A NEARLY DIAGNOSIS OF I SCHEMIC STROKE USING HYBRIDIZED MACHINE LEARNING ALGORITHM

Fathima S K[1], Kavipriya M[2], Haripriya V[3], Gokul [4]

[1]*AssistantProfessor CSE,SonaCollegeofTechnology, fathima.sk@sonatech.ac.in.*

[2]Student, *Sona College of Technology,Kavipriya.20cse@sonatech.ac.in.*

[3]Student,*SonaCollegeofTechnology,Haripriya.20cse@sonatech.ac.in.*

[4]Student,*SonaCollegeofTechnology,Gokul.20cse@sonatech.ac.in*

Abstract

The evolution of new technologies and approaches paves the way for non-invasive techniques in healthcare system. Speaking about strokes, it is one of the life-threateningdiseases, but there's a chance to save patient by predicting it early. Cardiovascular diseases are classified on the basis of MRI and CT scan images which are actually quite expensive. As there is a demand for non-invasivelow-cost methods for diagnosis, we propose an approach where forecasting of strokes is totally based on extracted EEG data. Classification of EEG signals is important to construct an accurate brain computer interface (BCI). There are several approaches for deep learning and machine learning techniques to classify EEG data, but dynamics of Brain is quite complex across it's mental task, hence it is a complicated task to built efficient algorithm with prior knowledge.so,2D Alex net in convolutional neural network(CNN) is used to analyse EEG covers different mental task in this study.

In this paper we are using machine learning algorithm like Random Forest, Logistic Regression and Synthetic Minority Oversampling Techniques. And going to develop a model called" Hybridized ML model to predict the strokes."

Keywords – Stroke, EEG, Machinelearning, Random forest, Logistic Regression.

1. INTRODUCTION

Nowadays, our lifestyle has changed a lot due to technical advancements and gadgets. Irrespective of age factor many are facing several health issues like hypertension, stroke, diabetes and cardiovascular diseases as a result of deleterious life. For a kind information, stroke has become a common disease which could be fatal or could cause long term disability. Approximately 20 million people experience by a year where 33% is left with disability and 40 to 45% result in death. Reportssays, by 2030 that will be a result of 200 million death cases globally. The cost for treatment & rehabilitation is extremely high which is difficult to afford.

In recent report, from 2014 -2015, the cost due to stroke incidents where around 50 billion dollars (1).Hence, stroke prediction is quite costly and it is highly desirable to reduce the risk. Predicting functional outgrowth after stroke would assist the doctors to take specific decisions.

Thus we are focusing on strokes in this paper and we attempted to build a system which uses bioelectrical images to predict. Machine learning has become a key factor in health technologies, which improves the quality and plays a vital role in improvisation of the system.

While use biomarkers and non-invasive technique to monitor the health. There are many other works which utilize ML to develop Strokes Risk Predictions (SRP) Models. These methods are classified as classical machine learning approaches like Decision Tree, Logistic Regression, Support Vector Machine (SVM)and deep neural network (DNN) best performance in stroke prediction can be achieved, but unfortunately those models rely on availability of large data whereas in reality such amount of data is not available.

SignalsfrombrainareusedtotakecontrolovermachinebyBrainComputerInterface(BCI). BCI system is enabled to translate brain activities into multiple task by the EEG signal recording by experiments conducted during training process. Neuro prosthetic based application such as improving vision and hearing impairment were mainly focused by BCI according to current trends.Likewise nervous system impairment brain related problems as well as damage in sensory organ can be replaced with the help of prosthesis [2]

The main challenges of BCI system in EEG signal processing is making a liable interface for a variety of BCI operations. The data which is give as input for training is come from continuous EEG data recording. Better performance will be provided by Deep learning if the size of the recorded data is high. Famouslyfor classification of MI signals for image representation CNN or Recurrent Neural Network (RNNs) are currently using technique.There are many studies available.Polar projection method was utilized to extract different frequency bands from power spectrum of each electrode which index is used to map the electrode location from 3D to 2D projection.For the given EEG time window each electrode is mapped to its location onto a skill like image, to construct an EEG topographic map,at each time index.

