https://doi.org/10.33472/AFJBS.6.10.2024.996-1003

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Neurovascular Crisis Prediction Using Machine Learning

Deepthi P
Assistant professor
Dept of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle,A P, India
deepthip@mits.ac.in

Kalyan B
Dept of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle,A P, India
kalyanboya269@gmail.com

Varun Choudary T
Dept of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle,A P, India
varunchowdary560@gmail.com

Yuva Prakash Reddy C
Dept of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle,A P, India
prakashreddyyuva999@gmail.com

Vishnu Sai Rao K
Dept of Computer Science & Engineering
Madanapalle Institute of Technology & Science
Madanapalle,A P, India
kvsr12032003@gmail.com

Volume 6, Issue 10, 2024

Received: 09 March 2024

Accepted: 10 April 2024

Published: 20 May 2024

doi:10.33472/AFJBS.6.10.2024.996-1003

Abstract- This project focuses on the development of a machine-learning model for the early prediction of neurovascular crises, commonly known as strokes. Neurovascular crises pose a significant threat to public health, and timely intervention is crucial for minimizing long-term damage and improving patient outcomes. Leveraging a diverse dataset comprising medical records, demographic information, and lifestyle factors, our model employs advanced machine learning algorithms to analyze and identify key risk factors associated with neurovascular crises. The predictive model utilizes a range of features, including age, blood pressure, cholesterol levels, smoking habits, and medical history, to establish correlations and patterns indicative of an increased risk of neurovascular crises. Our method seeks to improve prediction accuracy and reliability by utilizing techniques like feature engineering, data preprocessing, and model tuning. The suggested method helps identify people who are more likely to become sick but it also helps healthcare providers prioritize preventive care. Personalized patient care and focused risk reduction strategies are made possible by the integration of interpretable machine learning techniques, which enable the extraction of actionable insights.

Keywords:. Neurovascular crises, Stroke, Machine learning model, Predictive model, medical records, Demographic information, Data preprocessing, Model tuning, Feature engineering.

Introduction

Neurovascular crises present a significant challenge in the field of public health,[5]requiring creative approaches to lessen their effects. This project aims to advance the field by employing machine learning techniques to predict neurovascular crises.[8]The phrase "neurovascular crisis" refers to a range of disorders that impact the blood vessels in the brain, requiring a sophisticated strategy for early identification and treatment. This project uses machine learning to identify individuals who are more vulnerable by combining lifestyle factors, [17]medical histories, and demographic data to identify complex patterns within large datasets. Our method incorporates a wide range of characteristics, such as age, blood pressure, cholesterol, smoking status, and other pertinent factors, to create a predictive model that can identify minute signs of an imminent neurovascular emergency. [18]By means of methodical data preprocessing, feature engineering, and algorithmic fine-tuning, our system seeks to improve prediction accuracy and dependability.

Neurovascular crises, including strokes, are a prominent cause of disability and death globally, putting a significant strain on healthcare systems and individuals alike. [12],[15]Despite advances in medical research, correctly predicting and avoiding these catastrophes remains a difficult task. Machine learning is a promising method to solving this difficulty, harnessing the power of data analytics to identify subtle patterns and relationships that traditional diagnostic tools may miss. By leveraging the massive volumes of data collected in healthcare settings, we can create prediction models capable of identifying patients at high risk of neurovascular events, allowing for earlier intervention and potentially saving lives.

Neurovascular crises present a significant challenge in the field of public health,[10]requiring creative approaches to lessen their effects. This project aims to advance the field by employing machine learning techniques to predict neurovascular crises.[9]The phrase "neurovascular crisis" refers to a range of disorders that impact the blood vessels in the brain, requiring a sophisticated strategy for early identification and treatment. This project uses machine learning to identify individuals [14]who are more vulnerable by combining lifestyle factors, medical histories, and demographic data to identify complex patterns within large datasets. Our method incorporates a wide range of characteristics, such as age, blood pressure, cholesterol, smoking status, and other pertinent factors, to create a predictive model that can identify minute signs of an imminent neurovascular emergency. By means of methodical data preprocessing, feature engineering, and algorithmic fine-tuning, our system seeks to improve prediction accuracy and dependability.

