International Journal on Applications in Information and Communication Engineering
Volume 7 : Issue 1 : March 2021, pp 21 – 27 www.aetsjournal.com ISSN (Online) : 2394 - 6237

ISCHEMIC STROKE SUBTYPING USING RECURSIVE FEATURE ELIMINATION ALGORITHM

T.PREMAMALA M.E , AKSHYABHARATHI.M, DHARANI.P, DHARSHINI.B, MALATHI.M

Abstract— Ischemic stroke subtyping was not only highly valuable for effective intervention and treatment, but also important to the prognosis of ischemic stroke. The manual adjudication of disease classification was time-consuming, error-prone, and limits scaling to large datasets. In this study, an integrated machine learning approach was used to classify the subtype of ischemic stroke on The International Stroke Trial (IST) dataset. It considered the common problems of feature selection and prediction in medical datasets. Firstly, the importance of features were ranked by the Shapiro-Wilk algorithm and Pearson correlations between features were analyzed. Early diagnosis of stroke is essential for timely prevention and treatment. Investigation shows that measures extracted from various risk parameters carry valuable information for the prediction of stroke. This work investigates the various physiological parameters that are used as risk factors for the prediction of stroke.

Data was collected from International Stroke Trial database and was successfully trained and tested using Sequential Minimal Optimization. Then, used Recursive Feature Elimination with CrossValidation (**RFECV**), which incorporated linear SVC, RandomForestClassifier, ExtraTreesClassifier, AdobosClassifier, and MultinomialNaïveBayesClassifier as estimator respectively, to select robust features important to ischemic stroke subtyping. Furthermore, the importance of selected features was determined by Extra Trees-Classifier. Finally, the selected features were used by Extra-Trees-Classifier and a simple deep learning model to classify the ischemic stroke subtype on IST dataset. It was suggested that the described method could classify ischemic stroke subtype accurately. And the result showed that the machine learning approaches outperformed human professionals.

I. INTRODUCTION

Stroke had become a major cause of disability worldwide. It was predicted that by 2030, there could be almost 70 million stroke survivors, and more than 200 million disability adjusted life-years (DALYs) lost from stroke each year. Stroke burden in high-income countries was very serious, and the burden of stroke increases rapidly in low-income and middle-income countries in recent years with the rapid development of social economy. Classification of ischemic stroke subtype required synthesis of historical, examination, laboratory, electrocardiographic, and imaging data to infer a mechanism and assign causal, etiologic, or phenotypic classification.

II. REVIEW OF LITERATURE SURVEY

1) MACHINE LEARNING–BASED MODEL FOR PREDICTION OF OUTCOMES IN ACUTE STROKE

E.M. Arsava, K.L. Furie (2007) [1] prediction of long-term outcomes in ischemic stroke patients may be useful in treatment decisions. Machine learning techniques are being increasingly adapted for use in the medical field because of their high accuracy. This study investigated the applicability of machine learning techniques to predict long-term outcomes in ischemic stroke patients. Method this was a retrospective study using a prospective cohort that enrolled patients with acute ischemic stroke. Favorable outcome was defined as modified Rankin Scale score 0, 1, or 2 at 3 months. It developed 3 machine learning models (deep neural network,

T.Premamala M.E , Professor , Department of Information Technology , Velalar College of Engineering and Technology, Thindal, Erode.
Akshyabharathi.M , Department of Information Technology , Velalar College of Engineering and Technology, Thindal, Erode.
Dharani.P , Department of Information Technology , Velalar College of Engineering and Technology, Thindal, Erode.
Dharshini.B, Department of Information Technology , Velalar College of Engineering and Technology, Thindal, Erode.
Malathi.M , Department of Information Technology , Velalar College of Engineering and Technology, Thindal, Erode.

random forest, and logistic regression) and compared their predictability.

To evaluate the accuracy of the machine learning models, also compared them to the Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score. Result, A total of 2604 patients were included in this study, and 2043 (78%) of them had favorable outcomes. The area under the curve for the deep neural network model was significantly higher than that of the ASTRAL score (0.888 versus 0.839. P [1].

