iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84320

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

Heart Stroke Prediction

G. Manoj Kumar[1], Sheik Gafur[2]
[1]*Assistant Professor,* [2] *MCA Final Semester, Master of Computer Applications, Sanketika Vidya Parishad Engineering College, Vishakhapatnam, Andhra Pradesh, India*

Abstract: ***The Heart Stroke Prediction System is a machine learning-based application developed to predict the risk of heart stroke using patient health information. The system helps users identify potential stroke risks at an early stage through intelligent data analysis. It considers important health parameters such as age, gender, hypertension, heart disease, BMI, smoking status, glucose level, and work type. The collected data is preprocessed and analyzed to improve prediction accuracy. A Logistic Regression algorithm is used to classify whether a person is likely to experience a stroke or not. The model is trained using historical healthcare data and evaluated with testing datasets. The application is developed using Python and Flask to provide a simple and user-friendly web interface. Users can enter their health details and receive instant prediction results. The system also includes a BMI calculator and basic health guidance to improve awareness. It supports early diagnosis and encourages preventive healthcare measures. The proposed system reduces manual assessment time and provides quick decision support. It offers an efficient, accurate, and cost-effective solution for heart stroke risk prediction. Overall, this project demonstrates how machine learning can assist healthcare professionals and individuals in improving early stroke detection and promoting better health management.***

Keywords: Heart Stroke Prediction, Machine Learning, Logistic Regression, Healthcare Analytics, Flask, Python, Data Preprocessing.

I. INTRODUCTION

Heart stroke is one of the leading causes of death and disability across the world. Early identification of stroke risk can help reduce complications and improve patient survival. Traditional methods of diagnosis often require medical expertise and may consume significant time. With the advancement of Machine Learning, healthcare systems can predict diseases more accurately using patient data. The Heart Stroke Prediction System is designed to analyze various health parameters and estimate the possibility of a stroke. The system uses information such as age, gender, hypertension, heart disease, BMI, glucose level, smoking status, and work type. These factors are processed using data preprocessing and feature selection techniques. A Logistic Regression algorithm is applied to classify whether a person is at risk of stroke. The trained model provides quick and reliable prediction results. The application is developed using Python and Flask with a simple web-based interface. Users can easily enter their health details and receive instant predictions. The system also includes a BMI calculator to promote health awareness. This reduces manual effort and supports healthcare professionals in making informed decisions. Machine learning improves prediction accuracy by learning patterns from historical medical data. The proposed system is fast, user-friendly, and cost-effective. It encourages early detection and preventive healthcare measures. Overall, the Heart Stroke Prediction System demonstrates how artificial intelligence can enhance healthcare services and support better medical decision-making.

II. EXISTING SYSTEM

The existing heart stroke prediction systems mainly depend on traditional medical diagnosis and manual analysis. Most systems require expert consultation to assess the patient's health condition. They often consume more time and may delay the prediction process. Many existing applications use limited patient data, reducing prediction accuracy. Some systems do not provide real-time prediction through a user-friendly interface. In several cases, users cannot access the prediction service remotely.[13] Traditional methods also involve higher costs due to repeated medical examinations. Existing models may not effectively handle missing or inconsistent data. Many systems lack integrated features such as BMI calculation and health guidance. They also provide limited support for preventive healthcare decisions. The prediction accuracy of older techniques is comparatively lower than modern machine learning approaches. As a result, there is a need for an intelligent, fast, and reliable prediction system. These limitations motivate the development of an advanced machine learning-based Heart Stroke Prediction System.