Furthermore, integrating machine learning into clinical practice has the potential to transform the delivery of tailored treatment.[16] Healthcare practitioners can improve outcomes and resource allocation by adapting preventative tactics and treatment plans to each patient's unique characteristics and risk profiles. This study is a step toward attaining this objective by proving the revolutionary power of data-driven techniques in the field of neurovascular health. [11]We aim to pave the path for a future in which neurovascular crises are not only anticipated precisely but also averted entirely, thereby improving the quality of life for millions of people around the world.

Literature Review

[1] This initiative employs machine learning to address the critical challenge of stroke prevention and early detection. By analyzing data from participants, including lifestyle factors and medical histories, the project aims to predict and categorize strokes, such as thrombotic or embolic types. The goal is to identify individuals at risk for stroke and provide early warning signals to prevent or reduce the severity of these events, ultimately improving public health outcomes.

[2] Stroke is a significant global health issue, with high mortality and disability rates. This study provides an overview of stroke from a public health perspective, highlighting its economic and social impact. As populations age and risk factors increase, the burden of stroke is expected to rise, underscoring the need for effective prevention strategies.

[3] After heart disease, stroke is the second leading cause of mortality worldwide. This research leverages advancements in wireless communication and body area networks to facilitate early diagnosis and treatment of stroke. By combining artificial intelligence with body area networks, the study aims to expedite the diagnosis process and improve healthcare outcomes for stroke and related conditions.

[4] Despite declines in mortality rates, stroke remains a significant global health burden. The Global Burden of Diseases Study provides comprehensive estimates of stroke burden, informing healthcare planning and priority setting. By utilizing various data sources and methodologies, the study offers valuable insights into the prevalence and impact of stroke over time.

[5] Ischemic stroke poses a significant challenge due to its variable phenotypes. This study explores automated phenotyping approaches using electronic medical records, demonstrating the potential for improving stroke classification accuracy.

[6] Brain stroke incidence is increasing in India, posing a significant health challenge. This research proposes a mechanism for accurately predicting brain strokes using data from stroke patients, with promising results for early diagnosis and treatment.

[7] Rising mortality rates among adolescents and young adults highlight the urgency of addressing health issues in this population. By employing artificial intelligence techniques, this approach aims to predict and prevent brain strokes, emphasizing the importance of precise risk assessment and intervention.

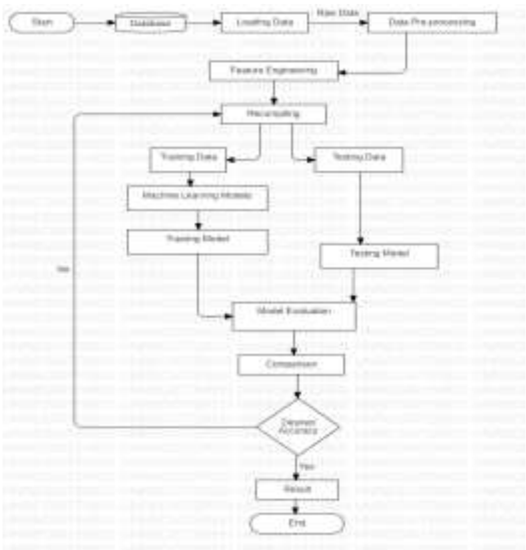
[8] Brain strokes result from reduced blood supply to the brain, leading to cell death. This study evaluates eight classifiers for brain stroke detection, with the Logistic Regression Model demonstrating the highest accuracy. Comparing these models provides valuable insights for stroke prediction and intervention.

Proposed Architecture

Several essential phases are commonly involved in the development of machine learning models: loading data, preparing data, selecting features, separating data, generating models, and evaluating them. Loading data is the initial phase, in which the dataset is imported into the programming environment or tool of choice. Once the data has been imported, preprocessing procedures are used to clean and prepare it for analysis. This may involve resolving missing values, scaling features, and encoding categorical variables. The most relevant and informative characteristics are then determined in order to improve model performance and reduce overfitting. After feature selection, the data is divided into training and testing sets to evaluate the model's performance on previously unknown datasets. Model development comprises choosing an effective algorithm and training it using training data. Finally, model assessment is carried out using several metrics to analyze the model's performance, such as accuracy, precision, recall, and F1-score, to indicate its ability to predict new data. Overall, these processes represent a disciplined method to creating and testing machine learning models.

Research Methodology

In contrast to previous approaches, employing a Random Forest model offers distinct advantages for neurovascular crisis [7] prediction. One key strength lies in its ability to handle complex and non-linear relationships within heterogeneous datasets effectively. Random Forest's ensemble nature, combining multiple decision trees, enhances robustness and generalizability, mitigating overfitting issues prevalent in simpler models.



4.1 Data Acquisition:

In the world of data science and machine learning, data collection is a critical first step that serves as the foundation for later analysis and predictive modeling attempts. The development of a good dataset is a critical prerequisite for our specific project targeted at forecasting neurovascular crises, namely brain strokes. To do this, we went to Kaggle, a well-known platform known for its collection of various datasets and ability to facilitate collaborative data science efforts. Our chosen dataset,[19] appropriately termed "Neurovascular Crisis Prediction (Brain Stroke Prediction)," proved to be a goldmine of vital information. Spanning a variety of aspects, including medical records, demographic information, and lifestyle characteristics of individuals, this dataset was rigorously collected with the sole intention of allowing the creation and enhancement of prediction models aimed at forecasting the beginning of strokes. The gathering of such a comprehensive dataset not only demonstrates the need for sound data sourcing procedures, but it also provides the framework for later phases of data exploration, preprocessing, and model building.

4.2. Data Preprocessing

Data preprocessing, the foundation of every data analysis or machine learning venture, emerges as a multidimensional undertaking with the ultimate goal of transforming raw data into a pure, analytically tractable format. In the context of our study based on neurovascular crisis prediction, the preprocessing of our collected dataset was critical. With a dedication to guaranteeing data quality, consistency, and preparedness for future analytical endeavors, our preprocessing operations included a bevy of rigorous chores. First and foremost, a thorough sweep was performed to discover and correct any instances of missing data, with null values being examined using specialized methods such as isnull() and isnull().sum(). Furthermore, attempts were made to discover and remove duplicate records, utilizing techniques such as duplicated() and duplicated().sum(). Furthermore, thorough attention was given to the handling of zero values within relevant columns, with tight methods in place to round off age values and remove rows containing zero age entries. Furthermore, to reduce the influence of outliers on subsequent analyses, robust outlier identification approaches such as the Interquartile Range (IQR) method were used, notably in columns like 'avg_glucose_level' and 'bmi.' By painstakingly navigating these pretreatment methods, our dataset emerged with increased integrity, paving the way for the subsequent phases of exploratory data analysis and predictive modeling efforts.

4.3 Exploratory Data Analysis (EDA):

Beginning exploratory data analysis (EDA) is a watershed moment in the data science journey, allowing for the discovery of latent insights and uncovering the underlying narratives hidden inside the dataset's labyrinthine folds. In our attempt to understand the complexities of neurovascular crises and the start of brain strokes, EDA acted as an invaluable compass, leading us through the dense thicket of data. Using an arsenal of descriptive statistics and visualization approaches, [20]we began on a journey of discovery, peeling back the layers of abstraction to unveil the hidden truths inherent in the dataset's fabric. Histograms, box plots, scatter plots, and correlation matrices evolved as reliable companions, providing a comprehensive view of the complicated tapestry of variables and their interrelationships. With a keen eye for patterns and anomalies, we navigated the data environment, uncovering subtle connections and predicting elements that remained hidden under the surface. Indeed, EDA not only facilitated data-driven disclosures, but it also established the framework for informed decision-making in the future phases of feature selection, model creation, and interpretive analysis. Thus, armed with the enlightening insights gained by EDA, we went forth with renewed zeal, ready to begin on the revolutionary journey of predictive modeling and neurovascular crisis prediction.

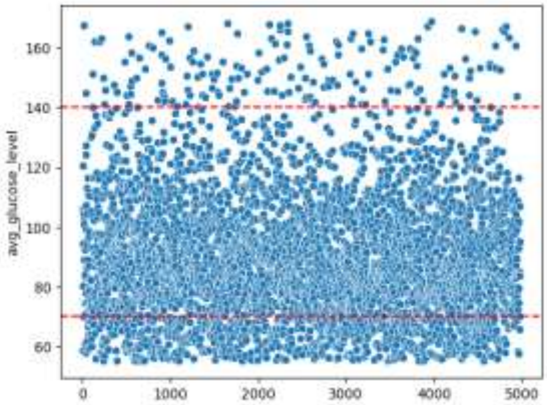


Fig.1 Plot Graph

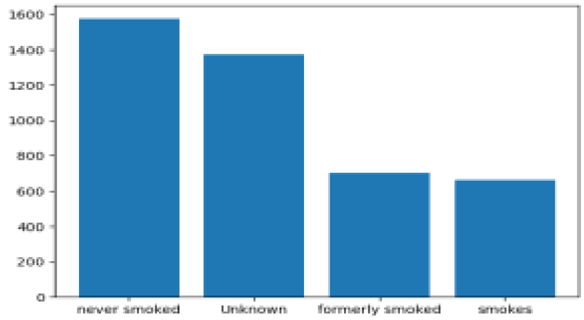


Fig.2 Bar graph

- Blood Pressure: Systolic and diastolic blood pressure readings.
- Cholesterol Levels: Total cholesterol levels in the blood.
- Smoking Habits: Categorized as non-smoker, formerly smoked.
  each class before and after applying smote technique.

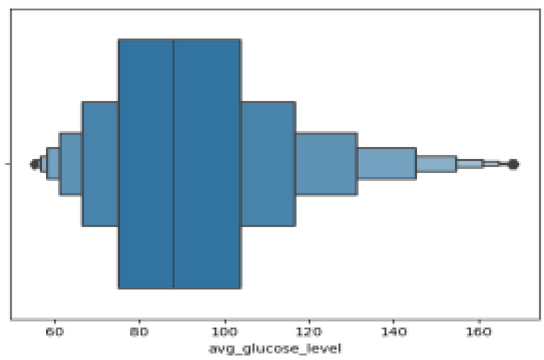


Fig.3 Glucose Level

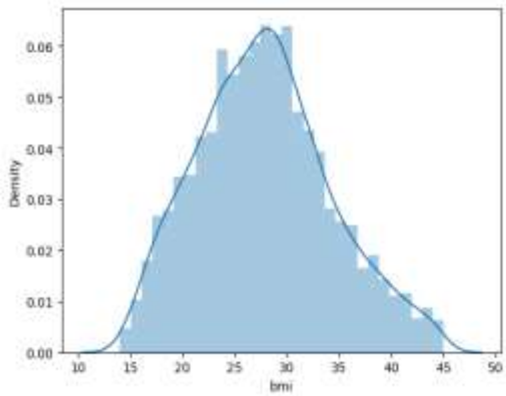


Fig.4 Density x BMI

4.4. Feature Engineering

Feature engineering was performed to derive meaningful predictors from the raw dataset,

enhancing the predictive power of the model. Key features considered in the analysis included.

4.5. Handling Imbalanced Data

To address the imbalance in the dataset between stroke and non-stroke cases, we applied the Synthetic Minority Over-sampling Technique (SMOTE). This oversampling technique generated synthetic instances of the minority class to achieve

a balanced dataset, thus preventing bias towards the majority class during model training. The images below give a clear picture of the count of records in

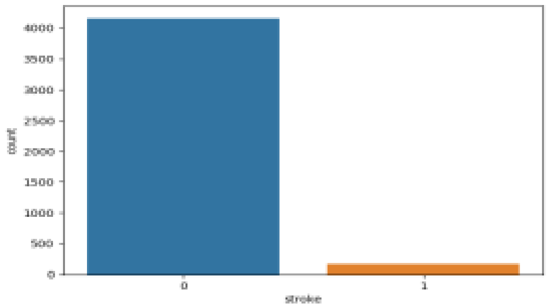


Fig.5 Stroke count

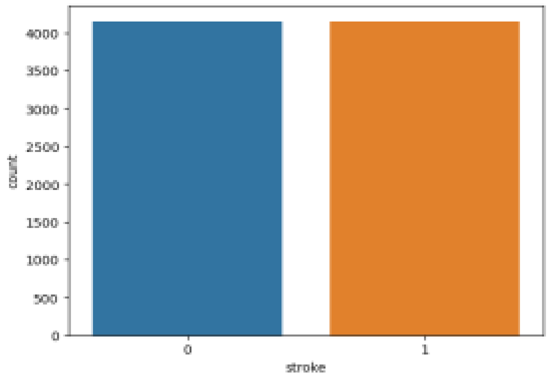


Fig.6 Stroke Count

4.6. Model Building

Several machine learning algorithms were employed to train predictive models on the preprocessed dataset. These algorithms included:

- Logistic Regression: Used for binary classification and interpretable results.
- Support Vector Machine (SVM): Suitable for high-dimensional data and nonlinear relationships.
- Decision Tree: Simple yet effective for capturing complex decision boundaries.
- Random Forest: Ensemble method for improved predictive performance and robustness.

Model training and evaluation were performed using the training and testing sets, with performance metrics such as accuracy scores, classification reports, and confusion matrices utilized to assess model effectiveness.

Results and Discussion

The model assessment method included a thorough study of several performance indicators in addition to accuracy. We thoroughly examined accuracy, recall, and F1-score values for each machine learning algorithm to acquire a more detailed picture of their prediction skills. Precision refers to the proportion of genuine positive forecasts among all positive predictions, whereas recall quantifies the proportion of true positive predictions among all real positive cases. The F1-score, which is a harmonic mean of accuracy and recall, gives a fair evaluation of a model's performance in both classes. By evaluating these parameters, we gained insight into the algorithms' ability to effectively categorize stroke and non-stroke patients, revealing their practical relevance in clinical situations.

Furthermore, the investigated effects of feature engineering strategies on model performance, with the goal of identifying the most informative predictors for stroke prediction. We aimed to simplify the input feature space while keeping crucial information related to stroke prediction by meticulously selecting features. Our findings demonstrated that careful feature engineering improved model performance across several methods, emphasizing the significance of feature selection in enhancing prediction accuracy. The Logistic Regression model has an accuracy of 81.57, closely followed by SVM with 77.12. The decision tree and random forest models also performed admirably, with accuracies of 92.77 and 95.36, respectively. Our technique not only enhanced prediction accuracy but also aided in the discovery of critical biomarkers and risk variables related to stroke incidence, adding to the larger field of stroke research and clinical practice.

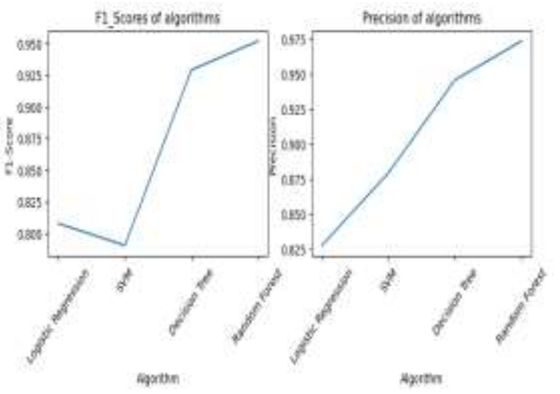


Fig.7 f1_scores and precisions

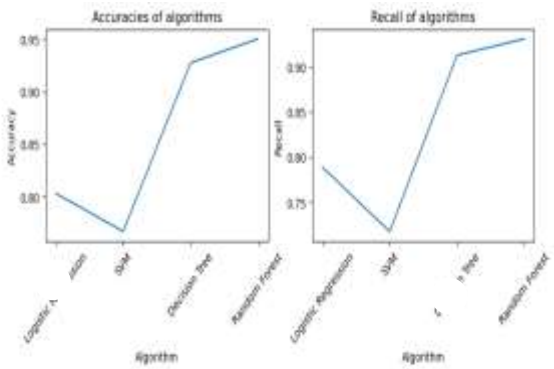


Fig.8 Accuracies and Recall

Conclusion

In conclusion, our study demonstrates the efficiency of machine learning techniques in early stroke prediction, highlighting the importance of leveraging data-driven approaches for proactive healthcare interventions. By harnessing the power of advanced analytics and predictive modeling, healthcare providers can identify individuals at high risk of strokes and implement targeted preventive measures to mitigate the impact of neurovascular crises. Moving forward, further research and validation of our predictive model on diverse datasets are warranted to enhance its robustness and generalizability in real-world clinical settings.

Our findings highlight the need of interdisciplinary collaboration in enhancing stroke prediction and preventive initiatives. We may construct comprehensive and contextually appropriate prediction models by promoting collaborations among data scientists, healthcare practitioners, and politicians. Integrating clinical insights with machine learning algorithms allows us to understand complicated disease causes and personalize therapies to particular patients' needs.

Furthermore, our research underlines the importance of continued investment in healthcare infrastructure and technology innovation to facilitate the wider use of predictive analytics in clinical practice. A strong data infrastructure, interoperable electronic health records, and secure data sharing platforms are critical enablers for using machine learning in healthcare delivery. Furthermore, investing in workforce training and capacity-building projects is critical for equipping healthcare workers with the necessary skills to properly use predictive modeling's potential.

References

1. Manjula Gururaj Rao, K Hemant Kumar Reddy, Sumathi Pawar, Karthik P, Kishan R Kotian, Dhanush P Hegde, "Employing Machine Learning in A Multimodal Approach for Brain Stroke Modelling and Identification", 2023 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), pp.219-224, 2023.
2. E. S. Donkor, "Stroke in the century: a snapshot of the burden epidemiology and quality of life", Stroke research and treatment, 2018.
3. M. Anand Kumar, A. Suresh Kumar, "A Body Area Network Approach for Stroke-Related Disease Diagnosis Using Artificial Intelligence with Deep Learning Techniques", Advances in Computing and Data Sciences, vol.1613, pp.243, 2022.
4. Johnson, Catherine Owens, et al. "Global, regional, and national burden of stroke, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016." The Lancet Neurology 18.5 (2019): 439-458.
5. S. -F. Sung, C.-Y. Lin and Y. -H. Hu, "EMR-Based Phenotyping of Ischemic Stroke Using Supervised Machine Learning and Text Mining Techniques", in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2922-2931, Oct. 2020.
6. Santhanakrishna C, Akash Kumar Singh, Chigilipalli Sriharsha, Ramya M, Srinivasan R, Jagadeesan S, "Early Cerebral Infarction Detection and Classification Using Machine Learning Approaches", 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), pp.1-7, 2023.
7. Mv Anuprabha, K Madhumitha, K Sowmyadevi, T Jaspar Vinitha Sundari, K Paramasivam, "Prediction of Stroke Severity in Brain Through Machine

Learning", 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), pp.461-467, 2023.

8. IZZET TURKALP AKBASLI, "Brain Stroke Prediction Dataset", Brain Stroke Prediction Dataset | Kaggle, July 2022
9. R. Mahesh et al., "Automated Prediction of Brain Stroke Disease Classification Using Machine Learning Algorithm Techniques", Automated Prediction of Brain Stroke Disease Classification Using Machine Learning Algorithm Techniques, June 2022.
10. R. Kanchana and R. Menaka, "Ischemic stroke lesion detection characterization and classification in CT-images with optimal features selection", Biomed. Eng. Lett, vol. 10, pp. 333-344, 2020.
11. T. D. Musuka, S. B. Wilton, M. Traboulsi and M. D. Hill, "Diagnosis and management of acute ischemic stroke: speed is critical", CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne, vol. 187, no. 12, pp. 887-893, 2015.
12. SATHYANARAYAN RAO, "Ensemble Models with Voting and XGBoost", Ensemble Models with Voting and XGBoost | Kaggle, Aug. 2022.
13. Samantha Lin, "Stroke Prediction. Constructing Prediction Model for the... | by Samantha Lin | Geek Culture | Medium", Medium, Mar. 2021.
14. R, K, A. Johnson, S. Anand, R and M, "Neuroimaging and deep learning for brain stroke detection - A review of recent advancements and prospects", ComputerMethods and Programs in Biomedicine, pp. 197, 2020.
15. C Pathanjali, G Monisha, T Priya, K Ruchita Sudarshan and Samyuktha Bhaskar, "Machine Learning for Predicting Ischemic Stroke", INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT), vol. 09, no. 05, May 2020.
16. A. G. Kolias, P. J. Kirkpatrick and P. J. Hutchinson, "Decompressive craniectomy: past present and future", Nat. Publ. Gr., pp. 1-11, 2013.
17. A. W. Unterberg, J. Stover, B. Kress and K. L. Kiening, "Edema and brain trauma", Neuroscience, vol. 129, no. 4, pp. 1019-1027, 2004.
18. C. Werner and K. Engelhard, "Pathophysiology of traumatic brain injury", Br. J. Anaesth., vol. 99, no. 1, pp. 4-9, 2007.
19. L. Gao et al., "An approach for estimation of the surface area of scalp flap based on the characteristic parameters after decompressive craniectomy", 2014 IEEE Int. Symp. Bioelectron. Bioinformatics IEEE ISBB 2014, pp. 8-11, 2014.
20. W. Hacke, S. Schwab, M. Horn, M. Spranger, M. De Georgia and R. von Kummer, "'Malignant' Middle Cerebral Artery Territory Infarction: Clinical Course and Prognostic Signs", Arch. Neurol., vol. 53, no. 4, pp. 309-315, Apr. 1996
21. B. Ruf et al., "Early decompressive craniectomy and duraplasty for refractory intracranial hypertension in children: results of a pilot study", Crit. Care, vol. 7, no. 6, pp. R133-8, 2003.