2) DIFFERENTIATION OF RENAL ANGIOMYOLIPOMA WITHOUT VISIBLE FAT FROM RENAL CELL CARCINOMA BY MACHINE LEARNING BASED ON WHOLE-TUMOR COMPUTED TOMOGRAPHY TEXTURE FEATURES

R.I. Lindley, C.P. Warlow (2008) [2] morphological findings showed poor accuracy in differentiating angiomyolipoma without visible fat (AMLwvf) from renal cell carcinoma (RCC). Purpose: To determine the performance of a machine learning classifier in differentiating AMLwvf from different subtypes of RCC based on whole-tumor slices of CT images. Material and Methods: In this retrospective study, 171 pathologically proven renal masses were collected from a single institution. Texture features were extracted from whole-tumor images in three phases including the pre-contrast (PCP), corticomedullary (CMP), and reprographic (NP) phases.

A support vector machine with the recursive feature elimination method based on fivefold cross-validation (SVM-RFECV) with the synthetic minority oversampling technique (SMOTE) was utilized to establish classifiers for differentiating AMLwvf from all subtypes of RCC (all-RCC), clear cell RCC (ccRCC), and non-ccRCC. The performances of the classifiers based on three-phase and single-phase images were compared with each other and morphological interpretations. Results: A machine learning classifier achieved the best performance in differentiating AMLwvf from all-RCC, ccRCC, and non-ccRCC.[2]

3)AUTOMATING ISCHEMIC STROKE SUBTYPE CLASSIFICATION USING MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING

A. Khosla, Y. Cao (2010) [3] manual adjudication of disease classification is time-consuming, error-prone, and limits scaling to large datasets. In ischemic stroke (IS), subtype classification is critical for management and outcome prediction. This study sought to use natural language processing of electronic health records (EHR) combined with machine learning methods to automate IS subtyping. Methods: Among IS patients from an observational registry with TOAST subtyping adjudicated by board-certified vascular neurologists, It analyzed unstructured text-based EHR data including neurology progress notes and neuroradiology reports using natural language processing. It performed several feature selection methods to reduce the high dimensionality of the features and 5-fold cross validation to test generalizability of our methods and minimize over fitting.

It is used several machine learning methods and calculated the kappa values for agreement between each machine learning approach to manual adjudication. Then performed a blinded testing of the best algorithm against a held-out subset of 50 cases. Results: Compared to manual classification, the best machine-based classification achieved a kappa of .25 using radiology reports alone, .57 using progress notes alone, and .57 using combined data. [3]

4)GLOBAL STROKE BELT GEOGRAPHIC VARIATION IN STROKE BURDEN WORLDWIDE

A.S. Kim, E.Cahill (2015) [4] Over the past several decades, much of the developed world has experienced a sustained reduction in age-standardized stroke mortality and morbidity rates.For many of these countries, these improvements have translated into declines in absolute stroke mortality and morbidity as well. For example, stroke had been the third-leading cause of death in the United States for >70 years until 2008, when it became the fourth leading cause of death. Just 5 years later, stroke became the fifth leading cause of death 4 a drop that reflected the >60%

decline in the age-adjusted mortality rate from stroke over just the past few decades.5,6 For most other developed countries, the recent experience and outlook for stroke has been similar.

Over the past 20 years, high-income countries as a group have experienced a 13% decline (95% confidence interval, 6–18%) in the age-standardized incidence rate, a 37% decline (95% confidence interval 19–39%) in the age-standardized mortality rate, and a 21% decline (95% confidence interval, 10– 27%) in the age-standardized disease burdenfrom ischemic stroke as measured in disability-adjusted life year (DALY) loss rates.8,9 (DALYs are a summary measure of disease burden that combines the impact of years of healthy life lost caused by premature death with the effect of long-term disability on quality of life over time.) For hemorrhagic stroke, there have been similar improvements in these indicators over the same period. [4]

5) THE ASCOD PHENOTYPING OF ISCHEMIC STROKE (UPDATEDASCO PHENOTYPING)

P. Amarenco, J. Bogousslavsky (2013) [5] ASCO phenotyping (A: atherosclerosis;

S: small-vessel disease ; C: cardiac pathology; O: other causes) assigns a degree of likelihood of causal relationship to every potential disease (1 for potentially causal, 2 for causality is uncertain, 3 for unlikely causal but the disease is present, 0 for absence of disease, and 9 for insufficient workup to rule out the disease) commonly encountered in ischemic stroke describing all underlying diseases in every patient.