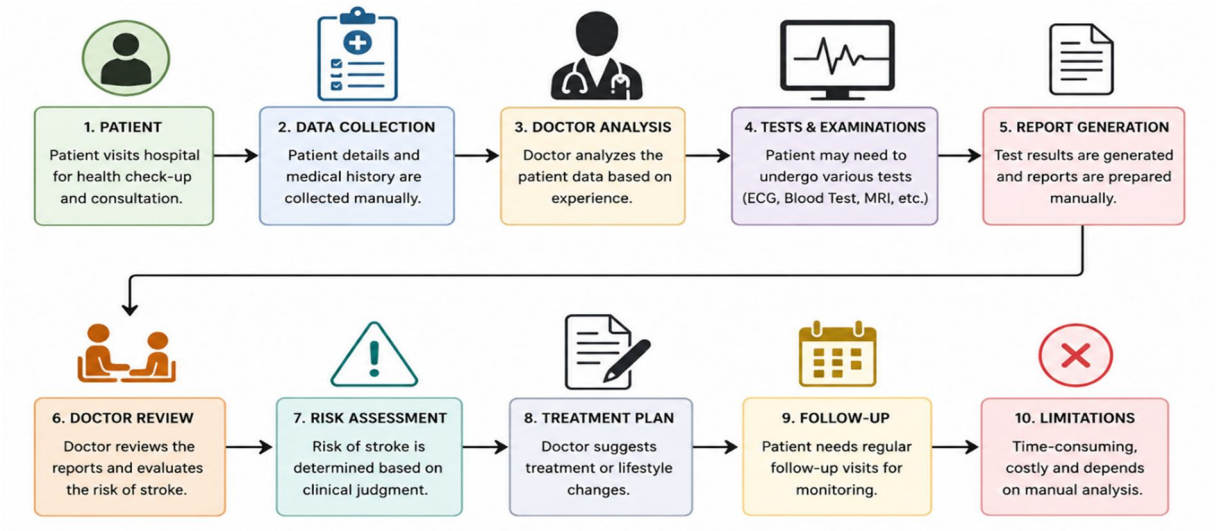


Figure1 : Flow Chart

A. *Challenges*

The existing heart stroke prediction systems face several challenges that affect their efficiency and reliability. [3]Most systems depend on manual medical analysis, which increases the time required for diagnosis. Accurate prediction requires experienced healthcare professionals, making the process costly. Many traditional methods are unable to provide real-time stroke risk assessment. The quality of prediction is highly dependent on the availability of complete patient data. Missing or inconsistent medical records reduce the accuracy of the prediction model. Existing systems often fail to process large healthcare datasets efficiently. Some applications are not user-friendly and require technical knowledge to operate. Limited accessibility prevents users from obtaining predictions from remote locations. Data privacy and security remain major concerns in healthcare applications. Many systems do not support continuous monitoring of patient health conditions. The absence of automated decision support reduces the effectiveness of early diagnosis. Traditional approaches may produce inconsistent results due to human error. Integrating multiple health parameters into a single prediction model is also challenging. [12] Maintaining high prediction accuracy for different patient groups is difficult. Existing models may not adapt well to newly available medical data. High computational requirements can reduce system performance. Lack of preventive health recommendations further limits the usefulness of conventional systems. These challenges highlight the need for an intelligent, accurate, and machine learning-based Heart Stroke Prediction System.

III. PROPOSED SYSTEM

The proposed Heart Stroke Prediction System is a machine learning-based application designed to predict the risk of stroke accurately and efficiently. It uses patient health information such as age, gender, hypertension, heart disease, BMI, glucose level, smoking status, and work type for prediction. [15] The system applies data preprocessing techniques to improve data quality and model performance. A Logistic Regression algorithm is used to classify whether a person is at risk of stroke. The model is trained using historical healthcare data and tested for prediction accuracy. The application is developed using Python and Flask, providing a simple and interactive web interface. Users can easily enter their health details and receive instant prediction results. The system also includes a BMI calculator to help users monitor their health status. It minimizes manual effort and reduces the time required for stroke risk assessment. [4]The application provides reliable decision support for both healthcare professionals and individuals. It improves prediction accuracy through intelligent data analysis. The proposed system is user-friendly, cost-effective, and easily accessible. Overall, it supports early disease detection and promotes preventive healthcare through machine learning technology..

IV. ALGORITHM

A. *Data Collection and Preprocessing*

The first step in the Heart Stroke Prediction System is collecting patient health-related data from reliable healthcare datasets. The dataset contains important attributes such as age, gender, hypertension, heart disease, average glucose level, BMI, smoking status, work type, residence type, and stroke status. [7]Before training the Machine Learning model, the collected data undergoes preprocessing to improve its quality. Missing values are handled, duplicate records are removed, and categorical features are

converted into numerical values using encoding techniques. The dataset is then normalized to ensure consistent feature scaling. Outliers and inconsistent values are identified and corrected to improve prediction accuracy. Finally, the processed data is divided into training and testing datasets, which helps evaluate the model's prediction performance accurately.

B. Logistic Regression Model Training

After preprocessing, the training dataset is provided to the Logistic Regression supervised Machine Learning algorithm. The model learns the relationship between patient health parameters and stroke occurrence by analyzing historical medical data. [6] During training, the algorithm identifies patterns that indicate whether a patient is likely to experience a stroke. Different health attributes are assigned appropriate importance based on their contribution to the prediction. The trained model is continuously evaluated using testing data to improve its accuracy and reduce prediction errors. Once the desired performance is achieved, the trained model is saved and integrated with the web application for real-time stroke prediction.

C. Stroke Risk Prediction and Classification

When a user enters patient health information through the application, the trained Logistic Regression model processes the input data and predicts the probability of stroke. The system classifies the patient as either "High Stroke Risk" or "Low Stroke Risk" based on the prediction result. Along with the prediction, the application provides an instant result that helps users understand their health condition. [11]The prediction is based on important health factors such as age, BMI, glucose level, hypertension, and smoking status. This intelligent classification supports early diagnosis and enables healthcare professionals to make informed medical decisions for timely treatment and preventive care.

D. Decision Support and Performance Evaluation

The final stage focuses on providing decision support for early stroke prevention and healthcare management. Patients identified as high-risk can consult healthcare professionals for further diagnosis and preventive treatment. The application also assists users by promoting health awareness through BMI analysis and risk assessment. Model performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score to ensure reliable predictions. Regular retraining with updated healthcare data helps maintain high prediction accuracy over time. Overall, the Logistic Regression algorithm provides a secure, scalable, and intelligent solution for predicting stroke risk, supporting early diagnosis, improving healthcare decision-making, and reducing the chances of severe medical complications.

V. TECHNIQUES

A. Machine Learning Classification

Machine Learning Classification is the core technique used in the Heart Stroke Prediction System. It is a supervised learning approach where the model is trained using historical patient health records containing both input features and the corresponding stroke labels. [10]The Logistic Regression algorithm learns the relationship between health parameters such as age, gender, hypertension, heart disease, BMI, glucose level, smoking status, and work type. After training, the model predicts whether a person is likely to experience a stroke or not. This technique provides accurate, fast, and data-driven predictions, helping healthcare professionals and individuals identify stroke risk at an early stage and take preventive measures.

B. Data Preprocessing and Feature Engineering

Data preprocessing is an essential technique that improves the quality of the healthcare dataset before model training. Patient records are cleaned by removing duplicate entries and handling missing or inconsistent values. Categorical attributes such as gender, work type, residence type, and smoking status are converted into numerical values using encoding methods. Feature scaling and normalization are applied to ensure all health attributes contribute equally to the prediction process.

C. Logistic Regression Prediction Technique

The proposed system uses Logistic Regression as the primary prediction technique for stroke risk classification. Unlike traditional statistical methods, Logistic Regression estimates the probability of stroke based on multiple patient health factors. The algorithm analyzes relationships between medical attributes and classifies patients into High Stroke Risk or Low Stroke Risk categories. It is simple, efficient, and highly suitable for binary classification problems.

