AUTOMATED DETECTION AND CLASSIFICATION OF BRAIN STROKE LESIONS IN MRI USING MACHINE LEARNING TECHNIQUES

A. Samsudin[1], N. Mohd Saad[2,*], A. R. Abdullah[1], N. S. Mohd Noor[2], A. S. Muda[3] and S. Kandaya[1]

[1] Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia.
[2] Faculty of Electronics and Computer Technology and Engineering, Universiti Teknikal Malaysia Melaka, 76100 Durian Tunggal, Melaka, Malaysia.
[3] Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia.

*Corresponding Author's Email: norhashimah@utem.edu.my

Article History: Received September 12, 2024; Revised April 8, 2025; Accepted May 5, 2025

ABSTRACT: Magnetic Resonance Imaging (MRI) is vital for diagnosing brain disorders, especially strokes. Although prompt treatment within six hours improves outcomes, manual diagnosis by neuroradiologist remain subjective and time-consuming, necessitating automated solutions for efficiency and accuracy. This study aims to develop an automated approach for detecting and classifying stroke lesions in diffusion-weighted imaging (DWI) MRI scans using machine learning. This study employed advanced segmentation techniques and classification models, including linear discriminant analysis (LDA), support vector machine (SVM), weighted k-Nearest Neighbor (k-NN), and a bagged tree classifier, to differentiate ischemic and hemorrhagic strokes at various stages. The findings show that k-Means is optimal for stroke lesion segmentation, achieving a Dice index of 0.85, while SVM demonstrates the highest classification accuracy of 98.5% with an average training time of 1.8 seconds, suggesting an efficient automated solution for timely and accurate stroke diagnosis. Results demonstrate improved diagnostic accuracy, which has significant implications for clinical practice.

KEYWORDS: *Magnetic Resonance Imaging; Stroke Diagnosis; Machine Learning; Diffusion-weighted image; Stroke Lesion.*

1.0 INTRODUCTION

Stroke can be classified into two types: ischemic and hemorrhagic, and further categorized into three main stages: acute, sub-acute, and chronic. The differentiation of stroke types and stages can be difficult due to overlapping features, making the diagnosis particularly challenging and requiring the expertise of professional neuroradiologists [1]. Stroke poses a significant health concern and is a leading cause of death and disability worlwide. Globally, 1 in 4 adults over the age of 25 will have a stroke in their lifetime [2,3]. In Malaysia, it ranks among the top five leading causes of death, affecting approximately 40,000 people anually [4]. The National Stroke Association of Malaysia (NASAM) reports that one in six individuals worldwide will be diagnosed with stroke, making it the third leading cause of adult disability [5]. Stroke occurs when a blood vessel is blocked or ruptured due to a blood clot, leading to a sudden onset of neurological symptoms or signs caused by vascular issues [6]. Prompt treatment is essential to mitigate the harmful effects of stroke and facilitate better recovery. Diagnosing brain stroke requires the expertise of professional neuroradiologists due to its critical nature. Early detection and diagnosis play a pivotal role in successful therapy and treatment planning for stroke patients [7].

In contemporary medical practice, Diffusion-Weighted Imaging (DWI) has emerged as a vital tool for the early detection of stroke [8]. As an MRI sequence, DWI offers high lesion contrast, enabling the detection of hyperacute and acute stroke areas. It measures the diffusion of water molecules within tissue structures on a pixel level [9]. Although DWI is adept at visualizing stroke lesions, manual detection and diagnosis are time-consuming, prompting the need for computer-aided diagnosis techniques [9]. Despite advancements in imaging techniques, accurate segmentation and classification of stroke lesions remain challenging due to the heterogeneity of stroke pathogenesis and variations in imaging modalities. This study aims to develop a machine learning-based approach to automate the detection and classification of stroke lesions in MRI, thereby assisting clinicians in early and accurate diagnosis.

To address this issue, a dedicated computer system like Computer-Aided Diagnosis (CAD) is essential to provide neuroradiologists with a second opinion or clinical validation of medical images. By leveraging advanced technology, CAD can assist neuroradiologists in improving the accuracy of their diagnosis [10]. Automatic image segmentation and classification can be developed within the CAD system to enhance the precision, accuracy, and computational speed of segmentation methods. This integration of computer-assisted techniques promises to streamline stroke analysis and facilitate more efficient and accurate diagnosis.

Currently, imaging analysis is a vital and emerging trend for the early detection and accurate diagnosis of stroke disease. Magnetic Resonance Imaging (MRI) plays a significant role in delineating various aspects of acute stroke brain ischemia, including the presence, size, location, extent, and effects. It can also identify hypo perfused tissue at risk of infarction and reveal other cerebrovascular pathology [11]. Advanced MRI technology, particularly Diffusion-Weighted Imaging (DWI), has achieved high accuracy in identifying ischemic lesions with 99% sensitivity and 92% specificity [12]. However, the analysis of these complex MRI datasets has become a time-consuming and challenging task for neuroradiologists, who must manually extract crucial information from the images [13]. Manual segmentation suffers from limitations in terms of cost, time, reliability, and reproducibility [14]. Additionally, the current imaging technique alone may not provide sufficient information to identify salvageable tissue or monitor tissue reorganization after a stroke [15].

Multispectral neuroimaging, such as MRI, is commonly used for stroke diagnosis and management. However, the current practice involves processing these studies individually through visual inspection, leading to time-consuming and non-quantitative assessments, especially when urgent decisions are needed ("time is brain"). Radiologists are thus seeking improved tools for accurate diagnosis. Various techniques for stroke detection and segmentation are proposed, but accurate segmentation remains challenging due to the diverse shapes, locations, and intensity inhomogeneity of stroke lesions [16]. Manual segmentation by trained neuroradiologists is possible but impractical due to its tedious, time-consuming, and non-reproducible nature. Therefore, an efficient and accurate computer-aided diagnosis system is crucial.

