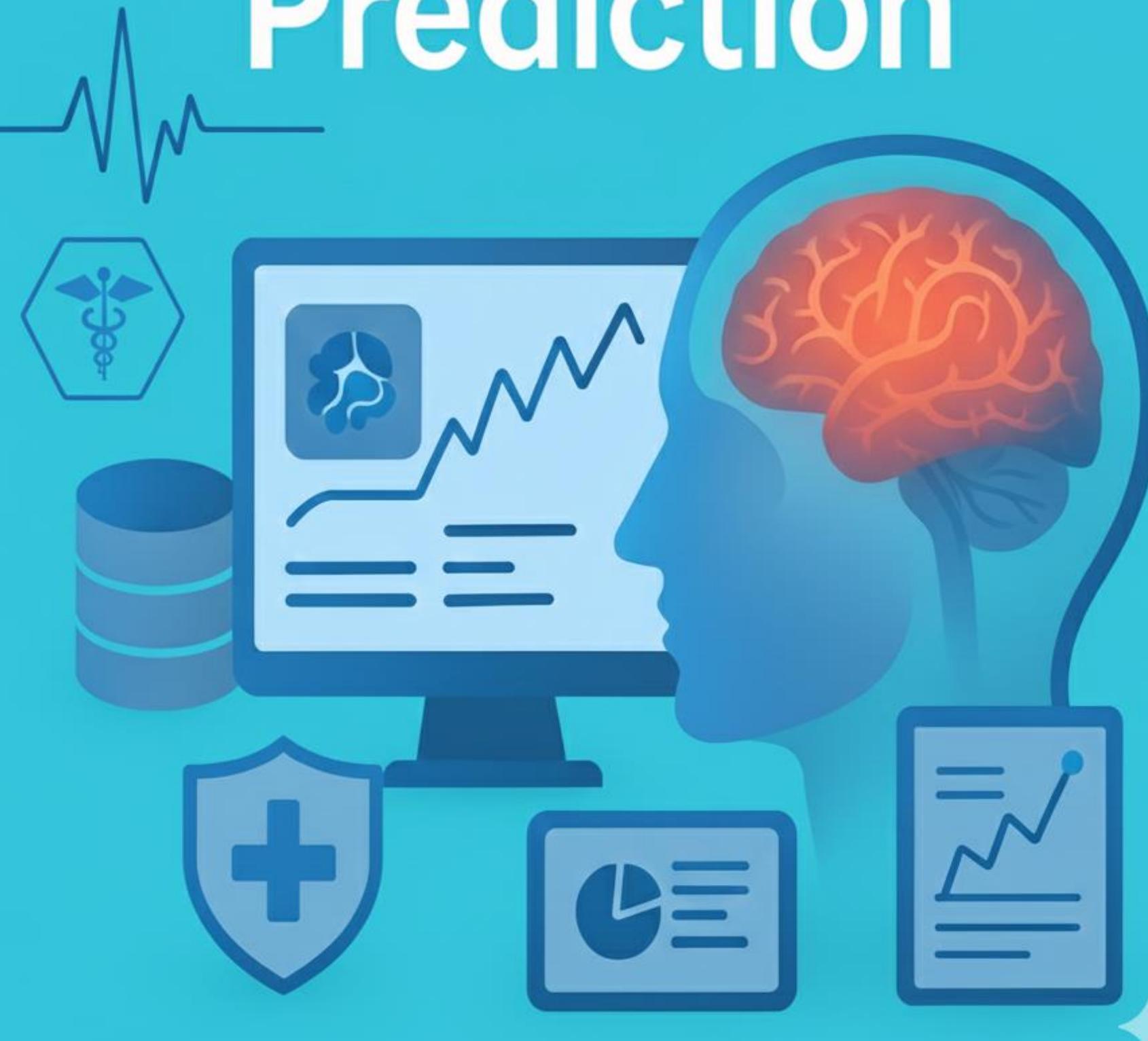


Stroke Risk Prediction



Stroke Risk Prediction Using Machine Learning



We help doctors provide the best possible patient care.

Team



Stroke Risk Prediction



Eng./ Ahmed Ghanem
Operation Team Leader at
Petrobel(Zohr Gas Field)



Captain Eng. /Mohamed Nasr
Computer Vision Engineer,
Electro Optics Engineer .



Doaa Gad-Allah
Data Analyst Lecturer
at Arab Academy for Science,
Technology and Maritime Transport



Tariq Elnaggar
Data Scientist

Ahmed Waled
Production Engineer



Ahmed Walid
Manufacturing Engineer



Real-World Scenario

- A 58-year-old man walked into a clinic with mild dizziness and slight numbness in his left hand.
- He seemed stable, his speech was normal, and his vital signs showed nothing alarming.
- He was reassured and sent home.
- 48 hours later, he returned to the emergency department — this time unable to speak or move his left side.
- He had suffered a major stroke. stroke happens.
- If his risk had been detected early, immediate preventive action could have changed his entire outcome.
- This is why early stroke prediction matters.
- Our project brings AI-driven risk detection to support clinicians — before the stroke happens.