D. Web Technologies and Database Integration

The Heart Stroke Prediction System combines modern web technologies with Machine Learning for efficient implementation. Flask is used to develop the backend application and integrate the trained Logistic Regression model with the web interface for real-time prediction services. HTML, CSS, and Bootstrap are used to create an interactive and responsive user interface that allows users to enter their health details easily. Python performs data processing, model training, and prediction tasks, while CSV-based healthcare datasets are used to store patient information for training and testing. This integration provides a fast, secure, and user-friendly platform for stroke risk prediction**.**

VI. METHODOLOGY

The methodology of the Heart Stroke Prediction System begins with collecting patient health data from reliable healthcare datasets. [3] The collected dataset includes important attributes such as age, gender, hypertension, heart disease, average glucose level, BMI, smoking status, work type, residence type, and stroke status. The data is preprocessed by removing missing values, duplicate records, and inconsistencies to improve data quality. Categorical attributes are encoded into numerical values, and feature scaling is applied where necessary. The cleaned dataset is divided into training and testing datasets for model development. A Logistic Regression supervised Machine Learning algorithm is trained using the prepared data to learn stroke prediction patterns. The trained model is evaluated using performance metrics such as accuracy, precision, recall, and F1-score. After successful training, the model is integrated with a Flask web application for real-time prediction. The frontend is developed using HTML, CSS, and Bootstrap to provide a simple and user-friendly interface. [4]Python is used for data preprocessing, model training, and prediction, while the healthcare dataset is used for training and testing purposes. Users can enter their health details through the web application for instant prediction. The Machine Learning model predicts whether a person is at High Stroke Risk or Low Stroke Risk. The system also provides a BMI calculator to help users monitor their health condition. Healthcare professionals can use the prediction results to support early diagnosis and preventive treatment. The application delivers quick and reliable predictions, reducing the time required for manual assessment. Regular retraining with updated healthcare data helps maintain prediction accuracy over time.

