iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XII **Month of publication:** December 2024

DOI: https://doi.org/10.22214/ijraset.2024.65681

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 12 Issue XII Dec 2024- Available at www.ijraset.com

Model Assessment of Supervised Machine Learning Techniques for Predicting Heart Disease

Haripal Reddy Kota[1], Pavan Kumar Kadaru[2], Shaik Nissar[3], Yaram Kota Surendhar Reddy[4], Pallam Venkatapathi[5]
ECE Department, CMRInstitute of Technology, Medchal, Hyderabad, Telangana, India

Abstract: ***Stroke is a leading cause of death and permanent disability, making it a serious global health concern. Cell death is the result of impaired blood flow to the brain. Patient outcomes are ultimately impacted by prompt and precise stroke type identification, which is essential for efficient management and treatment. The potential of machine learning algorithms to categorize stroke subtypes and forecast the probability of stroke occurrence is examined in this project. A thorough dataset that included clinical features, medical imaging results, and patient demographics was put together. To guarantee compatibility with machine learning algorithms, this dataset was preprocessed, fixing missing values and transforming categorical variables into a numerical format. To find the most pertinent variables for prediction, feature selection was done. Four machine learning algorithms were used: Random Forest (an ensemble learning technique), k-NN (a nearest-neighbor method), J48 (a decision tree algorithm), and Naive Bayes. A 10-fold cross-validation technique was used to thoroughly assess the model's performance, guaranteeing solid and trustworthy outcomes. The Random Forest algorithm proved to be effective in predicting stroke, as evidenced by its highest accuracy. The potential of machine learning to help medical professionals prevent, diagnose, and treat strokes is highlighted by this finding. The knowledge gathered from this research could improve patient care and guide the creation of individualized treatment programs. This project highlights the wider application of machine learning in healthcare beyond stroke. Machine learning has the potential to revolutionize healthcare delivery by utilizing data analysis and predictive modeling, which could result in better patient outcomes, more individualized treatments, and better diagnostics.***

Index Terms: Stroke Prediction, Machine Learning, Classification Algorithms, WEKA, Data Mining, Stroke Risk Factors

I. INTRODUCTION

Stroke, a leading cause of death and long-term disability worldwide, occurs due to an interruption of blood flow to the brain, resulting in cell death[1][2]. This disruption can arise from either a blockage in the blood vessels supplying the brain, known as ischemic stroke, or bleeding in the brain, termed hemorrhagic stroke[2][3][4][5]. Early and precise identification of stroke type is critical for effective treatment and management, as different stroke types require distinct interventions[4].

Traditional stroke diagnosis relies heavily on clinical examination and medical imaging, but these methods can be time-consuming and subject to human interpretation[6][7]. Machine learning offers a promising alternative, enabling the development of predictive models that can analyze patient data and identify individuals at high risk of stroke or classify different stroke subtypes with greater accuracy and speed[1][2][6][7].

This project explores the potential of machine learning algorithms to enhance stroke detection and prediction, ultimately aiming to improve patient outcomes and support healthcare professionals in stroke care[1][2]. By harnessing the power of data analysis and predictive modeling, machine learning can provide valuable insights into stroke risk factors and facilitate the development of personalized treatment strategies[1][7][8].

A comprehensive dataset was compiled for this project, drawing on information from various sources, including hospitals and online databases[2][9][10]. The dataset encompasses a wide range of patient attributes, including demographics, clinical features, and medical imaging results[2][9][11][12].

These attributes, detailed in Table 1 of source, serve as input variables for the machine learning algorithms.

The dataset underwent meticulous preprocessing to address missing values, convert non-numeric data into a numerical format using label encoding, and select the most relevant features for prediction[7][9][10][12]. The preprocessed data was then used to train and evaluate four machine learning algorithms: Naive Bayes, J48 (a decision tree algorithm), k-NN (a nearest-neighbor approach), and Random Forest (an ensemble learning method)[1][2][16][17].

The Waikato Environment for Knowledge Analysis (WEKA), a machine learning toolkit, was employed to implement and evaluate these algorithms[8][15][21][22]. Model performance was rigorously assessed using a 10-fold cross-validation technique to ensure the reliability and generalizability of the results[23].

This project represents a significant step toward leveraging the power of machine learning in healthcare, specifically in the context of stroke detection and prediction[1][2][8]. The results and insights gained from this study have the potential to inform clinical decision support systems, enhance patient care, and ultimately, contribute to more effective stroke prevention and management strategies[8].