CONTENTS



01

Overview

04

Model Workflow

07

Predictive Model
Performance

10

Real-World Application &
Impact

13

Future Enhancements

02

Project Goal

05

**Dataset Description,
CLEANING AND
PREPROCESSING**

08

Model Explainability

11

Cost

14

Conclusion

03

OBJECTIVES

06

Key Insights from Data

09

Deployment Strategy

12

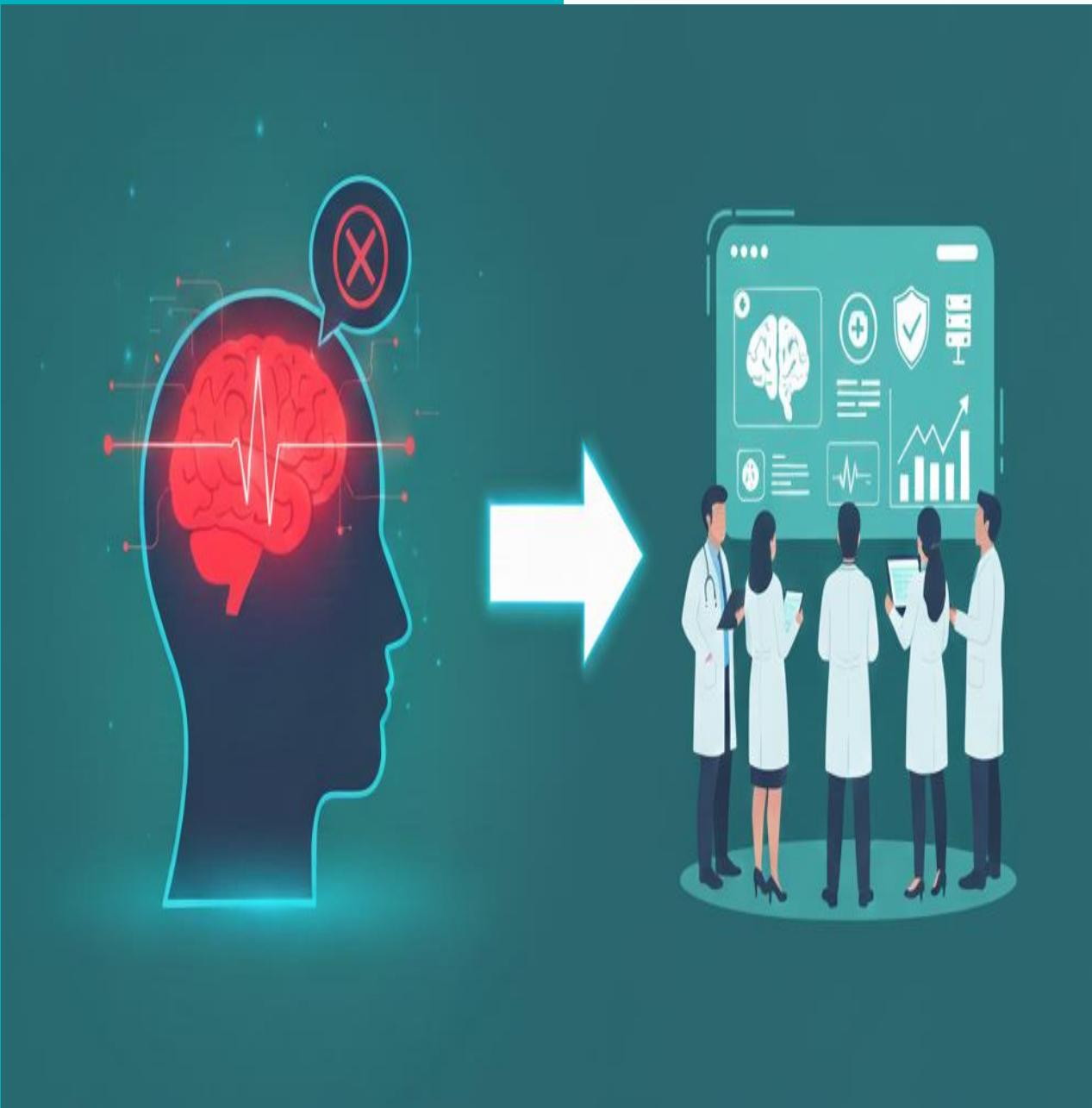
Business Model

01

Overview

- Stroke is one of the leading causes of death and long-term disability worldwide.
- Early identification of individuals at risk can significantly reduce complications through preventive care.
- This project aims to develop a machine learning model that predicts the likelihood of stroke risk based on various medical and lifestyle factors.
- The model utilizes both binary and continuous features (e.g., symptoms, age) to:
 - Generate a stroke risk percentage
 - Provide a clear risk classification → “At Risk” vs. “Not at Risk”





02 Project Goal

Objective: Build a predictive model that accurately identifies individuals at high risk of stroke.

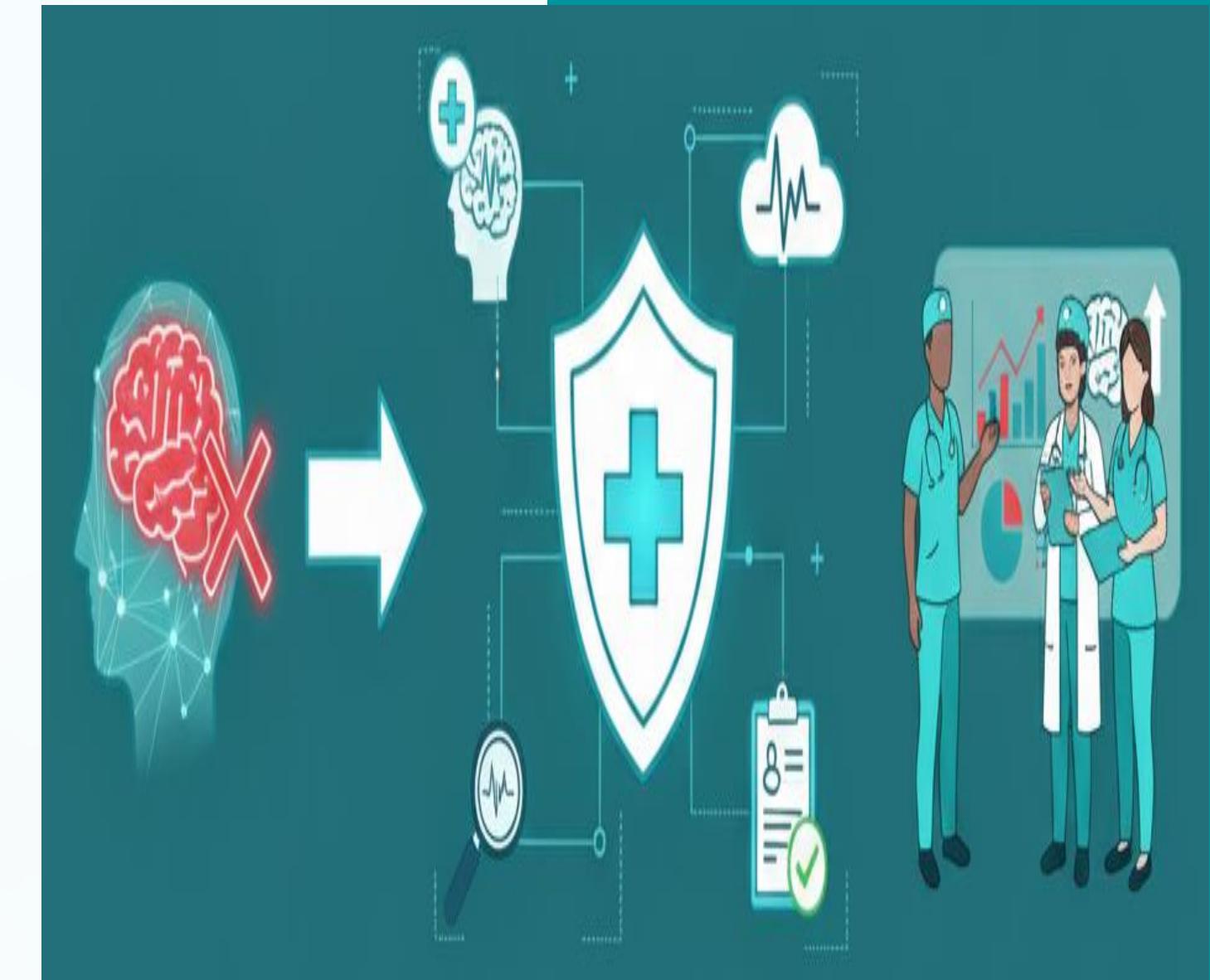
Expected Impact:

- Early intervention → reduces emergency admissions.
- Data-driven support → assists doctors in faster decision-making.
- Integration-ready → fits into electronic medical record (EMR) systems.



03 OBJECTIVES

- ✓ Build a clean, labeled dataset containing clinical and lifestyle variables relevant to stroke risk.
- ✓ Preprocess binary and numerical features for ML model input.
- ✓ Train and evaluate multiple algorithms (Logistic Regression, Random Forest, XGBoost).
- ✓ Optimize model performance through hyperparameter tuning.
- ✓ Provide visual explanations using feature importance plots.
- ✓ Develop a predictive dashboard or web interface for real-time risk estimation.



SMART Project Objectives

 S - Specific

Develop a predictive ML model to assess stroke risk using patient metrics such as age and medical history.

 M - Measurable

Target: 85% accuracy, 0.90 ROC-AUC, strong F1-Score.

 A - Achievable

Use available healthcare datasets, standard workflows, and proven ML algorithms.

 R - Relevant

Support healthcare with a data-driven tool for early stroke detection and improved clinical decision-making.

 T - Time-Bound

Complete the project within 6 weeks.