In this new evolution of ASCO called ASCOD, added a ‘D’ for dissection, recognizing that dissection is a very frequent disease in young stroke patients. Have also simplified the system by leaving out the ‘levels of diagnostic evidence’, which has been integrated into grades 9 and 0. Moreover, have also changed the cutoff for significant carotid or intracranial stenosis from 70% to more commonly used 50% luminal stenosis, and added a cardiogenic stroke pattern using neuroimaging. ASCOD captures and weights the overlap between all underlying diseases present in ischemic stroke patients. [5]

6) PREVIOUS USE OF ASPIRIN AND BASELINE STROKE SEVERITY

S. Ricci, S. Lewis (2006) [6] some studies suggest that taking aspirin regularly at the time of the onset of stroke reduces stroke severity. Other studies suggest the converse (ie., that previous aspirin therapy is associated with greater stroke severity). It sought to examine this question among the patients enrolled in the International Stroke Trial (IST). Methods Analysis of the associations of reported use of aspirin in the 3 days before randomization in IST with baseline stroke severity (as assessed by stroke clinical syndrome, predicted outcome at 6 months, and observed outcome at 6 months). This paper adjusted analyses for confounding factors. Results. This paper excluded those patients who were first scanned after trial entry and were found to have an intracerebral hemorrhage as the cause of the stroke leading to randomization. This paper performed analyses for all treatment groups combined.

For the 17 850 patients with ischemic stroke, data at baseline and follow-up were available for 100% and 99%, respectively. Among these patients, 3820 (21.4%) reported previous aspirin use. Previous aspirin use appeared, in univariate analyses, to be associated with greater baseline stroke severity, more severe stroke syndrome, and, in control subjects, worse observed outcome at 6 months. However, after adjustment, these associations were no longer significant.[6]

III.EXISTING SYSTEM

A stroke subtype classification should be useful both in daily clinical practice and in epidemiological and genetic studies, randomized acute clinical trials, and prevention studies of various types. The OCSP classification could be easily used to assess IS severity and predict the prognosis. Modern machine learning based model for prediction of stroke risk and prognosis. Random forest, gradient boosting machines and deep neural network were used and the accuracy of prediction was significantly increased. They had tested that advanced machine learning methods performed on unstructured textual data in the electronic health

International Journal on Applications in Information and Communication Engineering
Volume 7 : Issue 1 : March 2021, pp 21 – 27 www.aetsjournal.com ISSN (Online) : 2394 - 6237

record (HER) can identify TOAST subtype with high concordance and inter rater reliability.

DISADVANTAGES

- Time-consuming
- Error-prone
- Professional dependent
- Limits scaling to large datasets.

