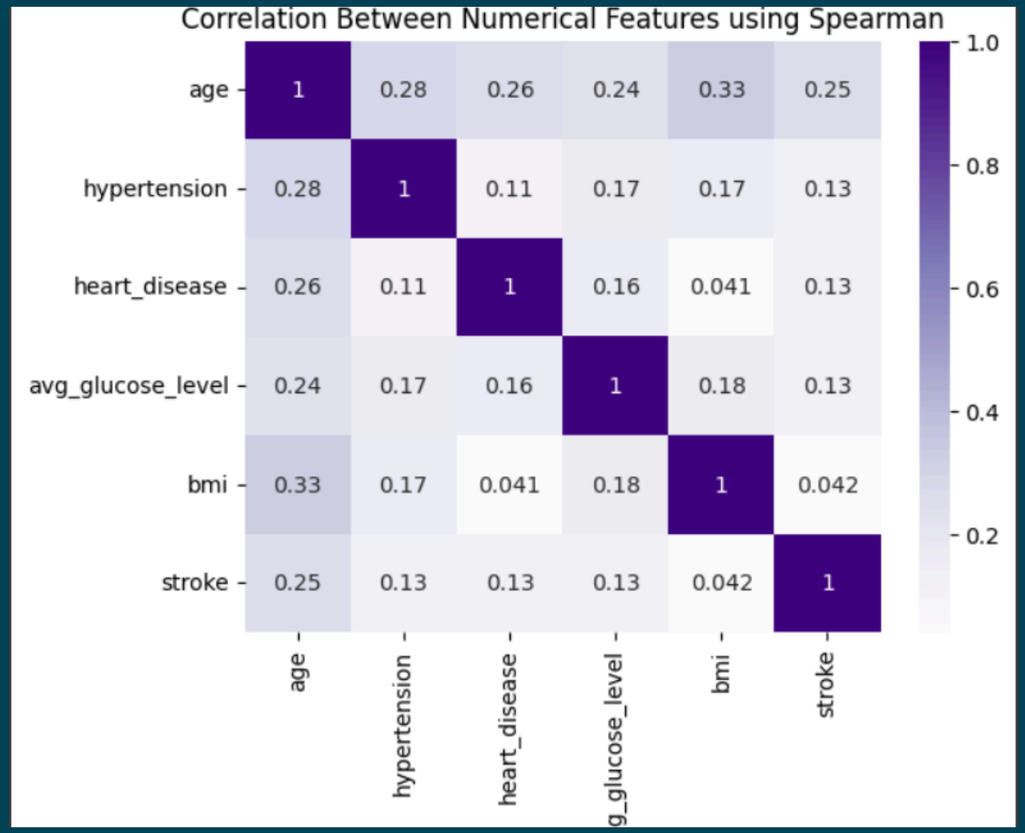
MRI IMAGE DATASET (DEEP LEARNING):

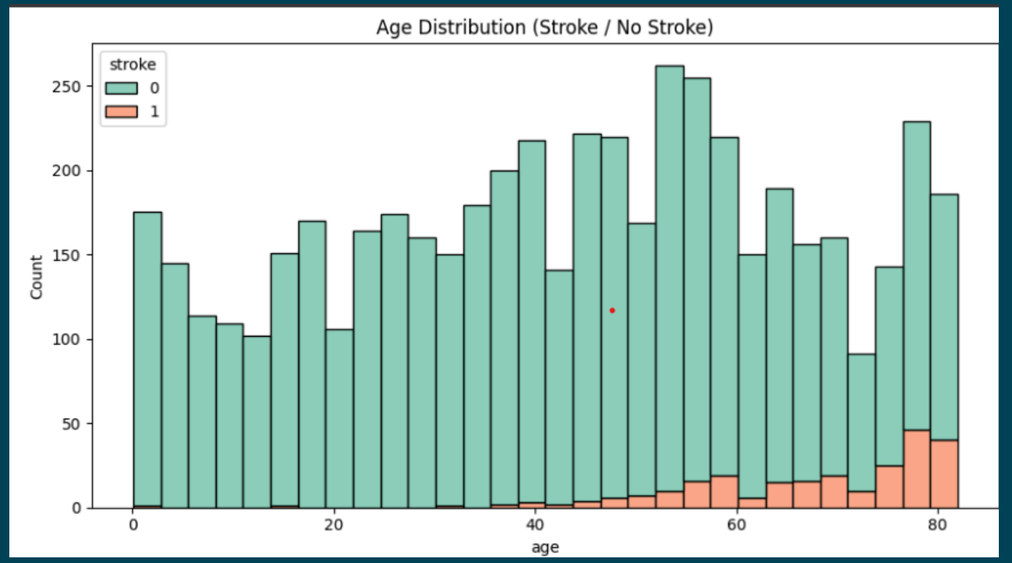
- Categories: Normal, Ischemic, Haemorrhagic.
- Goal: Classify MRI scans into stroke type.