A. Input / Output

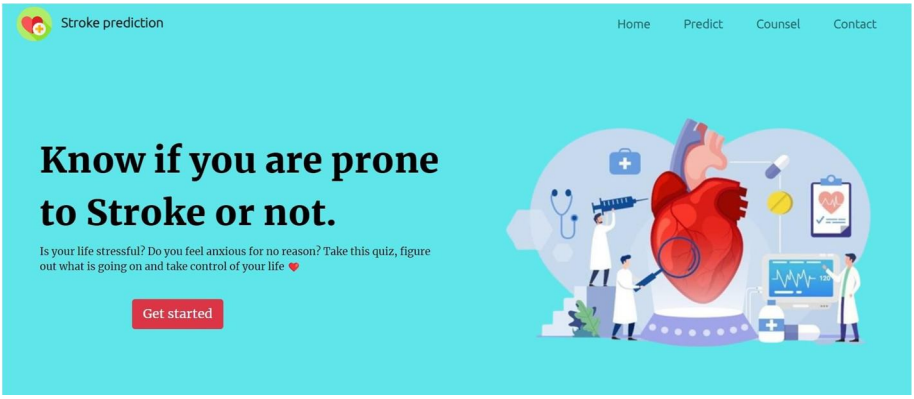


Figure 2: Dashboard

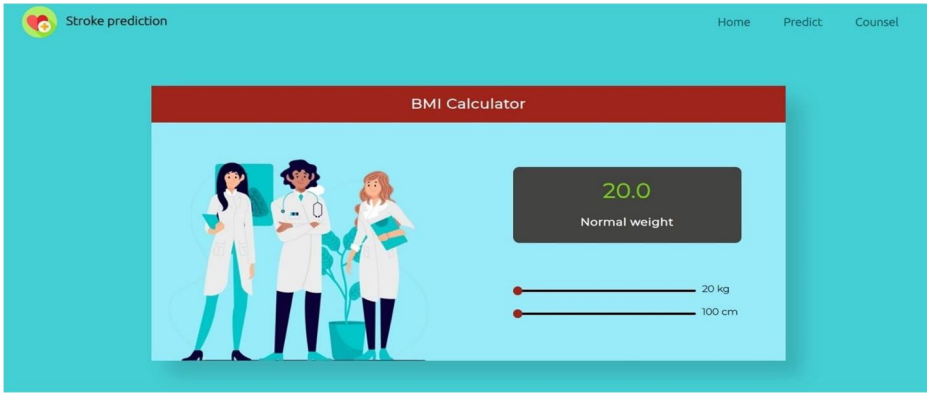


Figure 3: Bmi Calculator

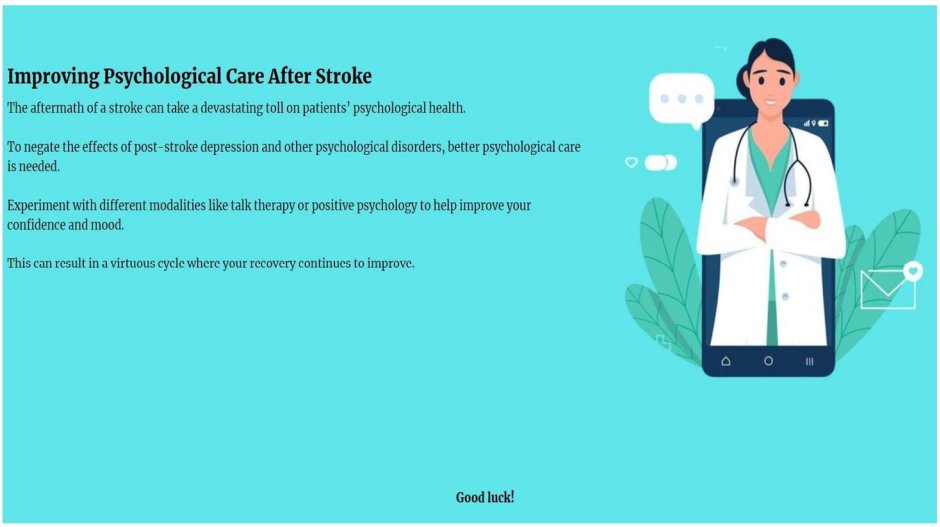


Figure 4:Result

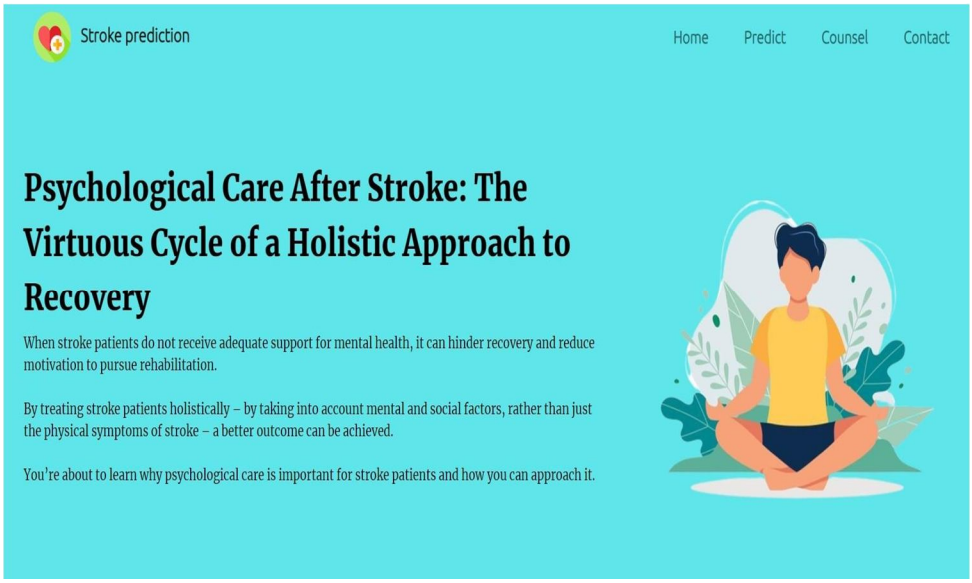


Figure 5: Stroke Predicition Result

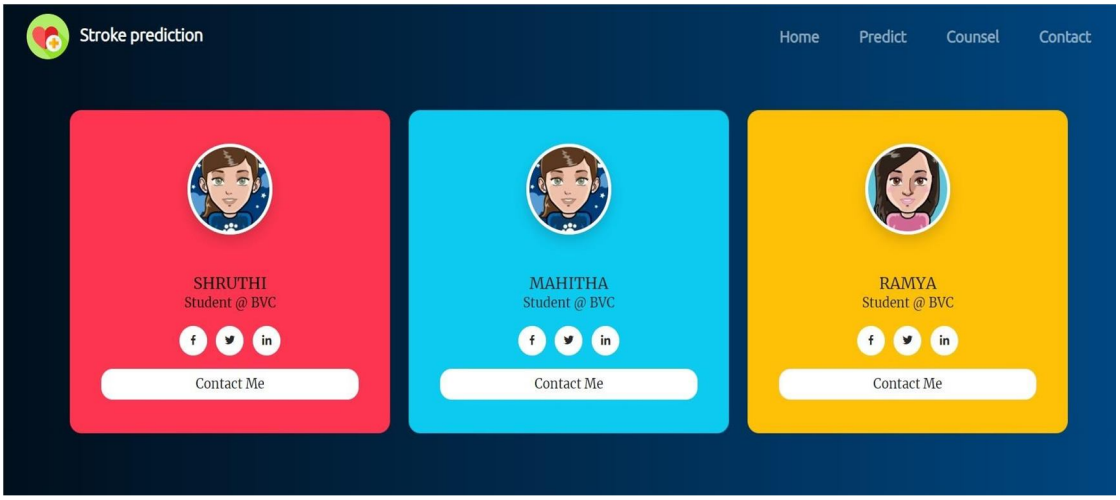


Figure 6: Stroke Predicition Profiles

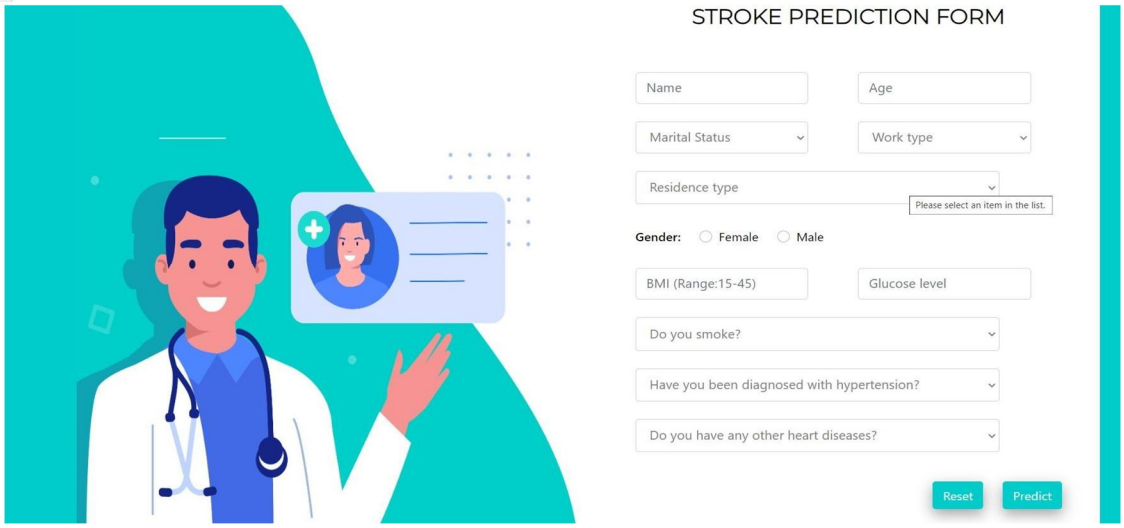


Figure 7: Stroke Predicition Form

VII. CONCLUSION

The **Heart Stroke Prediction System** is an effective machine learning application developed to predict the risk of stroke using patient health information. The system analyzes important medical parameters such as age, hypertension, heart disease, BMI, glucose level, and smoking status to provide accurate predictions. By using the **Logistic Regression** algorithm, the application delivers reliable and efficient stroke risk classification. The proposed system reduces the time and effort required for manual medical assessment. It provides a simple and user-friendly web interface developed using **Python** and **Flask**. The application supports early diagnosis by providing instant prediction results based on user input. It also includes a BMI calculator to promote health awareness and preventive care. The system helps healthcare professionals make informed decisions with better accuracy. It minimizes human errors and improves the efficiency of disease prediction. The model can be updated with new healthcare data to maintain high prediction performance. The proposed solution is cost-effective, secure, and easily accessible for users. It demonstrates the practical application of artificial intelligence in the healthcare sector.