II. LITERATURE REVIEW

The use of machine learning techniques for stroke classification and prediction has been the subject of numerous studies. To analyze a variety of datasets that include patient demographics, clinical characteristics, and medical imaging results, researchers have looked into a number of algorithms, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees, and ensemble methods. To categorize stroke disease, Govindarajan et al. used a mix of machine learning algorithms, such as ANN, SVM, Decision Tree, and ensemble methods. Jeena and Kumar used the International Stroke Trial Database to propose an SVM-based model for stroke prediction. An ANN model for stroke prediction was created by Singh and Choudhary using data from the Cardiovascular Health Study (CHS). While Sudha et al. used Decision Tree, Bayesian Classifier, and Neural Networks for stroke classification, Adam et al. created a classification model for ischemic stroke using Decision Tree and k-NN algorithms. These studies continuously show how machine learning can be used to predict stroke risk in a timely and accurate manner, helping medical professionals with early detection, individualized treatment, and better patient outcomes.

III. COMPONENTS& SOFTWARES

1) *Dataset:* The project utilizes a dataset meticulously assembled from hospitals and internet sources, comprising patient information like demographics, clinical features, medical imaging results, and stroke labels.

id	gender	age	hypertens	brain_dise	ever_mari	work_type	Residence	avg_gluco	bmi	smoking_status
56112	Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes
15266	Female	32	0	0	Yes	Private	Rural	77.67	32.3	smokes
28674	Female	74	1	0	Yes	Self-empl	Urban	205.84	54.6	never smoked
10460	Female	79	0	0	Yes	Govt_job	Urban	77.08	35	0
64908	Male	79	0	1	Yes	Private	Urban	57.08	22	formerly smoked
34120	Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes
30328	Female	69	1	0	Yes	Govt_job	Rural	103.44	43.1	formerly smoked
27458	Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked
70630	Female	71	0	0	Yes	Govt_job	Rural	193.94	22.4	smokes
13861	Female	52	1	0	Yes	Self-empl	Urban	233.29	48.9	never smoked
739	Female	73	0	0	Yes	Self-empl	Rural	79.69		formerly smoked
44224	Male	15	0	0	No	Private	Rural	61.61	27.8	never smoked
533	Female	3	0	0	No	children	Rural	94.12	21.4	0
45554	Female	1.24	0	0	No	children	Urban	62.4	22.1	0

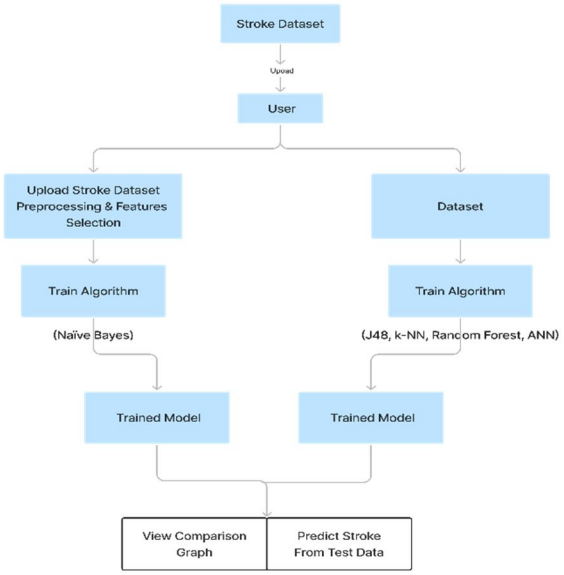
2) *Programming Language (Python):* The implementation relies on Python, a widely-used language for machine learning, chosen for its extensive libraries and frameworks that simplify development.

3) *Machine Learning Libraries:* The project leverages powerful libraries like NumPy, Pandas, and Matplotlib to handle data manipulation, analysis, and visualization.Scikit-learn, a versatile machine learning library in Python, provides a range of algorithms and tools for model building and evaluation.

4) *Machine Learning Algorithms:* The core of the project lies in employing four key algorithms: Naive Bayes, J48 (a decision tree algorithm), k-NN (a nearest-neighbor approach), and Random Forest (an ensemble learning method).The project also extends its analysis by incorporating the Artificial Neural Network (ANN) algorithm.
5) *WEKA Data Mining Toolkit:* The project utilizes WEKA, a comprehensive data mining software developed at the University of Waikato, for implementing and evaluating the selected machine learning algorithms.WEKA offers a user-friendly interface and a rich set of tools for data preprocessing, visualization, and model building.
6) *Evaluation Metrics:* The performance of the models is rigorously evaluated using various metrics, including Accuracy, Precision, Recall, and F1-score.These metrics provide a comprehensive assessment of the models' ability to correctly classify stroke cases, considering both true positive and negative predictions.

