Stroke Prediction using Machine Learning

John Smith[1], Jane Doe[2*]
[1]Department, Faculty, Institution, City, Country
[2]Department, Faculty, Institution, City, Country
[1]john.smith@email.com, [2*]jane.doe@email.com
[*]Corresponding author

Abstract

Stroke is a leading cause of long-term disability and death worldwide, with risks increasing with age and factors like hypertension and diabetes. The goal of this research is to develop an early and accurate prediction model for stroke to enable effective preventive healthcare interventions. This study employs Random Forest and Support Vector Machine (SVM) algorithms. Due to dataset imbalance, the Synthetic Minority Oversampling Technique (SMOTE) is applied to enhance minority data representation. Models are optimized via hyperparameter tuning using Bayesian Optimization. Model evaluation is conducted using accuracy, precision, recall, and F1-score metrics, with cross-validation ensuring reliability on unseen data. Experimental results show that the SVM algorithm with SMOTE and Bayesian optimization achieves the highest accuracy of 0.726404. These findings indicate that optimized machine learning models can significantly contribute to early stroke prediction and support decision-making in preventive healthcare systems.

Keywords: *keyword1; keyword2; keyword3; keyword4; keyword5*

1. Introduction

1.1 Background

1.2 Literature Review

1.3 Research Rationale

1.4 Research Questions and Objectives

2. Research Methods

2.1 Formulas

$$E = mc^2 \tag{1}$$

3. Results and Discussion

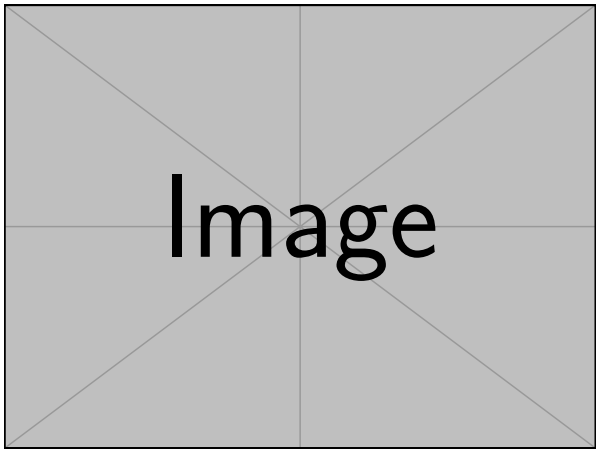

Figure 2: A single-column figure example.

4. Conclusion

Acknowledgements

References

[1]

[2]

[3]

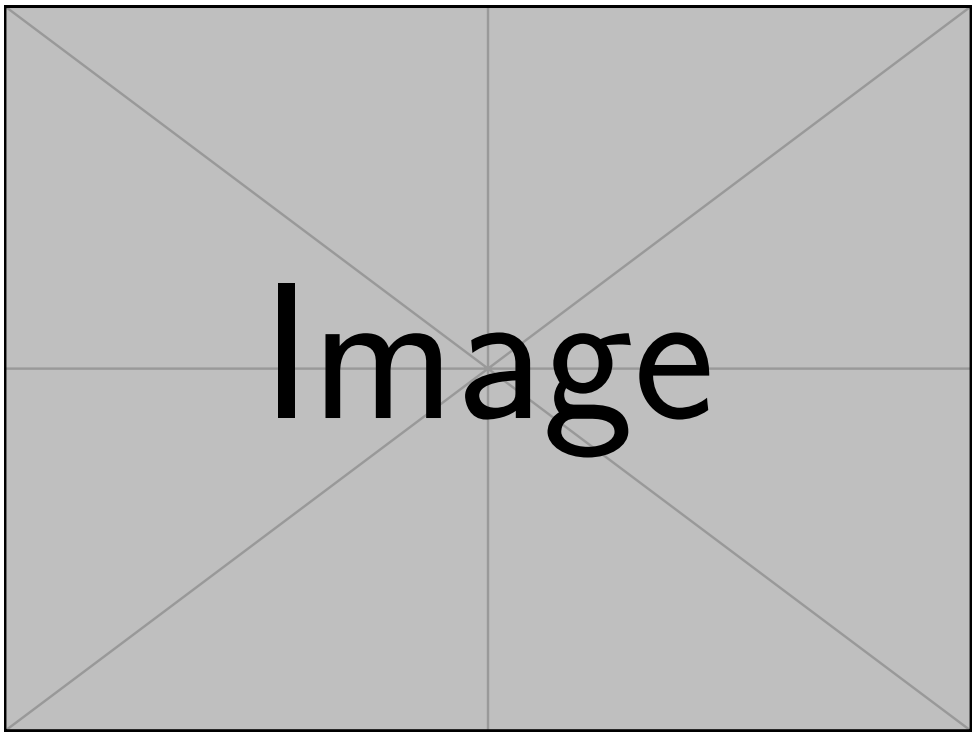

Figure 1: Description of your figure. Make sure to reference it in text using Fig. 1.

Table 1: Dataset characteristics for stroke prediction

Characteristic	Value	Percentage
Total Samples	5110	100%
Stroke Cases	249	4.87%
Non-stroke Cases	4861	95.13%