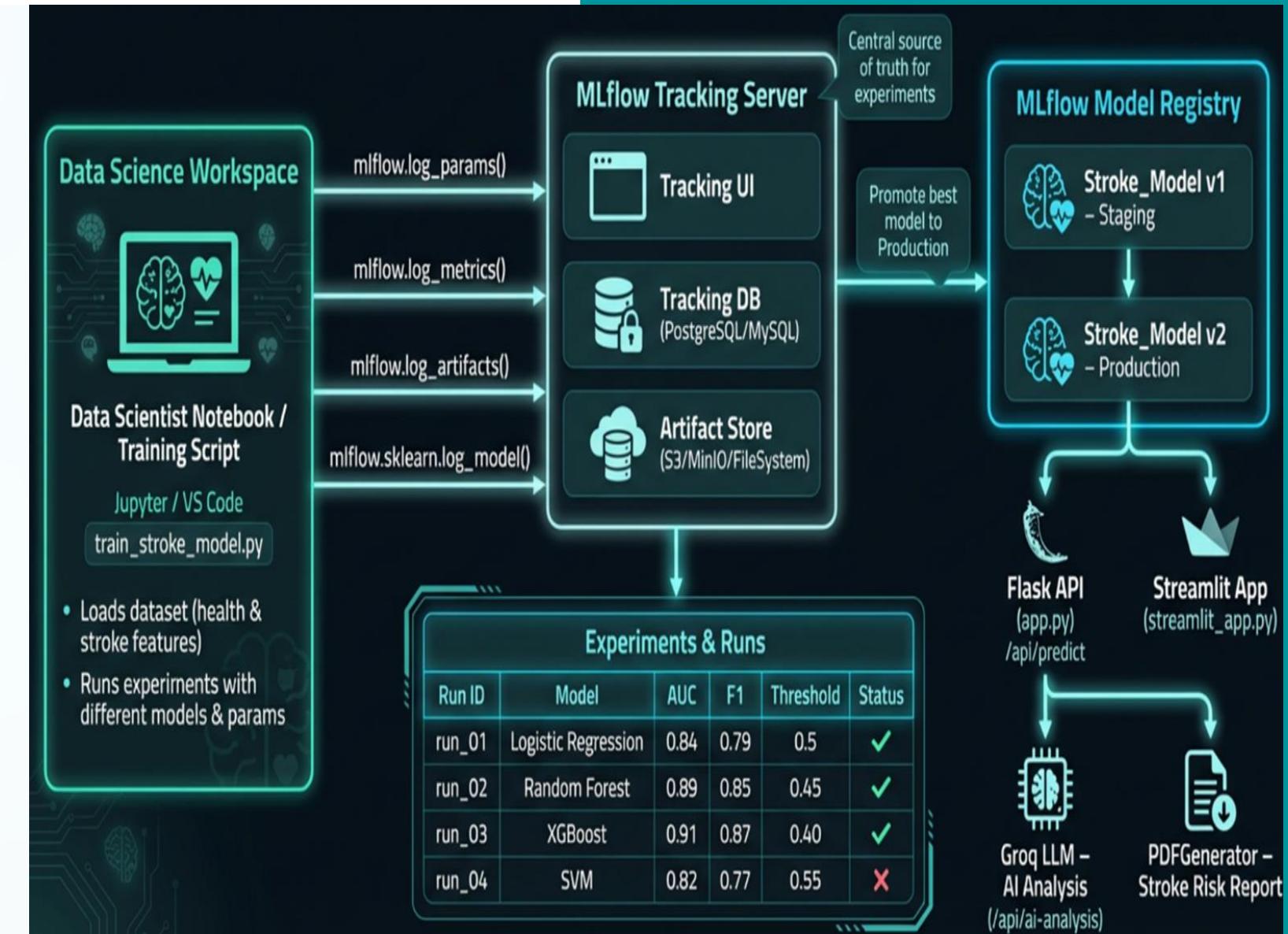


04

Model Workflow

Pipeline Summary:

- ✓ Data Cleaning.
- ✓ Exploratory Data Analysis (EDA)
- ✓ Model Training (Random Forest, Logistic Regression, XGBoost)
- ✓ Model Evaluation (Accuracy, ROC-AUC)
- ✓ Feature Importance And SHAP Interpretation
- ✓ Deployment via Flask as a Dashboard
- ✓ It analyzes the patient's condition using artificial intelligence.



05

Data Description, CLEANING AND PREPROCESSING

Source: Healthcare dataset with 70,000 records & 18 clinical + lifestyle features.

Key Variables:

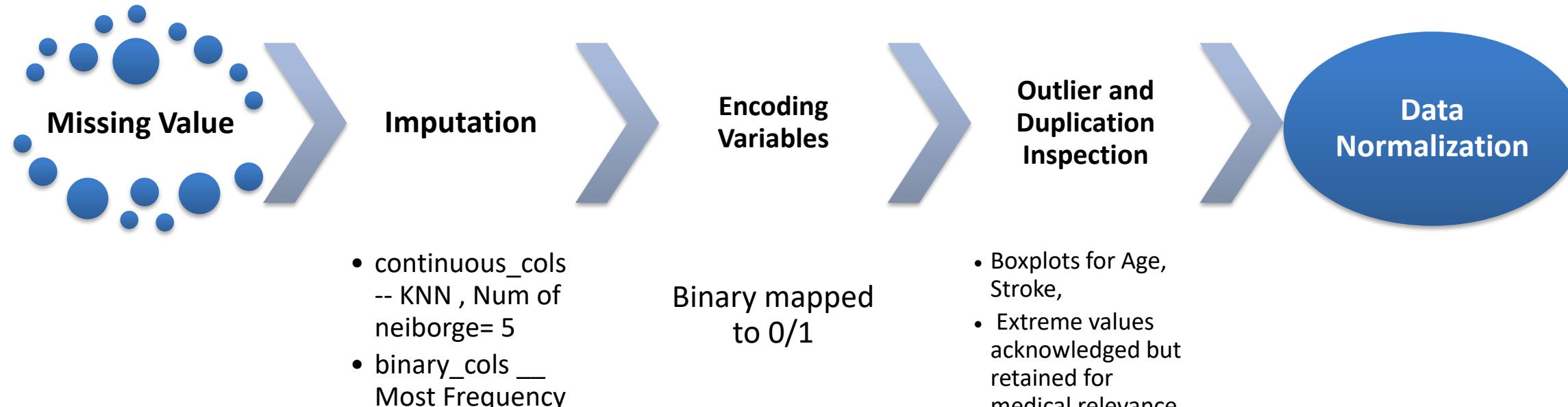
- **Numeric:** Age, Stroke Risk (%)
- **Binary Symptoms:** Chest Pain, Dizziness, Fatigue, Sleep Apnea, Anxiety
- **Target:** At Risk (Binary: 1 = At Risk, 0 = Not At Risk)



Dataset

	Feature	Dtype
0	Chest Pain	float64
1	Shortness of Breath	float64
2	Irregular Heartbeat	float64
3	Fatigue & Weakness	float64
4	Dizziness	float64
5	Swelling (Edema)	float64
6	Pain in Neck/Jaw/Shoulder/Back	float64
7	Excessive Sweating	float64
8	Persistent Cough	float64
9	Nausea/Vomiting	float64
10	High Blood Pressure	float64
11	Chest Discomfort (Activity)	float64
12	Cold Hands/Feet	float64
13	Snoring/Sleep Apnea	int64
14	Anxiety/Feeling of Doom	int64
15	Age	int64
16	Stroke Risk (%)	float64
17	At Risk (Binary)	int64

Dealing With Missing Values

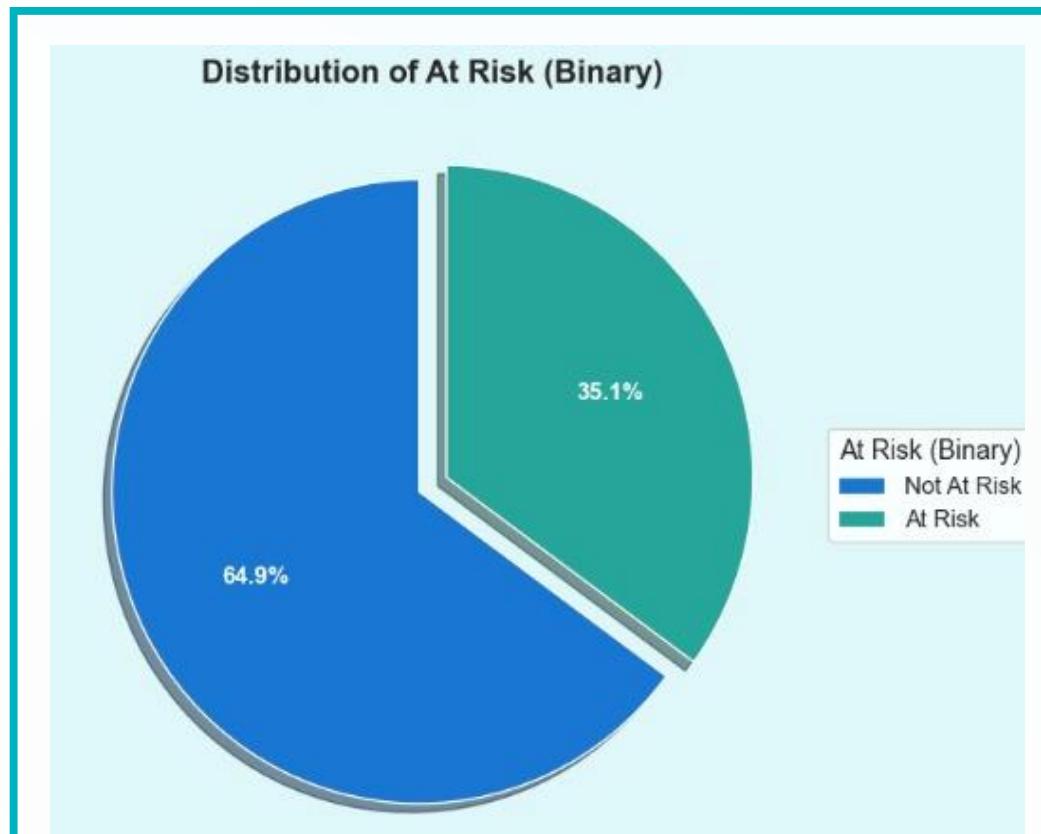


Removing Duplicates

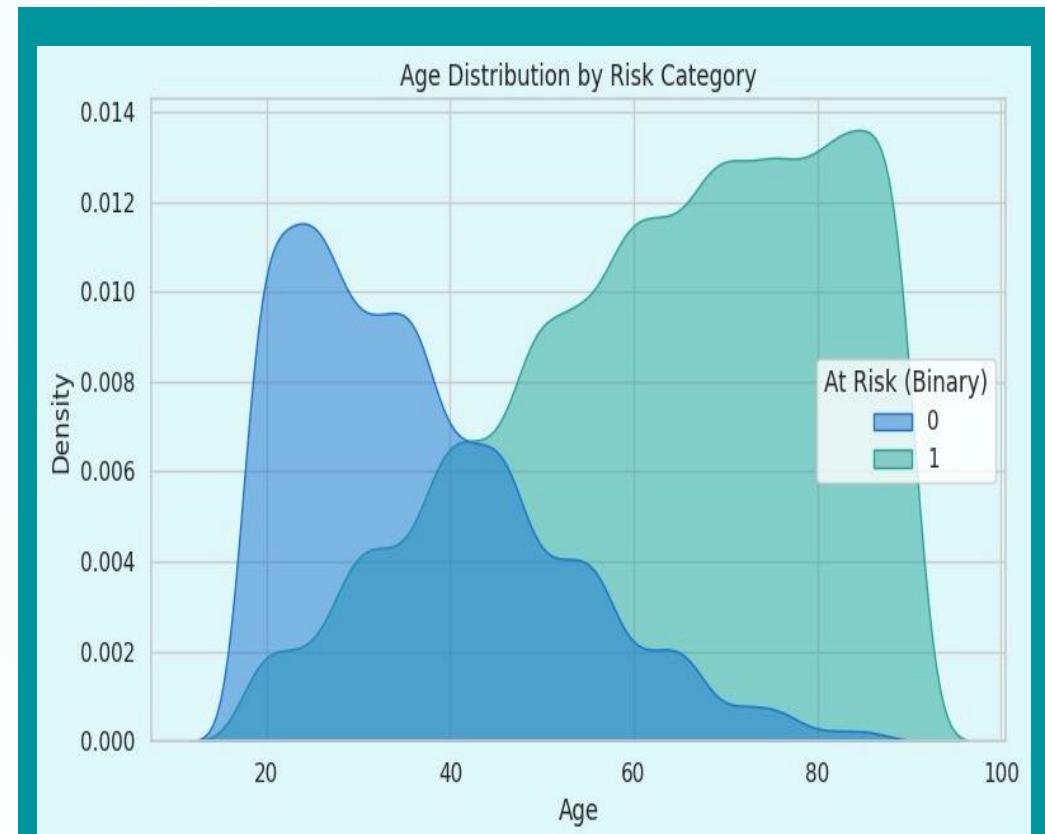
```
df.duplicated().sum()/len(df)
df.drop_duplicates(inplace=True)
df
```



Key Insights from Data



- Average age of patients: 54.1 years
- 35.1% labeled “At Risk” for stroke



“Stroke Risk (%)” and “Age” strongly correlated with actual stroke risk.



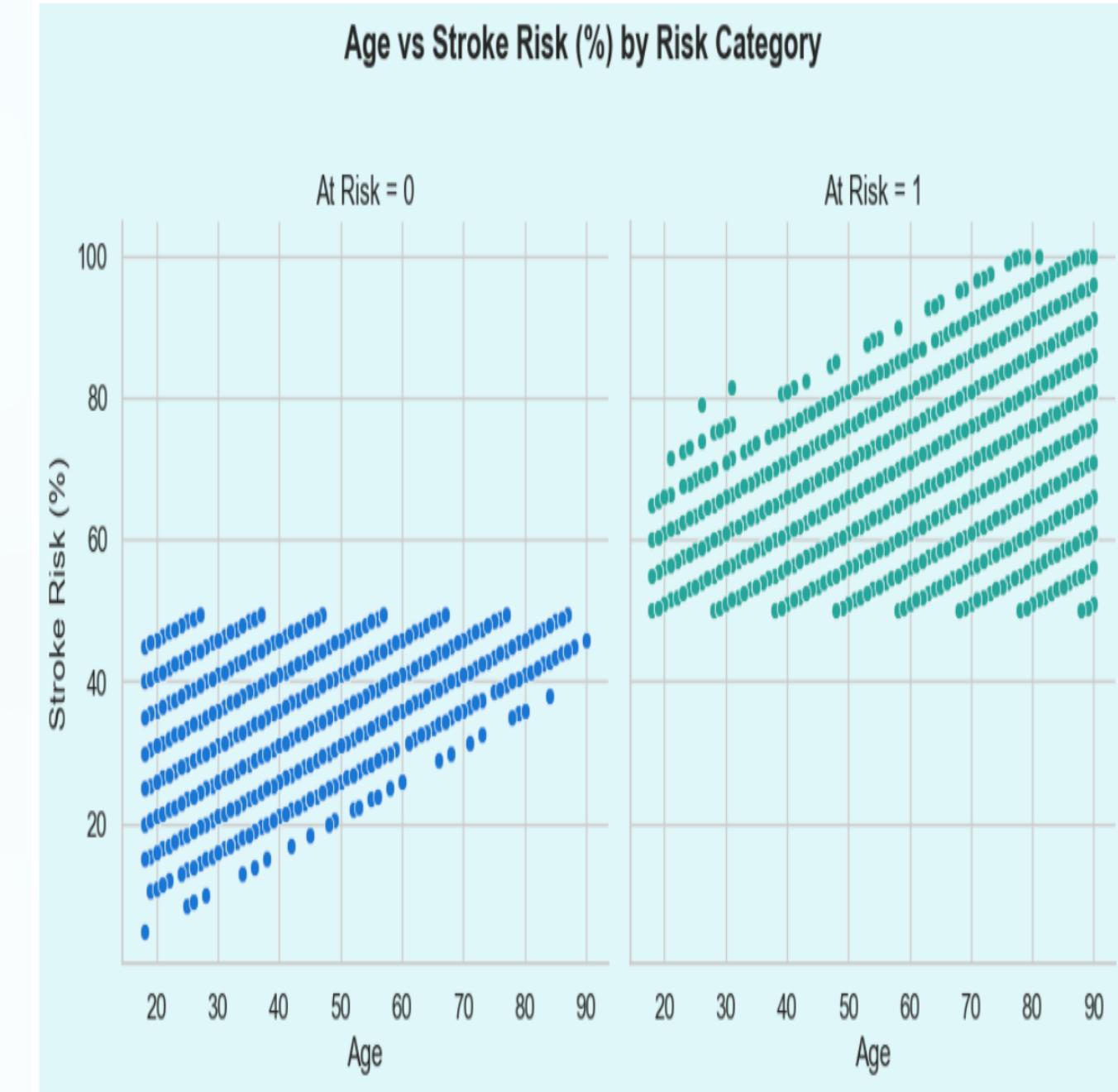
Most common symptoms: Excessive Sweating, Dizziness, Nausea, Chest Pain



Key Insights from Data

Age vs Stroke Risk

- Not At Risk (0): Stroke risk ranges from 10%-50%, with a slight increase with age and never exceeding 50%.
- At Risk (1): Risk is consistently higher (50%-100%) and clearly rises with age.
- Clear separation between groups:
 - Max risk in At Risk exceeds 100%, while Not At Risk stays below 50%.
 - Min risk in At Risk rarely drops below 50%.
 - Min risk in Not At Risk remains under 15%.
- Conclusion: Age combined with Stroke Risk shows strong class separation, supporting label quality and keeping Age as a continuous feature.