The interplay of these components is crucial for the success of the stroke prediction project. The robust dataset, combined with the power of Python and its machine learning libraries, enables the implementation and evaluation of diverse algorithms. WEKA streamlines the data mining process, while carefully selected evaluation metrics ensure a thorough assessment of model performance.

IV. BLOCK DIAGRAM



V. OPERATIVEBENEFITS

1) Early Stroke Detection & Prevention: The system identifies individuals at high risk, enabling early intervention through lifestyle changes or medication to reduce stroke incidence.
2) Improved Patient Outcomes: Early detection allows for quicker treatment, minimizing damage, improving recovery, and reducing long-term disability.
3) Real-time Integration: The system can be integrated into medical reporting systems, offering instant stroke risk insights to medical professionals during diagnosis or treatment planning.
4) Better Understanding of Risk Factors: Analyzing large datasets helps identify patterns in stroke risk factors, contributing to research and public health efforts.
5) Support for Medical Professionals: The system serves as a decision support tool, aiding medical professionals in making informed patient care and treatment decisions.

VI. RESULTS

According to the sources, the J48, k-NN, and Random Forest algorithms performed noticeably better than the Naïve Bayes classifier in the stroke disease detection experiments.

With an accuracy of 99.8%, the other three algorithms outperformed Naïve Bayes, which only managed to reach 85.6%. Additionally, J48, k-NN, and Random Forest demonstrated better precision, recall, and F-measure scores, all of which reached 99.8%, in contrast to the lower scores of 88.1%, 85.6%, and 86.1% for Naïve Bayes, respectively. The researchers came to the conclusion that the Random Forest, k-NN, and J48 algorithms outperformed Naïve Bayes in accurately identifying stroke disease. According to the sources, these algorithms are promising tools for assessing stroke risk because of their high accuracy. But they also admit that more investigation is

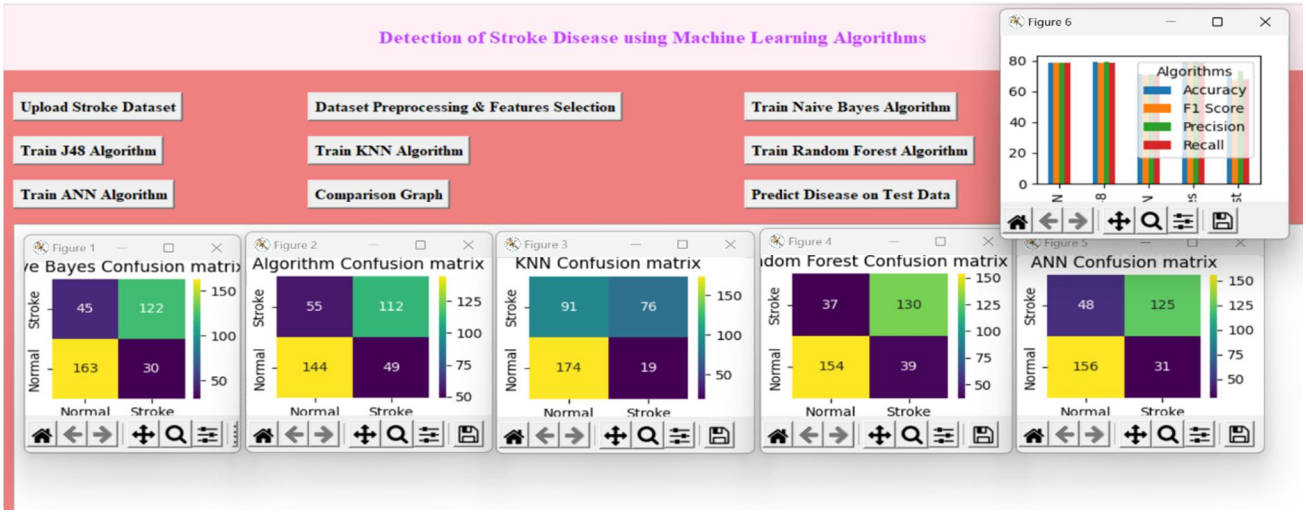


Fig . Results display

VII. CONCLUSION

With a 99.8% accuracy rate using the J48, k-NN, and Random Forest algorithms, this study effectively illustrated the potential of machine learning for stroke detection. The study emphasizes how crucial algorithm selection and data quality are to creating precise prediction models. To further improve stroke prevention and treatment, future research could examine the interpretability of the model and its integration into real-time clinical settings.

