Enhanced Fuzzy Score-Based Decision Support System for Early Stroke Prediction

MAYSSA BEN KAHLA, Higher Institute of Computer Science and Communication Techniques, University of Sousse, H. Sousse, Tunisia and RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba, Tunisia
DALEL KANZARI, Higher Institute of Applied Science and Technology, University of Sousse, Sousse, Tunisia and LARODEC, ISG Tunis, University of Tunis, Tunis, Tunisia
SANA BEN AMOR, University of Sousse, Sousse, Tunisia and Department of Neurology, Sousse, Tunisia
SONIA AYACHI GHANNOUCHI, RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba, Tunisia and Higher Institute of Management of Sousse, University of Sousse, Sousse, Tunisia
RICARDO MARTINHO, School of Technology and Management, Polytechnic of Leiria, Leiria, Portugal and CINTESIS, University of Porto, Porto, Portugal

According to the Global Health Observatory, stroke ranks second worldwide in causing dementia, right after Alzheimer's disease. The mortality rate linked to dementia resulting from stroke is high because symptoms are often recognized late, and stroke can be misinterpreted as other brain disorders. Early detection and diagnosis of stroke is crucial. Therefore, increasing awareness of stroke symptoms and implementing preventive measures becomes imperative. Prompt intervention by healthcare professionals can improve outcomes and reduce long-term complications of stroke.

The research introduces an innovative approach for early stroke prediction using a fuzzy scoring-based Decision Support System. This approach encompasses three main modules: Mind map-based Data Modeling, Fuzzy scoring computing, and Machine Learning (ML)-Based Decision System. By incorporating fuzzy logic, the approach extracts valuable knowledge from imprecise and uncertain data. Combining the fuzzy stroke risk model with a ML-based decision support system aims to enhance stroke prediction accuracy and improve preventive measures and patient outcomes. The approach's effectiveness was validated using real clinical data and tested with various ML classifiers, including K-Nearest Neighbor (KNN), Logistic Regression (LR), Decision Tree (DT), Artificial Neural Network (ANN), and Support Vector Machine (SVM). The results showed a strong correlation between stroke cases and computed risk-scoring values.

In comparison to predictions without fuzzy scoring and other related works, the stroke risk prediction using the proposed approach demonstrated higher accuracy, making it a promising method for early stroke detection and prevention.

Mayssa Ben Kahla contributed in the approach, wrote the main article text, and prepared figures. Dalel Kanzari contributed in the approach, wrote the main article text, and reviewed the article. Sana Ben Amor contributed in the approach and reviewed the article. Sonia Ayachi Ghannouchi contributed in the approach and reviewed the article. Ricardo Martinho reviewed the article.

Authors' Contact Information: Mayssa Ben Kahla (corresponding author), Higher Institute of Computer Science and Communication Techniques, University of Sousse, H. Sousse, Tunisia and RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba, Tunisia; e-mail: benkahla.mayssa@issatso.u-sousse.tn; Dalel Kanzari, Higher Institute of Applied Science and Technology, University of Sousse, Sousse, Tunisia and LARODEC, ISG Tunis, University of Tunis, Tunis, Tunisia; e-mail: kndalel@gmail.com; Sana Ben Amor, University of Sousse, Sousse, Tunisia and Department of Neurology, Sousse, Tunisia; e-mail: kaffelsana@yahoo.fr; Sonia Ayachi Ghannouchi, RIADI Laboratory, National School of Computer Sciences, University of Manouba, Manouba, Tunisia and Higher Institute of Management of Sousse, University of Sousse, Sousse, Tunisia; e-mail: sonia.ayachi.ghannouchi@gmail.com; Ricardo Martinho, School of Technology and Management, Polytechnic of Leiria, Leiria, Portugal and CINTESIS, University of Porto, Porto, Portugal; e-mail: rmartin.estg@gmail.com.

ACM 2637-8051/2025/1-ART7
https://doi.org/10.1145/3703461

CCS Concepts: • **Applied computing → Health informatics**;

Additional Key Words and Phrases: Fuzzy logic, Stroke disease, scoring modeling, prediction, mind-map, ML-based-Decision Support System

ACM Reference format:
Mayssa Ben Kahla, Dalel Kanzari, Sana Ben Amor, Sonia Ayachi Ghannouchi, and Ricardo Martinho. 2025. Enhanced Fuzzy Score-Based Decision Support System for Early Stroke Prediction. *ACM Trans. Comput. Healthcare* 6, 1, Article 7 (January 2025), 23 pages.
https://doi.org/10.1145/3703461

1 Introduction

Stroke is characterized by the sudden interruption of blood flow to the brain, which leads to oxygen deprivation in the affected brain areas. According to the World Health Organization,[1] 15 million people worldwide suffer a stroke each year, and 5 million of them die as a result. Another 5 million people are left permanently disabled due to stroke, making it a major public health concern. This pathology represents the second cause of dementia (after Alzheimer's disease) and the second cause of mortality with 20% of people who die annually.

The increased mortality rate is due to the very late detection of symptoms and the lack of knowledge of relevant risk factors, which are generally imprecise, approximate, and unclear, such as stress level, diabetes, and smoking.

A stroke occurs suddenly and the unawareness of risk factors can lead to incorrect prognosis by medical personnel [39]. The treatment success depends on the time of symptoms detection. For example, the treatment of "thrombolysis" [5] has proven to be effective in cerebral infarction. This treatment is possible only within 4.5 hours after the onset of symptoms and depending on the patients features.

Despite the considerable progress of new information and communication technologies, and in particular, those related to artificial intelligence [3, 5, 7, 35], there are few works focused on detecting the relevant risk factors, detection to help specialists properly predict stroke and initiate appropriate treatments, and save patients' lives. A significant number of research works are looking at this problem using **Machine Learning (ML)** techniques such as **Support Vector Machines (SVM)**, **Decision Trees (DTs)**, and **K-Nearest Neighbor (KNN)**, but no concrete results have been found regarding early detection and prevention of stroke, based on risk factors. For example, authors in [10] applied three classification approaches: **Neural Network (NN)**, DT, and the **Random Forest (RF)**. Also, the work in [22] employed ML techniques to classify **Ischemic Strokes (ISs)**. They implemented the DT and KNN methods.

Stroke risk prediction models rely on binary logic, which can lead to inaccuracies and uncertainties when dealing with imprecise or uncertain input variables. Fuzzy logic provides a more flexible and intuitive approach to handling uncertainty in data, which can lead to more accurate and reliable stroke risk prediction models. The objective of this research is to create an approach that combines a fuzzy stroke risk prediction model with a ML-Based Decision Support System to predict stroke disease. This combined system is designed to effectively handle both imprecise and inaccurate input data, along with accurate input data, resulting in significantly improved accuracy and reliability in predictions. While previous works have explored the use of ML algorithms in stroke risk prediction with only accurate data. We believe that our approach can significantly improve stroke risk prediction and provide valuable insights into the complex interactions between risk factors and symptoms.

Thus, the remaining issues are: How can fuzzy scoring help the prediction of a stroke? and Which attributes can be considered to predict the risk of stroke by using fuzzy logic? Does the fuzzy risk stroke scoring improve the prediction of the stroke?

[1] "World Health Organization (WHO)," Stroke, Cerebrovascular accident. [Online]. Retrieved May 5, 2023 from https://www.emro.who.int/health-topics/stroke-cerebrovascular-accident/index.html.

Our approach encompasses two primary objectives:

— Propose a solution utilizing the fuzzy logic paradigm to evaluate imprecise and uncertain data, enabling early-stage prediction of the risk of stroke without relying on clinical analyses like imaging treatments or biological analyses, and so on.
— Improved early prediction stroke accuracy.

The rest of this article is organized as follows: Section 2 presents related work. The methodology is explained in Section 3. The Fuzzy Scoring-Based Decision Support System to Predict Stroke is described in Section 4. Section 5 presents the experiments and validation. The last section displays the conclusion of the article.

2 Related Work

Several medical surveys have shown that there are determining factors in the occurrence of the disease stroke. For example, Kanase and Jhaveri [19] divided stroke into two types: ischemic and hemorrhagic. They grouped stroke risk factors into modifiable and non-modifiable risk factors. According to the authors in [2, 16, 19], major risk factors include age, history of cerebrovascular event, smoking, hypertension, dyslipidemia, diabetes mellitus, cardiovascular diseases, obesity, alcohol consumption, physical inactivity, and genetic risk factors. Recently, COVID-19 was considered as another stroke risk factor. For example, gender, age, and race are non-modifiable risk factors. On the other hand, **High Blood Pressure (HBP)**, smoking, and physical inactivity are considered modifiable risk factors. Also, the presented symptoms of stroke include speaking difficulty, headaches, numbness or sudden weakness of the face, arm, or leg, and cognitive decline.

Table 1 summarizes these three research works [2, 16, 19], in which stroke can be characterized by its type, risk factors, symptoms, and medical outcome (result).

Many works used ML in different fields. One significant contribution is the hierarchical graph-based text classification framework, which utilizes contextual node embeddings and BERT-based dynamic fusion to enhance classification accuracy [29]. Additionally, research on text augmentation has introduced a hybrid approach that combines semantic role labeling with ant colony optimization, proving effective in improving the quality of text data for ML models [28, 30].

Ensemble methods have also been a focus, with studies exploring the combination of keyword extraction techniques and ML classifiers to address various challenges [32]. Furthermore, the issue of imbalanced learning has been tackled through a consensus clustering-based undersampling approach [25], while genetic rank aggregation has been proposed as a method for improving sentiment classification performance [31].

In the domain of sentiment analysis, particularly concerning product reviews and sarcasm detection, **Deep Learning (DL)** models and topic-enriched word embeddings have been employed to achieve notable results [26, 27, 33]. These studies collectively highlight the significant impact of integrating advanced ML techniques combined with text processing methods, helping to improve the accuracy and efficiency of decision support models.

Other computer science works are based on the recognition of the disease's risk factors using ML approaches such as in the work of Cheon et al. [7], the authors combined **Principal Component Analysis (PCA)** and **Deep Neural Network (DNN)** to detect strokes based on medical service data.

In the work of Dev et al. [10], the authors applied three popular classification approaches: NN, DT, and RF, for the purpose of stroke prediction from only four risk factors: age, heart disease, the patient's average glucose level, and the presence of hypertension.

Almadani and Alshammari applied the PCA, J48 (C4.5) JRip, and NN algorithms to predict stroke [3]. According to their findings, patients with heart disease and hypertension, diabetes mellitus, renal disease, hyperlipidemia, or blood (platelet) abnormalities have an increased risk of getting a stroke.

Islam [17] proposed a stroke detection system using the fuzzy logic inference system and the fuzzy C-Means classifier, to create a detection model. This study is based on data from the Bangladeshi population.

Table 1. Comparison of Stroke Medical Research Literature

Reference	Stroke type	Risk factors	Symptoms	Result
[16]	Ischemic stroke: Embolic stroke	Potential heart conditions, Atrial fibrillation (AF), Heart attack, HBP Diabetes, Dyslipidemia, Smoking, Excessive alcohol consumption, COVID-19, Sedentary lifestyle	-	-
[2]	Ischemic stroke	Age, HBP, Diabetes, Smoking, Heart disease, History of a stroke, Gender	Sudden confusion, Difficulty speaking, Difficulty understanding speech, Sudden numbness or weakness of the face, arm or leg, Sudden difficulty seeing in one or both eyes, Sudden difficulty walking, Dizziness, Loss of balance or coordination, Sudden headache	The majority of respondents (78) identified sudden numbness of the face, arms, and legs. While 42% of them identified vision problems.
[19]	*Ischemic stroke:* Thrombotic stroke Embolic stroke *Hemorrhagic stroke*	*Not modifiable:* Age, gender, Race/ethnicity, genetics, Rural /urban areas *Modifiable:* HBP, Smoking, Diet, Obesity, Physical inactivity, Stroke risk factors are specific to women, such as early menopause, pregnancy, and so on	Headaches, Weakness, Sudden numbness or weakness of the face, arm, or leg, especially on one side of the body, Sudden confusion, Difficulty speaking or understanding, Sudden disturbance of vision in one or both eyes. Sudden difficulty walking, Dizziness, loss of balance or coordination, Crises, Ataxia	*The results show that:* In 50% (130) of the female subjects, Generalized weakness of 64%, Ataxia by 37%, Headaches of 60%, Language disorders of 42%, Weakness of 45%

Chantamit-o pas and Goyal [6] proposed predictive stroke analysis techniques using a DL model applied to a heart disease dataset. They used 10 risk factors: age, gender, blood pressure, chest pain, cigarettes, family history, hypertension, cholesterols, heart rate, and blood vessels.

Jeena and Kumar [36] provide a study of various risk factors to understand the probability of stroke. They used a regression-based approach to identify the relationship between a factor and its corresponding impact on stroke. They implemented SVM with different kernel functions. Symptoms and risk factors used as an entry of the SVM model are age, sex, walking symptoms, **Atrial Fibrillation (AF)**, face deficit, arm/hand deficit, visible infarction on CT, dyphasia, hemianopia, visuospatial disorder, and cerebellar signs.

Yahia et al. [43] employed ML algorithms to classify ISs. They chose to implement the DT and KNN techniques as algorithms of ML. The features and attributes, considering the following: age, sex, irritability, convulsions, left-side weakness, right-side weakness, mouth deviation, difficulty in speaking, if the patient is unable to walk, headache, difficulty in seeing, results of CT, results of **Magnetic Resonance Imaging (MRI)**.

Arslan et al. [4] intend to assess different medical data mining approaches to predict IS. They employed the SVM, stochastic gradient boosting, and penalized **Logistic Regression (LR)** as data mining approaches. The dataset of this study was collected from Turgut Ozal Medical Centre, Inonu University, Malatya, Turkey which contained 17 predictors: age, gender, educational status, marital status, alcohol consumption, white blood cell, hematocrit, hemoglobin, platelet, glucose, blood urea nitrogen, creatinine, sodium, potassium, chlorine, prothrombin time, and calcium.

Min et al. [23] developed an algorithm for predicting stroke from potentially modifiable risk factors, which used LR for model derivation. Modifiable stroke risk factors used by Min and colleagues in this study include hypertension, cardiac disease, diabetes, dysregulation of glucose, metabolism, AF, and lifestyle factors.

The main objectives of the research of Almadani and Alshammari [3] are to use data mining techniques to predict patients at risk of developing stroke and to find the patient who has a higher chance to develop a stroke. To achieve these objectives "Almadani and Alshammari" they implemented three algorithms: C4.5, Jrip, and **Multi-Layer Perceptron (MLP)**. As features of the input, they used 147 attributes including heart diseases, immunity diseases, diabetes, militias, kidney diseases, hyperlipidemia, and epilepsy.

The paper of Thammaboosadee and Kansadub [42] presents the data mining process that was used for building a stroke prediction model based on demographic information and medical screening data. The pre-processed demographic data characteristics included: hypertension, diabetes, heart disease, asthma bronchitis allergy, hyperlipidemia, accident, fracture, cancer, rheumatoid gout, tuberculosis, osteoporosis, weight change, urinary incontinence, vertigo, human immunodeficiency virus, liver disease, herpes zoster or psoriasis, systemic lupus erythematosus, depressive, pregnant, kidney, family cancer, family heart disease, family diabetes, family heredity, bleed, muscle, loss balance, sex, age, province, marital status, education, and occupation. In this work, the authors implement **Naive Bayes (NB)** algorithms, DT algorithm, in particular, the c4.5 algorithm, and also **Artificial Neural Network (ANN)**.

The research study of Arunkumar et al. [37] focuses to design and develop a prototype system by integrating data mining results with a knowledge-based system that facilitates diagnosis and treatment process as well as providing an opinion and a level of risk for the patient. In this research, the following techniques were used: Bayes Net, NB, Decision Table, JRip, J48, and RF.

The JRip classifier has generated 39 rules. The rules involved 10 features among the 11 features from the sample dataset. The 39 rules generated were distributed between 27 rules for the normal class, 11 rules for the hemorrhagic class, and only 1 rule for ischemic.

The paper of Sailasya and Kumari [38] is based on predicting the occurrence of stroke using ML. To achieve their objectives, they compared six ML algorithms: LR, DT, Classification, RF, KNN and NB. The input features for all these algorithms are: gender, age, hypertension, heart disease, ever married, work type, residence type, average glucose level, **Body Mass Index (BMI)**, smoking status.

In [18], a ML model is proposed to predict stroke occurrence for a patient. The authors implemented the RF, the LR, DT, and the KNN. They use gender, age, hypertension, heart disease, ever married, work type, residence type, average glucose level, BMI, and smoking status as features.

The research of Tazin et al. [41] uses a range of physiological parameters (age, hypertension, heart disease, ever married, average glucose level, and BMI) and ML algorithms, such as LR, DT, RF, and Voting Classifier, to train four different models for reliable prediction of stroke.

Dev et al. [10] proposed a predictive analytics approach for stroke prediction using ML and NNs. In this article, the authors benchmark four popular classification approaches: NN, DT, RF, and SVM for the purpose of stroke prediction from patient attributes. Other DL approaches such as **Convolutional Neural Networks (CNNs)**, LASSO, and Elastic Net are implemented for the prediction of stroke. In this article, the authors used gender, age, hypertension, heart disease, marital status, occupation type, residence (urban/rural) type, average glucose level, BMI, and the patient's smoking status as input features of all the algorithms.

In [21], a framework for the identification of bioelectrical signals combined with the DL approach enables the early detection and prediction of stroke disease. In this work, the authors implemented four algorithms: **Long-Short-Term-Memory (LSTM)**, Bi-directional LSTM, Gated Recurrent Unit, and Feedforward NN.

In [1], a dataset containing medical, physiological, and environmental tests for stroke was used to evaluate the efficacy of ML, DL, and a hybrid technique between DL and ML on the MRI dataset for cerebral hemorrhage. The features and metrics for the stroke dataset used in this work were: gender, age, hypertension, heart disease, ever-married, work type, residence type, average glucose level, BMI, and smoking status. The features are fed into the various classification algorithms, namely, SVM, KNN, DT, RF, and MLP. In the second dataset, the MRI images were evaluated by using the AlexNet model and AlexNet+SVM hybrid technique.

The work of [40] focused on ML model analysis to predict the early outcomes of IS and used model explanation skills to interpret the results. They compared four ML models, namely SVM, RF, **Light Gradient Boosting Machine (LGBM)**, and DNN.

The objective of the research in [12] is to apply three current DL approaches for 6-month IS outcome predictions, using the openly accessible International Stroke Trial dataset. Furthermore, another objective of this research is to compare these DL approaches with ML for performing clinical prediction. The authors implemented CNN, LSTM, residual neural network, deep Forest, RF, and SVM.

In the study referenced as [11], a robust framework for long-term risk prediction of stroke occurrence is designed using ML techniques. The researchers developed and evaluated several ML models for this purpose. These models include NB, RF, LR, KNN, **Stochastic Gradient Descent (SGD)**, DT, MLP, and **Majority Voting (MV)**. The researchers trained and tested these models to assess their performance in predicting the long-term risk of stroke.

We notice that most related works include risk factors in their studies (Table 2) to reduce the burden of stroke, but they do not consider imprecise and inaccurate variables such as sports level. Moreover, they do not model the semantic relationship between variables and do not provide a risk score for stroke. The majority of the works that have invested in disease detection have used advanced methods and algorithms based on the patients' real data, but not for the case of disease prevention for healthy subjects.

Therefore, to correct the mentioned limitations, we chose to model our data by a mind map [13] that can help us to make new connections between features that we may not have previously considered. By visually representing the data, we may be able to identify relationships and patterns that are not immediately apparent from looking at the data in a more traditional format. Also, we applied fuzzy logic to handling uncertainty and imprecision variables because fuzzy logic provides a way to handle such data by assigning degrees of membership to different categories. In addition, fuzzy logic allows for the incorporation of expert knowledge in the form of linguistic rules. The advantage of stroke risk prediction modeling is that it does not require large amounts of data or complex ML algorithms. Instead, it relies on expert knowledge and can be easily updated as new research on stroke risk factors becomes available. It also provides a transparent and interpretable model that can help clinicians and patients

Table 2. Comparison of the Literature on Computer Science Research Applied to Stroke

Reference	Risk scoring prediction	Training dataset	Methods	Stroke prediction stroke	Stroke detection
[7]	-	Real Korean patients	DNN, PCA	√	
[3]	-	Real Saudi Arabia patients	PCA, J48, NN	√	
[10]	-	Real Electronic Medical Record for patients	NN, DT, RF	√	
[17]	-	Real Bangladesh patients	Fuzzy logic, C-means		√
[6]	√	UCI Machine learning Web site Dataset	NB, SVM, DL	√	
[36]	-	International Stroke Trial database	SVM	√	
[43]	-	Real Sudanese patients	DT, KNN		√
[4]	-	Real Turkey patients	SVM, SGB, PLR	√	
[23]	√	Real Korean patients	LR	√	
[42]	-	Real Thailand patients	NB, DT, ANN	√	
[38]	-	Kaggle Dataset	LR, DT, RF, KNN, SVN, NB	√	
[18]	-	Kaggle Dataset	RF, LR, DT, KNN	√	
[41]	-	Kaggle Dataset	LR, DT, RF, VC	√	√
[10]	-	Kaggle Dataset	NN, DT, RF, SVM, CNN, LASSO, ElasticNet	√	
[1]	-	Kaggle Dataset	RF, DT, KNN, SVM, MLP, AlexNet, AlexNte+SVM		√
[40]	-	Real Taiwan patient	SVM, RF, LGBM, DNN	√	
[11]	-	Kaggle Dataset	NB, RF, KNN, SGD, DT, MLP, MV	√	

make informed decisions about stroke prevention and treatment. However, ML and DL algorithms rely heavily on accurate and precise input variables to generate accurate predictions or classifications. When the input variables are imprecise or inaccurate, these algorithms may encounter several disadvantages, including decreased accuracy because when the input data contain errors or inaccuracies, the ML or DL algorithms may produce inaccurate or unreliable results. These inaccuracies may propagate through the entire model and affect the overall accuracy of the predictions. Also when the input data are imprecise or inaccurate, the ML or DL algorithm may learn the noise and inconsistencies in the data, rather than the underlying patterns. This can lead to overfitting, where the algorithm performs well on the training data, but poorly on new data. Also, when the input data are imprecise or inaccurate, the model may have difficulty generalizing to new data that are different from the training data. This can result in poor performance on real-world data. In addition, we have used approximative input data processing, the numerical and categorical input variables to handle the uncertain data and identify and remove redundant or irrelevant features, or create new features that capture important relationships between the input variables. Also, to enhance the accuracy of stroke prediction we combine the fuzzy risk stroke model with an ML-Based Decision Support System. This combined system is specifically designed to handle various types of input data, including both imprecise and inaccurate data, as well as accurate data. By incorporating fuzzy logic techniques and ML algorithms, the system can effectively manage and interpret uncertain or incomplete information, leading to improved accuracy in early predicting stroke.

To summarize in this article, we present the following major contributions:

(1) Data modeling by mind maps to highlight hidden patterns that can also be generalized for other case studies; we utilize a mind mapping [13] technique to represent stroke and identify pertinent variables. This technique enables the extraction of relevant features and provides a semantic understanding of the data.
(2) Discrete data quantifying by fuzzy logic to have a **Stroke Risk Score (RSS)**.
(3) Fuzzy stroke risk prediction modeling independently of learning databases.
(4) Numeric and categorical input variable handling.
(5) Seamlessly integrates a fuzzy stroke risk model and an ML-based decision support system, to improve the accuracy of stroke prediction. Designed to efficiently process input data, including imprecise, inaccurate, and precise information, the combined system uses fuzzy logic techniques and ML algorithms to effectively manage the uncertainties associated with variables such as age and gender. By leveraging these techniques,

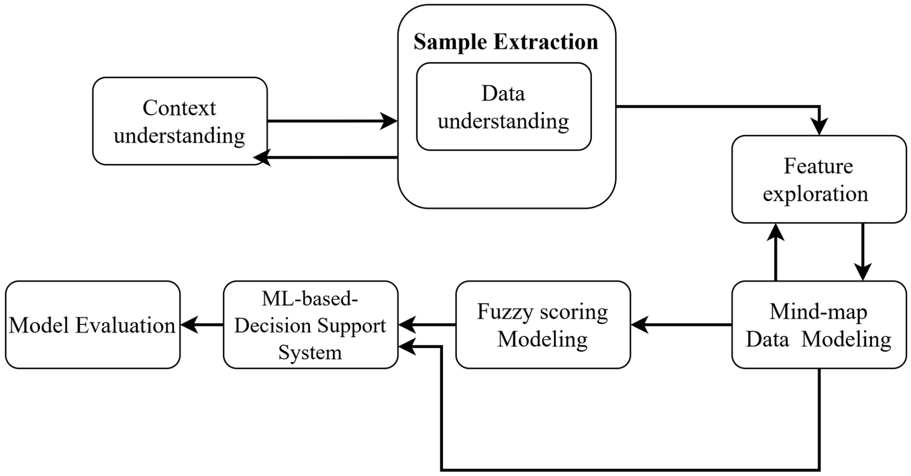

Fig. 1. Process of our methodology.

the combined system effectively manages inconsistencies and uncertainties in the input data, thereby improving data quality.

As a result, the comprehensive approach will provide reliable and accurate predictions, even in the presence of imperfect or uncertain data.

3 Methodology

The main focus of our research is to present convincing empirical evidence of the effectiveness achieved through the integration of fuzzy logic and ML-based decision support systems. The main objective is to demonstrate how this fusion can be exploited to develop highly accurate models for predicting and modeling stroke risk. To achieve our goal, we adopted a holistic approach combining Crisp-DM and SEMMA methodology. Firstly, we focused on understanding the context and then extracting and modeling the relevant data, despite the limitations of the data size. Secondly, we designed and evaluated a robust prediction model. Figure 1 illustrates our overall methodology, composed of seven main blocks:

(1) *Context Understanding*: We identified the problem to be solved, the related variable and we defined the scope of our research.
(2) *Sample Extraction*: We integrated a representative sample of data suitable for our study from "Cerebral Vasoregulation in Elderly with Stroke, 2018" [24] and real clinical data from Sahloul hospital. To better understand the data, we referred to relevant works such as [14, 15, 34], and also sought advice from valued professor Sana Ben Amor.[2]
(3) *Feature Exploration*: We used data graphs to visually represent relationships between features and identify dependencies. This process enabled us to make informed decisions about the appropriate features to include in our approach.
(4) *Mind Map Data Modeling*: By leveraging insights from relevant works and benefiting from the expertise of Prof. Sana Ben Amor, we designed a mind map to succinctly summarize the key risk factors associated with strokes and employed it as a framework to structure and model our data.
(5) *Fuzzy Scoring Modeling*: We developed our approach named "fuzzy stroke risk prediction," employing the fuzzy logic algorithm for predicting strokes.

[2]Head of the neurology department at the Sahloul hospital: https://www.researchgate.net/profile/Sana-Amor-2.

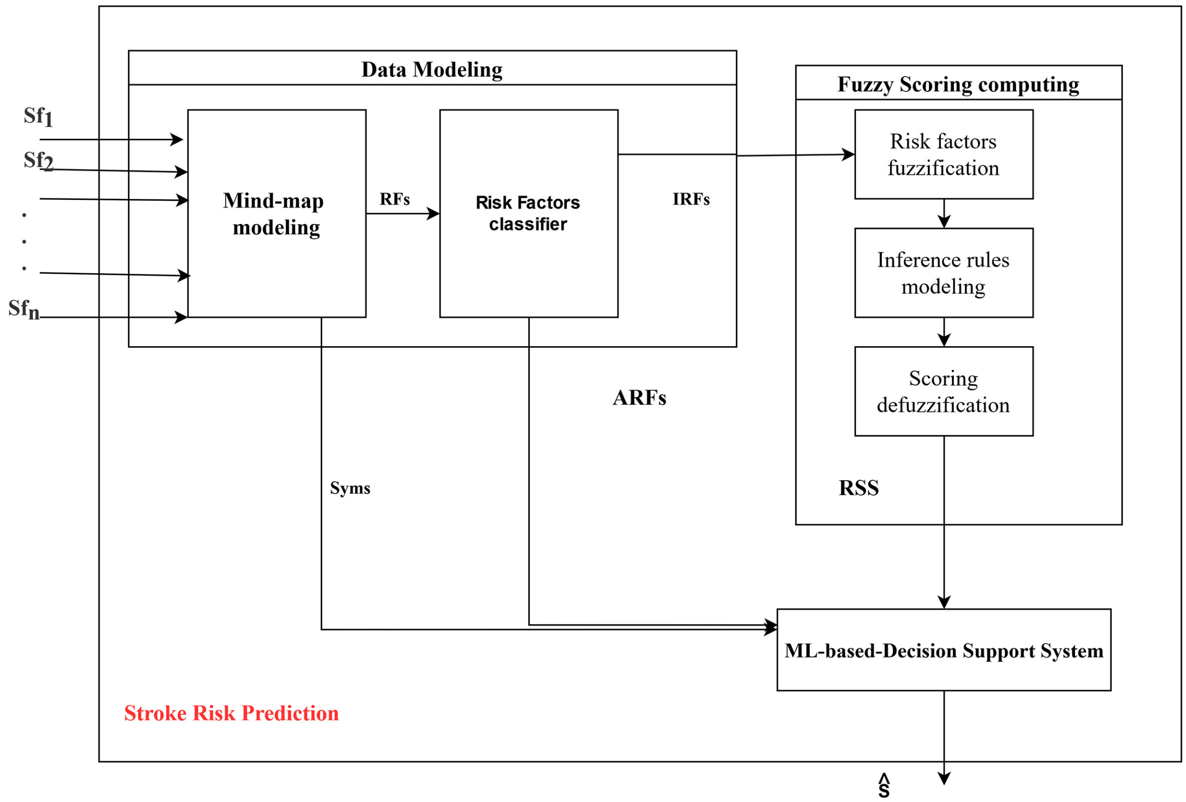

Fig. 2. Fuzzy scoring-based decision support system to predict stroke model.

(6) *ML-Based Decision System*: We designed a set of ML classifiers for stroke prediction in our ML-Based Decision Support System. These classifiers included KNN, LR, DT, ANN, and SVM. Inputs included stroke risk scoring obtained from the fuzzy stroke risk prediction model, symptoms, and **Accurate Risk Factors (ARFs)**.

(7) *Model Evaluation*: To assess the performance of our model, we carried out an in-depth evaluation, comparing it to a decision support system that predicts stroke, but does not have a fuzzy stroke risk scoring system. In addition, we compared the accuracy of our model with other relevant studies in this field. Thanks to this comprehensive evaluation, we were able to determine the effectiveness of our approach in delivering accurate stroke predictions.

4 Fuzzy Scoring-Based Decision Support System to Early Predict Stroke

Our proposed model for early stroke prediction is centered around a fuzzy score decision support system, which consists of three crucial elements: a data modeling process, a fuzzy score calculation process, and an ML-based decision support system. These components are illustrated in Figure 2.

The input data for the system include a set of stroke characteristics (sf1...sfn) and risk factors (RFs) associated with stroke. These risk factors are further classified into two distinct categories: **Imprecise Risk Factors (IRFs)** and precise risk factors (ARF). Additionally, the prediction process considers stroke symptoms (Syms).

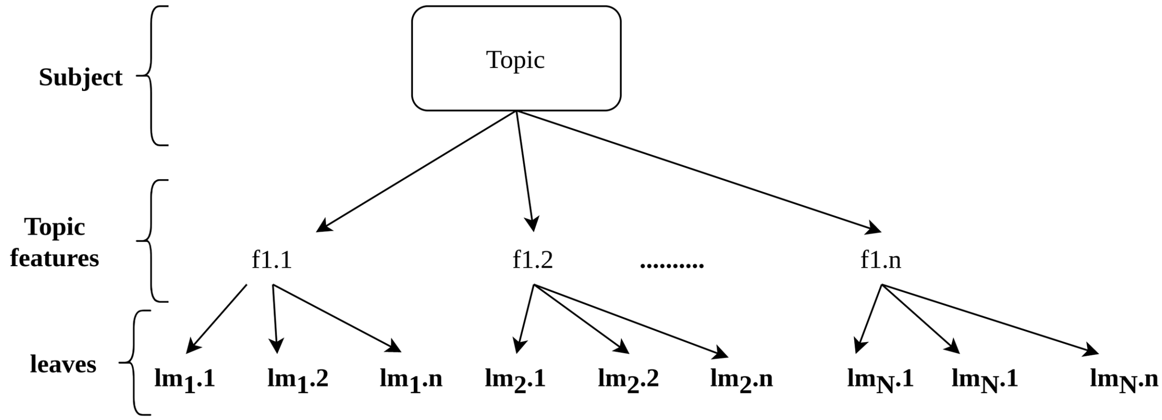

Fig. 3. Global levels of mind map.

Using fuzzy scoring techniques, the system calculates a RSS, which takes into account input characteristics, risk factors, and symptoms. The model's outcome is represented by the output $\hat{S}$, which indicates the predicted occurrence of stroke.

In this study, we propose to use a mind map for knowledge representation and fuzzy logic rules to assess stroke risk factors, thus improving the accuracy of ML algorithms used in stroke prediction.

4.1 Data Modeling

4.1.1 Mind Map Modeling. To organize and comprehend the data along with their interconnections, we have employed a mind map structure. The construction of this mind map (Figure 4) is derived from [14, 15, 34] and the information presented in Table 1. The mind map consists of three primary levels, depicted in Figure 3:

— *Subject*: This level corresponds to the disease under consideration.
— *Topic Features*: It encompasses various axes of analysis related to the subject.
— *Leaves*: These correspond to complex atomic data, offering more detailed, finer-grained information.

The mind map depicted in Figure 4 provides a comprehensive overview of stroke risk factors, relevant variables, and symptoms. Within this mind map, certain risk factors can be modified and may vary among individuals, while others remain stable and unchangeable. Modifiable risk factors, such as smoking, diabetes, and obesity, can appear or disappear over time. Conversely, non-modifiable risk factors, including gender, age, and race, cannot be altered.

In our mind map, the subject is represented by the stroke disease, while the topic features are represented by the risk factor axis. The leaves within the mind map symbolize specific variables such as gender, age, cardiac disease, and more.

4.1.2 Risk Factors Classifier. In this section, we classify the risk factor features into two categories: inaccurate variables and accurate variables of risk factors. As shown in Figure 2, the risk factor classifier produces two sets of variables: IRFs representing the imprecise risk factors of stroke, and ARFs representing the precise risk factors. The IRFs are used as input variables for the fuzzy scoring computing module, while the ARFs serve as input variables for the ML-Based Decision Support System module.

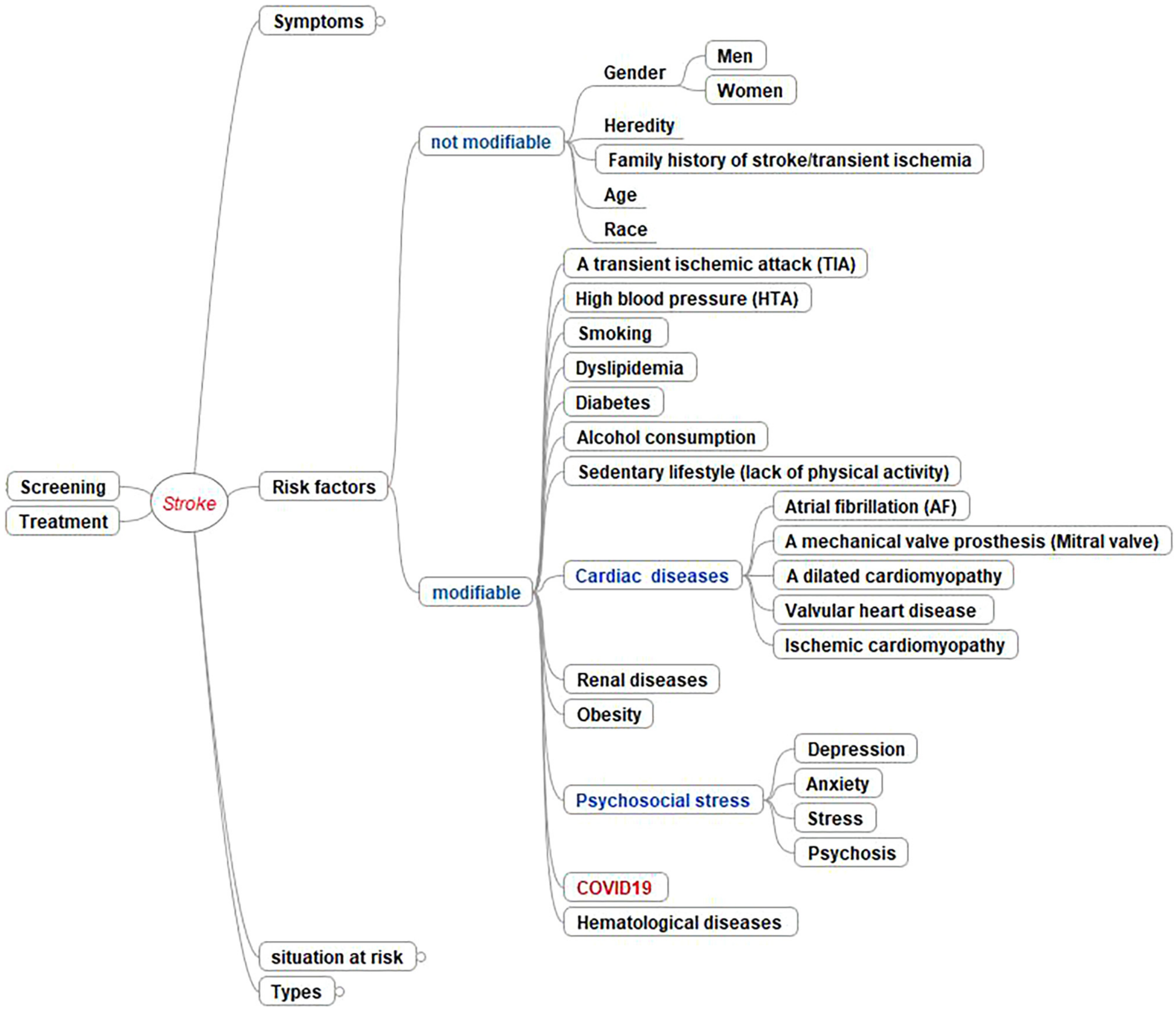

Fig. 4. Detailed levels of Mind-map summarizing the risk factors of stroke.

4.2 Fuzzy Scoring Computing Process

The main goal of the fuzzy scoring model is to incorporate non-linear input data, unveil concealed patterns, and accurately depict the uncertainty associated with variables to achieve precise predictions of the RSS. Fuzzy Scoring Modeling comprises three key processes: risk factors fuzzification, inference rules modeling, and scoring defuzzification.

(1) *Risk Factors Fuzzification.* The risk factors fuzzification phase aims to convert numerical data into linguistic variables. In Figure 5, the fuzzification module takes numerical values of stroke risk factors as input and produces fuzzified variables based on low, medium, and high-class membership degree functions.

    As an illustration, the fuzzification process of the risk factor “age” involves categorizing it into four classes: low, medium, high, and very high. The membership degrees of the risk factor “age” for each

Fig. 5. Risk factors fuzzification.

category, namely "low," "medium," "high," and "very high," are respectively denoted by Equations (1) to (4), as exemplified in [20].

$$\mu_{age-R-low}(age) = \begin{cases} 1 & \text{for age } \leq 40 \\ \frac{age-40}{9} & \text{for } 40 < \text{age} < 49 \end{cases} \tag{1}$$

$$\mu_{age-R-Medium}(age) = \begin{cases} 0 & \text{for age } \leq 49 \\ \frac{age-49}{2} & \text{for } 49 < \text{ age } < 51 \\ 1 & \text{for } 51 \leq \text{ age } \leq 60 \\ \frac{65-age}{5} & \text{for } 60 < \text{ age } < 65 \\ 0 & \text{for age} \geq 65 \end{cases} \tag{2}$$

$$\mu_{age-R-high}(age) = \begin{cases} 0 & \text{for age} \leq 55 \\ \frac{age-55}{10} & \text{for } 55 < \text{age} < 65 \\ 1 & \text{for } 65 \leq \text{ age } \leq 75 \\ \frac{85-age}{10} & \text{for } 75 < \text{age} < 85 \\ 0 & \text{for age } \geq 85 \end{cases} \tag{3}$$

$$\mu_{age-R-very-high}(age) = \begin{cases} 0 & \text{for age} \leq 70 \\ \frac{age-70}{15} & \text{for } 70 < \text{age} < 85 \\ 1 & \text{for age } \geq 85 \end{cases}. \tag{4}$$

Additionally, let's consider another example of fuzzification for the risk factor "smoking cigarettes per day." This variable is categorized into three classes: low, high, and very high. The membership degrees of the risk factor "smoking cigarettes per day" for each category, namely "low," "high," and "very high," can be

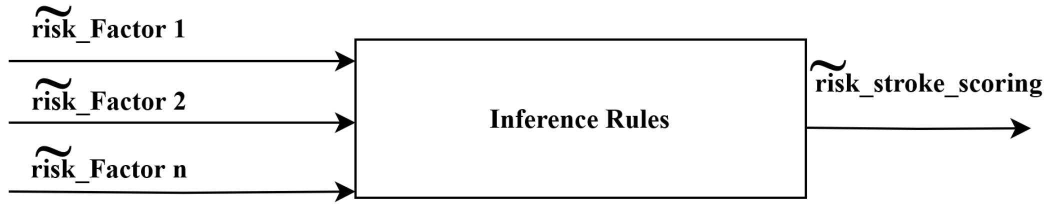

Fig. 6. Inference rules modeling.

respectively represented by Equations (5) to (7):

$$\mu_{Smoking-R-low}(smoking) = \begin{cases} 1 & \text{for smoking } \leq 9 \\ \frac{smoking-9}{2} & \text{for } 9 < \text{smoking} < 11 \end{cases} \tag{5}$$

$$\mu_{Smoking-R-high}(smoking) = \begin{cases} 0 & \text{for smoking } \leq 8 \\ \frac{smoking-8}{4} & \text{for } 8 < \text{smoking} < 12 \\ 1 & \text{for } 12 \leq \text{smoking } \leq 25 \\ \frac{30-smoking}{5} & \text{for } 25 < \text{smoking} < 30 \\ 0 & \text{for smoking } \geq 30 \end{cases} \tag{6}$$

$$\mu_{Smoking-R-very-high}(smoking) = \begin{cases} 0 & \text{for smoking } \leq 15 \\ \frac{smoking-15}{20} & \text{for } 15 < \text{smoking} < 35 \\ 1 & \text{for smoking } \geq 35 \end{cases}. \tag{7}$$

(2) *Inference Rules Modeling.* The concept of inference rules involves setting up rules based on fuzzy stroke risk factors, rating and aggregating these rules to obtain a single, comprehensive fuzzy RSS (see Figure 6).
The inference rules of fuzzy stroke risk prediction follow the structure of Equation (8) [22]:

$$\begin{gathered}\textbf{if } \mathbf{A} \text{ is } s_A \text{ and } \mathbf{D} \text{ is } s_D \text{ and } \mathbf{Sm} \text{ is } s_{sm} \text{ and } \mathbf{Al} \text{ is} \\ s_{Al} \text{ and } \mathbf{PA} \text{ is } s_{PA} \textbf{ Then scoring} \text{ is } s_{scoring}, \end{gathered} \tag{8}$$

where:
— **A**: represents the age
— s_A: is the membership function class of age
— **D**: represents the Diabetes
— s_D: is the membership function class of Diabetes
— **Sm**: represents the smoking
— s_{sm}: is the membership function class of smoking
— **Al**: represents the alcoholism

Fig. 7. Scoring defuzzification.

- s_{Al}: is the membership function class of alcoholism
- **PA**: represents the physical activity
- s_{PA}: is the membership function class of physical activity.

(3) *Scoring Defuzzification.* The scoring defuzzification process is the reverse of fuzzification, where mapping is performed to convert the fuzzy stroke risk scoring into precise outcomes, as illustrated in Figure 7.

We used the **Center of Gravity (COG)** method defuzzification. This method determines a definite value by considering the COG of the fuzzy set. To perform this task, the total area of the membership function distribution representing the combined control action is divided into several sub-areas. The area and centroid of each sub-area are calculated and their sum is used to obtain the defuzzified value of a discrete fuzzy set [9]. Equation (9) [20], [8] represents our scoring defuzzification:

$$risk\_stroke\_Scoring = \frac{\int_U (risk\_factor \mu(risk\_factor)\, drisk\_factor)}{\int_U \mu(risk\_factor)\, drisk\_factor}, \tag{9}$$

where:

- $\mu(risk\_factor)\, drisk\_factor$: denotes the area of the region bounded by the curve risk_factor.

4.3 ML-Based Decision Support System

The ML-Based Decision Support System serves as the ultimate component of our model, holding a critical role in stroke prediction and decision-making. This module utilizes ML classifiers that take inputs from other modules, namely the "RSS," "ARFs," and "Syms." By processing this collective information, the system generates an output denoted as $\hat{S}$, representing the stroke prediction.

5 Experiments and Validation

5.1 Experiments

5.1.1 Experiment Setup. The fuzzy scoring computing process developed using scikit-fuzzy is a Python library that provides tools for fuzzy logic and fuzzy control. Fuzzy logic is a mathematical approach that deals with uncertainty and imprecision in data and decision-making.

The ML-Based Decision Support System developed using Keras whose high-level NNs API is written in Python.

5.1.2 Fuzzy Scoring Computing. For the validation of our approach, we considered age, smoking, diabetes, physical activity, and alcoholism as input variables for the fuzzy scoring modeling. The output parameter focused on the risk scoring of stroke.

We implemented the memberships of each feature with the Python programming language.

The memberships of each feature were implemented using the Python programming language. Tables 3 and 4 provide two examples of the membership functions for the variables "age" and "smoking," respectively. In both fuzzification processes, trapezoidal membership functions were utilized. The visual representation of these implemented memberships can be observed in Figures 8 and 9. In the inference rule modeling phase, we

Table 3. Risk Level Classification for Age

Age value	Risk
<50 years	Low
[50, 65] years	Medium
[66, 75] years	High
>75 years	Very high

Table 4. Risk Level Classification for Smoking

Number of cigarettes/day	Risk
<10 cigarettes	Low
[10, 20] cigarettes	High
>20 cigarettes	Very high

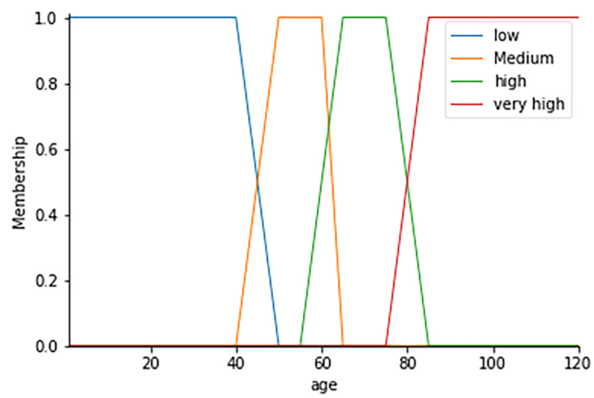

Fig. 8. Fuzzification of age risk.

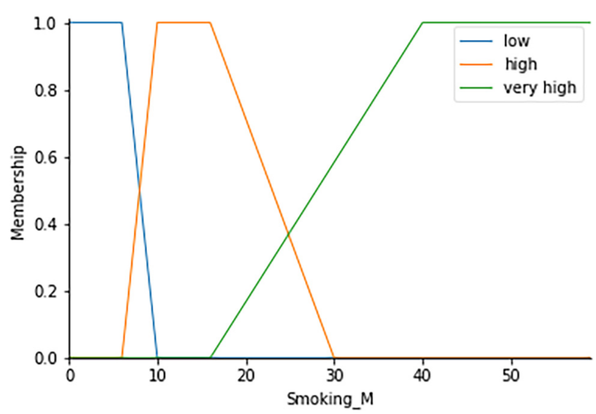

Fig. 9. Fuzzification of smoking for men.

```
RuleBlock:
  R1: if sport is low and age is Medium and alcoholism is Medium and diabetes is Medium and smoking is very high then scoring stroke risk is high
  R2: if sport is low and age is very high and alcoholism is low and diabetes is very high and smoking is very high Then scoring stroke risk is very high
  R3: if sport is low and age is Medium and alcoholism is Medium and diabetes is low and smoking is low Then scoring stroke risk is Medium
  R4: if sport is low and age is low and alcoholism is low and diabetes is low and smoking is low Then scoring stroke risk is low

      .
      .
      .


END_RuleBlock
```

Fig. 10. An example of inference rules.

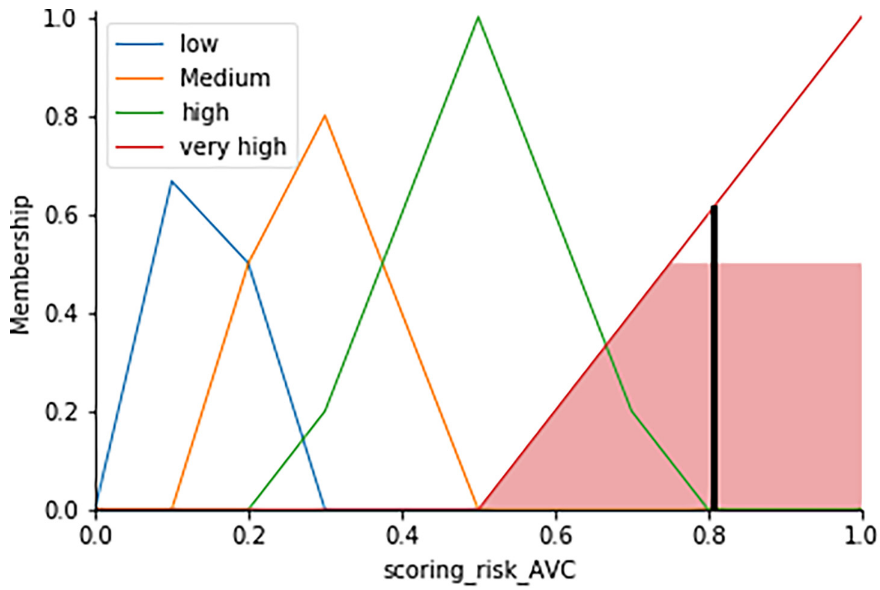

Fig. 11. Example scoring stroke risk classification.

implemented a total of 767 rules, encompassing all possible combinations of features and classes. Figure 10 shows four illustrative examples of these inference rules.

Defuzzification provides the simulation of the scoring and its corresponding class. The inputs for this process include age, the number of cigarettes per day, the number of diabetic years, the number of drinks of alcohol consumed per day, and the number of hours per week dedicated to physical activity. For instance, we conducted a test of our model using the following parameters:

— *Number of cigarettes per day*: 20
— *Age*: 80 years
— *Number of diabetic years*: 15 years
— *Number of glasses of alcohol per day*: 2.5 glasses/day
— *Number of hours per week of physical activity*: 0

The resulting RSS was 0.80, classified as "very high" according to Figure 11.

5.1.3 ML-Based Decision Support System. In our experiment, we implemented the ML-Based Decision Support System using two different architectures. The first architecture, which we proposed and illustrated in Figure 2, takes as input the risk scoring of stroke derived from the fuzzy scoring computing model, along with the ARFs

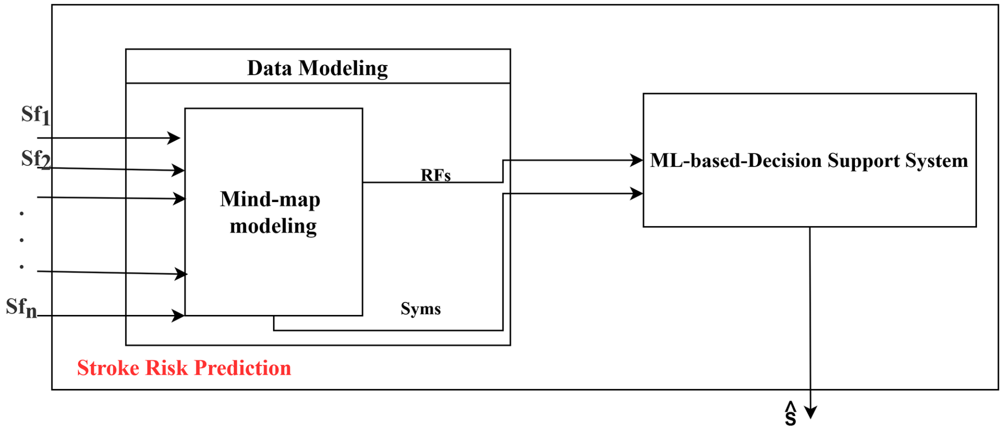

Fig. 12. Decision support system without fuzzy scoring model.

and symptoms. The second architecture closely resembles the first but omits certain processes, including risk factors classifiers, fuzzy scoring computing, and the risk scoring of stroke. In the second architecture, as shown in Figure 12, the inputs for the second architecture are limited to the risk factors of stroke and the symptoms. Both architectures in our experiment incorporate a range of ML classifier algorithms, including KNN, LR, DT, ANN, and SVM. These algorithms analyze the input data and make predictions or decisions within the ML-Based Decision Support System.

To select the optimal value of k in the KNN algorithm, we begin by defining a range of possible k values, usually spanning from 1 to a few dozen. We then apply k-fold cross-validation to assess the model's performance for each k. By plotting these results, we can identify the k that provides the best balance between accuracy and computational efficiency. While smaller k values can cause overfitting, larger k values may lead to underfitting.

Table 5 presents the accuracy metrics of each classifier for both architectures. It provides a comprehensive overview of the accuracy achieved by each classifier in their respective architectures.

(1) *Architecture Including Fuzzy Scoring Computing*: Upon analyzing Table 5, we observe that the KNN, LR, and SVM classifiers achieved a perfect accuracy rate of 100%. The ANN classifier demonstrated an accuracy of 98.33%. In contrast, the DT classifier exhibited the lowest accuracy among all classifiers, achieving 96.66% accuracy.
(2) *Architecture without Fuzzy Scoring Computing*: In this architecture, the results obtained by the ML classifiers are summarized in Table 5. The latter illustrates that the LR and ANN classifiers achieved the highest accuracy, at 96.66%. The DT classifier achieved an accuracy of 95%, while the SVM classifier achieved an accuracy of 91.66%. The KNN classifier had the lowest accuracy at 86.66%.

5.2 Validation

5.2.1 Validation of Fuzzy Scoring Computing. We tested our model with a dataset extracted from [24] which contains the age, diabetes, smoking, alcoholism, and physical activity risk factors.

We obtained the simulation results presented in Figure 13. We noticed that the stroke risk scoring obtained from our approach is very close to the real value. From the correlation matrix of Figure 14, we can see that the value of the correlation between the real cases and our risk scoring is 0.72. Also, we can see that diabetes and smoking have a strong correlation with the risk scoring.

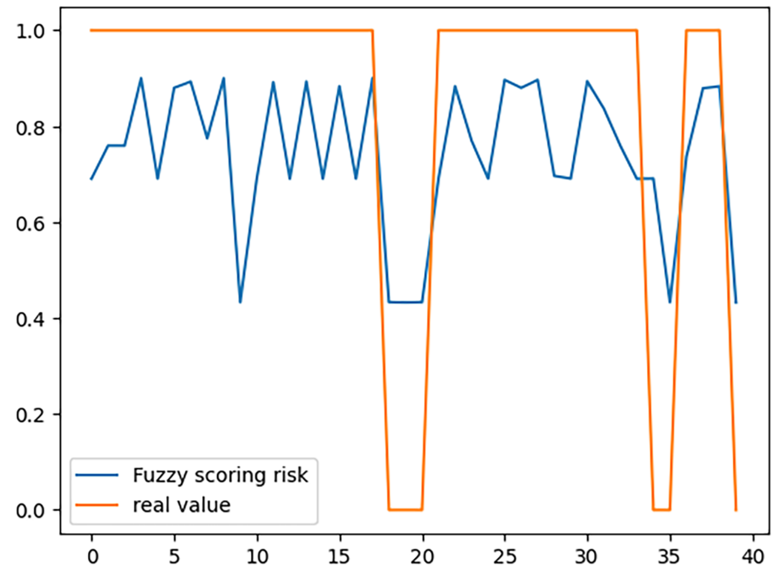

Fig. 13. The stroke scoring risk vs. real cases.

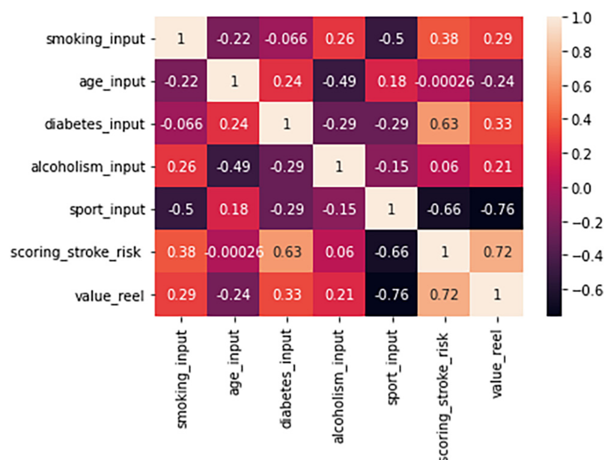

Fig. 14. Correlation matrix.

5.2.2 *Validation of ML-Based Decision Support System.* To validate our approach, we conducted three distinct comparisons:

— Firstly, we compared our Fuzzy Scoring-Based Decision Support System with a Decision Support System that does not utilize fuzzy scoring for stroke prediction. This allowed us to assess the effectiveness and added value of incorporating fuzzy scoring techniques in our model.
— Secondly, we validated our model using out-of-sample data. By testing our system on data that was not used during the model development phase, we could evaluate its performance in real-world scenarios and ensure its generalizability.

Table 5. Accuracy Test between Fuzzy Scoring-Based Decision Support System and Decision Support System without Fuzzy Scoring to Predict Stroke

Model	Accuracy of fuzzy scoring-based decision support system	Accuracy of decision support system without fuzzy scoring
KNN	100	86.66
LR	100	96.66
DT	96.66	95.00
ANN	98.33	96.66
SVM	100	91.66

Table 6. Precision Test between Fuzzy Scoring-Based Decision Support System and Decision Support System without Fuzzy Scoring to Predict Stroke

Model	Precision of fuzzy scoring-based decision support system	Precision of decision support system without fuzzy scoring
KNN	100	95
LR	100	97
DT	97	97
ANN	100	98
SVM	100	92

Table 7. Recall Test between Fuzzy Scoring-Based Decision Support System and Decision Support System without Fuzzy Scoring to Predict Stroke

Model	Recall of fuzzy scoring-based decision support system	Recall of decision support system without fuzzy scoring
KNN	100	87
LR	100	97
DT	100	95
ANN	98	100
SVM	100	97

— Thirdly, we compared our work with other relevant studies in the field of stroke prediction. This comparison aimed to evaluate the performance of our approach against existing methods and determine its competitiveness and advancements in the field.

By conducting these comparisons, we were able to validate the efficacy of our Fuzzy Scoring-Based Decision Support System for stroke prediction and demonstrate its comparative advantages over alternative approaches.

From analyzing Tables 5 to 7, we observed the following results:

— Our model using KNN achieved 100% accuracy, while the model without fuzzy scoring reached 86.66%. In terms of precision, our model also attained 100%, compared to 95% for the model without fuzzy scoring. For recall, our model achieved 100%, whereas the model without fuzzy scoring attained 87%.
— Similarly, for LR, our model achieved 100% accuracy, while the model without fuzzy scoring achieved 96.66% accuracy. In terms of precision, our model also attained 100%, compared to 97% for the model without fuzzy scoring. For recall, our model achieved 100%, whereas the model without fuzzy scoring attained 97%.

Table 8. Our Model Validation on Out-of-Sample Data Using Accuracy Metric

Model	Accuracy of this approach with insample data	Accuracy of this approach with outsample data
KNN	100	98.33
LR	100	95
SVM	100	96.66

Table 9. Model Validation on Out-of-Sample Data Using Precision Metric

Model	Precision of this approach with insample data	Precision of this approach with outsample data
KNN	100	100
LR	100	94.23
SVM	100	96.18

— With the DT, our model achieved 96.66% accuracy, whereas the other model achieved 95.00% accuracy. In terms of precision, our model attained 97%, compared to 97% for the model without fuzzy scoring. For recall, our model achieved 100%, whereas the model without fuzzy scoring attained 95%.
— The ANN model provided us with 98.33% accuracy, whereas the model without fuzzy scoring obtained 96.66% accuracy. In terms of precision, our model attained 100%, compared to 98% for the model without fuzzy scoring. For recall, our model achieved 98%, whereas the model without fuzzy scoring attained 100%.
— Lastly, for SVM, our model achieved 100% accuracy, while the model without fuzzy scoring obtained 91.66% accuracy. In terms of precision, our model also attained 100%, compared to 92% for the model without fuzzy scoring. For recall, our model achieved 100%, whereas the model without fuzzy scoring attained 97%.

Based on these results, we can conclude that our model consistently outperformed the model without fuzzy scoring, demonstrating higher accuracy in all cases. Furthermore, KNN, LR, and SVM achieved 100% accuracy, precision, and recall, suggesting that these algorithms performed exceptionally well.

After observing that KNN, LR, and SVM achieved the same accuracy, the same precision, and recall in our initial comparison, we sought to determine the best classifier by evaluating our approach with out-of-sample data. According to Tables 8 and 9, KNN achieved an accuracy of 98.33% and 100% precision on the out-of-sample data, while SVM recorded an accuracy of 96.66% and precision of 96.18%. LR, however, had the lowest performance, with an accuracy of 95% and precision of 94.23%. Based on these results, we concluded that KNN performed the best among the three classifiers in our approach.

According to the data presented in Table 10, our approach demonstrates superior accuracy compared to other methods. Our approach achieves a remarkable 98.33% accuracy, while the highest accuracy achieved by other approaches is 96%.

By considering the three comparisons, it becomes evident that our Fuzzy score_based approach consistently achieves the highest accuracy regarding others related works.

This achievement is mainly attributed to the pre-processing and fuzzy modeling of the input variables, representing the major risk factors for the disease, notably age, diabetes, smoking, alcoholism, and physical activity. A significant contribution of this work lies in the extraction of critical knowledge from independent and heterogeneous rough data, enabling the development of a valuable model for accurately determining RSSs using fuzzy logic.

6 Conclusion and Future Work

We proposed a Fuzzy Scoring-Based Decision Support System approach that offers a new and innovative way to early predict stroke. The fuzzy scoring-based Decision Support System is composed of three main modules: Data

Table 10. A Comparative Analysis of This Work and Existing Works

Reference	Mind map data modeling	Applied classifier	Best classifier	Accuracy
Our approach	√	KNN, LR, DT, ANN, SVM	KNN	98.33%
[7]	-	DNN, PCA	DNN	83.48%
[3]	-	PCA, DT, NN	DT	95.25%
[10]	-	NN, DT, RF	NN	78%
[17]	-	Fuzzy logic+cmeans	Fuzzy logic+cmeans	95.1%
[6]	-	NB, SVM, DL	SVM	49%
[36]	-	SVM	SVM	90%
[4]	-	SVM, SGB, PLR	SVM	95%
[23]	-	LR	LR	62%
[42]	-	NB, DT, ANN	ANN	84%
[38]	-	LR, DT, RF, KNN, SVN, NB	NB	82%
[41]	-	LR, DT, RF, VC	RF	96%
[10]	-	NN, DT, RF, SVM, CNN, LASSO, ElasticNet	NN	80%
[40]	-	SVM, RF, LGBM, DNN	RF	82.9%

Modeling, Fuzzy Scoring Computing, and ML-Based Decision System. The Data Modeling module comprises two major processes: Mind Map modeling and the risk factor classifiers. The use of a mind map enables us to model the semantic relationships among independent input variables and extract meaningful features. This approach provides a deeper understanding of the intricate connections between various risk factors and their influence on stroke risk. Significant risk factors are thus identified and relevant features are extracted to develop the scoring system. Moreover, the risk factor classifiers module plays a crucial role in categorizing the risk factors as accurate or inaccurate. The fuzzy scoring computing module encompasses three primary processes: risk factor fuzzification, inference rules modeling, and scoring defuzzification. These processes involve converting the input risk factors into fuzzy sets, formulating inference rules, and applying defuzzification techniques to obtain a risk score.

The resulting risk score can be used to identify people at high risk of stroke enabling targeted interventions to prevent or mitigate the risk.

The ML-based decision system module is designed to predict strokes. It includes ML classifiers that consider input data such as fuzzy stroke risk assessment, specific risk factors, and symptoms. Based on this information, the module produces a score indicating the stroke prediction.

The experimental and validation results demonstrated a high level of conformity with real detected cases. Moreover, our approach achieved the highest accuracy when compared to architectures without fuzzy scoring and other related works.

We conclude that the approach proposed by the fuzzy score-based support system holds great promise for advancing the field of stroke risk assessment and prevention. With further research and development, this approach could lead to a better understanding of the complex relationships between different risk factors and their influence on stroke risk. Ultimately, this could lead to improved stroke prevention and treatment strategies.

Statements and Declarations

We hereby declare that we have no conflicts of interest to disclose pertaining to the submission of our article to ACM Transactions on Computing for Healthcare. We understand the importance of transparency and the need to uphold ethical standards in academic publishing. To the best of our knowledge, there are no financial, professional, or personal relationships that could be perceived as potential conflicts of interest that may influence the objectivity, integrity, or impartiality of our research or its publication.

We affirm that the following statements accurately represent our situation regarding conflicts of interest:

— *Financial Relationships*: We have not received any financial support, funding, grants, or sponsorships from any organization that may have a direct or indirect interest in the research presented in this article.

—*Professional Relationships*: We have no professional relationships or affiliations with any individuals, organizations, or companies that could be seen as potential conflicts of interest with the research or its publication.
—*Personal Relationships*: We have no personal relationships with any individuals involved in the review, editorial, or publication process of ACM Transactions on Computing for Healthcare that could influence the fairness, objectivity, or impartiality of the evaluation or decision-making process.

References

[1] Zeyad Al-Mekhlafi, Ebrahim Senan, Taha Rassem, Badiea Al-Shaibani, Nasrin Makbol, Adwan Alanazi, Tariq Almurayziq, and Fuad Ghaleb. 2022. Deep learning and machine learning for early detection of stroke and haemorrhage. *Computers, Materials and Continua* 72 (02 2022), 775–796. DOI: https://doi.org/10.32604/cmc.2022.024492

[2] Abdullah Albohari, Abdul Kareem Alshami, Abdul Nahas, Adeel Aslam, Mohammed Zawiah, and Shazia Jamshed. 2020. Awareness towards signs and symptoms of stroke and assessment of factors associated with knowledge of five symptoms of stroke among general public Malaysia: A regression analysis. (03 2020). DOI: https://doi.org/10.21203/rs.3.rs-17583/v1

[3] Ohoud Almadani and Riyad Alshammari. 2018. Prediction of stroke using data mining classification techniques. *International Journal of Advanced Computer Science and Applications* 9 (01 2018). DOI: https://doi.org/10.14569/IJACSA.2018.090163

[4] Ahmet Arslan, Cemil Colak, and Ediz Sarihan. 2016. Different medical data mining approaches based prediction of ischemic stroke. *Computer Methods and Programs in Biomedicine* 130 (03 2016). DOI: https://doi.org/10.1016/j.cmpb.2016.03.022

[5] Stephen Bacchi, Toby Zerner, L. Oakden-Rayner, T. Kleinig, S. Patel, and J. Jannes. 2019. Deep learning in the prediction of ischaemic stroke thrombolysis functional outcomes: A pilot study. *Academic Radiology* (2019).

[6] Pattanapong Chantamit-o pas and Madhu Goyal. 2017. Prediction of stroke using deep learning model. In *Proceedings of the International Conference on Neural Information Processing*, 774–781. DOI: https://doi.org/10.1007/978-3-319-70139-4_78

[7] Songhee Cheon, Jungyoon Kim, and Jihye Lim. 2019. The use of deep learning to predict stroke patient mortality. *International Journal of Environmental Research and Public Health* 16 (05 2019), 1876. DOI: https://doi.org/10.3390/ijerph16111876

[8] Amine Chohra, Kurosh Madani, and Dalel Kanzari. 2010. Fuzzy cognitive and social negotiation agent strategy for computational collective intelligence. In *Proceedings of the 1st International Conference on Computational Collective Intelligence: Semantic Web, Social Networks and Multiagent Systems (ICCCI '09)*, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 143–159. DOI: https://doi.org/10.1007/978-3-642-15034-0_9

[9] Franck Dernoncourt. 2011. Introduction à La Logique Floue. (08 2011).

[10] Soumyabrata Dev, Hewei Wang, Chidozie Shamrock Nwosu, Nishtha Jain, Bharadwaj Veeravalli, and Deepu John. 2022. A predictive analytics approach for stroke prediction using machine learning and neural networks. *Healthcare Analytics* 2 (2022), 100032. DOI: https://doi.org/10.1016/j.health.2022.100032

[11] Elias Dritsas and Maria Trigka. 2022. Stroke risk prediction with machine learning techniques. *Sensors* 22 (06 2022), 4670. DOI: https://doi.org/10.3390/s22134670

[12] Gang Fang, Zhennan Huang, and Zhongrui Wang. 2022. Predicting ischemic stroke outcome using deep learning approaches. *Frontiers in Genetics* 12 (01 2022). DOI: https://doi.org/10.3389/fgene.2021.827522

[13] Paul Farrand, Fearzana Hussain, and Enid Hennessy. 2002. The efficacy of the 'mind map' study technique. *Medical Education* 36 (06 2002), 426–431. DOI: https://doi.org/10.1046/j.1365-2923.2002.01205.x

[14] Jose Gutierrez and Charles Esenwa. 2015. Secondary stroke prevention: Challenges and solutions. *Vascular Health and Risk Management* 11 (08 2015), 437. DOI: https://doi.org/10.2147/VHRM.S63791

[15] Valerie Hill and Amytis Towfighi. 2017. Modifiable risk factors for stroke and strategies for stroke prevention. *Seminars in Neurology* 37 (06 2017), 237–258. DOI: https://doi.org/10.1055/s-0037-1603685

[16] F. Ibrahim and N. I. Murr. 2020. *Embolic Stroke*. StatPearls Publishing.

[17] Farzana Islam. 2018. A fuzzy logic based predictive model for early detection of stroke. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 1841–1844. DOI: https://doi.org/10.1145/3267305.3277838

[18] Md Islam, Sharmin Akter, Md Rokunojjaman, Jahid Hasan Rony, and Al Amin. 2021. Stroke prediction analysis using machine learning classifiers and feature technique. *International Journal of Electronics and Communications Systems* 1 (12 2021), 1–7. DOI: https://doi.org/10.24042/ijecs.v1i2.10393

[19] Suraj Kanase and Nikita Jhaveri. 2020. Gender wise difference in presenting signs and symptoms of stroke: Observational study. *Indian Journal of Public Health Research and Development* 11 (05 2020), 270–273.

[20] Dalel Kanzari. 2013. Fuzzy psychological behavior for computational bilateral negotiation. In *Proceedings of the 2013 International Conference on Computer Applications Technology (ICCAT '13)*, 1–5. DOI: https://doi.org/10.1109/ICCAT.2013.6522004

[21] Mandeep Kaur, Sachin Sakhare, Kirti Wanjale, and Farzana Akter. 2022. Early stroke prediction methods for prevention of strokes. *Behavioural Neurology* 2022 (04 2022), 1–9. DOI: https://doi.org/10.1155/2022/7725597

[22] Yahya Lambat, Nick Ayres, Leandros Maglaras, and Mohamed Amine Ferrag. 2021. A Mamdani type fuzzy inference system to calculate employee susceptibility to phishing attacks. *Applied Sciences* 11, 19 (2021). DOI: https://doi.org/10.3390/app11199083

[23] Seungnam Min, Se Jin Park, Dong Kim, Murali Subramaniyam, and Kyung-Sun Lee. 2018. Development of an algorithm for stroke prediction: A national health insurance database study in Korea. *European Neurology* 79 (04 2018), 214–220. DOI: https://doi.org/10.1159/000488366

[24] Vera Novak. 2018. Cerebral vasoregulation in elderly with stroke. DOI: https://doi.org/10.13026/C2DW96

[25] Aytug Onan. 2019. Consensus clustering-based undersampling approach to imbalanced learning. *Scientific Programming* 2019 (03 2019), 1–14. DOI: https://doi.org/10.1155/2019/5901087

[26] Aytuğ Onan. 2019. Topic-enriched word embeddings for sarcasm identification. In *Software Engineering Methods in Intelligent Algorithms*. Radek Silhavy (Ed.), Springer International Publishing, Cham, 293–304.

[27] Aytug Onan. 2020. Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks. *Concurrency and Computation: Practice and Experience* 33 (06 2020). DOI: https://doi.org/10.1002/cpe.5909

[28] Aytuğ Onan. 2023. GTR-GA: Harnessing the power of graph-based neural networks and genetic algorithms for text augmentation. *Expert Systems with Applications* 232 (2023), 120908. DOI: https://doi.org/10.1016/j.eswa.2023.120908

[29] Aytuğ Onan. 2023. Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion. *Journal of King Saud University - Computer and Information Sciences* 35, 7 (2023), 101610. DOI: https://doi.org/10.1016/j.jksuci.2023.101610

[30] Aytuğ Onan. 2023. SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization. *Journal of King Saud University - Computer and Information Sciences* 35, 7 (2023), 101611. DOI: https://doi.org/10.1016/j.jksuci.2023.101611

[31] Aytuğ Onan and Serdar Korukoğlu. 2017. A feature selection model based on genetic rank aggregation for text sentiment classification. *Journal of Information Science* 43, 1 (02 2017), 25–38. DOI: https://doi.org/10.1177/0165551515613226

[32] Aytuğ Onan, Serdar Korukoğlu, and Hasan Bulut. 2016. Ensemble of keyword extraction methods and classifiers in text classification. *Expert Systems with Applications* 57 (2016), 232–247. DOI: https://doi.org/10.1016/j.eswa.2016.03.045

[33] Aytug Onan and Mansur Alp Toçoğlu. 2021. A term weighted neural language model and stacked bidirectional LSTM based framework for sarcasm identification. *IEEE Access* 9 (2021), 7701–7722. DOI: https://doi.org/10.1109/ACCESS.2021.3049734

[34] Jeyaraj Pandian, Seana Gall, Mahesh Kate, Gisele Silva, Rufus Akinyemi, Bruce Ovbiagele, Pablo Lavados, Dorcas Gandhi, and Amanda Thrift. 2018. Prevention of stroke: A global perspective. *The Lancet* 392 (10 2018), 1269–1278. DOI: https://doi.org/10.1016/S0140-6736(18)31269-8

[35] Teresa Podsiadły-Marczykowska, Bogdan Ciszek, and Artur Przelaskowski. 2014. Development of diagnostic stroke ontology - Preliminary results, 261–272. DOI: https://doi.org/10.1007/978-3-319-06596-0_24

[36] Jeena R. S. and Sukesh Kumar. 2016. Stroke prediction using SVM. In *Proceedings of the 2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT '16)*, 600–602. DOI: https://doi.org/10.1109/ICCICCT.2016.7988020

[37] Anusuya Ramasamy Arunkumar, Addisu Mulugeta, and Gergito Duba. 2020. Integrated datamining with knowledge management framework for stroke disease. *International Journal of Current Research* 12 (04 2020). DOI: https://doi.org/10.24941/ijcr.38446.04.2020

[38] Gangavarapu Sailasya and Gorli L. Aruna Kumari. 2021. Analyzing the performance of stroke prediction using ML classification algorithms. *International Journal of Advanced Computer Science and Applications* 12, 6 (2021). DOI: https://doi.org/10.14569/IJACSA.2021.0120662

[39] Raúl Soto-Cámara, Jerónimo González-Bernal, Josefa González-Santos, Jose Aguilar-Parra, Rubén Trigueros, and Remedios López-Liria. 2020. Knowledge on signs and risk factors in stroke patients. *Journal of Clinical Medicine* 9 (08 2020), 2557. DOI: https://doi.org/10.3390/jcm9082557

[40] Po-Yuan Su, Yi-Chia Wei, Hao Luo, Chi-Hung Liu, Wen-Yi Huang, Kuan-Fu Chen, Hung-Yu Wei, and Tsong-Hai Lee. 2022. Machine learning models for predicting influential factors of early outcomes in acute ischemic stroke: Registry-based study. *JMIR Medical Informatics* (03 2022).

[41] Tahia Tazin, Md Nur Alam, Nahian Nakiba Dola, Mohammad Sajibul Bari, Sami Bourouis, and Mohammad Monirujjaman Khan. 2021. Stroke disease detection and prediction using robust learning approaches. *Journal of Healthcare Engineering* 2021 (2021). DOI: https://doi.org/10.1155/2021/7633381

[42] Sotarat Thammaboosadee and Teerapat Kansadub. 2019. Data mining model and application for stroke prediction: A combination of demographic and medical screening data approach. *Interdisciplinary Research Review* 14, 4 (10 2019), 61–69. Retrieved from https://ph02.tci-thaijo.org/index.php/jtir/article/view/221565

[43] Selma Yahiya, Adil Yousif, and Mohammed Bashir. 2016. Classification of ischemic stroke using machine learning algorithms. *International Journal of Computer Applications* 149 (09 2016), 26–31. DOI: https://doi.org/10.5120/ijca2016911607

Received 17 August 2023; revised 16 September 2024; accepted 28 October 2024