EXPLORATORY DATA ANALYSIS

EXPLORE DATASET MORE

- Stroke patients are often older (points cluster in higher age range).





- Age has the highest correlation with stroke (0.25).
- BMI also shows moderate correlation (0.33 with age, 0.042 with stroke).
- Hypertension & heart disease increase stroke risk, though correlation values are weaker (0.13).
- Glucose level plays a role (0.13 correlation with stroke).
- Overall: multiple factors contribute, but stroke is a complex, multi-factorial disease (not explained by a single variable).

PROBLEMS IN TABULAR DATASET

IMBALANCED DATASET

Use SMOTE (oversampling)

MISSING VALUES IN FEATURES LIKE BMI.

Imputation (mean/mode).

NOISE IN CATEGORICAL FEATURES.

One-hot encoding, normalization.

PROBLEMS IN IMAGES

DIFFERENT IMAGE SIZES.

Resize to standard
(e.g., 224x224).

SMALL DATASET.

Data augmentation
(rotation, flipping).

WHY WE USE F1-SCORE FOR CLASS 1

- In our dataset, stroke cases (class 1) are very rare compared to non-stroke cases (class 0).
- If we only look at accuracy, a model could predict "No Stroke" for everyone and still be ~95% accurate — but it would completely miss stroke patients.
- For healthcare, missing a stroke case (false negative) is far more dangerous than having extra false

MODEL SELECTION

Logistic Regression

- Simple and interpretable.
- Achieved the highest F1-score for Class 1 (stroke cases) → most important since it's the minority class.
- Selected as the final ML model

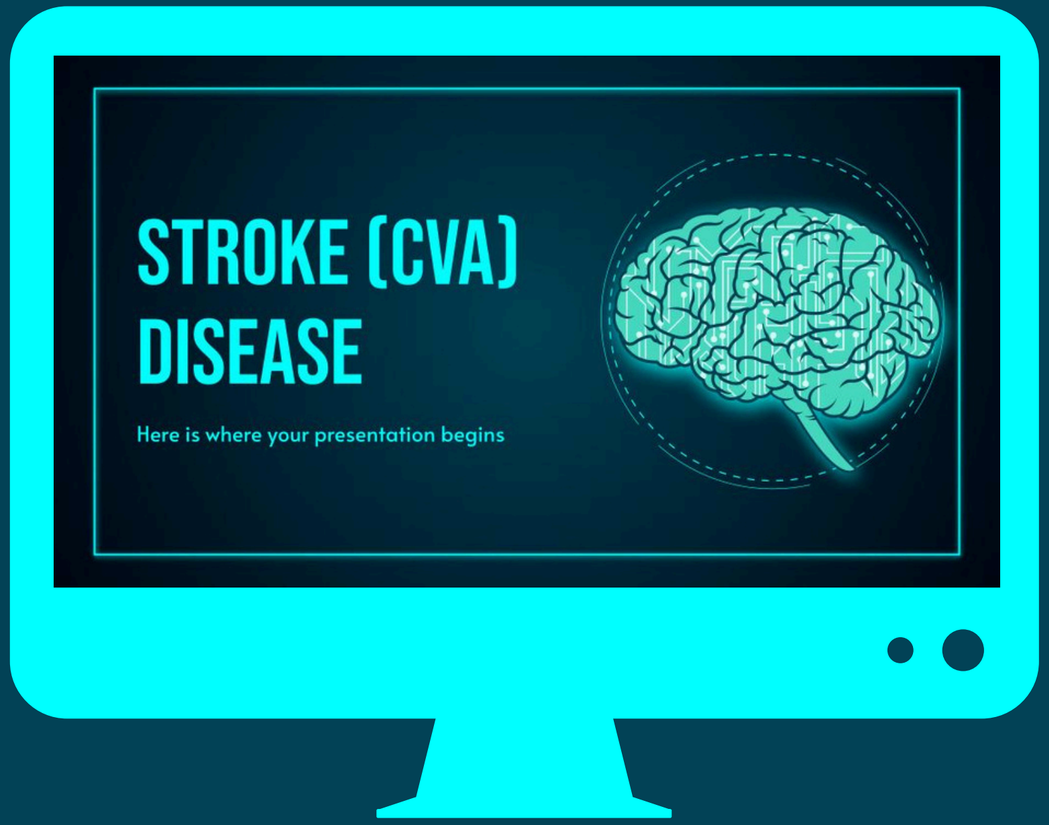
MODEL SELECTION

Deep Learning (MRI Images)