Emerging methodologies continually introduce fresh approaches to stroke detection and segmentation. Nonetheless, due to the diverse array of potential shapes, locations, and variations in intensity, precise segmentation remains a formidable challenge [17]. While manual segmentation conducted by trained neuroradiologists is feasible, it is a laborious, time-intensive, and non-repeatable process [18]. Consequently, a computer-aided diagnosis system's success hinges on achieving both speedy processing and accurate segmentation.

Segmentation techniques can be broadly categorized into three groups: edge-based, region-based, and clustering-based. Region-based techniques are hampered by susceptibility to noise and intensity inhomogeneities, which can lead to holes or disconnection in extracted regions [19]. Clustering, a technique employed in segmentation, groups pixels based on their similarities. However, it falters when confronted with images corrupted by noise, outlines, or other artifacts [20]. Edge-based segmentation falls short in scenarios with fuzzy boundaries. Consequently, an algorithmic framework is necessary to accurately segment stroke lesions, while also mitigating noise, computational time, and complexity.

Conventional pattern recognition approaches necessitate intricate image processing tasks for noise reduction and feature enhancement before classification can yield high accuracy [21]. Recent studies have demonstrated that machine learning techniques, encompassing fuzzy logic and support vector machines (SVMs), hold promise for addressing the aforementioned challenges. This is due to the robustness of machine learning in handling noisy images, particularly in medical contexts, to solve intricate pattern recognition dilemmas [22].

2.0 RESEARCH METHODOLOGY

This study presents an automated machine learning-based approach for detecting, segmenting, and classifying stroke lesions in Diffusion-Weighted Imaging (DWI) MRI scans. The image analysis process begins with preprocessing steps that include normalization, background removal, and enhancement, aimed at preparing the images for more effective segmentation and classification. Segmentation is performed using the k-Means clustering technique combined with the Fast Marching Method (FMM), which enables accurate delineation of the Region of Interest (ROI) by removing cerebrospinal fluid (CSF) regions that share similar intensity with hypointense lesions. This segmentation process involves assigning pixel data to clusters based on Euclidean distance, updating cluster centroids iteratively, and validating performance against manual segmentations from neuroradiologists. Evaluation metrics include the Dice coefficient, Jaccard index, False Positive Rate (FPR), and False Negative Rate (FNR).

Following segmentation, spatial features are extracted from the ROI to characterize stroke lesion types. These features include the mean, median, and mode for general classification; standard deviation for hyperintense lesions; and the mean of the boundary for hypointense lesions. These extracted features are then used in the classification process. Four classification algorithms: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), weighted k-Nearest Neighbor (k-NN), and the bagged tree classifier are employed to categorize the stroke types. A 5-fold cross-validation strategy is used to divide the dataset into training and testing subsets, ensuring balanced representation for each stroke type. The classification results are evaluated based on accuracy, sensitivity, and specificity, with the SVM classifier achieving the best performance, showing 98.5% accuracy. These findings demonstrate that the proposed method is effective and reliable for supporting timely stroke diagnosis using DWI MRI images.

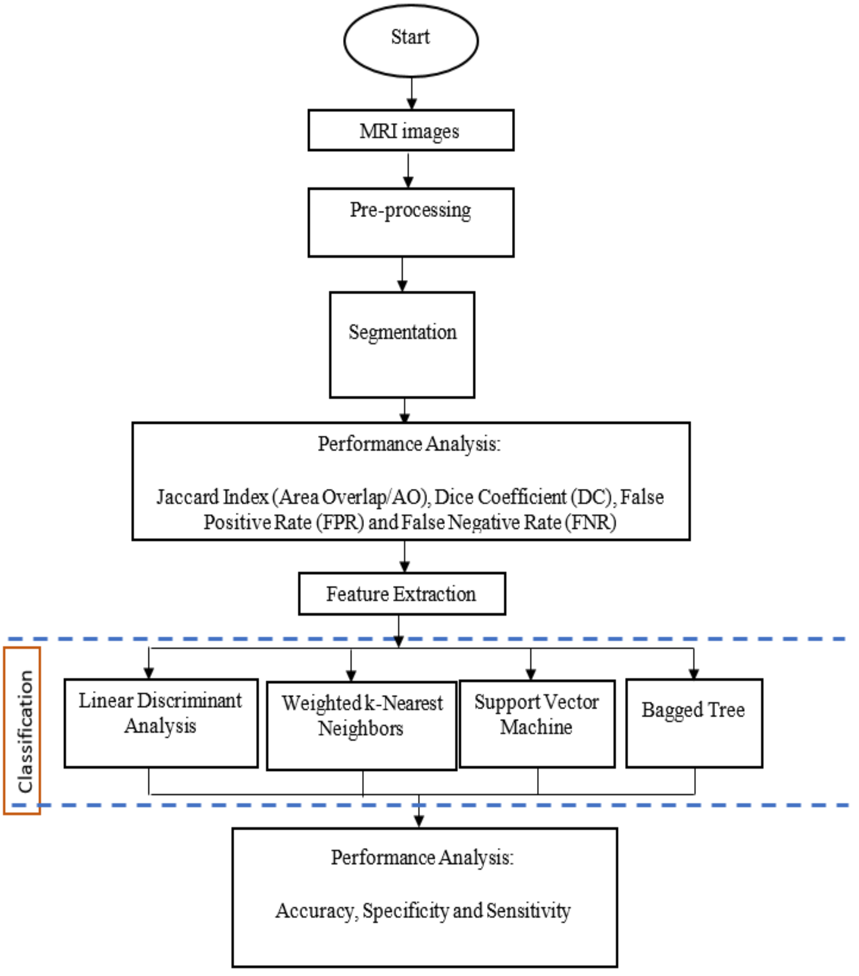


Figure 1: Flowchart of DWI image analysis using machine learning technique

2.1 Image segmentation using K-means with fast marching technique (FMM)

The flowchart in Figure 2 illustrates the utilization of the k-Means technique within this study. The process commences by designating the desired number of clusters, denoted as "k," derived from the input image of the image pre-processing phase. The cluster set is established to capture the spectrum of intensities present. Following the identification of this intensity range, the centroid value is derived, after which the disparity between the data point and the centroid is calculated. Subsequently, the clusters are organized based on the smallest difference value attained. In the context of hypointense lesions, the algorithm employs the FMM to segment the cerebrospinal fluid (CSF) that shares a comparable grey level intensity range with the hypointense image.