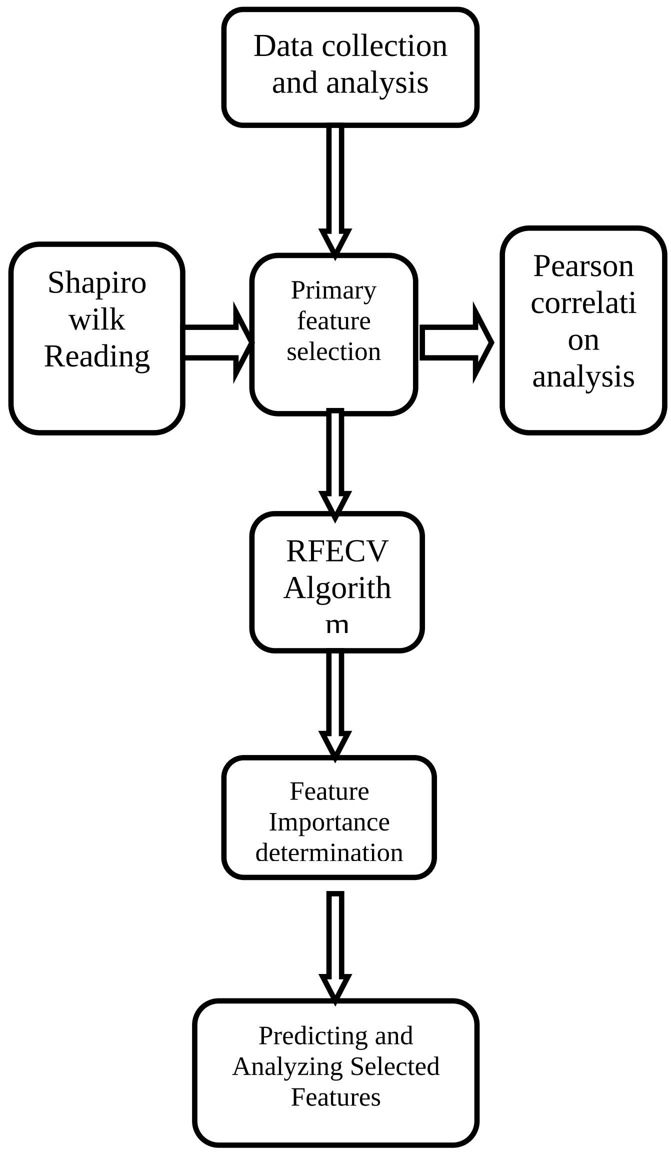
IV. PROPOSED SYSTEM

The point of the preliminary was to set up whether early organization of ibuprofen, heparin, both or neither impacted the clinical course of intense ischemic stroke. To start with, the assessor was prepared on the underlying arrangement of highlights and the significance of each component was gotten either through a coef attribute or through a feature importance attribute. At that point, the most un-significant highlights were pruned from current arrangement of highlights. That technique was recursively rehashed on the pruned set until the ideal number of highlights to choose was in the long run reached. RFECV performed RFE in a cross-approval circle to locate the ideal number of highlights. The incorporated AI approach of RFECV utilized in the examination received straight as its assessor separately. In this examination, a Recursive Feature Elimination (RFE) calculation was completed with programmed tuning of the quantity of highlights chose with cross-approval.

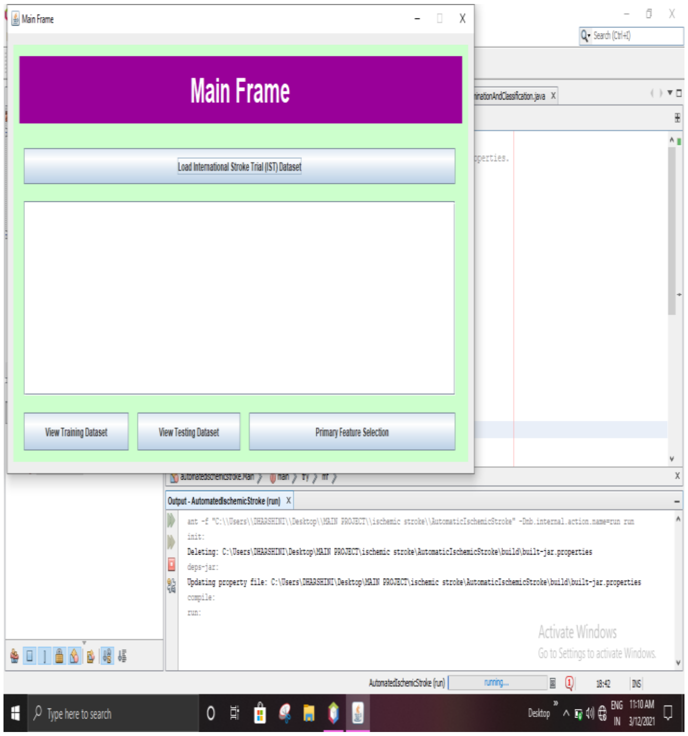
ADVANTAGES:

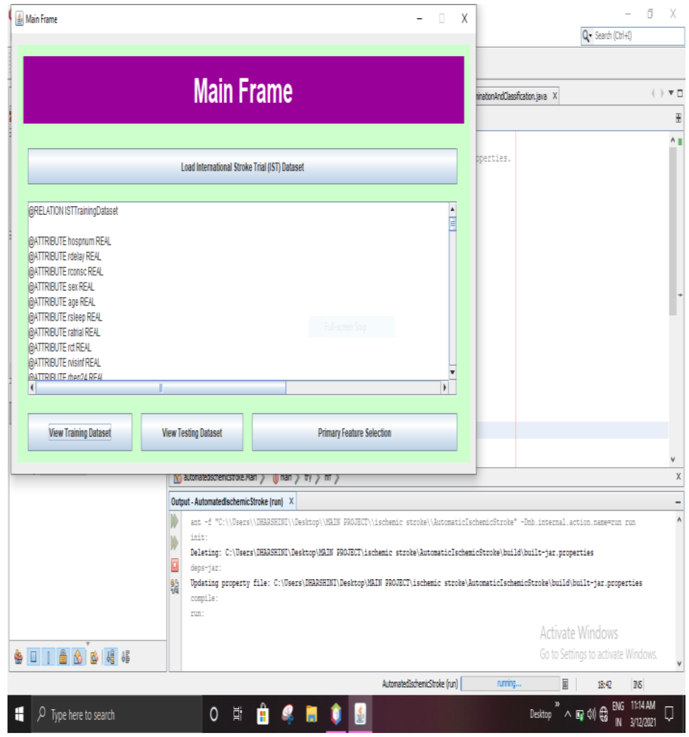
- The outcome likewise indicated that AI approaches beat human experts by subtyping IS.
- In this investigation OCSP IS subtype framework was utilized.
- This framework was only sometimes used to subtype and arrange IS.
- However the framework had the benefits of effectively to utilize and surveying IS seriousness immediately in crisis.
- In the examination this paper utilized highlights in early IST, following stage some new highlights would be gathered to subtype IS as indicated by other progressed IS characterization framework.
- Its more, more complex AI approach would be utilized to research new possible danger factors or reasons for stroke.