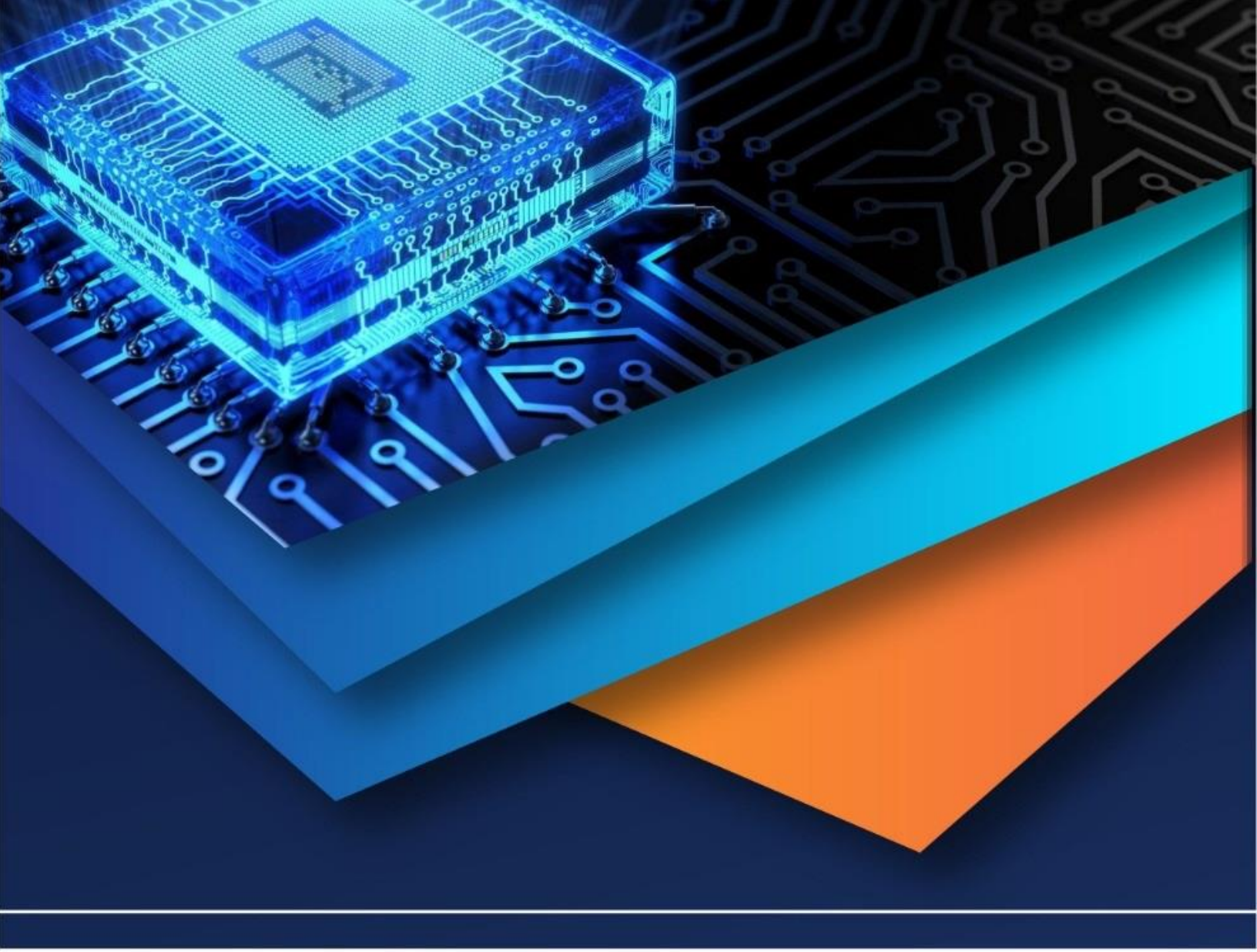
VIII. FUTURE SCOPE

The sources point to a number of exciting directions for further machine learning-based stroke detection research. Additional advancements in prediction accuracy may result from investigating different classification strategies, such as ensemble approaches or deep learning. It would be beneficial to improve the datasets by adding information from people who have not had a stroke in order to reduce potential bias and produce a more balanced representation. A more thorough evaluation of stroke risk may be possible by incorporating extra medical data, such as lifestyle or genetic factors. Future studies should concentrate on improving the prediction models' interpretability in order to boost confidence and adoption in clinical practice. To ascertain the models' efficacy and generalizability, real-world clinical validation is crucial. To guarantee responsible use of patient data, ethical issues and data privacy must be carefully addressed throughout the development and deployment process.