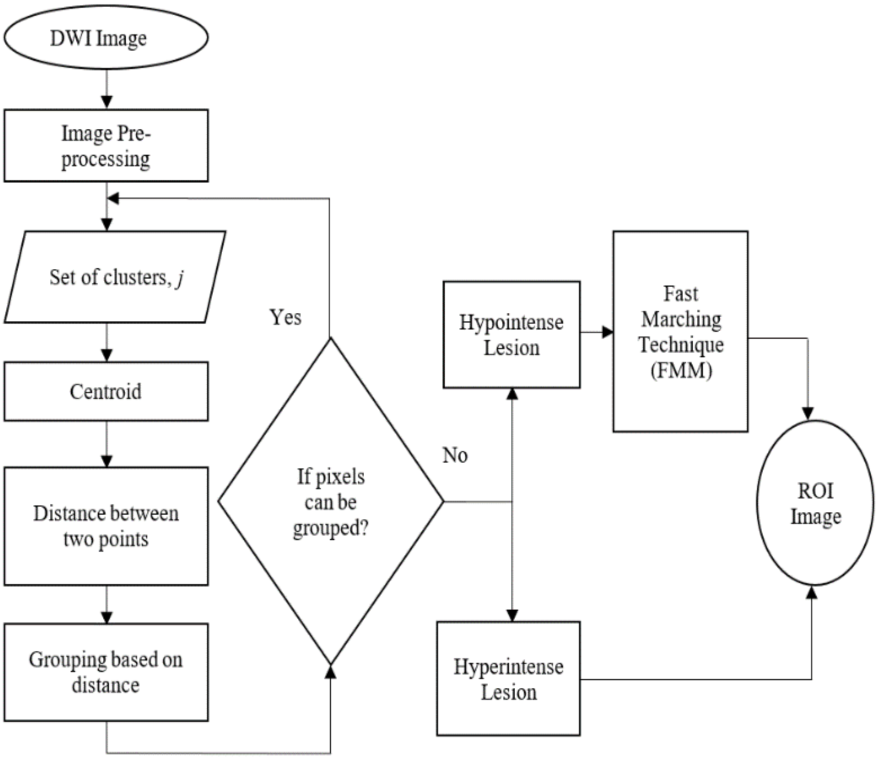


Figure 2: Flowchart of k-Means with FMM segmentation technique

The procedure of this method initiates by choosing the number of clusters, denoted as "j," from the set of data points, x. The selection process for cluster "j" involves assessing the mean distance between the data points, x, and their respective cluster centroids, denoted as s_j. This assessment generates a diverse range of metrics. As the value of "j" increases, the distance between the data point and the metric generated diminishes, until the set of clusters, j, reaches a zero value corresponding to the total number of data points, x. This outcome

demonstrates that the metric acquired can effectively segment the region of interest (ROI) within the stroke lesion. According to the results of this experiment, the optimal number of iterations required for the clusters and data points to achieve the best outcome is determined to be 4.
Once the appropriate "k" value is chosen, the algorithm iterates through the clusters, j, and data points, employing the data assignment and centroid update process. Each data point within a given dataset is assigned to the nearest centroid using the squared Euclidean distance formula. The distance between two points, "p" and "q," is defined as the square root of the sum of the squares of the differences between their corresponding coordinates. The Euclidean distance between points "p=(px, py)" and "q=(qx, qy)" is defined by Equation (1) [23].

$$d(p,q) = \sqrt{(qx - px)^2 + (qy - py)^2} \tag{1}$$

The Euclidean distance algorithm calculates the smallest distance between a column vector "x" and a set of column vectors within the codebook matrix. This algorithm determines the minimum distance to "x" and identifies the column vector within the codebook matrix that is nearest to "x," as defined by Equation (2) [24].

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{2}$$

The centroids are subsequently calculated by obtaining the average of all data points, "x," that are assigned to the cluster represented by centroid "s_j." Let the collection of data point assignments for each "i"-th cluster centroid be denoted as "s_j." The formula is presented in Equation (3).

$$\frac{1}{|s_j|} \sum x_i \in {s_j}^{x_i} \tag{3}$$

The algorithm repeatedly alternates between data assignment and centroid updates until a specified stopping criterion is met. At this juncture, the clusters derived from the data points remain unchanged, and the summation of distances is minimized. The algorithm begins with initial estimations for the "k" centroids, which can be either randomly generated or selected from the dataset in a random manner. In order to enhance the performance of the k-Means technique for chronic stroke lesions, the Fast Marching Method (FMM) is implemented to eliminate the cerebrospinal fluid (CSF) area in the brain image. The concept underlying FMM involves obtaining numerical approximations for solutions to the nonlinear eikonal equation, where this equation pertains to non-linear partial differential equations encountered in problems of wave propagation, as delineated in Equations (4) and (5) [25].

$$|\nabla u(x)| = f(x) \text{ in } \Omega, \;\; f(x) > 0 \tag{4}$$

$$u(x) = 0 \text{ for } x \; \in \partial\Omega \tag{5}$$

In this context, a function of time, denoted as "u," is associated with a speed function "f" in the normal direction at a specific point "x" situated on the advancing surface of the CSF region. This point "x" is designated as the seed point and comprises both a row and column vector derived from the CSF intensity value. The speed function is precisely determined, and the time at which the contour intersects point "x" is derived by solving the equation. Alternatively, "u(x)" can be conceptualized as the minimal time required to reach the boundary ∂Ω, commencing from point "x." The Fast Marching Method (FMM) capitalizes on this interpretation of optimal control to construct a solution outward, commencing from the known information of the ROI boundary values. Following the establishment of the seed point, the distinct weights of CSF pixels are computed based on the observed intensity differences between the brain and CSF regions.