Atfirst,thestudyaddsspatialandtemporaldimensions of EEG signals to a 2D EEG topographicmap then, topographic maps at different time indiceswere cascaded to populate a 2D image for a giventimewindow. At last, Alex Net was enabled by topographic maps to learn features from the spatial and temporal dimension of the brain signals. The classification accuracy for BCI system was improved by showing the result by converting the EEG classification problem from 1D static to 2D image. Therandomforestalgorithm with accuracy 95.44% and the model wedeveloped gave a best accuracy rate of approximately91%.Also, to colonize 2D topographic image the illustrated topographical characters were converted to grayscale, cascaded sequentially per each subject and resized.The signal activation changes in both temporal and spatial dimensions are represented by the constructed EEG topographical images. Also, the completely different EEG MI takes are classified from the constructed topographical images by Alex Net CNN architecture.

Large scale studies show the result in favor of classierapproaches like Random Forest (RF). RF has beenused successfully in various biomedical applications,suchastheautomatedpulsedetectionduringelectrocardiogram-based cardiopulmonaryresuscitation.
According to stroke, most of the researches neuroimaging and outcome estimation to detect (IS) ischemic stroke lesions with the use of ML methods.If has only been recently, however in a group of non-traumatic intracerebral hemorrhage (ICH) patients a study estimated stroke outcome prediction at 3 months. Using a nationwide disease registry.
Previous studies concluded that ML techniques can beeffective topredict functionaloutcomeofIS long-termpatientsorforpredictionofsymptomaticintracranialhemorrhage following thrombolysis from CT images, stillall works agree on the need to carry out further studies inorder to confirm results, incorporate new variables andresolve their limitations/weaknesses.The chance of recovery of the patients ,the accurate prediction of the stroke is essential for that we are taking into account the frequency of cerebrovascular diseases .The famousmachinelearningmethodstogetthebestresultofprediction of stroke or any other diseases are RandomForest (RF), Logistic Regression (LR) Synthetic MinorityOversamplingTechnique (SMOTE).
The control of remedial homeostasis need to be increased and select and follow up for reperfusion has to be possible hence each patient needs should be addressed by constructing predictive model which find stroke patient at worst case. Likewise, identifying the most suitable cases is essential to respond exactly to treatments regarding to new regenerative cellular or molecular therapies.

2. LITERATURESURVEY

In[4]theauthorshaveusedtwoalgorithmsnamelydecision tree and random forest to classify and analysisstroke. Decision tree is for feature extraction and randomforestalgorithmisforpatternclassification.Theyusedindependent variables like age, hypertension, blood sugarlevel, body mass index (BMI), married status and history ofheart disease. In the article [5], the deep learning-basedclassificationanddecisiontreealgorithmhaveusedtoexhibits strong classification performances with a little runtimeandcomputationalcostandnotsuitableforonline prediction, butbetterforoffline.In[6],theauthorhadhatched one device to predict the disease of stroke alongwithlocations,timeanddurationusingtherawEEGsignals.Thestudyinthearticle[7]representstheresultsforchangingtheEEGclassificationproblemfromonedimensiontimeseriestoatwo-dimensionimageclassification problem with average accurate value of 81.09%. In[8], the author has developed a prediction model in whichtheEEGsignalsaretransformedtoimagesandclassification of signals to predict the early stage of strokewith accuracy of 70.64%. The article [9] is describing thesolution and model to deal with small data and imbalanceddata in the prediction of early stroke by using ElectronicHealthRecord(EHR)asdatafortheirmodel.Thestudyin
[10] is concocting the software-based model where there isaclassificationofdepressionbrainsignalsandnormalbrain signals which used to predict or induce stroke byusing EEG signals. There is no requirement for a set offeatures to fed a classifier for classification. In [11,22], theauthorhas given an improvement for the diagnosis ofstrokebyconsolidatingtheelectroencephalogram(EEG)andgalvanic skin response (GSR) signals by using theopendatasetforemotionanalysis,buttheaccuracyis73.4%. In [12], the automated detection of ischemic strokeby using EEG data which is gathered from EEG sensorsand the overall performance of the developed model is wasmeasured by the value of accuracy, sensitivity, specificity.In[13,23],theauthorhasusedmachinelearningtechniques for classifying the stroke signals and normalsignals and also the bioelectrical signals are combined withnatural processing language (NLP) for the better output forunderstanding.In[14],fordetectionEEGbasedimageclassificationtheauthorshaveusedbi-directionaldeeplearningframeworks.Theclassificationisinabnormalsignals and normal signals and this gives the distinguishbetween two hemispheres (right and left). In the article [15],theauthors aretryingtopredict thecerebralstroke fordiagnosisbyusingthephysiologicaldata ofincompleteandimbalanceddatasetusingthedeeplearningapproaches.Itreduces the false negative rate effectively.In [16, 17, 18], there are many approaches that can dealwith strokes to avoid the human's death and damage ofheart and brain tissues. The already available techniqueshave some pitfall the prediction like producing the correctresults, time and space complexities, if we are using inputdata in the form of MRI and CT scans means it costs highwhich is notsuitable foralleconomic levelpeople. Andalsofindingtheaccurateresultforclassificationofministroke (TIA) and actual stroke is somewhat difficult tofind.So,tosolveallthesestruggles,thispaperproposedahybridizedmachinelearningtechniquestopredictthe strokewith bestaccuracy.