Predictive Model Performance

Models Evaluated: Logistic Regression, Random Forest, XGBoost.

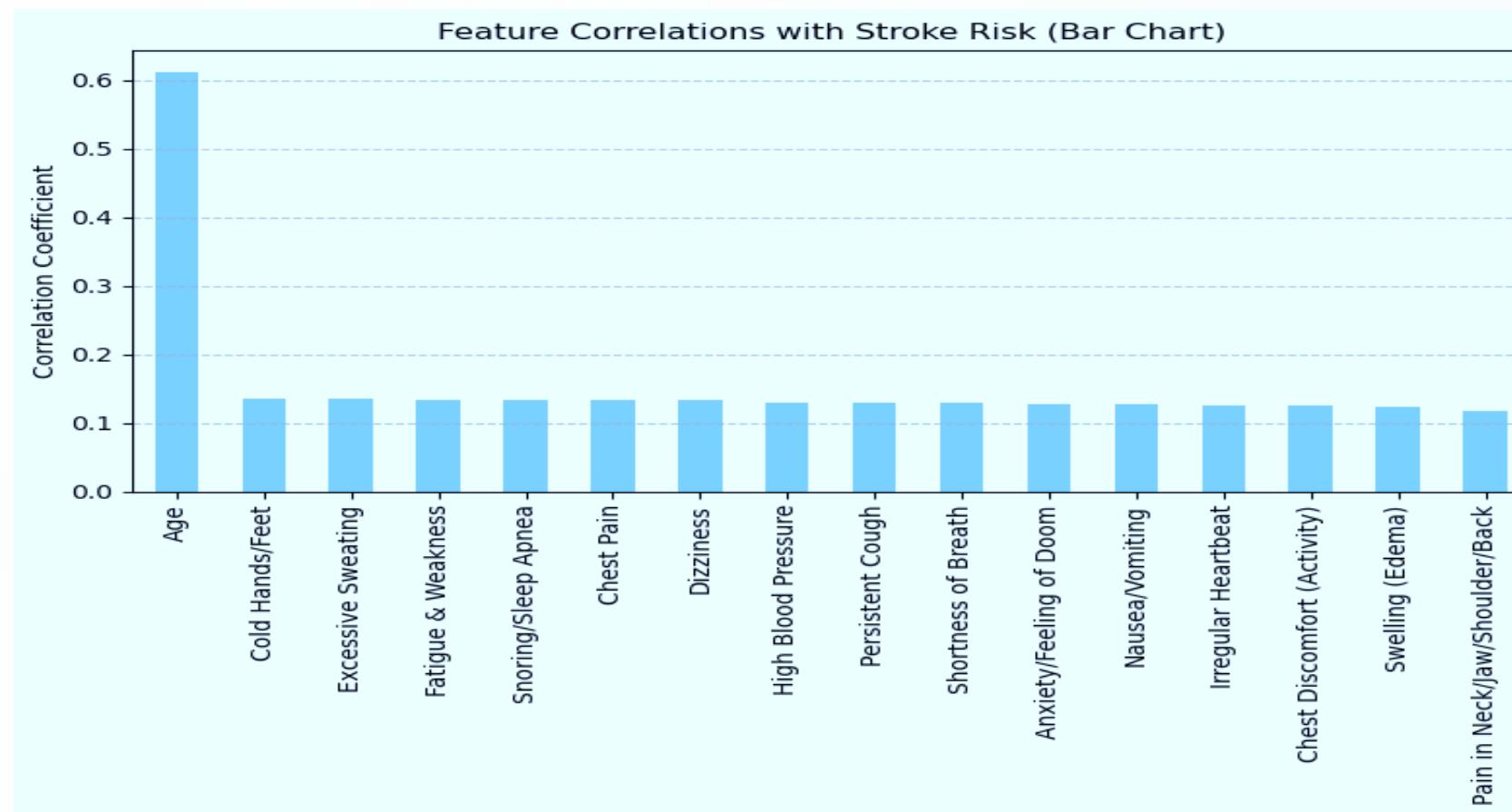
Result: Model can reliably predict patients at high stroke risk with near-perfect recall — crucial for healthcare triage.

	Model	Accuracy	Precision	Recall	F1 Score	AUC
0	Logistic Regression	0.998212	0.997697	0.999553	0.998624	0.999801
1	Decision Tree	0.882833	0.907188	0.912927	0.910049	0.870025
2	Random Forest	0.946031	0.953809	0.963534	0.958646	0.990025
3	XGBoost	0.994782	0.994509	0.997470	0.995987	0.999636



08

Model Explainability



Feature Importance (Top 5):

- ✓ Age
- ✓ Cold Hands/Feet
- ✓ Excessive Sweating
- ✓ Fatigue & Weakness
- ✓ Snoring/Sleep Apnea

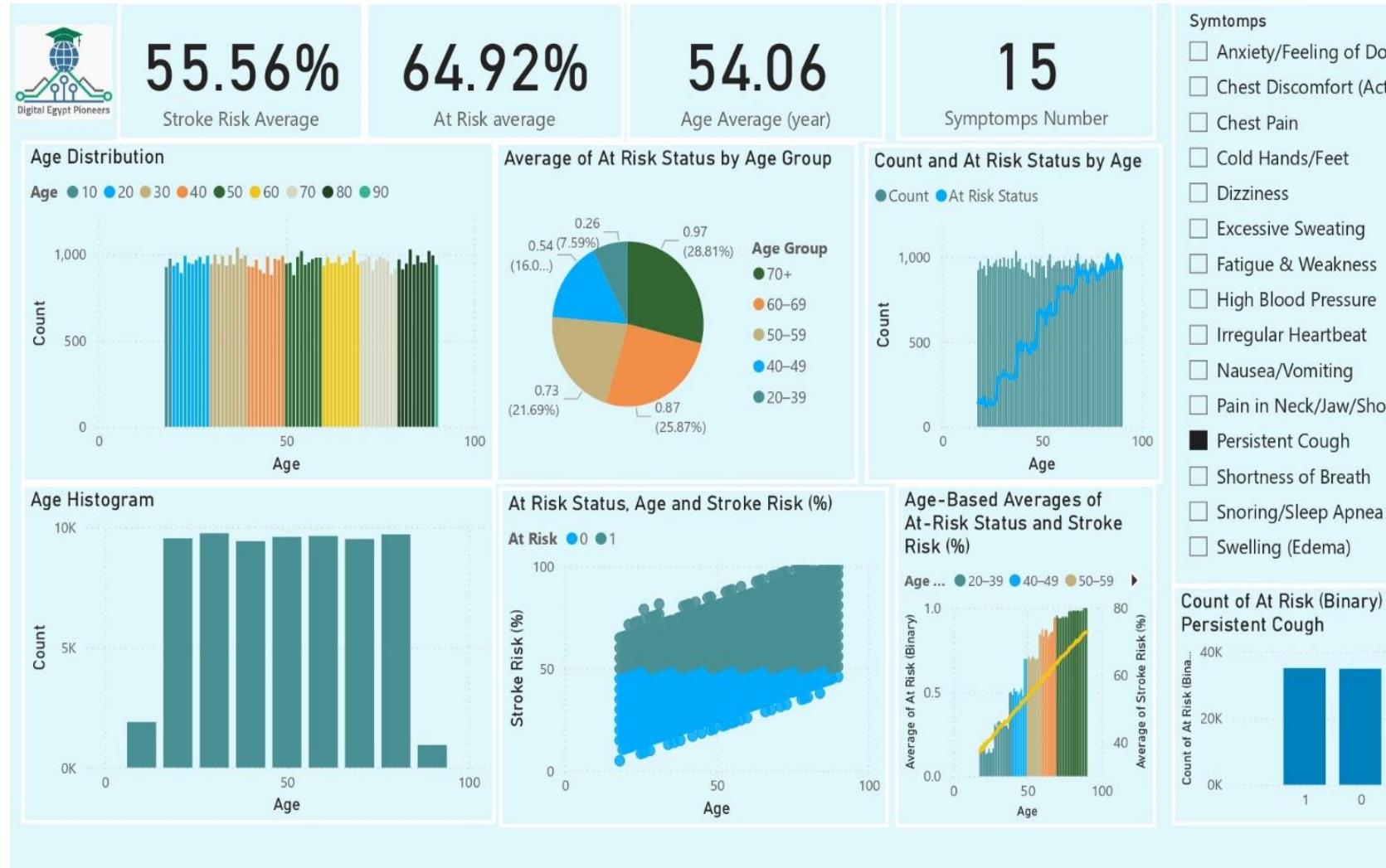
→ Clinicians can easily understand why a patient is classified as high-risk.