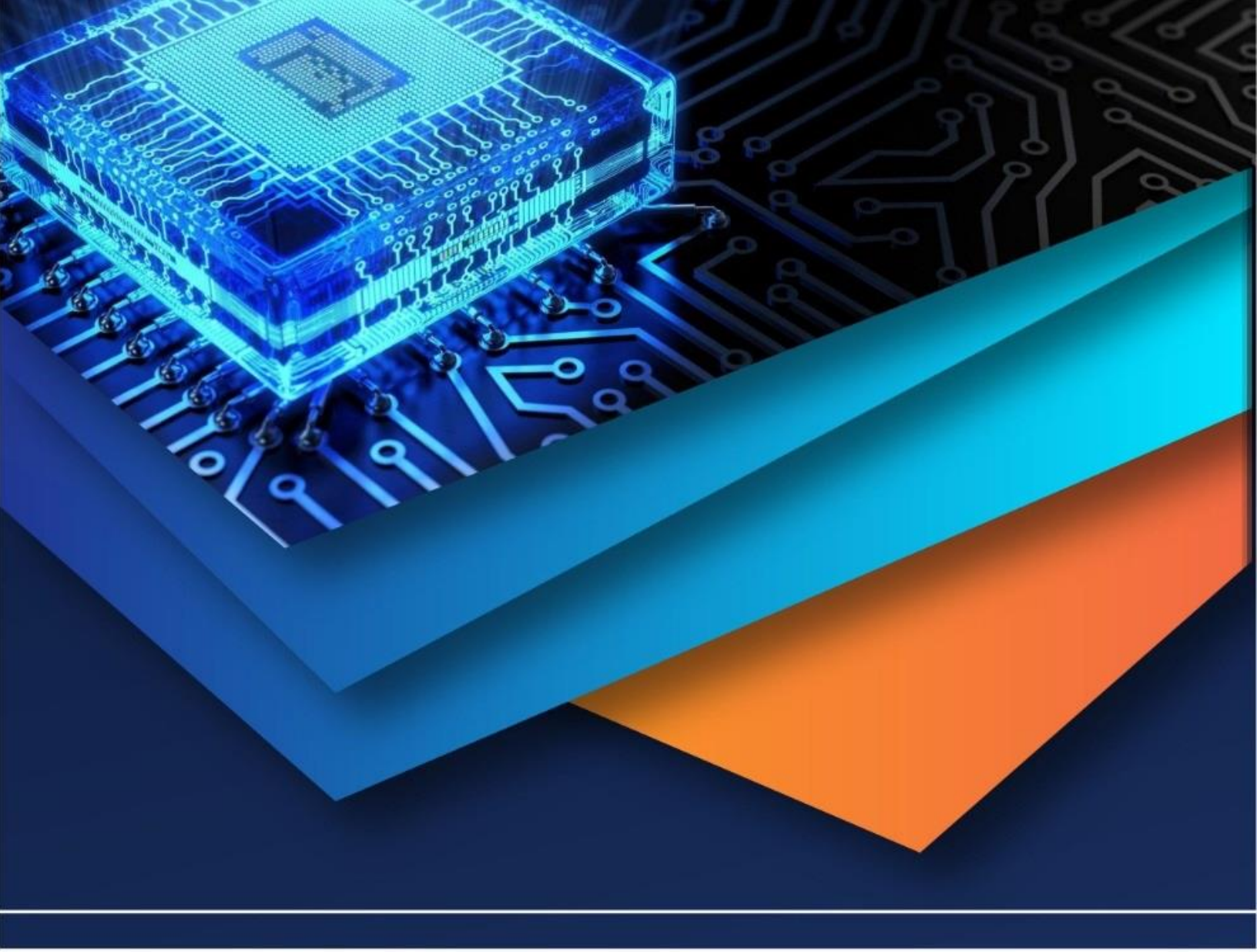
REFERENCES

[1] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied Logistic Regression (3rd Edition). John Wiley & Sons, Hoboken, New Jersey, USA. ISBN: 978-0470582473.
[2] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research (JMLR), Vol. 12, pp. 2825–2830. Available: https://scikit-learn.org/
[3] Grinberg, M. (2018). Flask Web Development: Developing Web Applications with Python (2nd Edition). O'Reilly Media, Sebastopol, California, USA. ISBN: 978-1491991732.
[4] McKinney, W. (2022). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter (3rd Edition). O'Reilly Media. ISBN: 978-1098104030.
[5] Harris, C. R., Millman, K. J., van der Walt, S. J., et al. (2020). "Array Programming with NumPy." Nature, Vol. 585, No. 7825, pp. 357–362. DOI: https://doi.org/10.1038/s41586-020-2649-2
[6] Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment." Computing in Science & Engineering, IEEE Computer Society, Vol. 9, No. 3, pp. 90–95. DOI: 10.1109/MCSE.2007.55.
[7] World Health Organization (WHO). Stroke: Key Facts and Global Health Statistics. World Health Organization, Geneva, Switzerland. Available: https://www.who.int/news-room/fact-sheets/detail/stroke
[8] Centers for Disease Control and Prevention (CDC). About Stroke and Stroke Prevention. U.S. Department of Health & Human Services. Available: https://www.cdc.gov/stroke/
[9] Kaggle. Stroke Prediction Dataset by Fedesoriano. Kaggle Inc. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset
[10] Python Software Foundation. Python 3 Official Documentation. Python Software Foundation. Available: https://docs.python.org/3/