2.2 Performance evaluation for segmentation technique

The segmentation method's effectiveness is assessed using metrics including the Jaccard index (area overlap, AO) as described in

Equation (6), the dice coefficient (DC) also in Equation (7), the false positive rate (FPR) as in Equation (8), and the false negative rate (FNR) following Equation (9). The evaluation of performance involves comparing the segmented ROI image with the results of manual visual inspection conducted by a neuroradiologist. The computation of these metrics is carried out according to the methods outlined by Saad et al. (2017) [9].

$$AO = \frac{A \cap G^C}{A \cup G} \tag{6}$$

$$DC = \frac{2|A \cap G^C|}{|A| \cup |G|} \tag{7}$$

$$FPR = \frac{A \cap G^C}{A \cup G} \tag{8}$$

$$FNR = \frac{A^C \cap G}{A \cup G} \tag{9}$$

where A signifies the segmentation outcomes acquired through the segmentation method, while G represents the manual reference segmentation. These metrics are visually depicted in Figure 3.

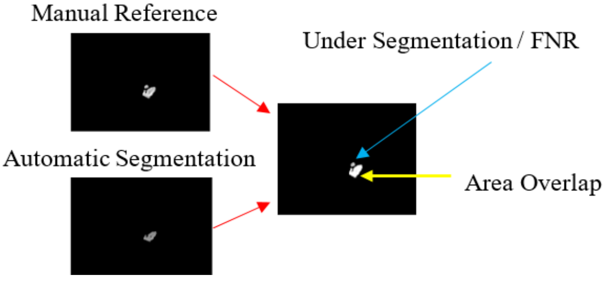


Figure 3: Segmentation assessment indices

Jaccard and Dice metrics calculate the similarity of the segmented areas by evaluating the overlap between the manually generated reference and the automatically produced segmentation. FPR and FNR serve to measure excessive and insufficient segmentation, respectively. Elevated Jaccard and Dice scores, along with diminished FPR and FNR values, indicate minimal errors or high accuracy in the measurement.

2.3 Feature extraction

In this phase, pertinent information regarding the Region of Interest (ROI) obtained from the segmentation process is extracted, with the aim of analyzing and illustrating the attributes of each brain stroke lesion. The purpose of this feature extraction lies in providing input to the classifier, enabling it to discern the stroke type based on the available information. The present study scrutinizes the region pixels within DWI images using a first-order statistical method. Given that this ROI relies on intensity signal, it's divided into two segments within the image: hyperintense and hypointense lesions. Mean, median, and mode are utilized for segregating the image into these categories. For hyperintense lesions, standard deviation is employed, while for hypointense lesions, the mean of the region boundary is used [26]. The standard deviation and mean of the boundary serve to distinguish the unique characteristics of each stroke lesion. Figure 4 outlines the feature extraction process for DWI stroke lesions.

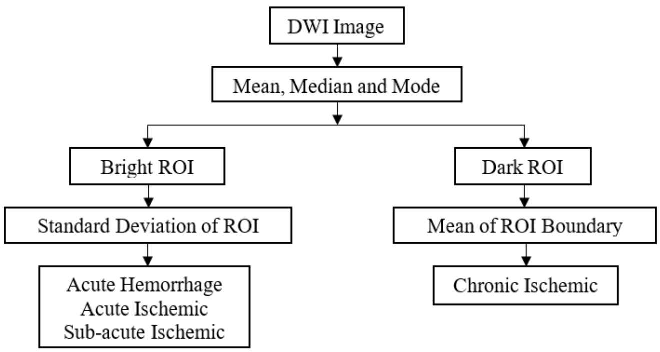


Figure 4: Flowchart of feature extraction technique

2.4 Classification technique

A classification approach is introduced to categorize stroke types based on the features extracted from the optimal segmentation outcome. This study introduces four distinct classification methods for analysis: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), weighted k-Nearest Neighbors (k-NN), and the bagged tree classifier. For the DWI image analysis utilizing machine learning techniques, the process involves the independent establishment of both a testing set and a training set to classify stroke types. This division is executed through the application of cross-validation technique. Through this technique, the algorithm for stroke lesion features is partitioned into subsets, with some utilized for training and the remainder reserved for testing. Within this study, a 5-fold cross-validation strategy is adopted. The features of the stroke lesions are randomly allocated into five distinct subsets, with each subset containing approximately equal samples of each stroke lesion type. Consequently, each set of stroke lesion features undergoes five rounds of analysis, employing distinct training and testing subsets on each occasion.

1) Linear discriminant analysis (LDA): LDA constitutes one of the supervised machine learning techniques initially introduced by Fisher in 1936. This method accomplishes stroke type classification by assigning predictive equations grounded in the attributes of the Region of Interest (ROI). The characteristics of the ROI, acting as discrete dependent variables, are represented through scatter plots. The objective of employing LDA is to identify a concise set of features that lead to an effective predictive model capable of categorizing stroke types distinctly. This can be illustrated using axes that optimize the separation between different stroke types [27]. The technique maps the feature space onto a reduced subspace, all while retaining the distinctive discriminatory information for each stroke type intact.
2) Weighted k-nearest neighbor (k-NN): The k-Nearest Neighbors (k-NN) stands as a nonparametric classification method within the learning set. It accomplishes stroke type classification by assigning new feature observations to the nearest stroke lesion type based on the utilized covariates [28].
3) Support vector machine (SVM): SVM is known as a classifier that is best in classifying two classes or types. Since this research focuses on four types of strokes, SVM assigns an error-correcting output codes (ECOC) model [29]. ECOC model requires a one-versus-one, which determines the type of stroke lesion, t_s that the binary learners train on, and a decoding scheme that uses loss, g, which determines how the results from the features of stroke, f_s of the binary classifiers are aggregated [30]. Coding design refers to a matrix element where each binary learner is assigned to each type of stroke lesion, t_s.

