IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# PREDICTING STROKES USING A HYBRID STROKE PREDICTION TECHNIQUE BASED ON ENSEMBLE LEARNING (HSPEL)

**[1]Deepan L, [2]Sivarajan**
[1]PG, Embedded system Technologies,[2] Assistant Professor, Department of EEE
[1]Department of Electrical and Electronics Engineering
[1]Vel Tech Multi Tech Dr.RangarajanDr.Sakunthala Engineering College, Chennai, Tamilnadu

*Abstract* : The negative impact of stroke in society has led to concerted efforts to improve the management and diagnosis of stroke. With an increased synergy between technology and medical diagnosis, caregivers create opportunities for better patient management by systematically mining and archiving the patients' medical records. Therefore, it is vital to study the interdependency of these risk factors in patients' health records and understand their relative contribution to stroke prediction. This paper systematically analyzes the various factors in electronic health records for effective stroke prediction. Using various statistical techniques and ensemble learning, the most important factors for stroke prediction are assessed and the patients' dataset samples classified. The proposed scheme called HSPELs (Heart Stroke Prediction with Ensemble Learning) predicts the risks of patients being affected with heart strokes. The schema achieved an accuracy of approximately 93 percent in classifying samples that are like to be affected by stroke..

## 1.1INTRODUCTION:

Stroke is the sixth leading cause of mortality in the United States according to the Centers for Disease Control and Prevention (CDC). Stroke is a non-communicable disease that kills approximately 11% of the population. In the United States, approximately 795,000 people suffer from the disabling effects of strokes on a regular basis. It is India's fourth leading cause of death. Strokes are classified as ischemic or hemorrhagic. In a chemical stroke, clots obstruct the drainage; in a hemorrhagic stroke, a weak blood vessel bursts and bleeds into the body . Stroke may be avoided by leading a healthy and balanced lifestyle that includes abstaining from unhealthy behaviors, such as smoking and drinking, keeping a healthy body mass index (BMI) and an average glucose level, and maintaining an excellent heart and kidney function.

Stroke prediction is essential and must be treated promptly to avoid irreversible damage or death. With the development of technology in the medical sector, it is now possible to anticipate the onset of a stroke by utilizing MLTs as they allow for accurate prediction and proper analysis. Majority of previous stroke-related research has focused on, among other things, the prediction of heart attacks. Amazing developments in the field of medicine with the aid of technology have been wirnessed. With the advent of annotated dataset of medical records, data mining techniques can be used to identify trends in the dataset. Such analysis has helped the medical practitioners to make an accurate prognosis of any medical conditions. It has led to improved healthcare conditions and reduced treatment costs. The use of data mining techniques in medical records has great impact on the fields of healthcare and bio-medicine. This assists the medical practitioners to identify the onset of disease at an earlier stage. This study focuses on strokes, and identifies key factors that are associated with its occurrence. Several studies have analyzed the importance of lifestyle types, medical records of patients on the probability of the patients to develop stroke. MLTs are also now employed to predict the occurrence of stroke. However, there is no study that attempts to analyze all the conditions related to patient, and identify the key factors necessary for stroke prediction. This research work attempts to bridge this gap by providing a systematic analysis of the various patient records for the purpose of stroke prediction. Using a publicly available dataset, key factors that cause strokes are identified. The proposed scheme called HSPELs is also benchmarked with several popular MLTs for evaluating its classifications on the dataset of patient records.

## 1.2 STROKES

Stroke occurs when the blood flow to various areas of the body is disrupted or diminished, resulting in the cells in those areas of the body not receiving the nutrients and oxygen they require and dying. A stroke is a medical emergency that requires urgent medical attention. Early detection and appropriate management are required to prevent further damage to the affected area of the body and other complications in other parts of the body. There are three primary types of strokes (refer Figure 1):