Deployment

USING POWER BI

Featuring



Model evaluation



09 Deployment Strategy

USER INTERFACE – UI/UX AND AI AGENT

The screenshot shows the mobile application interface for stroke risk prediction. At the top, there are language and theme settings (Arabic, Dark Mode). The main title is "تقييم خطر السكتة الدماغية" (Stroke Risk Assessment) with a subtitle "أدخل البيانات واحصل على التقييم فوراً". Below this is a form with fields for age (59), gender (Male), and symptoms. The symptoms section contains 12 items with binary answers (Yes or No). The results summary at the bottom indicates a high risk level (high risk score: 95.15%, high risk level: high).

البيانات المدخلة:

- الاسم: السيد سيد جدا
- العمر: 59
- الجنس: ذكر

الأعراض:

- اضطراب نبضات القلب: لا
- صعوب في التنفس: لا
- آلام في الصدر: نعم
- دوخة: لا
- التعب و ضغط مرتفع: نعم
- آلام في الرقبة / الكتف / الظهر: لا
- ارتجاع في الصدر (النساط): لا
- دوخة: نعم
- تورم (وذمة): لا
- سعال المستمر: لا
- السعال الزائد: لا
- ضغط دم مرتفع: لا
- الغثيان/القيء: لا
- أرجاع في الصدر: نعم
- التشخير/توقف التنفس أثناء النوم: لا
- القلق / الشعور بالهلالك: لا
- آلام في الباردة / القدمين: لا
- النوم: نعم

تحليل الحالة (بالذكاء الاصطناعي):

تحاليل طبية متعلقة بخطورة السكتة الدماغية:

- مستوى الخطير: high
- الاحتمالي: 95.15%

تحاليل طبية متعلقة بخطورة السكتة الدماغية:

- النتيجة: حسنة.
- خطورة السكتة الدماغية التي تم الحصول عليها تشير إلى أن خطر حدوث سكتة دماغية لديك هو عال، حيث بلغت نسبة الخطير 95.63%. هذا يعني أن هناك عوامل معينة قد تزيد من خطر حدوث سكتة دماغية لديك، مثل الآلام في الصدر، التعب والضغط المرتفع، دوخة، السعال المستمر، ضغط الدم المرتفع، الأيدي الباردة أو القدمين، والتشخير أو توقف التنفس أثناء النوم. من الجدير بالذكر أن هذه النتيجة مبدئية ولا تعتبر تشخيصاً نهائياً، ويجب عليك زيارة الطبيب أو الطوارئ لتحديد الخطير الحقيقي ووضع خطة علاجية مناسبة.

PDF report

The PDF report header includes the logo and name of the project: "Road to Egypt's digital transformation". The report title is "报 告 书" (Report). The content includes the user information and the detailed symptom analysis from the mobile app, followed by a summary of the AI-generated diagnosis.

بيانات المستخدم:

- الاسم: depi سيس
- العمر: 55
- الجنس: أنثى

الأعراض:

- آلام في الصدر: نعم
- صعوب في التنفس: لا
- اضطراب نبضات القلب: لا
- ضغط دم مرتفع: نعم
- ارتجاع في الصدر (النساط): لا
- دوخة: نعم
- تورم (وذمة): لا
- آلام في الرقبة / الكتف / الظهر: لا
- السعال المستمر: نعم
- السعال الزائد: لا
- ضغط دم مرتفع: نعم
- الغثيان/القيء: لا
- أرجاع في الصدر: نعم
- التشخير/توقف التنفس أثناء النوم: نعم
- القلق / الشعور بالهلالك: لا
- آلام في الباردة / القدمين: لا
- النوم: نعم

تحليل الحالة باستخدام الذكاء الاصطناعي:

تحاليل طبية متعلقة بخطورة السكتة الدماغية التي تم الحصول عليها تشير إلى أن خطر حدوث سكتة دماغية لديك هو عال، حيث بلغت نسبة الخطير 95.63%. هذا يعني أن هناك عوامل معينة قد تزيد من خطر حدوث سكتة دماغية لديك، مثل الآلام في الصدر، التعب والضغط المرتفع، دوخة، السعال المستمر، ضغط الدم المرتفع، الأيدي الباردة أو القدمين، والتشخير أو توقف التنفس أثناء النوم. من الجدير بالذكر أن هذه النتيجة مبدئية ولا تعتبر تشخيصاً نهائياً، ويجب عليك زيارة الطبيب أو الطوارئ لتحديد الخطير الحقيقي ووضع خطة علاجية مناسبة.



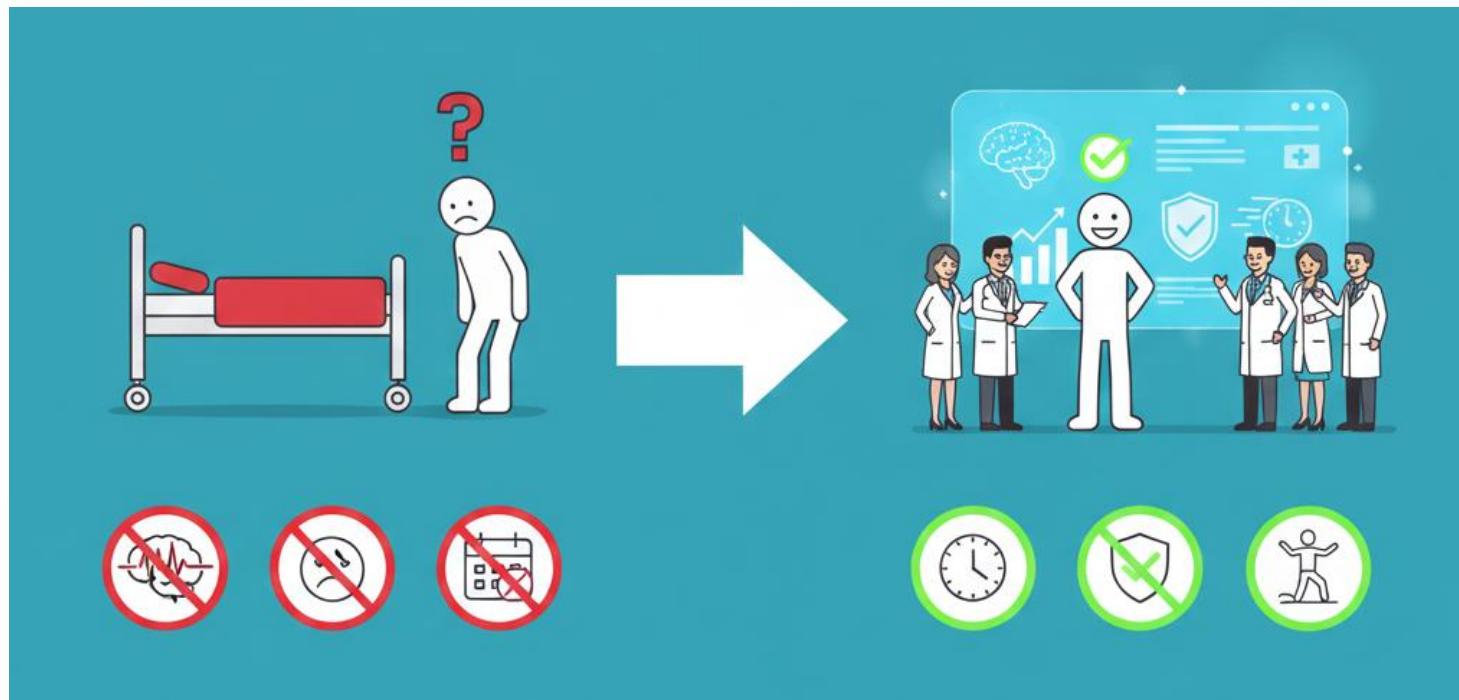
10 Real-World Application & Impact

How It Helps Healthcare Professionals:

- Identifies at-risk patients before symptoms worsen.
- Prioritizes urgent cases for preventive care.
- Assists in hospital resource planning and early treatment decisions.