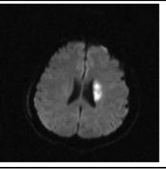
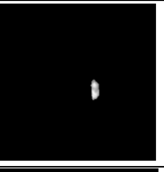
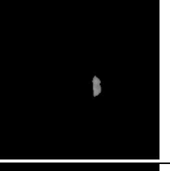
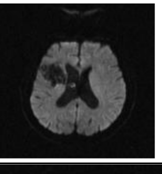
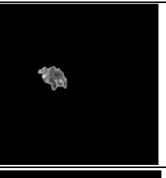
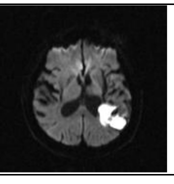
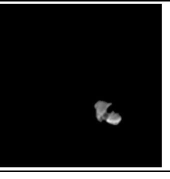
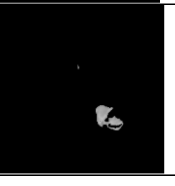
4) Bagged tree: Bagged tree is an ensemble learning technique that operate by constructing a multitude of decision trees at training time and outputting the type of stroke that is the mode of the type of stroke of the individual trees. The training algorithm for bagged tree applies the general technique of bagging learner. Given a training set $t_s = t_{s1}, \dots, t_{sn}$ which is the type of stroke with response $f_s = f_{s1}, \dots, f_{sn}$ the features of stroke lesion, bagging repeatly, B selects a random sample with replacement of the training set [31].

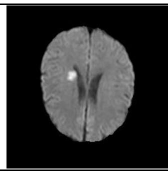
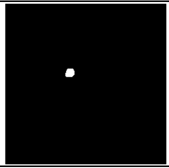
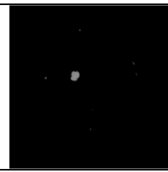
3.0 RESULTS AND DISCUSSION

3.1 Image segmentation using K-means with fast marching technique (FMM)

The k-Means method represents an unsupervised learning algorithm employed for image segmentation through clustering techniques. The incorporation of the Fast Marching Method (FMM) into the k-Means technique enables the removal of cerebrospinal fluid (CSF) and facilitates the segmentation of the Region of Interest (ROI) boundary within the DWI image. The outcomes of the stroke lesion segmentation from the original image are illustrated in Table 1.

Table 1: The segmentation results of k-Means with FMM

Type of Stroke	DWI Image	Manual Reference	Segmentation Result
Acute Ischemic Stroke			
Chronic Ischemic Stroke			
Acute Hemorrhage Stroke			

Sub-acute Ischemic Stroke			

As indicated in Table 1, the segmentation outcomes reveal that the k-Means combined with FMM effectively delineates accurate Regions of Interest (ROIs) for both hyperintense and hypointense lesions. The k-Means segmentation methodology accomplishes this by segmenting the ROI into distinct cluster values, thereby classifying different pixel groups within the DWI image. However, in the case of hypointense lesions, the ROI shares a cluster value with the cerebrospinal fluid (CSF) and other noise components. Consequently, to properly define the ROI and eliminate the CSF and undesired noise, FMM is integrated into the k-Means segmentation technique.

3.2 Classification technique

A classification methodology has been devised utilizing the Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), weighted k-Nearest Neighbors (k-NN), and the bagged tree classifier. This technique serves the purpose of categorizing stroke lesion types within DWI images. The input features employed for these classifiers consist of the Region of Interest (ROI) image, which is derived from the optimal segmentation technique proposed through k-Means with FMM segmentation. Scatter plots are employed to visualize the stroke patient samples, showcasing both correct and incorrect classifications. Consistently, across all scatter plot diagrams representing distinct features, the results depict samples that have been both correctly and incorrectly classified. To further assess classifier performance, scatter plot diagrams of mean boundary and standard deviation are presented. Figure 5 shows the model prediction of the ROI sample. The model prediction predicts the correct and incorrect sample of ROI. The scatter plot diagram in Figure 5 illustrates that two samples were wrongly categorized into a different stroke type. As per the scatter plot diagram, both of these misclassified samples, initially belonging to acute hemorrhage, were erroneously predicted to be of the acute ischemic stroke type.

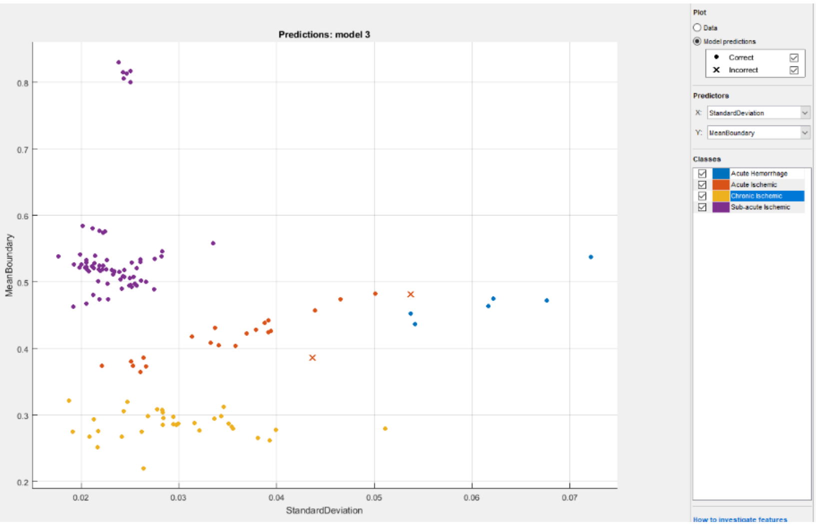
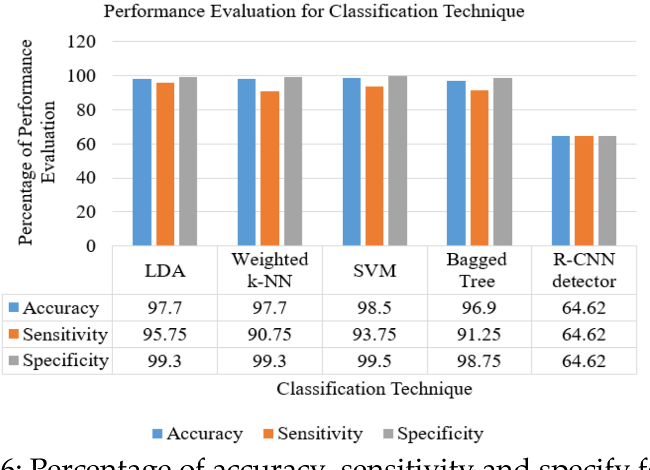


Figure 5: Model prediction of standard deviation and mean boundary for SVM classifier

3.3 Performance evaluation for classification technique

In this section, the performance of the accuracy, sensitivity and specificity of each machine learning classification technique and R-CNN detector is represented. Figure 6 illustrates the accuracy, sensitivity and specificity of the proposed classification techniques.