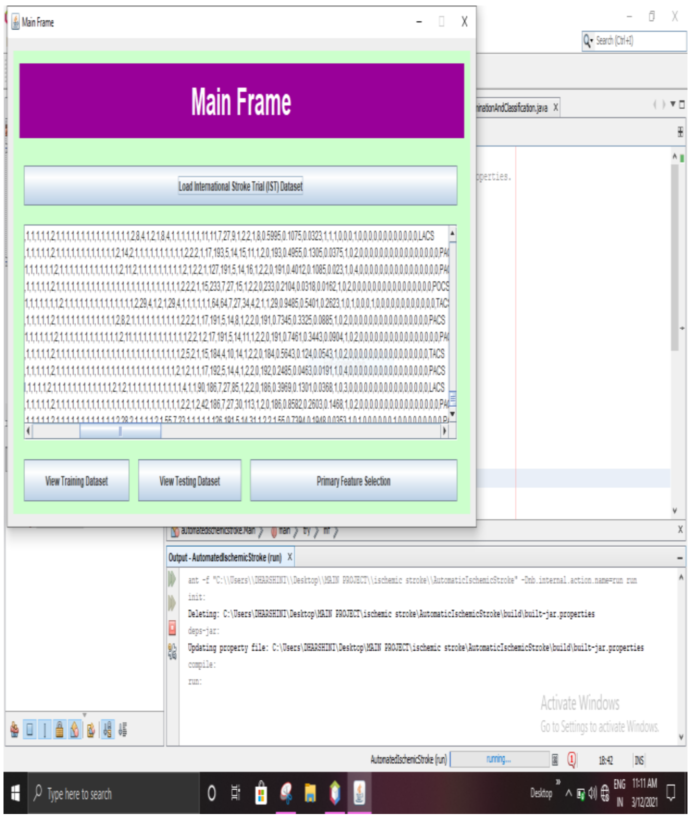
V. BLOCK DIAGRAM

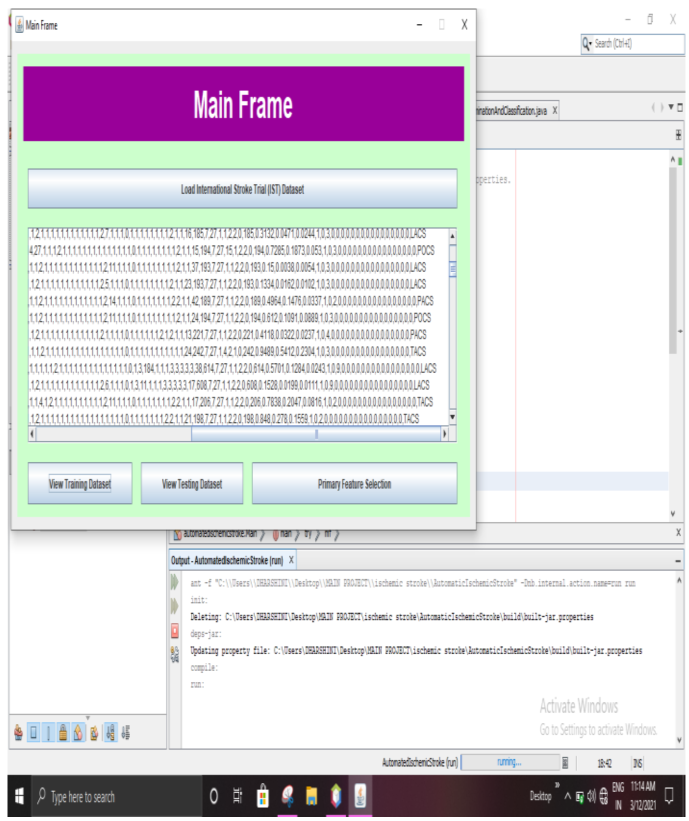


VI.SCREENSHOTS

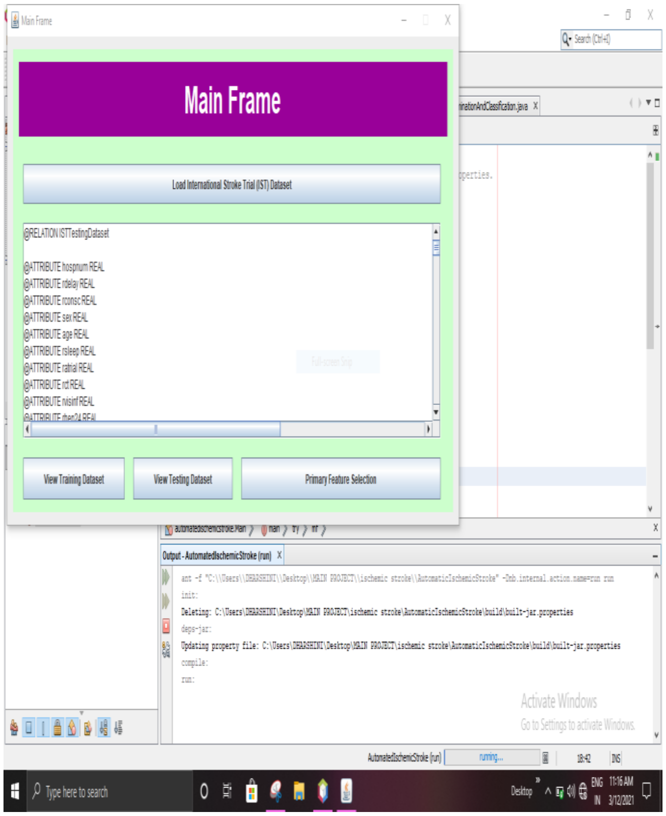


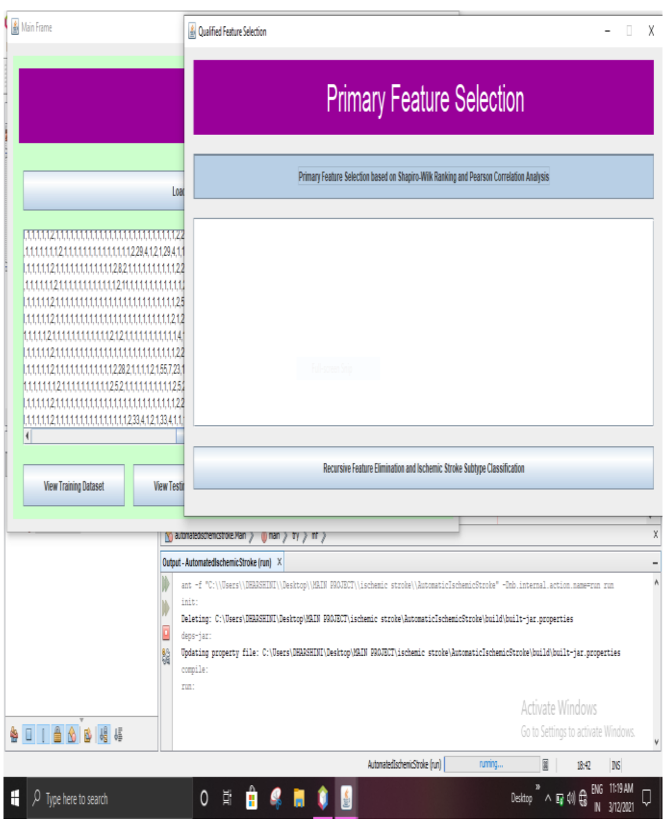






International Journal on Applications in Information and Communication Engineering
Volume 7 : Issue 1 : March 2021, pp 21 – 27 www.aetsjournal.com ISSN (Online) : 2394 - 6237





VII. CONCLUSION

Stoke was an enormous, planned, randomized controlled preliminary, with 100% complete standard information and over 99% complete subsequent information. When gathering information, we just erased sections with missing information without ascribing the missing information in the dataset. Since the dataset generally comprised of discrete worth, information preprocessing was not did. Regardless of whether information preprocessing was completed with normalization, standardization, and the classifiers, for example, straight SVC, MultinomialNaïveBayes and Ada Boost didn't perform better. The RFECV technique functioned admirably in different fields, for example, picture handling, monetary information investigating, and was at that point utilized in clinical exploration. The classifiers utilized in the investigation; aside from Extra Trees, Random Forest and the basic profound learning model, didn't function admirably (with most elevated exactness of 0.815) to subtype ischemic stroke (IS) with 8 neurological deficiencies. However, the basic profound learning model and Extra Trees could subtype IS precisely with just 5 chosen neurological deficiencies.

REFERENCES

[1] V. L. Feigin, M. H. Forouzanfar, R. Krishnamurthi, G. A. Mensah, M. Connor, D. A. Bennett, A. E. Moran, R. L. Sacco, L. Anderson, T. Truelsen, M. O'Donnell, N. Venketasubramanian, S. Barker-Collo, C. M. M. Lawes, W. Wang, Y. Shinohara, E. Witt, M. Ezzati, M. Naghavi, and C. Murray, ''Global and regional burden of stroke during 1990–2010: Findings from the global burden of disease study 2010,'' Lancet, vol. 383, pp. 245–255, Jun. 2014.

[2] A.S. Kim, E. Cahill, and N.T. Cheng, ''Global stroke belt: Geographic variation in stroke burden worldwide,'' Stroke, vol. 46, no. 12, pp. 3564–3570, Dec. 2015.