Impact on Patient

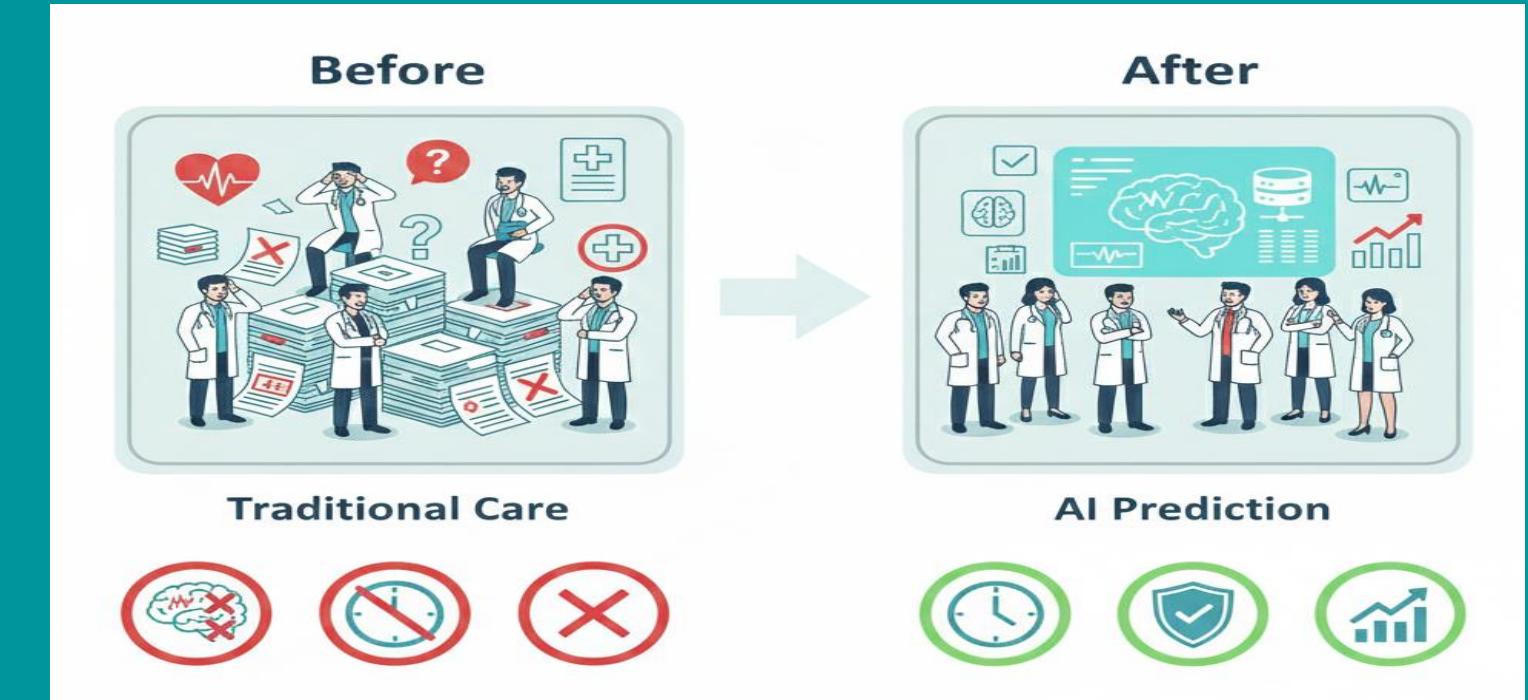


Delayed diagnosis of neurological disorders leads to significant costs and preventable complications. Addressing these issues is crucial for improving patient outcomes and optimizing healthcare resources.

Medical cost per person Avrg Range 250000 - 500000

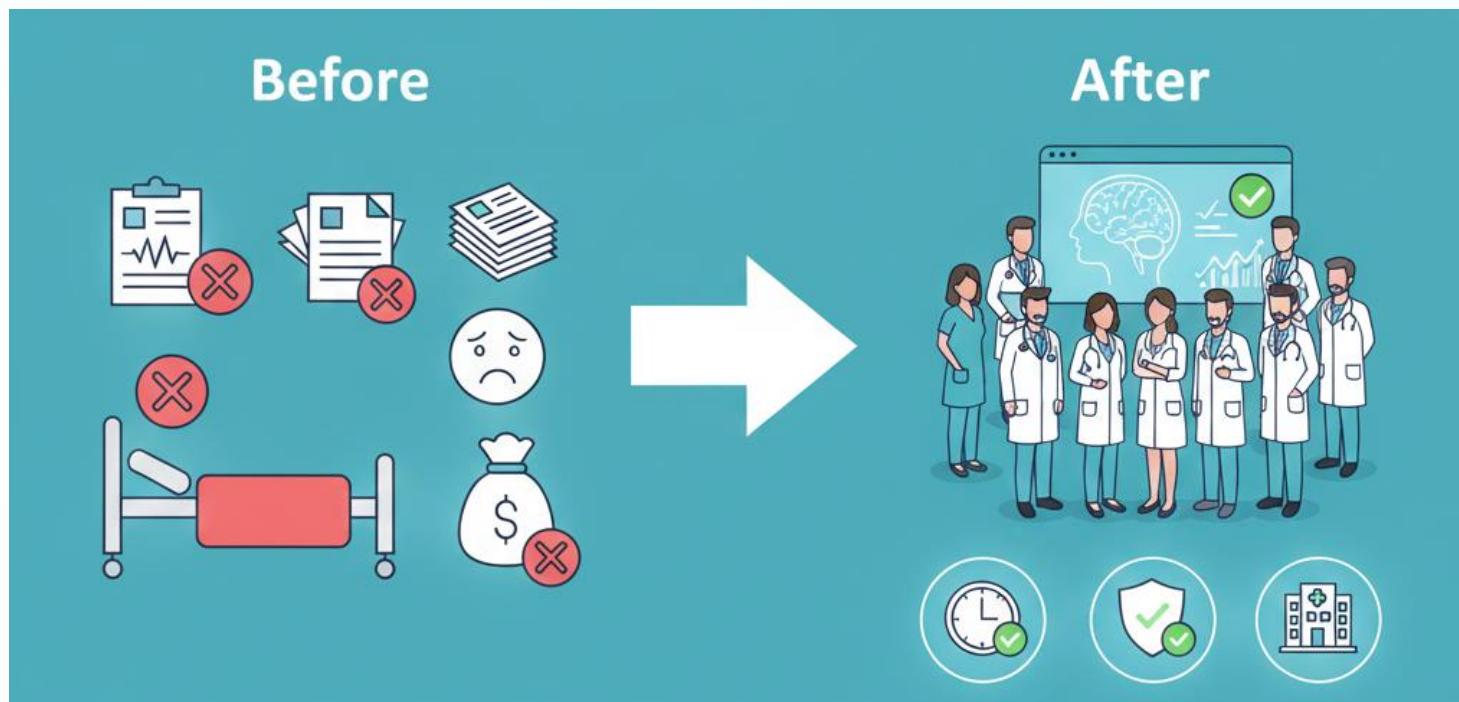


Impact on Doctors



provides valuable insights that assist physicians in interpreting Labs results. By analyzing complex data patterns, it highlights key features and potential issues, allowing doctors to make faster, more informed decisions. This support reduces the cognitive load on healthcare providers.