- Custom CNN (Convolutional Neural Network)
  - Designed to learn features directly from brain MRI scans.
  - Achieved good performance without needing transfer learning.
  - Chosen as the final DL model due to accuracy and simplicity.

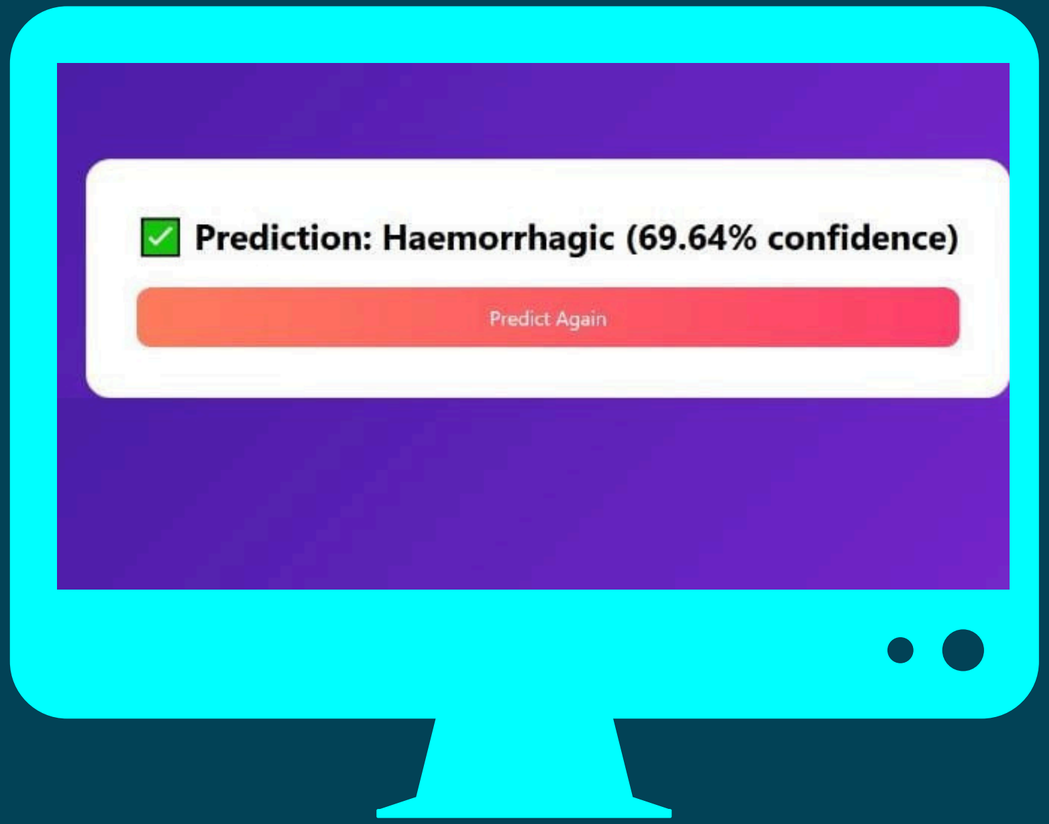
DEPLOYMENT

- Built a web app for doctors & patients.
- User uploads MRI scan or enters patient details.
- System predicts risk of stroke or type of stroke.



SAMPLE OUTPUT

- After uploading the image of MRI
- As we see the patient has stroke and its type is Heamorrhagic



CONCLUSION

- Stroke is highly imbalanced in the dataset → very few stroke cases compared to non-stroke.
- Age, hypertension, and glucose level are strong risk indicators.
- No single feature alone is enough → prediction requires combining multiple features.
- Machine learning models can help identify at-risk patients early and support preventive care.

THANKS!

DO YOU HAVE ANY QUESTIONS ?