3. PROPOSEDMETHODOLOGY

Inthissectionwearegoingtoseethedetailedprocessofthebuildingmodel.Themethodologywe

proposedinthispaperisbasedontimeseriesprediction methods because in this we are making thescientific predictions. The flow of the proposed modelisgivenintheFigure1,thefirststepisdatacollection, for this we are collecting the raw data fromEEG signals, MI (motor imaginary) captures the signalsin the form of1D. For converting 1D image to 2Dimage we are using Alex Net CNN). We will convertthe image as a value using in integer to binary vectors"One Hot Coding" is used. We decrease the value oftherangeofthepixelsfrom0to1.Nextstepisfeature extraction for this are using pooling layers inconvolutionalneuralnetworks.Andthenwewillextract the main features of characteristics from thesignals data, and store the data for input data. Basedoncomparison wewill predictthe stroke.

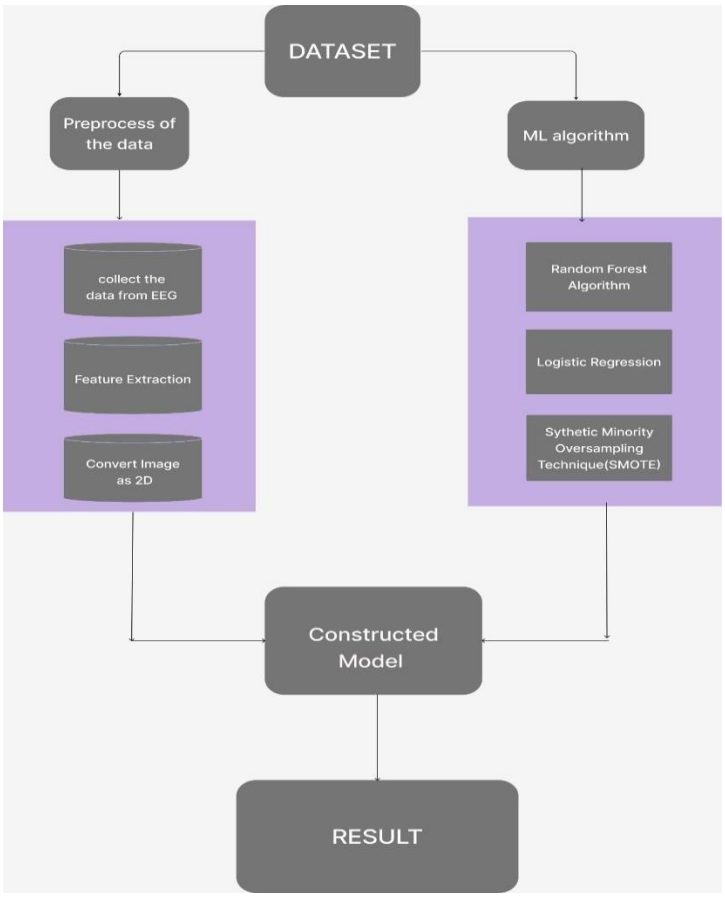

Figure1:Architectureofproposedmethodology

DataCollection

EEG was used to collect dataset for few models in the form of bioelectrical signal. EEG is a componentof ECG+EMG by using this method we are collectingraw data. Sometimes we also get imbalanced data, sotoavoidthoseissuesherewareusingSMOTEtechnique for balancing the dataset in order for bestaccuracy.

Pre-Processing

ThegatheredEEGinformationis from6channels(alpha,beta,gamma,theta,anddelta)atthefrequency1000HZ.Thedatawillbeportedtotheserver as the raw data and the server will create adatabase for the given data and then it will prepare forpreprocessing.

variety of classification methods is used to train the model after the splitting process.Once the model training on offline stroke based data in initiated, the training phase also get initiated.Thetestingisontheonlinetochecktheefficacyofthemodel.

Randomforestandlogisticregressionaretheclassificationalgorithmusedinthisstudy.

Proposedalgorithm

Usingthemachinelearningmethods,weareusingthefollowing algorithmsforthismodel:

1) Randomforest
2) LogisticRegression
3) SMOTE
4) CNN

In CNN algorithm, an input image was taken by the model and is differentiated from one another by assigning the aspects/objects in the image with importance (weight and biases). It is mainly used inimage analysis process, as we are giving image alsoan input, we are using this. It has four layers namelythe convolutional layer, pooling layer, Rectifier LinearActivation Unit (ReLu) correction layer and the fully-connected layer.

Random forest:

Random sample of data is used to train numerous independent decision tress individually which resides on RFs.During the training process,these trees are created and the result of the decision trees are collected as output.The final results are determined by voting process.Each DT must has vote for any one of the output classesAccording to the majority votes,the final predication is made with RF.TheRFclassificationdiagramisshownin Figure 2. The most allowing feature of the RFs is its flexibility and us addition it provides us with

unambiguous expectation since its employs default hyperparameters.

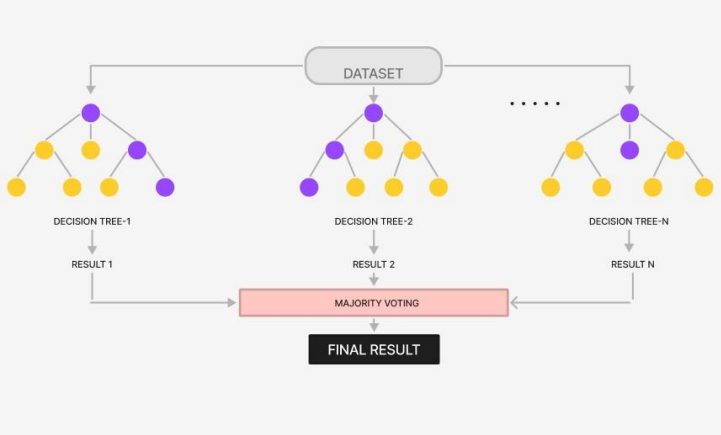

Figure 2:RandomForest

Logisticregression

The most commonly used ML algorithm [20] in the supervised learning approach is LR.Logistic regression used to predict the categorical dependent variable's output.The categorical dependent variable's output is predicted by using logistic regression.The nature of the output should be distance or categoricalthe result should be either 0 or 1. true or false etc.Logical regression & linear regression are more alike.

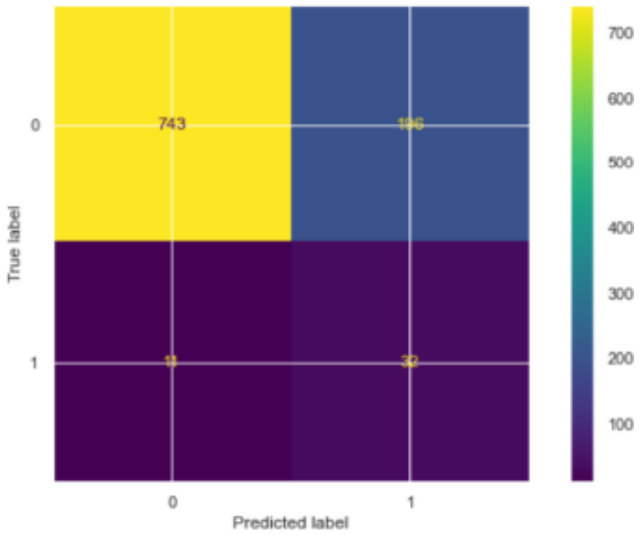

Figure.3. PredictionofLogisticRegression

The linear regression is used to address regression problems &LR is used to address the classification problems.The S-Shaped logistic function is used to forecast the two maximum values (0 or 1)

LRequation,

$$ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 x$$

$$=> P = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

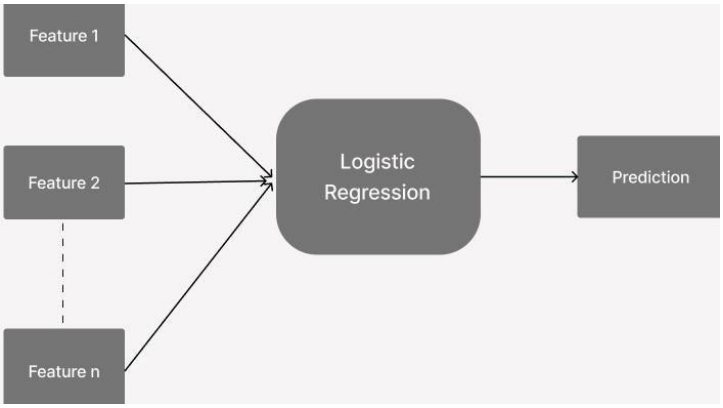

Figure4:Logistic Regression

Balanced dataset after applying SMOTE

Synthetic Minority Over-SamplingTechnique(SMOTE):

TheSMOTEmethod[21]isusedtobalancetheimbalancedataset.Inrealtimeworldgettingdataset,theremaybeapossibilityofgettingimbalancedorsmall data, so to solve this we are using this method.SyntheticMinorityOversamplingTechnique(SMOTE)isastatistical Technique is used to rise the numberof cases in a balanced way by smote.From the existing minority cases, the new instance is generated based on which the components works. The number of majority cases cannot be changed by the implementation of the smote.

SMOTE takes the whole dataset and processing if it is a minority cases, the percentage will be increased.

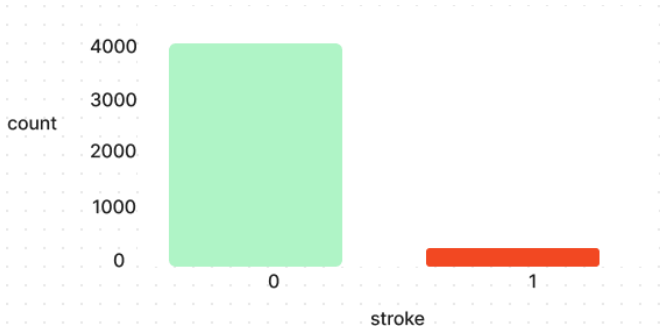

Figure.5.DataImbalance

4. RESULT ANALYSIS

From our analysis we conclude thatRandom Forest model is more efficient and easier of among algorithm by our analysis.Ithas more precision andbetter recall. Even though we have lotsof algorithm, we prefer because of its higher level ofprecise.Thispaperachieved95.44%precise by using SMOTEthe imbalanced datasetalso balancedin order to achieve best accuracyinpredictionofstroke.

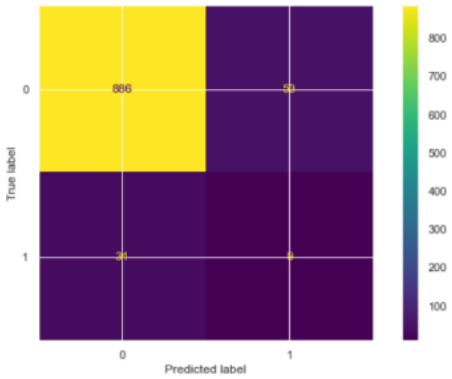

Figure.6. PredictionofRandomForest

5. CONCLUSION

Stroke is a very serious, life threatening medical condition in our society that must behealed before it harms or affect the human living. Fabricating aneffective machinelearningmodelcanhelpintheearlyforecasting and prediction of stroke and downgrade the serious impact ofthefuture. Based on multiple physiologicalvariables investigated in this study, we had constructed a predictive model to predict the stroke (Ischemic) in the early stage using various ML algorithms. According to theresearch studies,theRandomForestmethodwas testedwithbest accuracyof95.44%andduring cross validation the forecasting of brain stroke also found. Infuture we can enhance our model to use larger dataset by using an advancedmachine learning we can also build a model with moreaccuracy than we predictednow.

REFERENCES

[1].E. J. Benjamin, P.Muntner, and. S. Bittencourt, "Heart disease and stroke statistics—2019 update: A report from theAmericanHeartAssociation,"Circulation,vol.139,no.10,pp.e56–e528,2019.].

[2].J. Syed, H. Tehreem, M. Rabia, A. Muhammad, R. Talha, U. Muhammad and M. Muhammad, Wireless BrainComputer Interface for Smart Home and Medical System. Wireless Personal Communications. (2018). doi:10.1007/s11277-018-5932-x.

[3].P. Zhang, X. Wang, W. Zhang and J. Chen, Learning SpatialSpectral-Temporal EEG Features With Recurrent 3DConvolutional Neural Networks for Cross-Task Mental Workload Assessment. IEEE Trans Neural SystRehabil Eng. 2019Jan;27(1):31-42.doi:10.1109/TNSRE.2018.2884641.

[4] D. I. Puspitasari, A. F. Riza Kholdani, A. Dharmawati, M. E. Rosadi, and W. M. P. Dhuhita, "Stroke disease analysisand classification using decision tree and random forest methods,"in 2021 Sixth International Conference on InformaticsandComputing (ICIC),pp.1–4,Jakarta,Indonesia,2021.

[5] S.Chambon,M.N.Galtier,P.J.Arnal,G.Wainrib,andA.Gramfort,"Adeeplearningarchitecturefortemporalsleepstage classification using multivariate and multimodal timeseries," IEEE Transactions on Neural Systems andRehabilitation Engineering,vol.26,no.4,pp.758–769,2018

[6] S.Chambon,V.Thorey,P.Arnal,E.Mignot,andA.Gramfort,"DOSED:adeep learningapproachto detectmultiplesleepmicro-eventsinEEGsignal,"Journal ofNeuroscienceMethods,vol.321,pp.64–78,2019.

[7] A.M.AnwarandA.M.Eldeib,"EEGsignalclassification usingconvolutionalneuralnetworksoncombinedspatialandtemporaldimensionsforBCI systems,"in202042ndAnnualInternationalConferenceoftheIEEEEngineeringinMedicine &Biology Society(EMBC),pp.434–437,Montreal,QC,Canada,2020.

[8] W. Fadel, C. Kollod, M. Wahdow, Y. Ibrahim, and I. Ulbert,"Multi-class classification of motor imagery EEG signalsusing image-based deep recurrent convolutional neural network," in 2020 8th Int. Wint. Confe. on Brain-ComputerInterface(BCI),pp.1–4,Gangwon,Korea (South),2020.