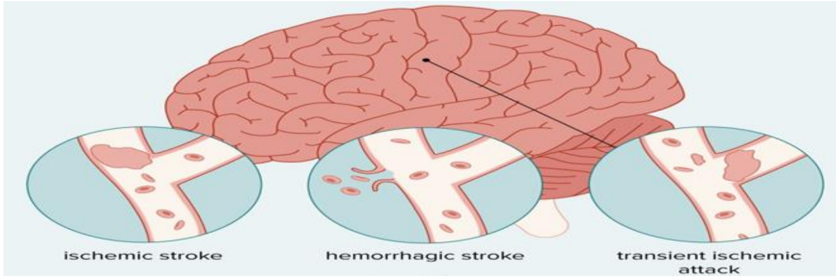



Figure 1 - Types of Strokes

The type of stroke you have affects your treatment and recovery process. In Ischemic strokes, the arteries supplying blood to the brain narrow or become blocked. Blood clots or severely reduced blow flow to the brain causes these blockages. Pieces of plaque breaking off and blocking a blood vessel can also cause them. There are two types of blockages Trusted Source that can lead to ischemic stroke: a cerebral embolism and cerebral thrombosis. The clot gets stuck, and stops the flow of blood and causes a stroke. Cerebral thrombosis (often referred to as thrombotic stoke) occurs when a blood clot develops at the fatty plaque within the blood vessel. According to the CDC, 87 percent Trusted Source of strokes are ischemic strokes. In a Transient ischemic attack (TIA) or mini strokes, occur when blood flow to the brain is blocked temporarily. Symptoms are similar to those of a full stroke. However, they' re typically temporary and disappear after a few minutes or hours, when the blockage moves and blood flow is restored. A blood clot usually causes a TIA. While it' s not technically categorized as a full stroke, a TIA serves as a warning that an actual stroke may happen. Because of this, it' s best not to ignore it. Seek the same treatment you would for a major stroke and get emergency medical help. According to the CDC Trusted Source, more than one-third of people who experience a TIA and don' t get treatment have a major stroke within a year. Up to 10 to 15 percent of people who experience a TIA have a major stroke within 3 months. Finally, hemorrhagic strokes happen when an artery in the brain breaks open or leaks blood. The blood from that artery creates excess pressure in the skull and swells the brain, damaging brain cells and tissues. The two types of hemorrhagic strokes are intracerebral and subarachnoid: (1) An intracerebral hemorrhagic stroke is the most common type of hemorrhagic stroke. It happens when the tissues surrounding the brain fill with blood after an artery bursts and (2) A subarachnoid hemorrhagic stroke is less common. It causes bleeding in the area between the brain and the tissues that cover it. According to the American Heart Association, about 13 percentTrusted Source of strokes are hemorrhagic. The loss of blood flow to the brain damages tissues within the brain. Symptoms of a stroke show up in the body parts controlled by the damaged areas of the brain. The sooner a person having a stroke gets care, the better their outcome is likely to be. For this reason, it' s helpful to know the signs of a stroke quickly. Stroke symptoms can include: Paralysis; Numbness or weakness in arms, face, and legs, especially on one side; Trouble speaking or understanding others; Slurred speech; Confusion, disorientation, or lack of responsiveness; Sudden behavioral changes, especially increased agitation; Vision problems, such as trouble seeing in one or both eyes with vision blackened or blurred, or double vision; Trouble walking; Lloss of balance or coordination; Dizziness, severe, sudden headache with an unknown cause; Seizures and nausea or vomiting. Certain risk factors make people susceptible to stroke. According to the National Heart, Lung, and Blood Institute Trusted Source, risk factors for stroke include: (1) Diet: An unbalanced diet can increase the risk of stroke. This type of diet is high in salt, saturated fats, trans fats and cholesterol. (2) Inactivity: Inactivity, or lack of exercise, can also raise the risk of stroke. Regular exercise has a number of health benefits. The CDC recommends that adults get at least 2.5 hours Trusted Source of aerobic exercise every week. This can mean simply a brisk walk a few times a week. (3) Heavy alcohol use: The risk of stroke also increases with heavy alcohol use. Drink in moderation means no more than one drink a day for women, and no more than two drinks a day for men. (4) Heavy alcohol use can raise blood pressure levels. It can also raise triglyceride levels, which can cause atherosclerosis. This is plaque buildup in the arteries that narrows blood vessels. (5) Tobacco use: Using tobacco in any form also raises the risk of stroke, since it can damage the blood vessels and heart. Nicotine also raises blood pressure. (6) Personal background: There are some risk factors for stroke that can' t be controlled like family history( Stroke risk is higher in some families because of genetic health factors, such as high blood pressure), age (The older you are, the more likely you are to have a stroke), race and ethnicity (African Americans, Alaska Natives, and American Indians are more likely to have a stroke than other racial groups) and health history. Certain medical conditions are also linked to stroke risk which include previous stroke or TIA, high blood pressure, high cholesterol, carrying too much excess weight, heart disorders, such as coronary artery disease, heart valve defects, enlarged heart chambers and irregular heartbeats, sickle cell disease, diabetes, blood clotting disorder and patent foramen ovale. The complications after stroke Trusted Source can vary. They may occur because of either a direct injury to the brain during the stroke, or because abilities have been permanently affected including, seizures, loss of bladder and bowel control, cognitive impairments, reduced mobility or range of motion, or ability to control certain muscle movements, depression and mood or emotional changes which can be managed by methods using medications, physical therapies and reservations