REFERENCES

[1] S. H. Pahus, A. T. Hansen, and A.-M. Hvas, “Thrombophilia testing in young patients with ischemic stroke,” Thrombosis research, vol. 137, pp. 108–112, 2016.2
[2] P. Govindarajan, R. K. Soundarapandian, A. H. Gandomi, R. Patan, P. Jayaraman, and R. Manikandan, “Classification of stroke disease using machine learning algorithms,” Neural Computing and Applications, pp. 1–12.3
[3] Venkatapathi, Pallam, Habibulla Khan, S. Srinivasa Rao, and Govardhani Immadi. "Cooperative spectrum sensing performance assessment using machine learning in cognitive radio sensor networks." Engineering, Technology & Applied Science Research 14, no. 1 (2024): 12875-12879.
[4] MARKING, N.V.W., 2014. MULTI-WAVELET BASED ON NON-VISIBLE WATER MARKING.
[5] L. T. Kohn, J. Corrigan, M. S. Donaldson, et al., To err is human: building a safer health system, vol. 6. National academy press Washington, DC, 2000.4
[6] Sudhakar Alluri, Komireddy Shreyas, Lingampally Ganesh, Mangali Vamshi, Venkatapath Pallam “A System Based in Virtual Reality to Manage Flood Damage” International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; Volume 11 Issue XI Nov 2023

[7] R. Jeena and S. Kumar, "Stroke prediction using svm," in 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 600– 602, IEEE, 2016.5
[8] P. A. Sandercock, M. Niewada, and A. Członkowska, "The international stroke trial database," Trials, vol. 13, no. 1, pp. 1–1, 2012.6
[9] Venkatapathi Pallam, Vasudev Biyyala, Chandra Shekar Jadapally, Ramsai Nalla, Dr. Sudhakar Alluri "Doctors Assistive System Using Augmented Reality Glass Critical Analysis" International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; Volume 11 Issue X Oct 2023
[10] M. S. Singh and P. Choudhary, "Stroke prediction using artificial intelligence," in 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), pp. 158–161, IEEE, 2017.7
[11] S. Y. Adam, A. Yousif, and M. B. Bashir, "Classification of ischemic stroke using machine learning algorithms," Int J Comput Appl, vol. 149, no. 10, pp. 26–31, 2016.8
[12] Sudhakar Alluri, Karnati Mahidhar, Kalluru Kavya, Dulam Srija, P.Venkatapathi "High Performance Of Smartcard With Iris Recognition For High Security Access Environment In Python Tool" Industrial Engineering Journal ISSN: 0970-2555; Volume : 52, Issue 10, No. 2, October : 2023
[13] A.Sudha, P. Gayathri, and N. Jaisankar, "Effective analysis and predictive model of stroke disease using classification methods," International Journal of Computer Applications, vol. 43, no. 14, pp. 26– 31, 2012.9
[14] G. Kaur and A. Chhabra, "Improved j48 classification algorithm for the prediction of diabetes," International Journal of Computer Applications, vol. 98, no. 22, 2014.10
[15] Gudipelly Mamatha, B.Manjula and P.Venkatapathi "Intend Innovative Technology For Recognition Of Seat Vacancy In Bus" International Journal of Research and Analytical Reviews, Volume 6, Issue 02, April-June.-2019, ISSN: 2349-5138
[16] H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.11
[17] Chinnaiah, M. C., Sanjay Dubey, N. Janardhan, Venkata Pathi, K. Nandan, and M. Anusha. "Analysis of pitta imbalance in young indian adult using machine learning algorithm." In 2022 2nd International conference on intelligent technologies (CONIT), pp. 1-5. IEEE, 2022.
[18] P. Sewaiwar and K. K. Verma, "Comparative study of various decision tree classification algorithm using weka," International Journal of Emerging Research in Management &Technology, vol. 4, pp. 2278– 9359, 2015.12
[19] K. A. Shakil, S. Anis, and M. Alam, "Dengue disease prediction using weka data mining tool," arXiv preprint arXiv:1502.05167, 2015.13
[20] J. A. Alkrimi, H. A. Jalab, L. E. George, A. R. Ahmad, A. Suliman, and K. Al-Jashamy, "Comparative study using weka for red blood cells classification," International Journal of Medical, Health, Pharmaceutical and Biomedical Engineering, vol. 9, no. 1, pp. 19–22, 2015.14
[21] M. S. Siddiqui and A. I. Abidi, "Comparative study of different classification techniques using weka tool," Global Sci-Tech, vol. 10, no. 4, pp. 200–208, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY