iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IX **Month of publication:** September 2023

DOI: https://doi.org/10.22214/ijraset.2023.54857

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 11 Issue IX Sep 2023- Available at www.ijraset.com

Deep Transfer Learning Based Stroke Risk Prediction

Sushma Reddy K B[1], Dr. Manjula G R[2], Nikhil G[3]
Department of CSE Jawaharlal Nehru New College of Engineering Shimoga, Karnataka, India

Abstract: ***Stroke is a medical condition that occurs when there is any blockage or bleeding of the blood vessels either interrupts or reduces the supply of blood to the brain. It causes the disability of multiple organs or unexpected death, if that patient's recognise and address risks at the right time, up to 80% of stroke occurrences can be averted. With the advancement of machine learning in medical science, the early recognition of stroke is very much possible that plays a vital role in diagnosis and getting read of this life taking disease. But it requires large well-labeled data. Due to the strict privacy protection policy in health-care systems, stroke data is usually distributed among different hospitals in small pieces. In addition, the positive and negative instances of such data are extremely imbalanced. Transfer learning can solve small data issue by exploiting the knowledge of a correlated domain, especially when multiple sources of data are available. In this work, deep Transfer Learning based Stroke Risk Prediction scheme is proposed to exploit the knowledge structure from multiple correlated sources and used bayesian optimization for selecting the best parameter. This model is tested in synthetic and real-world scenarios and the random forest gives more accuracy than other models like support vector machines (SVM), decision trees (DT), random forests (RF) and voting classifiers. It also shows the potential of real-world deployment among multiple hospitals aided with 5 G/B5G infrastructures.***
Keywords: Stroke Risk Prediction, Deep Learning, Transfer Learning, Accuracy, Bayesian Optimization.

I. INTRODUCTION

One of the most common illnesses that can cause death or permanent impairment in older individuals worldwide is stroke. For those who survived, the cost of treatment and rehabilitation placed an unnecessarily enormous burden on their families and the healthcare system because one in five stroke sufferers would pass away within a year. Since early interventions to delay the onset of and reduce the risks of stroke are expensive, accurate stroke prediction is highly desired. Clinical investigations have demonstrated that a number of common chronic illnesses, like diabetes and hypertension, are strongly associated with the risk of strokes, despite the small size of the stroke data. When numerous connected sources are available, Deep Transfer Learning techniques offer an acceptable framework to address the issue of minimal data. SRP models may be produced more effectively using the recommended framework. However, the performance of the model is strongly influenced by the quantity of transferred layers and the sequence in which different source domains are transferred. Typical parameter tuning methods like grid and random search typically fail due to the search space's size. The most popular model that is Bayesian Optimisation (BO), a technique for model-based global optimisation of black-box functions, is a Gaussian process because it is simple and adaptable in building a probabilistic model of the objective function. In order to find the best parameter, BO is used

II. LITERATURE SURVEY

Stroke Risk Prediction (SRP) Models have been created using a variety of works that use medical data. Classical machine learning methodologies and deep learning-based approaches can be used to broadly classify these techniques. The best performance in stroke prediction is reportedly achieved by deep neural networks (DNN). Its dependence on the availability of significant amounts of well-labeled data is, however, a well-known problem. It's possible that the amount of trustworthy data needed in a real-world scenario won't be easily accessible. It is often challenging to share stroke data between institutions due to the rigorous privacy protection policies in place in the healthcare system. As a result, tiny portions of the entire collection of stroke data are typically dispersed throughout several institutions. The both positive and negative examples in the stroke data are also incredibly unequal. Thus, the DNN-based SRP models could work poorly in real-world deployment.

Clinical risk prediction has a low-resource medical records challenge that has been addressed by the useful framework MetaPred. MetaPred uses labelled medical records from a high-resource domain in conjunction with deep predictive modelling and model-neutral meta-learning.

©IJRASET: All Rights are Reserved

In order to create a learning process that is more transferrable, developed an objective-level adaptation for MetaPred that not only benefits from quick adaptation at the optimization-level but also takes into consideration the supervision of the high-resources domain. Real-world EHR data are extensively evaluated for risk prediction tasks using 5 cognitive illnesses using a variety of source/target combinations. [1] The goal of this endeavour is to build a machine learning system that can forecast brain strokes using real-time EMG data. The SVM classifier is then trained on simulated data and evaluated on real data. The idea of data augmentation is used to produce samples, which makes it simpler to train with a large amount of data. When the simulation is conducted to determine whether the model behaves as predicted, The suggested model obtained 91.77% accuracy, 90.28% precision, 91.44% recall, and 91.12% F1-Score at the cutoff point.Compared to previous models, the suggested model was doing quite well. Furthermore, compared to competing techniques, the SVM's accuracy, recall, and f-measure are all greater. [2]
Presented a partitioning-Stacking prediction fusion approach based on enhanced attention of U-net network in order to accurately segment 3D stroke lesions on a T1-weight 3D MRI without adding an exorbitant computational cost. First, a more detailed segmentation ability may be achieved by the partition operation, which can also substantially enhance the representing capacity of comparable visuals. The accuracy of segmentation is considerably increased by the use of the 3D environment in the fusion procedure. The total segmentation performance is substantially enhanced by our improved attention model. A high degree of segmentation performance for our technique is shown by experiments on the complete ATLAS and BRATS 2017 data sets compared to other recent studies in the field. [3] The prediction of the stroke type is a growing problem for world health. The current technique is manual, time-consuming, costly, and requires greater experience from the doctor. The majority of studies on the robotized analysis of stroke and its subtypes focused on image preparation techniques, CT scan, and MRI. An artificial neural network provides a generic approach to problems. It uses a neural order computation approach to predict stroke infections. A stroke sickness prediction based on an artificial neural network improves the analytical precision with more consistency. The proposed approach automates stroke prediction using AI or machine learning methods. [4] A problem of great medical relevance was examined to see if deep neural network models could predict ischemic stroke. The first trials produced great results on the test set and were encouraging.Due to the fact that stroke-positive pictures come from a single source, however, dataset ablation and feature separation via vascular tree extraction were attempts to measure the impact of these parameters on performance. It was discovered that whereas vascular pictures alone were still sufficient to distinguish between ischemic stroke and non-ischemic stroke, model performance was not entirely generalizable to images from unknown sources. [5]

III. METHODOLOGY

A. Deep Transfer Learning

A machine learning approach called deep transfer learning makes use of the information gained from one activity to enhance the performance of another related task. This method, which is based on the idea that the lower layers of a neural network, like the convolutional layers, learn general features that are useful for a wide range of tasks while the higher layers learn task-specific features is particularly helpful when there is a dearth of labelled data for the target task. It allows the model to leverage the knowledge learned from a similar task with a larger dataset. A model may be modified to do a new job while keeping the broad features discovered while performing the initial task by fine-tuning the weights of the higher layers.

B. Deep Neural Networks

Deep learning's (DL) main goal is to create massive neural networks with complex input-output changes. The linking of a photo to DL is one modern implementation, the name of the person or individuals in the picture, similar to how online entertainment is put together. The identification of pictures using verbal cues is an innovative use of DL.

C. Convolutional Neural Network

Convolutional neural networks, one type of deep learning neural network design, are often used in computer vision. The subject of artificial intelligence known as computer vision enables a computer to comprehend and interpret an image or other visual data. Amazingly good in machine learning are artificial neural networks. Text, audio, and image data are just a few of the datasets that neural networks are applied to. For varied purposes, multiple neural network types are used. For instance, categorising pictures requires the use of convolutional neural networks, and this blog will construct a basic component for CNN. Identifying the sequence of words requires the use of recurrent neural networks, especially an LSTM.

D. CNN + LSTM

A CNN-LSTM (Convolutional neural networks with long short term memory) model gathers qualities from the data it receives, while the LSTM layers forecast trends. For data that are consecutive, a brief information structure called a period series is typically used. Given LSTM's performance with patterns, it was decided to use it as the DNN technique. CNN is frequently useful when seeking for information about a location, such as in a picture.

E. Bayesian Optimization

In the domain of machine learning, Bayesian optimisation has been widely applied for the purpose of hyperparameter tweaking. Despite the complexity of the terminology and mathematical calculations, the underlying idea is actually rather straightforward. The way that Bayesian Optimisation differentiates from Random Search and Grid Search is that it accelerates search times based on past results, whereas the other two approaches are uniform (or devoid of) historical judgments.

IV. IMPLEMENTATION

In this work, the learning objective in the target domain is to estimate the probability of stroke in a nearby hospital (see Fig. 1). Since the target domain's stroke data is frequently insufficient, the learning job can be enhanced by using a variety of source domains associated with multiple conditions (such as hypertension and diabetes). Medical datasets are gathered and maintained in various hospitals, however direct data sharing between hospitals is impractical because of the healthcare system's privacy protection policies. A reasonable address is to define a source domain as a particular illness at a particular hospital. The next step is to transfer knowledge structures from various source domains to both local and other hospitals in order to enhance the learning task in the target domain (i.e., stroke risk prediction at local hospital). One local hospital and K outside hospitals are taken into consideration in a scenario that illustrates the target and source domains. The suggested Hybrid Deep Transfer Learning-based Stroke Risk Prediction (HDTL-SRP) architecture serves as the system's brain, using all pertinent data sources to develop an accurate SRP model and the models accuracy is compared as shown in Fig 2. This paper suggests a Hybrid Deep Transfer Learning (HDTL) technique that transfers knowledge structure from several source domains dispersed throughout different hospitals to the target domain of stroke in order to address the problems brought on by tiny and unbalanced stroke data. The proposed HDTL-SRP infrastructure operates in a distributed manner without requiring hospitals to directly share patient records. It consists of four parts:

1) Generative Instance Transfer (GIT), which applies GAN to external data to generate synthetic instances for model training purposes
2) Network Weight Transfer (NWT), which uses data from highly correlated diseases like diabetes or hypertension
3) Bayesian Optimisation (BO), which seeks out the best transferred parameters
4) Active Instance Transfer (AIT), which chooses more informative synthetic stroke instances to create a balanced model.

A. Advantages of proposed system

1) The suggested framework can create SRP models more successfully.
2) Due to its ease of use and flexibility in creating a probabilistic model of the objective function, the Gaussian process is the most often used model for Bayesian Optimisation (BO), a method for model-based global optimisation of blackbox functions.

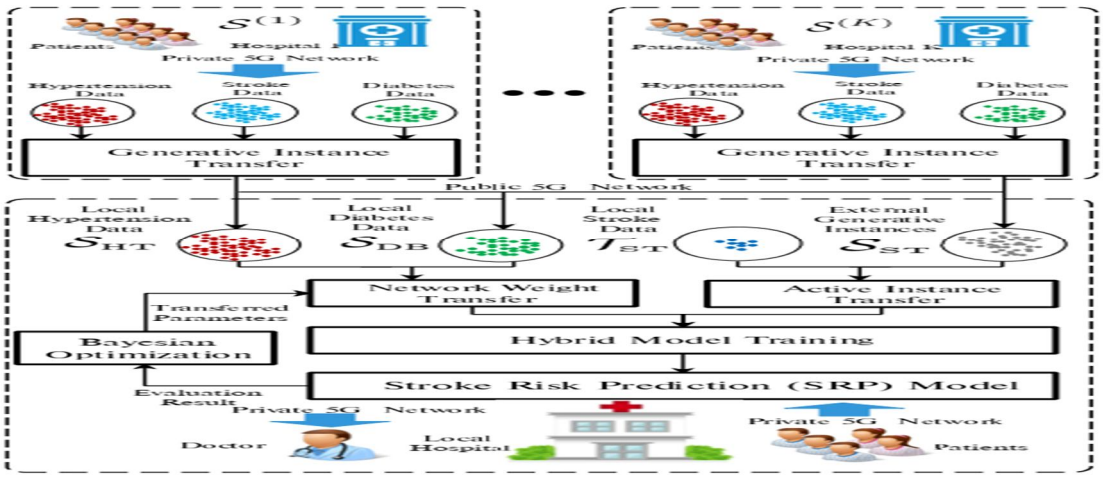


Fig. 1 System architecture

B. Algorithms

1) *Support Vector Machine (SVM):* Regression and classification problems may both be solved using SVM, a supervised machine learning approach. Even though refer to them as regression concerns, classification is the most appropriate term. Finding a hyperplane in an N-dimensional space that categorises the input points with clarity is the aim of the SVM method.
2) *Decision trees (DT):* They are supervised learning methods that are non-parametric and may be applied to classification and regression problems. A root node, branches, internal nodes, and leaf nodes make up its hierarchical tree structure.
3) *Random Forest(RF):* The Random Forest technique, a directed machine learning strategy, is commonly used when dealing with classification and regression problems. It is aware that forests include a lot of trees and that when new trees are added, the forest becomes larger.
4) *Voting Classifier:* Kagglers frequently use Voting Classifier, a machine learning technique, to enhance the performance of their model and advance up the rank ladder. Voting Classifier has major limitations, yet it may be used to improve performance on real-world datasets.



Fig. 2 Graph plotted based on algorithms accuracy

V. RESULTS

The Stroke risk prediction model results as shown below. The user have to signup Firstly and fill the Register Form with some details like username, emailid, mobile number , password and the snapshort of register page is shown in the fig. 3 after register the user can signin by using username and password, than submit the patients information as shown in fig. 4 and the predicted results will be displayed if the person having stroke risk or not as shown in fig. 5 and 6 the results will be predicted based on the person's health status.

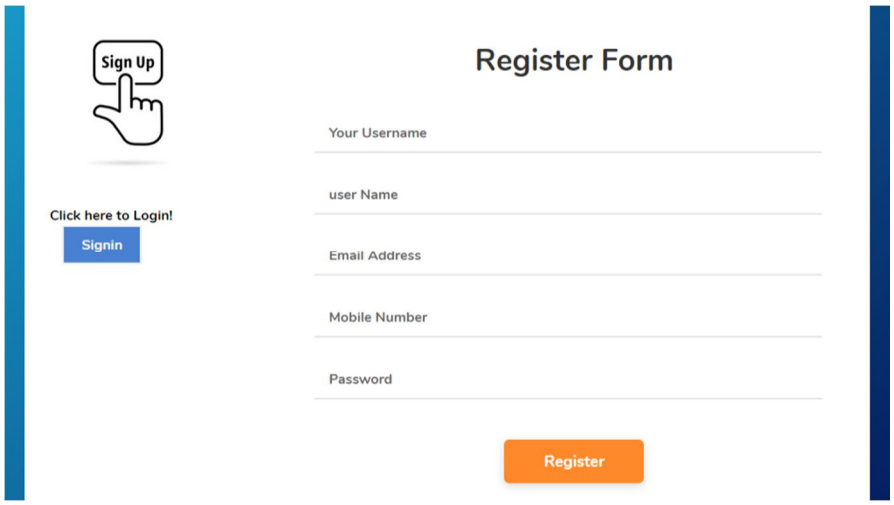


Fig. 3 user registration



Fig. 4 User input



Fig. 5 Prediction result

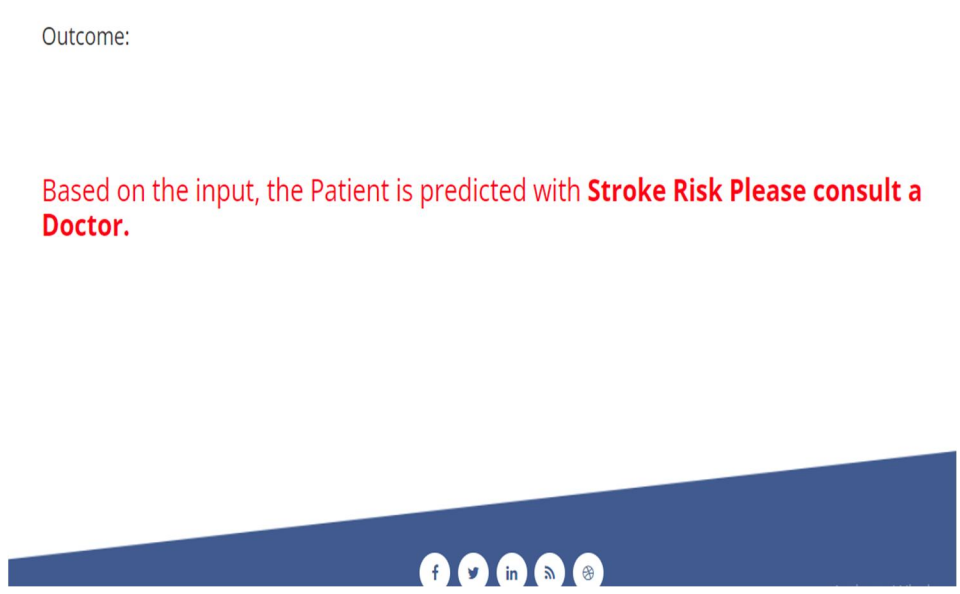


Fig. 6 Prediction result

©IJRASET: All Rights are Reserved

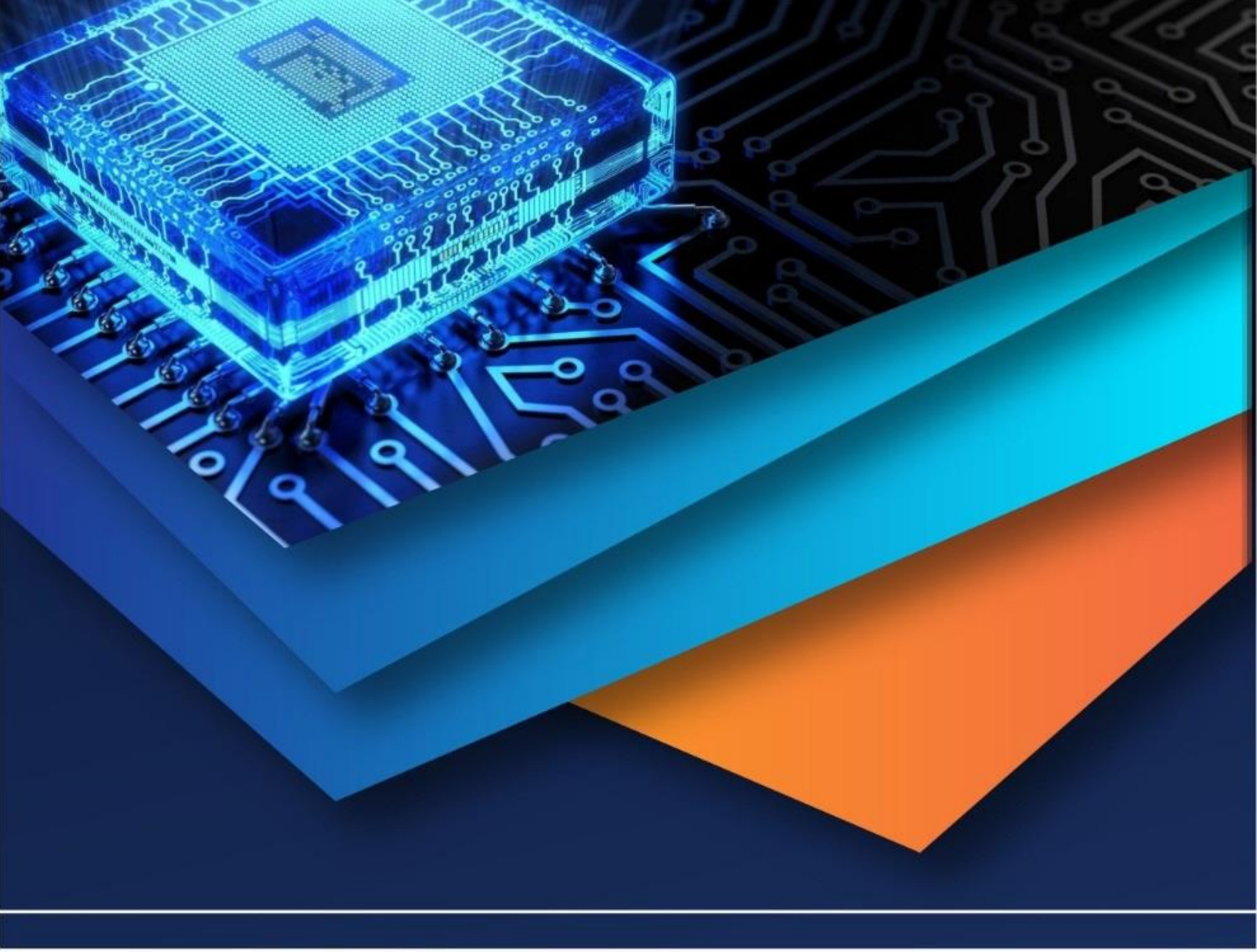
VI. CONCLUSION

Stroke is a life-threatening medical illness that should be treated as soon as possible to avoid further complications, the creation of a machine learning (ML) model could benefit in the early diagnosis of stroke and the subsequent reduction of its severe repercussions. The effectiveness of several ML algorithms in properly predicting stroke based on a number of physiological variables is analyzed. The experiment is performed over health care dataset collected from Kaggle. The different classification models like support vector machines, decision trees , random forests , and voting classifiers and a deep transfer learning model is developed to predict the possibility of whether a person will having a stroke risk or not and evaluating the model's performance. The experimental result shows that the proposed model is more effective than other existing models because it gives 99% of accuracy and the random forest is best suited in this case, this model can be used to diagnose a patient to determine the possibility of getting a stroke risk at an early stage. Finally, the analysis of stroke concerning different attributes discovered and it can be developed in hospitals having 5G/B5G infrastructure.

REFERENCES

[1] Soumyabrata Dev, Hewei Wang, Chidozie Shamrock Nwosu, Nishtha Jain , Bharadwaj Veeravalli and Deepu John, "A predictive analytics approach for stroke prediction using machine learning and neural networks", Elsevier, pp. 2772-4425, 2022.
[2] Elias Dritsas and Maria Trigka, "Stroke Risk Prediction with Machine Learning Techniques", MDPI, pp. 2105-4123, 2022.
[3] Md. Ashrafuzzaman, Suman Saha, and Kamruddin Nur, "Prediction of Stroke Disease Using Deep CNN Based Approach", Journal of Advances in Information Technology, vol. 13, no. 6, pp 604-614, 2022.
[4] R Pitchai, Bhasker Dappuri, P V Pramila, M Vidhyalakshmi, S Shanthi, Wadi B Alonazi, Khalid M A Almutairi, R S Sundaram and Ibsa Beyene, "An Artificial Intelligence-Based Bio-Medical Stroke Prediction and Analytical System Using a Machine Learning Approach", Hindawi, 2022.
[5] Tahia Tazin , Md Nur Alam, Nahian Nakiba Dola, Mohammad Sajibul Bari, Sami Bourouis , and Mohammad Monirujjaman Khan, "Stroke Disease Detection and Prediction Using Robust Learning Approaches", Hindawi, 2022.
[6] Gangavarapu Sailasya1 and Gorli L Aruna Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms", IJACSA, vol. 12, no. 6, pp 539-545, 2021.
[7] Adriano Pinto, Joana Amorim, Arsany Hakim, Victor Alves, Mauricio Reyes and Carlos A. Silva, "Prediction of Stroke Lesion at 90-Day Follow-Up by Fusing Raw DSC-MRI With Parametric Maps Using Deep Learning", IEEE Access, vol. 9, pp 26260- 26270, 2021.
[8] Kunder Akash Mahesh, Shashank H N, Srikanth S and Thejas A M, "Prediction Of Stroke Using Machine Learning", 2020.
[9] Haisheng Hui , Xueying Zhang, Fenglian Li , Xiaobi Mei and Yuling Guo, "A stroke is a Partitioning-Stacking Prediction Fusion Network Based on an Improved Attention U-Net for Stroke Lesion Segmentation", IEEE Access, vol 8, pp 47419- 47432, 2020.
[10] Songhee Cheon, Jungyoon Kim and Jihye Lim, "The Use of Deep Learning to Predict Stroke Patient Mortality", MDPI, pp 1-12, 2019.
[11] Xi Sheryl Zhang, Fengyi Tang, Hiroko Dodge, Jiayu Zhou, Fei Wang, "MetaPred: Meta Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records", ACM, 2019.
[12] Jahnavi V, Ankitha S, Harshavardhan N and Deepthi M, "An Artificial Intelligence of Approach for Predicting Different Types of Stroke", IJERT, vol. 8, pp 200-204, 2019.
[13] Gilbert Lim, Zhan Wei Lim, Dejiang Xu, Daniel S.W. Ting, Tien Yin Wong, Mong Li Lee and Wynne Hsu, "Feature Isolation for Hypothesis Testing in Retinal Imaging: An Ischemic Stroke Prediction Case Study", 2019.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)