**1.3 MOTIVATION**

Strokes may occur at all ages and is generally considered as only a disease of the elderly. The identification of strokes in young adults is an issue as classical cardiovascular risk factors are usually absent. On the other hand, the numerous alternative causes are rare conditions that are difficult to diagnose developing countries. Most cases are identified based on medical modalities including CT and MRI scans which need to be screened manually by clinicians for confirmations. The lack of available neurologists reduces the frequency of these diagnostics in addition to being time consuming. Though dizziness is one of the most common presenting symptoms for strokes, Physicians traditionally seek to differentiate inner ear conditions from strokes or other central causes of vertigo when evaluating dizziness by focusing on symptom quality or " type" of dizziness. This traditional approach to dizziness evaluation is flawed and may be a source of diagnostic errors. Minor attacks of strokes are also an issue. Minor stroke patients, stroke with minimal or subtle deficits, are frequently misdiagnosed when they present emergently, though they require diagnostic evaluations to identify their stroke mechanism and get appropriate secondary stroke preventions. Thus, strokes need to be categorized early for proper therapies which is the main aim of this work and working in line with this research work.

**1.4 OBJECTIVE**

Stroke is the second leading cause of mortality and a major cause of morbidity worldwide. Failure to correctly diagnose strokes may preclude initiation of secondary stroke preventions, resulting in adverse patient outcomes. While a great deal of research in stroke misdiagnosis has focused on the emergency setting, accurate diagnosis of strokes is essential to assure appropriate stroke treatment and prevention strategies are initiated regardless of the care setting. The main objectives of stroke predictions in this work include

- Improving diagnosis of strokes using patients' data
- Reducing risks for patients and enhancing preventive therapies for strokes.
- Effectively managing early diagnostics of strokes in humans by the use of technology and DMTs (data Mining Techniques) .
- Using DLTs to predict future and probabilities of strokes based on available data/datasets.

**2.1 HSPELs FRAMEWORK**

Stroke is a life-threatening medical illness that should be treated as soon as possible to avoid further complications. de-velopment of an ML model could aid in the early detection of stroke and the subsequent mitigation of its severe conse-quences. effectiveness of several ML algorithms in properly predicting stroke based on a number of physiological variables is investigated in this study. The proposed system aids clinicians by predicting strokes in datasets. Figure 2 depicts the flow of this work.

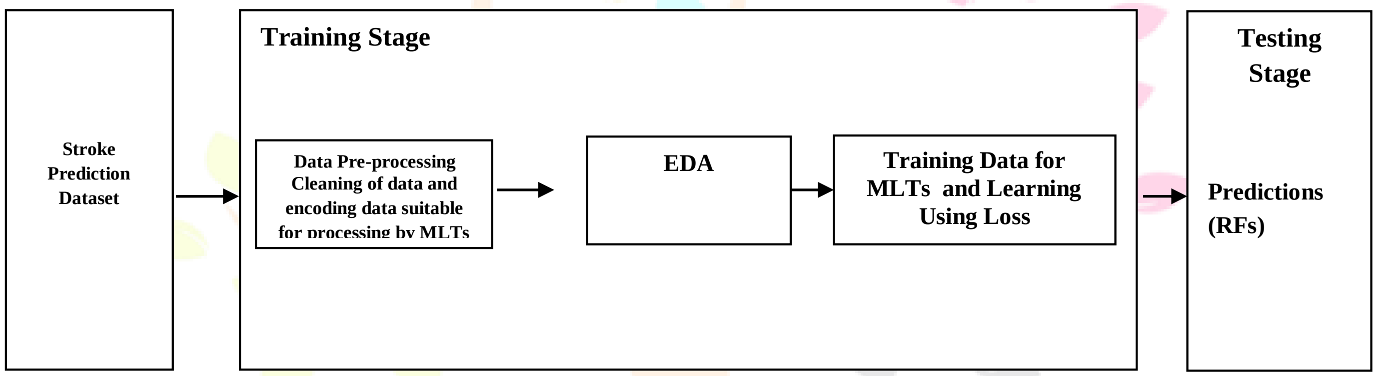


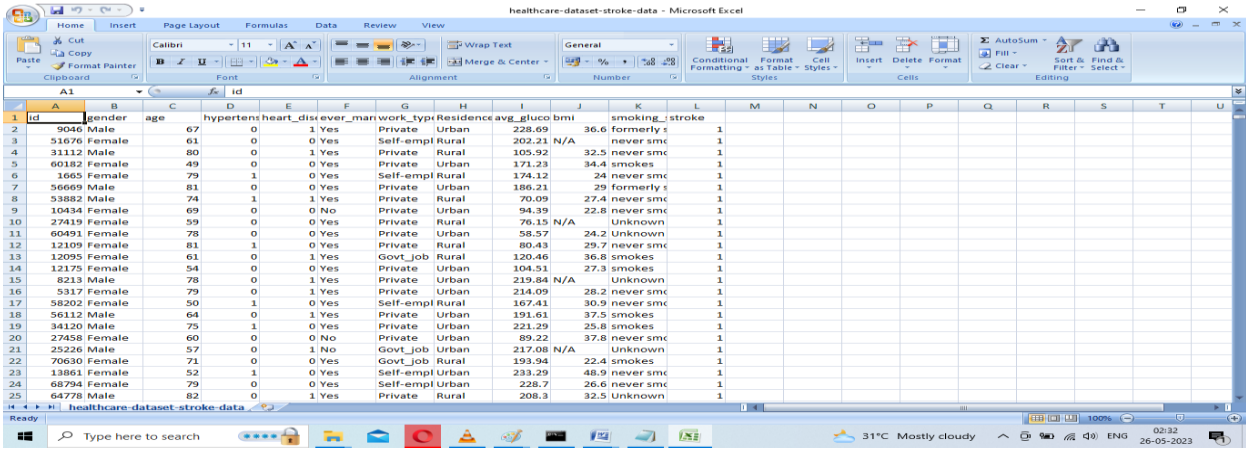

Figure. 2 – HSPELs Framework

**2.2 DATA SET**

Stroke Prediction Dataset has 12 clinical features for predicting stroke events. This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relavant information about the patient. Attribute Information is detailed below:

- id: unique identifier
- gender: "Male", "Female" or "Other"
- age: age of the patient
- hypertension: 0 if the patient doesn't have hypertension, 1 if the patient has hypertension
- heart_disease: 0 if the patient doesn't have any heart diseases, 1 if the patient has a heart disease
- ever_married: "No" or "Yes"
- work_type: "children", "Govt_jov", "Never_worked", "Private" or "Self-employed"
- Residence_type: "Rural" or "Urban"
- avg_glucose_level: average glucose level in blood
- bmi: body mass index
- smoking_status: "formerly smoked", "never smoked", "smokes" or "Unknown"*
- stroke: 1 if the patient had a stroke or 0 if not.

*Note: "Unknown" in smoking_status means that the information is unavailable for this patient. Figure 3 depicts a snapshop of the collected dataset.



| id | gender | age | hypertens | heart_dis | ever_mari | work_typ | Residence | avg_gluco | bmi | smoking_ | stroke |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 9046 | Male | 67 | 0 | 1 | Yes | Private | Urban | 228.69 | 36.6 | formerly s | 1 |
| 51676 | Female | 61 | 0 | 0 | Yes | Self-empl | Rural | 202.21 | N/A | never smc | 1 |
| 31112 | Male | 80 | 0 | 1 | Yes | Private | Rural | 105.92 | 32.5 | never smc | 1 |
| 60182 | Female | 49 | 0 | 0 | Yes | Private | Urban | 171.23 | 34.4 | smokes | 1 |
| 1665 | Female | 79 | 1 | 0 | Yes | Self-empl | Rural | 174.12 | 24 | never smc | 1 |
| 56669 | Male | 81 | 0 | 0 | Yes | Private | Urban | 186.21 | 29 | formerly s | 1 |
| 53882 | Male | 74 | 1 | 1 | Yes | Private | Rural | 70.09 | 27.4 | never smc | 1 |
| 10434 | Female | 69 | 0 | 0 | No | Private | Urban | 94.39 | 22.8 | never smc | 1 |
| 27419 | Female | 59 | 0 | 0 | Yes | Private | Rural | 76.15 | N/A | Unknown | 1 |
| 60491 | Female | 78 | 0 | 0 | Yes | Private | Urban | 58.57 | 24.2 | Unknown | 1 |
| 12109 | Female | 81 | 1 | 0 | Yes | Private | Rural | 80.43 | 29.7 | never smc | 1 |
| 12095 | Female | 61 | 0 | 1 | Yes | Govt_job | Rural | 120.46 | 36.8 | smokes | 1 |
| 12175 | Female | 54 | 0 | 0 | Yes | Private | Urban | 104.51 | 27.3 | smokes | 1 |
| 8213 | Male | 78 | 0 | 1 | Yes | Private | Urban | 219.84 | N/A | Unknown | 1 |
| 5317 | Female | 79 | 0 | 1 | Yes | Private | Urban | 214.09 | 28.2 | never smc | 1 |
| 58202 | Female | 50 | 1 | 0 | Yes | Self-empl | Rural | 167.41 | 30.9 | never smc | 1 |
| 56112 | Male | 64 | 0 | 1 | Yes | Private | Urban | 191.61 | 37.5 | smokes | 1 |
| 34120 | Male | 75 | 1 | 0 | Yes | Private | Urban | 221.29 | 25.8 | smokes | 1 |
| 27458 | Female | 60 | 0 | 0 | No | Private | Urban | 89.22 | 37.8 | never smc | 1 |
| 25226 | Male | 57 | 0 | 1 | No | Govt_job | Urban | 217.08 | N/A | Unknown | 1 |
| 70630 | Female | 71 | 0 | 0 | Yes | Govt_job | Rural | 193.94 | 22.4 | smokes | 1 |
| 13861 | Female | 52 | 1 | 0 | Yes | Self-empl | Urban | 233.29 | 48.9 | never smc | 1 |
| 68794 | Female | 79 | 0 | 0 | Yes | Self-empl | Urban | 228.7 | 26.6 | never smc | 1 |
| 64778 | Male | 82 | 0 | 1 | Yes | Private | Rural | 208.3 | 32.5 | Unknown | 1 |

Fig. 3 – Snap shot of Stroke Prediction Dataset

## 2.3 PRE-PROCESSING

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task. Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format. Some common steps in data preprocessing include:

- Data cleaning: this step involves identifying and removing missing, inconsistent, or irrelevant data. This can include removing duplicate records, filling in missing values, and handling outliers.
- Data integration: this step involves combining data from multiple sources, such as databases, spreadsheets, and text files. The goal of integration is to create a single, consistent view of the data.
- Data transformation: this step involves converting the data into a format that is more suitable for the data mining task. This can include normalizing numerical data, creating dummy variables, and encoding categorical data.
- Data reduction: this step is used to select a subset of the data that is relevant to the data mining task. This can include feature selection (selecting a subset of the variables) or feature extraction (extracting new variables from the data).

This work checks for null values and normalizes data columns. There are many ways to deal with null values. They can be dropped, filled with mean/ median values or even simply with values from the record before or after the missing values. The missing values of BMI were filled using Decision Trees. Figure 4 displays HSPELs Schema' s Pre-processing output.

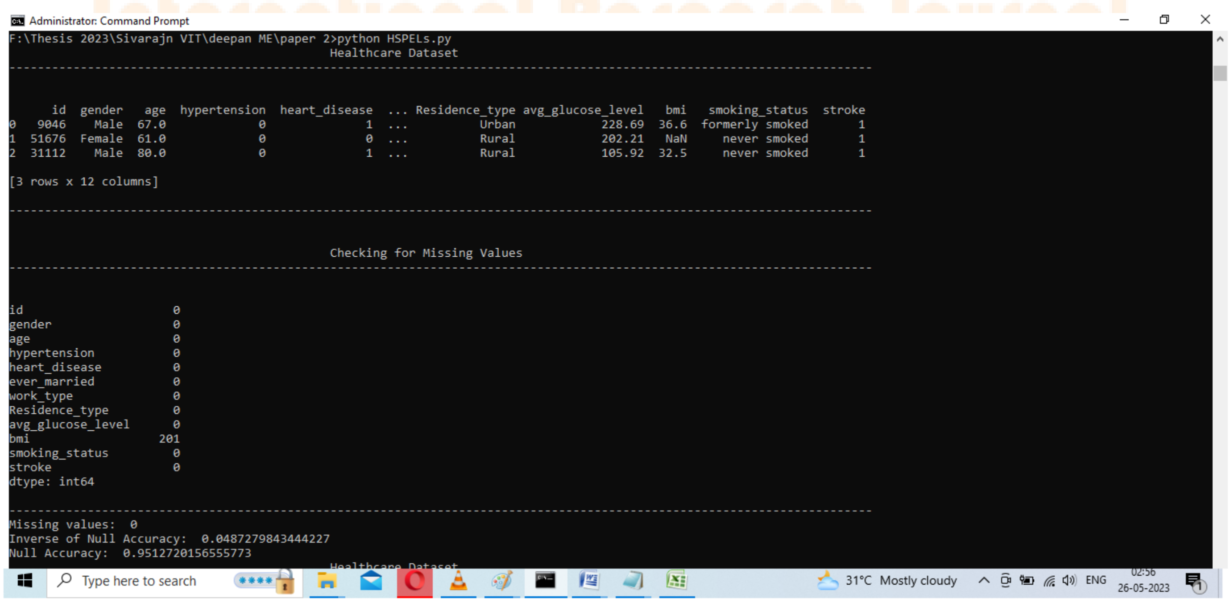



Figure 4 - HSPELs Schema' s Pre-processing output

## 2.4 EXPLORING THE DATA

EDA approaches, which were initially created by American mathematician John Tukey in the 1970s, are still a frequently employed strategy in the data discovery process. It helps examine and study data sets and summarise their key properties, frequently using data visualisation techniques. It makes it simpler to find patterns, identify anomalies, or test assumptions by figuring out how to alter data sources to achieve desired answers. EDA helps with a better understanding of the variables in the data collection and their relationships, and is generally used to investigate what data might disclose beyond the formal modelling or hypothesis testing assignment. It can also assist in determining the suitability of the statistical methods you are contemplating using for data analysis. Specific statistical functions and techniques can be performed with EDA. Selecting Most Important features based on Correlations by HSPELs: The basic maxim "whatever goes in, comes out" is followed by MLTs. Massive amounts of data are required to train MLT models in order to improve model learning. The majority of data are noisy, and some dataset columns may not have a big impact on how well a model works. More data columns also slow down the training procedures, which causes training from invalid and irrelevant data. Features that are chosen distinguish useful data from the rest. Understanding the significance of feature selection is vital when modelling MLTs for evaluations. It's quite uncommon for all of the variables in a dataset to be helpful for creating a model in real-world data science situations. Repetitive factors decrease a classifier's capacity to generalise while also potentially decreasing its overall accuracy. A model's overall complexity is also increased by including more variables. Finding the optimal combination of features that enables the construction of optimised models of the researched phenomena is the aim of feature selection. Feature selection MLTs may be roughly categorised as either supervised or unsupervised. In order to improve the effectiveness of supervised models like classification and regression, supervised techniques involve labelled data and the identification of pertinent features. Unlabeled data can be processed using unsupervised techniques. Fisher's Score, Correlation Coefficient, and Filter Methods are a few examples of feature selection method types. Instead of focusing on cross-validation performance, filter approaches capture the inherent characteristics of the features assessed by univariate statistics. The entropy decrease caused by a dataset modification is calculated using information gain. By assessing each variable's information gain in relation to the target variable, it may be utilised for feature selection. For categorical characteristics in a dataset, the Chi-square test is employed. It is possible to compute the Chi-square between each feature and the objective and then choose the appropriate number of features with the highest Chi-square values. One of the most popular supervised feature selection techniques is the Fisher score. According to Fisher's score, the method produces variable rankings in decreasing order. Feature selections based on correlations between variables is used by the proposed HSPELs schema. Figures 5 depicts EDA of the healthcare dataset.

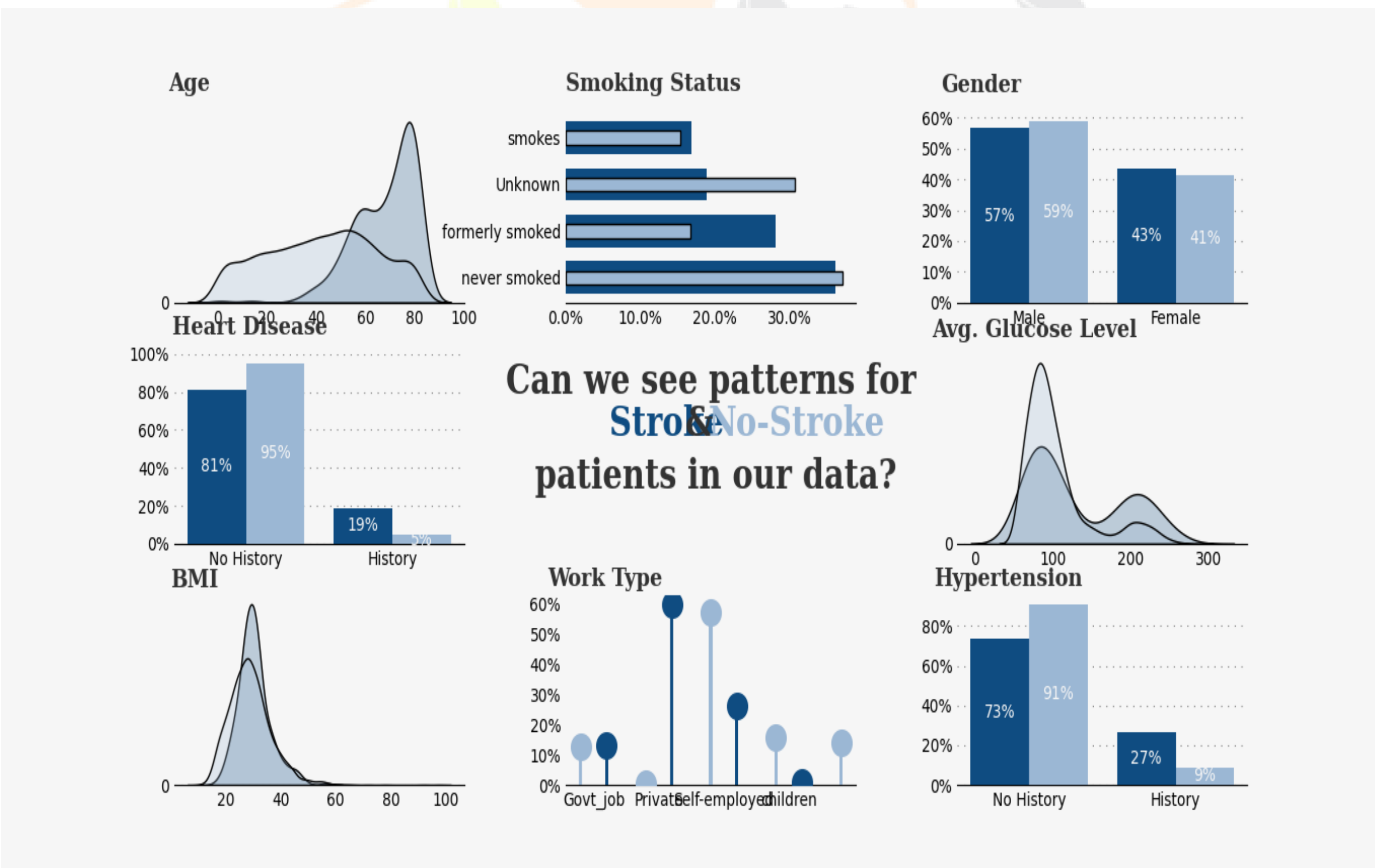



Figure 5 - EDA of the healthcare dataset.

### 2.5 HSPELs TRAINING AND IMPLEMENTATION

The train-test split is a technique for evaluating the performance of a machine learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a dataset and dividing it into two subsets. The first subset is used to fit the model and is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset. The Train Dataset is used to fit the machine learning model while the Test Dataset is used to evaluate the fit machine learning model. The objective is to estimate the performance of the machine learning model on new data: data not used to train the model. Model has to fit it on available data with known inputs and outputs, then make predictions on new examples in the future where we do not have the expected output or target values. The train-test procedure is appropriate when there is a sufficiently large dataset available. The idea of " sufficiently large" is specific to each predictive modeling problem. It means that there is enough data to split the dataset into train and test datasets and each of the train and test datasets are suitable representations of the problem domain. This requires that the original dataset is also a suitable representation of the problem domain. A suitable representation of the problem domain means that there are enough records to cover all common cases and most uncommon cases in the domain. This might mean combinations of input variables observed in practice. It might require thousands, hundreds of thousands, or millions of examples. Conversely, the train-test procedure is not appropriate when the dataset available is small. The reason is that when the dataset is split into train and test sets, there will not be enough data in the training dataset for the model to learn an effective mapping of inputs to outputs. There will also not be enough data in the test set to effectively evaluate the model performance. The estimated performance could be overly optimistic (good) or overly pessimistic (bad). The split percentage considerations should include: Computational cost in training the model., Computational cost in evaluating the model, Training set representativeness and Test set representativeness.. Common split percentages include: Train: 80%, Test: 20% or Train: 67%, Test: 33% or Train: 50%, Test: 50%. Figure 6 depicts test/Train Splits snapshot and classification output of HSPELs.

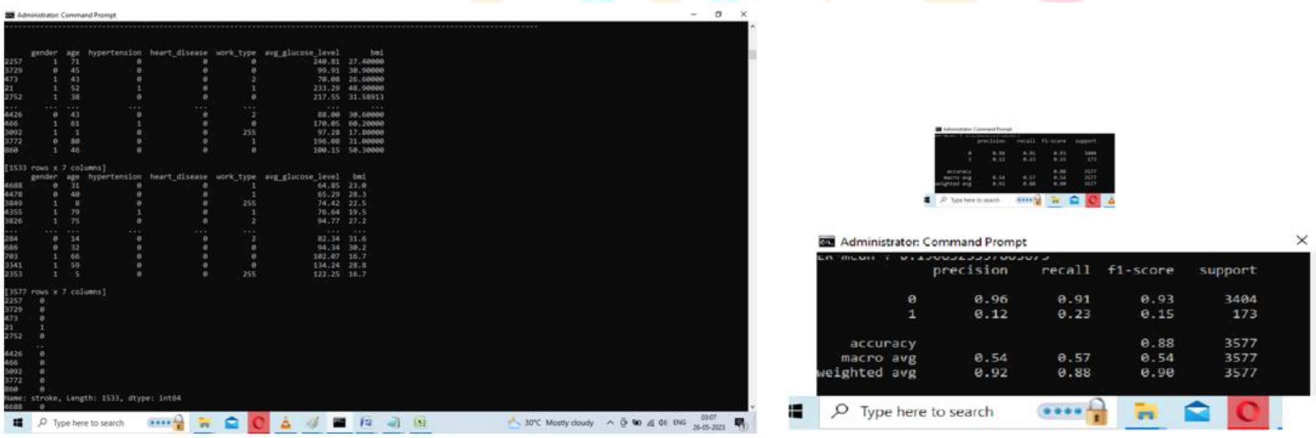



Figure 6 - EDA of the healthcare dataset.

### CONCLUSION:

In this paper, we presented a detailed analysis of patients' attributes in electronic health record for stroke prediction. We systematically analysed different features. This work performed feature correlation analysis and a step wise analysis for choosing an optimum set of features. We found that the different features are not well-correlated and a combination of only 4 features ( , , and ) might have good contribution towards stroke prediction. The analysis showed that almost all principal components are needed to explain a higher variance. The variable loadings however showed that the first principal component which has the highest variance might explain the underlying phenomenon of stroke prediction. We observed that most of the existing features in the EHR dataset are highly correlated to each other, and therefore do not add any additional information to the original feature space. Furthermore, a larger dataset will enable us to train our deep neural networks more efficiently. We plan to collect institutional data in our planned future work. The systematic analysis of the different features in the electronic health records will assist the clinicians in effective archival of the records. Instead of recording and storing all the features, the data management team can archive only those features that are essential for stroke prediction. Finally, the proposed schema based on Random forest classification achieves classification accuracy of 93%. According to the research, the random forest method outperforms other processes when cross-validation metrics are used in body stroke forecasting. future scope of this study is that using a larger dataset and machine learning models, such as AdaBoost, SVM, and Bagging, the framework models may be enhanced. will enhance the dependability of the framework and the framework' s presentation..

## REFERENCES


[1] R.S. Jeena, S. Kumar, Stroke prediction using SVM, in: Proc. International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2016, pp. 600– 602, http://dx.doi.org/10.1109/ICCICCT. 2016.7988020.

[2] S.-M. Hanifa, K. Raja-S, Stroke risk prediction through non-linear support vector classification models, Int. J. Adv. Res. Comput. Sci. 1 (3) (2010).

[3] J.K. Luk, R.T. Cheung, S. Ho, L. Li, Does age predict outcome in stroke rehabilitation? A study of 878 Chinese subjects, Cerebrovasc. Dis. 21 (4) (2006) 229– 234.

[4] S.N. Min, S.J. Park, D.J. Kim, M. Subramaniyam, K.-S. Lee, Development of an algorithm for stroke prediction: a national health insurance database study in Korea, Eur. Neurol. 79 (3– 4) (2018) 214– 220.

[5] M.S. Singh, P. Choudhary, Stroke prediction using artificial intelligence, in: 2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON), IEEE, 2017, pp. 158– 161.

[6] P. Chantamit-o, Prediction of stroke disease using deep learning model.

[7] A. Khosla, Y. Cao, C.C.-Y. Lin, H.-K. Chiu, J. Hu, H. Lee, An integrated machine learning approach to stroke prediction, in: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, 2010, pp. 183– 192.

[8] C.-Y. Hung, C.-H. Lin, T.-H. Lan, G.-S. Peng, C.-C. Lee, Development of an intelligent decision support system for ischemic stroke risk assessment in a population-based electronic health record database, PLoS One 14 (3) (2019) e0213007.

[9] D. Teoh, Towards stroke prediction using electronic health records, BMC Med. Inform. Decis. Mak. 18 (1) (2018) 1– 11.

[10] C.-Y. Hung, W.-C. Chen, P.-T. Lai, C.-H. Lin, C.-C. Lee, Comparing deep neural network and other machine learning algorithms for stroke prediction in a large-scale population-based electronic medical claims database, in: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2017, pp. 3110– 3113.

[11] X. Li, H. Liu, X. Du, P. Zhang, G. Hu, G. Xie, S. Guo, M. Xu, X. Xie, Integrated machine learning approaches for predicting ischemic stroke and thromboembolism in atrial fibrillation, in: AMIA Annual Symposium Proceedings, 2016, American Medical Informatics Association, 2016, p. 799.

[12] S. García, J. Luengo, F. Herrera, Tutorial on practical tips of the most influential data preprocessing algorithms in data mining, Knowl.-Based Syst. 98 (2016) 1– 29.

[13] M. Monteiro, A. C. Fonseca, A. T. Freitas et al., " Using machine learning to improve the prediction of functional outcome in ischemic stroke patients," IEEE/ACM Transac-tions on Computational Biology and Bioinformatics, vol. 15, no. 6, pp. 1953– 1959, 2018.

[14] T. Kansadub, S.S. Kiattisin, and C. Jalayondeja, " Stroke risk prediction model based on de-mographic data," in Proceedings of the 2015 8th Biomedical Engineering International Conference (BMEiCON), pp. 1– 3, Pattaya, November 2015.

[15] S. Y. Adam, A. Yousif, and M. B. Bashir, " Classification of ischemic stroke using machine learning algorithms," Inter-national Journal of Computer Application, vol. 149, no. 10, pp. 26– 31, 2016.


Research Through Innovation