[3] H.P. Adams, B.H. Bendixen, L.J. Kappelle, J. Biller, B.B. Love, D.L. Gordon, and E.E. Marsh, ''Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of org 10172 in acute stroke treatment,'' Stroke, vol. 24, no. 1, pp. 35–41, Jan. 1993.

[4] H.Ay, T. Benner, E.M. Arsava, K.L. Furie, A.B. Singhal, M.B. Jensen, C. Ayata, A.Towfighi, E.E. Smith, J.Y. Chong, W.J. Koroshetz, and A.G. Sorensen, ''A computerized algorithm for etiologic classification of ischemic stroke: The causative classification of stroke system,'' Stroke, vol. 38, no. 11, pp. 2979–2984, Nov. 2007.

[5] J. Bamford, P. Sandercock, M. Dennis, C. Warlow, and J. Burn, ''Classification and natural history of clinically identifiable subtypes of cerebral infarction,'' Lancet, vol. 337, no. 8756, pp. 1521–1526, Jun. 199.

[6] R.I. Lindley, C.P. Warlow, J.M. Wardlaw, M.S. Dennis, J. Slattery, and P.A.Sandercock, ''Interobserver reliability of a clinical classification of acute cerebral infarction.,'' Stroke, vol. 24, no. 12, pp. 1801–1804, Dec. 1993.

[7] P. Amarenco, J. Bogousslavsky, L.R. Caplan, G.A. Donnan, M.E. Wolf, and M.G.Hennerici, ''The ASCOD phenotyping of ischemic stroke (updated ASCO phenotyping),'' Cerebrovascular Diseases, vol. 36, no. 1, pp. 1–5, 2013.

[8] P. Amarenco, J. Bogousslavsky, L.R. Caplan, G.A. Donnan, and M.G. Hennerici, ''Classification of stroke subtypes,'' Cerebrovascular Diseases, vol. 27, no. 5, pp. 493–501, 2009.

[9] S. Ricci, S. Lewis, and P. Sandercock, ''Previous use of aspirin and baseline stroke severity: An analysis of 17 850 patients in the international stroke trial,'' Stroke, vol. 37, no. 7, pp. 1737–1740, Jul. 2006.

[10] A. Khosla, Y. Cao, C.C.-Y. Lin, H.K. Chiu, J.Hu, and H.Lee, ''An integrated machine learning approach to stroke prediction,'' in Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Washington, DC, USA, Jul. 2010, pp. 25–28.

[11] M.W. Kattan, ''Comparison of cox regression with other methods for determining prediction models and nomograms,'' J. Urol., vol. 170, pp. S6–S10, Dec. 2003.

[12] S.F. Weng, J. Reps, J. Kai, J.M. Garibaldi, and N. Qureshi, ''Can machine-learning improve cardiovascular risk prediction using routine clinical data?'' PLoS ONE, vol. 12, no. 4, Apr. 2017, Art. no. e0174944

[13] J. Heo, J.G. Yoon, H. Park, Y.D. Kim, H.S. Nam, and J.H. Heo, ''Machine learning-based model for prediction of outcomes in acute stroke,'' Stroke, vol. 50, no. 5, pp. 1263–1265, May 2019.

[14] R. Garg, E.Oh, A. Naidech, K. Kording, and S. Prabhakaran, ''Automating ischemic stroke subtype classification using machine learning and natural language processing,'' J. Stroke Cerebrovascular Diseases, vol. 28, no. 7, pp. 2045–2051, Jul. 2019.

[15] P.Xu, G. Zhao, Z. Kou, G. Fang, and W. Liu, ''Classification of cancers based on a comprehensive pathway activity inferred by genes and their interactions,'' IEEE Access, vol. 8, pp. 30515–30521, 2020.

[16] W. Liu, D.Li, and H.Han, ''Manifold learning analysis for allele-skewed DNA modification SNPs for psychiatric disorders,'' IEEE Access, vol. 8, pp. 33023–33038, 2020.