[9] J.Chen,Y.Chen,J.Li,J.Wang,Z.LinandA.K.Nandi,"StrokeRisk PredictionWithHybridDeepTransfer LearningFramework,"inIEEE JournalofBiomedicalandHealthInformatics,vol.26,no.1,pp.411-422,Jan.2022.

[10] U.R.Acharya,S.L.Oh,Y.Hagiwara,J.H.Tan,H.Adeli,andD.P.Subha,“AutomatedEEG-basedscreeningofdepression using deep convolutional neural network,” Computer Methods and Programs in Biomedicine, vol. 161, pp.103–113,2018.

[11] Y. H.Kwon,S.B.Shin, andS. D.Kim,“Electroencephalographybasedfusiontwo-dimensional(2D)-convolutionneuralnetworks(CNN) modelforemotionrecognitionsystem,”Sensors,vol.18,no.5,p.1383,2018.

[12] H. A. Adhi, S. K. Wijaya, C. Badri, and M. Rezal, “Automatic detection of ischemic stroke based on scaling exponentelectroencephalogram using extreme learning machine,” Journal of Physics: Conference Series, vol. 820, pp. 12005–12013,2017.

[13] E.C.Djamal, R.I.Ramadhan, M.I.Mandasari, andD. Djajasasmita,“Identificationof post-strokeEEGsignalusing wavelet and convolutional neural networks,” Bulletin of Electrical Engineering and Informatics, vol. 9, no. 5,pp.1890–1898,2020.

[14] A. Fares, S. H. Zhong, and J. Jiang, “EEG-based image classification via a region-level stacked bi-directional deeplearningframework,”BMCMedical Informaticsand DecisionMaking,vol.19,no.268,pp.1–11,2019.

[15] TianyuLiu,WenhuiFan,ChengWu,“A hybridmachinelearningapproachtocerebralstrokepredictionbasedonimbalancedmedicaldataset”,ArtificialIntelligence inMedicine,Volume101,2019.

[16] T. J. Kleinig and R. Vink, “Suppression of inflammation in ischemic and hemorrhagic stroke: therapeutic options,”CurrentOpinionin Neurology,vol.22,no.3,pp.294–301,2009.

[17] J. Kamtchum-Tatuene and G. C. Jickling, “Blood biomarkers for stroke diagnosis and management,” NeuroMolecularMedicine,vol.21,no.4,pp.344–368,2019.

[18] N. H. Chen, Y. M. Zhang, F. P. Jiang et al., “FLAIR vascular hyperintensity predicts early neurological deterioration inpatientswithacuteischemicstroke receivingendovasculartherapy,”NeurologicalSciences,2022.

[19]Yanyu Chen, Wenzhe Zheng, Wenbo Li, YimiaoHuang,“Large group activity security risk assessment and risk earlywarningbased onrandom forestalgorithm”,PatternRecognition Letters,Volume144,2021,

[20]Simon Nusinovici, Yih Chung Tham, Marco Yu Chak Yan, Daniel Shu Wei Ting, Jialiang Li, CharumathiSabanayagam, Tien Yin Wong, Ching-Yu Cheng ,Logistic regression was as good as machine learning for predictingmajor chronic diseases,JournalofClinical Epidemiology,Volume 122,2020.

[21] K.Cheng,C.Zhang,H.Yu,X.Yang,H.ZouandS.Gao,"GroupedSMOTEWithNoiseFilteringMechanismforClassifyingImbalanced Data,"inIEEEAccess,vol.7,pp.170668-170681,2019.

[22] Fathima, S. K., Velammal, B. L., Shanmugam, K., &Jayareka, K. S. (2021). AnintegratedIot based ApproachEnabled in UAV for the Early Prediction of Forest Fires. Annals of the Romanian Society for Cell Biology, 25(6), 11042-11054.

[23] Fathima,S.K.,Velammal,B.L.,Shanmugam,K.,&Vinodhini,V.AFuturistic ApproachOnEffectiveImplementationOf Enabling Technologies In The Internet Of Things (Iot) For The Establishment Of Smart Health Care ManagementSystem.Turkish Journal ofPhysiotherapyand Rehabilitation,32,3.