Impact on Hospitals



improving turnaround times for neurological diagnoses. By automating analysis.

hospitals can allocate resources more effectively, ensuring that specialists focus on complex cases while routine evaluations are expedited.

This leads to enhanced patient care and optimized operational efficiency.



Impact on Egypt



Our AI model presents a transformative opportunity for investors, aiming to revolutionize healthcare delivery in Egypt. By aligning with the national strategy for digital health, this solution enhances diagnostic capabilities, reduces costs, and ensures a sustainable return on investment through improved patient outcomes and operational efficiencies.

11 Cost



Category	Budget Allocation	Strategic Impact
R&D & Data Acquisition	35 %	Core technology and IP development
Cloud & Cybersecurity	15 %	Ensures scalability and trust
Talent & Collaboration	30 %	Drives innovation and clinical credibility
Regulatory & Compliance	10 %	Enables market access and certification
Marketing & GTM	10 %	Accelerates adoption and revenue generation



12 BUSINESS MODEL

Stroke Risk Prediction



KEY PARTNERS



- Ministry of Health and Population (MoHP)
- Public & private hospitals (Kasr Al-Ainy, Dar El Fouad, As-Salam International)
- Cloud providers with local data centers (Vodafone Egypt, Etisalat Misr Cloud)
- Universities and research centers (Cairo University, Ain Shams University)
- Pharmaceutical and insurance partners (Misr Insurance, AXA Egypt)

KEY ACTIVITIES



- Data collection & preprocessing from Egyptian hospitals
- Localized feature engineering & AI model training
- Arabic-language dashboard & visualization development
- Partnerships with Egyptian healthcare regulators for validation

KEY RESOURCES



- Egyptian patient health datasets (MoHP, hospitals)
- Localized machine learning models
- Cloud infrastructure within Egypt
- Data scientists & local medical professionals

VALUE PROPOSITIONS



- AI-driven stroke prediction tailored to Egyptian demographics
- Supports early detection & preventive care across Egypt's governorates
- Integrates with local hospital systems (HIS/EHR)
- Complies with Egyptian data privacy & healthcare regulations

CUSTOMER RELATIONSHIP



- Strategic partnerships with hospitals & health authorities
- Training sessions for Egyptian clinicians
- Arabic-language customer support & user community
- Awareness campaigns for preventive healthcare

CHANNELS



- Direct sales to hospitals & medical centers
- Partnerships with MoHP and university hospitals
- Medical conferences & exhibitions (e.g., Cairo Health Expo)
- Online professional networks

CUSTOMER SEGMENTS



- Egyptian hospitals & clinics (public & private)
- Insurance providers & healthcare ministries
- Medical universities & research organizations
- Doctors, health analysts, and early diagnosis programs

COST STRUCTURE



- AI model R&D and Egyptian data acquisition
- Cloud hosting & cybersecurity compliance
- Personnel, medical expert collaboration, & Arabic localization
- Marketing, regulatory, and operational costs in Egypt

REVENUE STREAM

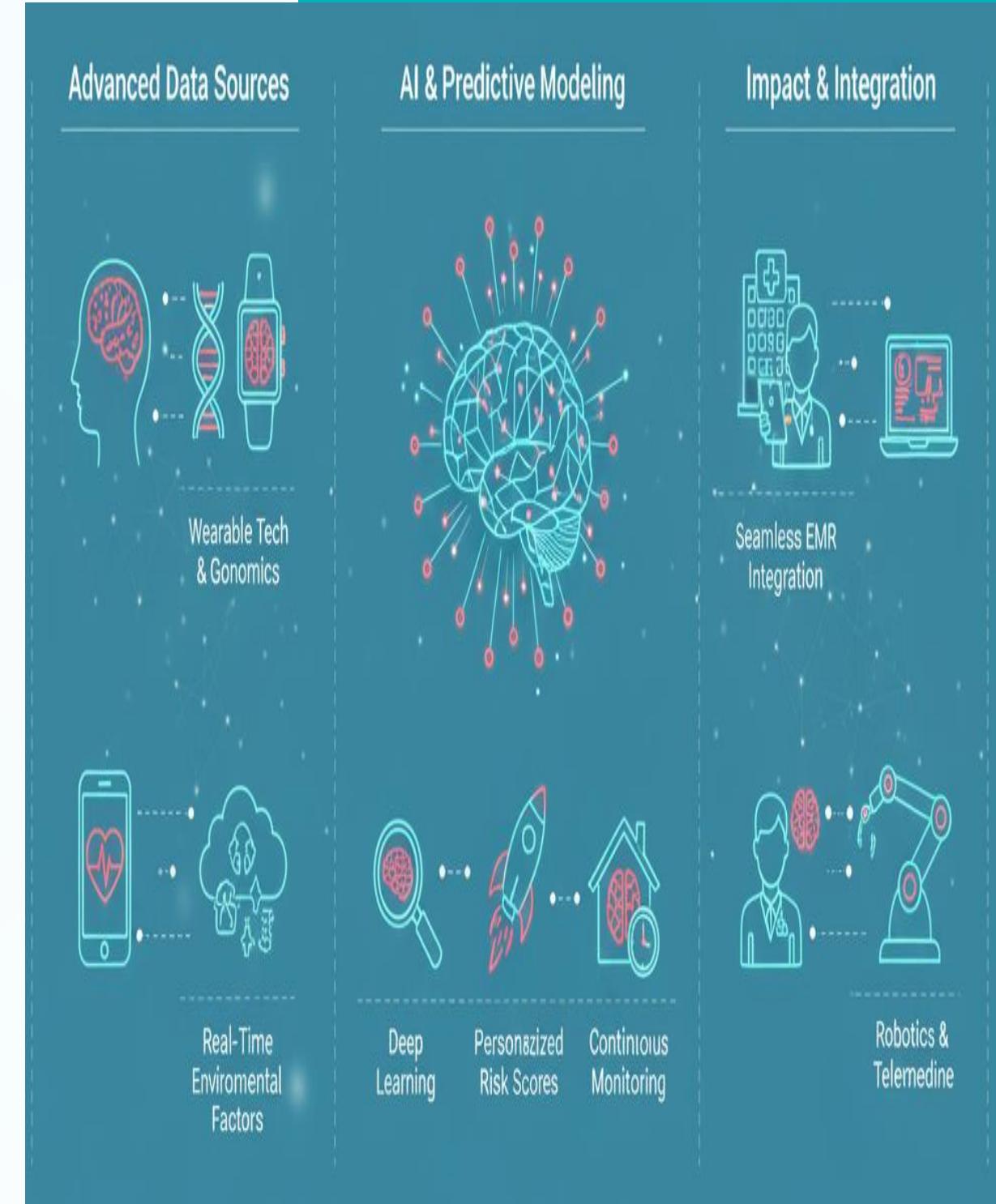


- Subscription SaaS model for Egyptian hospitals
- Government contracts for national health programs
- Custom AI analytics solutions for insurers
- Training & research collaboration fees

13 Future Enhancements

Next Steps:

- Integrate with EHR systems for real-time use.
- Add more clinical variables (blood sugar, cholesterol, BMI, Smoke statuses, marital Statuses etc.).
- Deploy a Streamlit web dashboard for clinicians.
- Use SHAP visualizations for transparent model decisions.
- Continuous improvement as more hospital data becomes available.
- The possibility of adding a service called "My Health" OR (صحتي) to the Egypt Digital Platform, which will contain all data and medical history and will include predictions of all health problems.



14

Conclusion

- Built an interpretable ML model predicting stroke risk with >95% accuracy.
- Model supports preventive healthcare and data-driven decision-making.
- Potential to reduce stroke incidence through early alerts and clinical adoption.



THANK
YOU

Together for a Stroke-Free Future





Any Questions?