	LDA	Weighted k-NN	SVM	Bagged Tree	R-CNN detector
Accuracy	97.7	97.7	98.5	96.9	64.62
Sensitivity	95.75	90.75	93.75	91.25	64.62
Specificity	99.3	99.3	99.5	98.75	64.62

Figure 6: Percentage of accuracy, sensitivity and specify for the classification technique

The percentage accuracy range surpasses 60%, with the R-CNN detector having the lowest rate at 64.62%, and the SVM displaying the highest rate at 98.5%. Within the classification procedure, the SVM classifier demonstrates the lowest False Negative Rate (FNR) among the techniques, as depicted in Figure 6, effectively categorizing stroke types within its class. Concerning the percentage sensitivity range, values remain above 60%, with the R-CNN detector registering the lowest rate at 64.62%, and the LDA exhibiting the highest at 95.75%. The greatest sensitivity value indicates accurate identification of brain stroke images in DWI scans. As for the percentage specificity range, it remains above 60%, with the R-CNN detector again recording the lowest rate at 64.62%, while the SVM achieves the highest at 99.5%. Notably, the SVM stands out for its exceptional specificity in classifying DWI images accurately within their respective classes. In conclusion, the SVM classifier proves its competence in terms of sensitivity, specificity, and accuracy. A test can excel in specificity but lack sensitivity, or vice versa. However, optimal results stem from high values in both aspects of the classifier. The SVM emerges as a robust classifier, demonstrating elevated sensitivity and specificity concurrently. Hence, the SVM is selected as the most proficient technique for stroke classification.

3.4 Stroke lesion classification benchmarking

From the performance analysis and evaluation of classification techniques, SVM is chosen as the best classification technique that is suited for the DWI stroke lesion classification. Table II concluded the results for the best classification technique obtained for each lesion from this section. For the SVM classifier, the accuracy obtained is 98.5%.

Table 2: The performance verification for stroke lesion classification

Actual	Classification Performance		
	Sensitivity	***Specify***	***Accuracy***
Acute Hemorrhage	0.75	1	98.5%
Acute Ischemic	1	0.98	
Chromic Ishemic	1	1	
Sub-acute Ischemic	1	1	

In comparison with the previous studies, Table 3 summarizes the results of other researchers in similar studies. The SVM classification technique has shown a higher accuracy rate with a higher specificity and sensitivity rate compared to other studies. The accuracy obtained was 98.5% with a specificity rate of 100% for acute hemorrhage, chronic ischemic and sub-acute ischemic. For the sensitivity, the highest rate is 100% for acute ischemic, chronic ischemic and sub-acute ischemic. Gupta et al. (2014) presents accuracy rate with 98% by using Neural Network. Adam et al. (2016) presents the second highest sensitivity rate with 98% for ischemic stroke. Subudhi et al. (2018) presents specificity rate with 86.73% for ischemic stroke. Toldre et al. (2016) presents accuracy rate with 80% by using SVM. Then, Kiranmayee et al. (2017) presents 66.7% by using decision tree and 96.3% by using random forest. Batra and Kaushik et al. (2017) presents accuracy rate with 98.2%, sensitivity rate with 97.5% and the highest specificity rate with 100% by using SVM. Other than that, Ruba et al. (2020) presents accuracy meningioma rate with 99.57%, glioma rate with 99.78% and pituitary tumors rate with 99.56% by using CNN. At last, Díaz-Pernas et al. (2021) presents accuracy rate with 97.30% using CNN.

Table 3: The comparison result of the performance verification for the stroke lesion classification benchmarking

Researcher	Techniques	Stroke Lesion	Results
Saad et al. (2017, [7])	Rule Based Classifier	Acute Stroke Chronic Stroke	Sensitivity = 84.38% Specificity = 83.33 % Accuracy = 84%
Adam et al. (2016, [32])	Decision Tree	Ischemic Stroke	Sensitivity = 98%
Gupta et al. (2014, [33])	Neural Network	Ischemic Stroke	Accuracy = 98%
Subudhi et al. (2018, [34]	SVM	Ischemic Stroke	Sensitivity = 87.37% Specificity = 86.73 % Accuracy = 90.22%
Telrandhe (2016, [35])	SVM	Tumor	Accuracy = 80%
Kiranmayee (2017, [36])	Decision Tree Random Forest	Tumor	Decision tree = 66.7% Random forest = 96.3%
Batra and Kaushik (2017, [37])	SVM	Tumor	Accuracy = 98.2% Sensitivity = 97.5% Specificity = 100%
Ruba (2020, [38])	GoogLeNet CNN	Tumor	Accuracy Meningioma = 99.57% Glioma = 99.78% Pituitary Tumors.= 99.56%
Díaz-Pernas (2021, [39])	Deep CNN	Tumor	Accuracy = 97.30%
Proposed Technique	SVM	Acute Hemorrhage Stroke Acute Ischemic Stroke Chronic Ischemic Stroke Sub-acute Ischemic Stroke	Sensitivity = 75% Specificity = 100 % Sensitivity = 100% Specificity = 98 % Sensitivity = 100% Specificity = 100 % Sensitivity = 100% Specificity = 100 % Accuracy = 98.5 %

As conclusion, SVM proves a superior technique for the stroke classification using DWI images. Automatic segmentation from k-Means with FMM is also performed at higher performance on this analysis.

4.0 CONCLUSION

The findings of this study affirm that integrating k-Means with the Fast Marching Method (FMM) provides accurate and efficient segmentation of stroke lesions in DWI MRI images. This technique successfully delineates both hyperintense and hypointense regions of interest, even in the presence of cerebrospinal fluid (CSF) interference. The segmentation accuracy, measured through the Dice coefficient and Jaccard index, indicates strong agreement with manual annotations by neuroradiologists, suggesting the proposed method's suitability for clinical application.In terms of classification, the Support Vector Machine (SVM) classifier outperformed other methods including Linear Discriminant Analysis (LDA), weighted k-Nearest Neighbors (k-NN), and the bagged tree classifier. It achieved a classification accuracy of 98.5 percent, the highest sensitivity and specificity values, and the lowest false negative rate among all tested models. These findings demonstrate the SVM's ability to accurately distinguish between different types of stroke lesions, including acute ischemic, chronic ischemic, sub-acute ischemic, and acute hemorrhagic strokes. This highlights the classifier's robustness in handling variation in lesion characteristics across different stroke types. When benchmarked against previous studies, the proposed model yielded higher or comparable performance, particularly in terms of specificity for multiple stroke categories. This reinforces the model's reliability and shows promise for supporting clinical decision-making, especially in scenarios where time-sensitive and reproducible assessments are critical, such as emergency stroke diagnosis.

From a clinical perspective, the implementation of this automated machine learning pipeline could serve as an effective Computer-Aided Diagnosis (CAD) tool. It offers potential to reduce reliance on time-consuming manual segmentation and classification by neuroradiologists while ensuring high diagnostic accuracy. This is particularly valuable in enhancing diagnostic workflow in healthcare settings with limited access to stroke imaging specialists. For future research, the focus should be on expanding the model's validation across more diverse patient datasets to assess its generalizability. Additionally, investigating real-time integration into clinical systems and improving model robustness for low-quality or incomplete scans are important next steps. Such advancements could further support the goal of deploying a reliable, fast, and automated stroke diagnostic tool in routine clinical practice.

ACKNOWLEDGMENTS

The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM) for financial support in this research form grant PJP/2022/FTKEE/S01887.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest and fully endorse the contents of this manuscript.

REFERENCES

[1] V. Velayudhan and P. Pawha, "Stroke Imaging" [Online]. Available: https://emedicine.medscape.com/article/338385-overview .

[2] K. W. Loo, and S. H. Gan, "Burden of Stroke In Malaysia", International Journal of Stroke, vol. 7, no. 2, pp. 165-167, 2012.

[3] M. H. Jali, T. A. Izzuddin, Z. H. Bohari, H. Sarkawi, M. F. Sulaima, M. F. Baharom, and W. M. Bukhari, "Joint Torque Estimation Model of sEMG Signal for Arm Rehabilitation Device Using Artificial Neural Network Techniques," in Advanced Computer and Communication Engineering Technology, pp. 671-682, 2015.

[4] World Stroke Organization, "Impact of Stroke," World Stroke Organization, Feb. 28, 2025. [Online]. Available: . .

[5] Smajlović, "Strokes in Young Adults: Epidemiology and Prevention," Vascular Health and Risk Management, vol. 11, pp. 157, 2015.

[6] J. Peter and A. H. Justus, "Knowledge and Practices of Stroke Survivors Regarding Secondary Stroke Prevention, Khomas Region, Namibia," Journal of Medical Biomedical and Applied Sciences, vol. 1, pp. 1-13, 2016.

[7] N. M. Saad, N. S. M. Noor, A. R. Abdullah, S. Muda, A. F. Muda, and H. Musa, "Segmentation and Classification Analysis Techniques for Stroke based on Diffusion-Weighted Images," IAENG International Journal of Computer Science, vol. 44, no. 3, 2017.

[8] O. Maier, M. Wilms, J. von der Gablentz, U. Krämer, and H. Handels, "Ischemic Stroke Lesion Segmentation in Multi-Spectral MR Images with Support Vector Machine Classifiers," in Medical Imaging 2014: Computer-Aided Diagnosis, vol 9035, pp. 21-32, 2014.

[9] A. Shyna and J. Deepa, "Automatic Detection and Segmentation of Ischemic Stroke Lesion from Diffusion Weighted MRI using Contourlet

Transform Technique," International Journal of Engineering Research & Technology (IJERT), vol. 3, no. 4, pp. 2335-2341, 2014.

[10] K. Nagenthiraja, K. Mouridsen, and L. R. Ribe, "Method for Delineation of Tissue Lesions," United States Patent 9,036,878, 2015.

[11] A. R. Mathew and P. B. Anto, "Tumor Detection and Classification Of MRI Brain Image Using Wavelet Transform And SVM," in International Conference on Signal Processing and Communication (ICSPC'17), pp. 75-78. 2017.

[12] K. Nael and W. Kubal, "Magnetic Resonance Imaging of Acute Stroke," Magnetic Resonance Imaging Clinics of North America, vol. 24, no. 2, pp. 293-304, 2016.

[13] K. Malhotra and D.S. Liebeskind, "Overview of Neuroimaging of Stroke" in Primer on Cerebrovascular Diseases (Second Edition), pp. 676-685, Academic Press, United Kingdom, 2017.

[14] C. L. Hung and Y. H. Wu, "Parallel Genetic-Based Algorithm On Multiple Embedded Graphic Processing Units for Brain Magnetic Resonance Imaging Segmentation," Computers and Electrical Engineering, vol. 61, pp. 373-383, 2017.

[15] H. Y. Wey, V. R. Desai, and T. Q. Duong, "Review of Current Imaging Methods in Stroke Research" Neurological Research, vol. 35, no. 10, pp. 1092-1102, 2013.

[16] A. Bunevicius, H. Yuan, and W. Lin, "The Potential Roles of 18F-FDG-PET in Management of Acute Stroke Patients" BioMed Research International, vol. 2013, pp. 1-14, 2013.

[17] C. Feng, D. Zhao, and M. Huang, "Segmentation of Stroke Lesions in Multi-spectral MR Images Using Bias Correction Embedded FCM and Three Phase Level Set" in International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries 2015, pp. 1-6, 2015.

[18] X. Lladó, A. Oliver, M. Cabezas, J. Freixenet, J. C. Vilanova, A. Quiles, L. Valls, T. L. Ramió, and À. Rovira, "Segmentation of Multiple Sclerosis Lesions in Brain MRI: A Review of Automated Approaches" Information Sciences, vol. 186, pp. 164-185, 2012.

[19] I. Despotović, B. Goossens, and W. Philips, "MRI Segmentation of the Human Brain: Challenges, Methods, and Applications," Computational and Mathematical Methods in Medicine, vol. 2015, pp. 1–23, 2015, doi: https://doi.org/10.1155/2015/450341.

[20] D. B. Birnale and S. N. Patil, "A Hybrid Technique based on Fuzzy-C-Means and SVM for Detection of Brain Tumor in MRI Images," International Journal of Engineering Science and Computing, vol. 6, no. 8, pp. 2571-2574, 2016.

[21] S. Ahmad Radzi, M. Khalil-Hani, and R. Bakhteri, "Finger-Vein Biometric Identification Using Convolutional Neural Network," Turkish Journal of Electrical Engineering & Computer Sciences, vol. 24, pp. 1863-1878, 2016.

[22] K. I. Qazi, H. K. Lam, B. Xiao, G. Ouyang, and X. Yin, "Classification Of Epilepsy Using Computational Intelligence Techniques," CAAI Transactions on Intelligence Technology, vol. 1, pp. 137-149, 2016.

[23] I. Dokmanic, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: essential theory, algorithms, and applications," IEEE Signal Processing Magazine, vol. 32, no. 6, pp. 12-30, 2015.

[24] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas, "The Euclidean distance degree of an algebraic variety," Foundations of computational mathematics, vol. 16, no. 1, pp. 99-149, 2016.

[25] J. Yang and F. Stern, "A highly scalable massively parallel fast marching method for the Eikonal equation," Journal of Computational Physics, vol. 332, pp. 333-362, 2017.

[26] M. A. Azeez, S. N. Mazhir, and A. H. Ali, "Detection and Segmentation of Lung Cancer using Statistical Features of X-Ray Images," Int J Comput Sci Mob Comput, vol. 4, no. 2, pp. 307-313, 2015.

[27] T. K. Kim et al., "Incremental linear discriminant analysis using sufficient spanning sets and its applications," Int. J. Comput. Vision, vol. 91, no. 2, pp. 216-232, Jan. 2011, doi: 10.1007/s11263-010-0381-3.

[28] A. H. Ali et al., "Classification of Brain Lesion using K-Nearest Neighbor technique and Texture Analysis," J. Phys. Conf. Ser., vol. 1178, no. 1, 012018, 2019, doi: 10.1088/1742-6596/1178/1/012018.

[29] A. Sage and P. Badura, "Intracranial hemorrhage detection in head CT using double-branch convolutional neural network, support vector machine, and random forest," Appl. Sci., vol. 10, no. 21, pp. 1-13, Oct. 2020, doi: 10.3390/app10217577.

[30] S. A. Taie and W. Ghonaim, “A new model for early diagnosis of alzheimer’s disease based on BAT-SVM classifier,” Bulletin of Electrical Engineering and Informatics, vol. 10, no. 2, pp. 759–766, Apr. 2021, doi: https://doi.org/10.11591/eei.v10i2.2714.

[31] N. M. Saad et al., "Automated Classification of Stroke Lesion Using Bagged Tree Classifier," IOP Conf. Ser. Mater. Sci. Eng., vol. 884, no. 1, 012078, 2020, doi: 10.1088/1757-899X/884/1/012078.

[32] S. Y. Adam, A. Yousif, and M. B. Bashir, “Classification of Ischemic Stroke using Machine Learning Algorithms”, International Journal of Computer Applications, 149(10), pp. 26-31, 2016.

[33] S. Gupta, A. Mishra and Menaka R, "Ischemic Stroke detection using Image processing and ANN," 2014 IEEE International Conference on

Advanced Communications, Control and Computing Technologies, Ramanathapuram, India, 2014, pp. 1416-1420, doi: 10.1109/ICACCCT.2014.7019334.

[34] A. Subudhi, S. Sahoo, P. Biswal, and S. Sabut, “Segmentation and Classification of Ischemic Stroke using Optimized Features in Brain MRI,” Biomedical Engineering: Applications, Basis and Communications, vol. 30, no. 03, p. 1850011, May 2018, doi: https://doi.org/10.4015/s1016237218500114.

[35] S. R. Telrandhe, A. Pimpalkar, and A. Kendhe, “Implementation of Brain Tumor Detection using Segmentation Algorithm & SVM Image Classification View project Sentiment Analysis View project Swapnil R Telrandhe Implementation of Brain Tumor Detection using Segmentation Algorithm & SVM,” Int. J. Comput. Sci. Eng., vol. 8, no. 7, pp. 278–284, 2016, [Online]. Available: . Last Accessed: September 2024.

[36] B. Kiranmayee, “Effective Analysis of Brain Tumor Using Hybrid Data Mining Techniques,” Int. J. Adv. Res. Comput. Sci., vol. 8, no. 7, pp. 286–293, 2017.

[37] A. Batra and G. Kaushik, “SECTUBIM : Automatic Segmentation And Classification of Tumeric Brain MRI Images using FHS (FCM, HWT and SVM),” Int. J. Eng. Sci. Comput., vol. 7, no. 6, pp. 13190–13194, 2017.

[38] T. Ruba, R. Tamilselvi, M. Parisa Beham, and N. Aparna, “Accurate Classification and Detection of Brain Cancer Cells in MRI and CT Images using Nano Contrast Agents,” Biomed. Pharmacol. J., vol. 13, no. 3, pp. 1227–1237, 2020.

[39] F. J. Díaz-Pernas, M. Martínez-Zarzuela, D. González-Ortega, and M. Antón-Rodríguez, “A deep learning approach for brain tumor classification and segmentation using a multiscale convolutional neural network,” Healthc., vol. 9, no. 2, p. 153, 2021.