[11] Pallets Project. Flask Official Documentation. Available: https://flask.palletsprojects.com/en/latest/
[12] Scikit-learn Developers. Scikit-learn User Guide and API Reference. Available: https://scikit-learn.org/stable/user_guide.html
[13] The Pandas Development Team. Pandas: Python Data Analysis Library Documentation. Available: https://pandas.pydata.org/docs/
[14] NumPy Developers. NumPy Official Documentation. Available: https://numpy.org/doc/stable/
[15] Bootstrap Team. Bootstrap 5 Documentation: Build Responsive Web Applications. Available: https://getbootstrap.com/docs/5.3/
[16] W3Schools. HTML, CSS, JavaScript, and Bootstrap Web Development Tutorials. Refsnes Data AS, Norway. Available: https://www.w3schools.com/

BIBLOGRAPHY

G.Manoj Kumar Working As An Assistant Professor In Masters Of Computer Applications (MCA) Sanketika Vidya Parishad Engineering College, Which Is Accredited With An 'A' Grade By NAAC, Affiliated To Andhra University. Completed Post Graduation In Andhra University College Of Engineering (AUCE). With Accredited By NAAC With His Areas Of Interest In Java,Python,DBMS,Web Technology, OS.

Sheik Gafur Is Currently Pursuing His Final Semester Of Master Of Computer Applications (MCA) At Sanketika Vidya Parishad Engineering College, Which Is Accredited With An 'A' Grade By NAAC, Affiliated To Andhra University, And Approved By AICTE. With A Keen Interest In Machine Learning And Artificial Intelligence, He Has Undertaken His Postgraduate Project Titled "**HEART STROKE PREDICTION**" The Project Has Been Successfully Carried Out Under The Guidance Of G.Manoj Kumar, Assistant Professor, SVPEC.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY