Automated Classification of Stroke Lesion Using Bagged Tree Classifier

N M Saad¹, A R Abdullah², A R Syafeeza³, N S M Noor⁴, N H Ali⁵ and A S Muda⁶
¹,³,⁴ Fakulti Kejuruteraan Elektronik dan Kejuruteraan Komputer, Universiti Teknikal Malaysia Melaka, Melaka, MALAYSIA.
² Fakulti Kejuruteraan Elektrik, Universiti Teknikal Malaysia Melaka, Melaka, MALAYSIA.
⁵ Faculty of Engineering and Technology, Multimedia University, Melaka, MALAYSIA.
⁶ Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, MALAYSIA.

E-mail: norhashimah@utem.edu.my

Abstract. Stroke is a “brain attack” that often causes paralysis, resulted from either bleeding in the brain (hemorrhagic) or the blockage of blood flow to the brain (ischemic). It posed a big challenge to Malaysian healthcare services with at least 32 deaths per day, while survivors were burdened with multiple problems. Conventionally, the diagnosis is performed manually by neuroradiologists during a highly subjective and time consuming task. Therefore, this paper intends to diagnose and classify stroke by investigating diffusion-weighted imaging (DWI) of brain stroke images using Bagged Tree classification. Stroke is classified into three main types which are acute stroke, chronic stroke and hemorrhage stroke. The performance of the proposed method is then verified using accuracy and Area Under the Curve (AUC). Based on the results, the overall accuracy for the classification is 96.7%. The AUC of each type of stroke for acute stroke, chronic stroke and hemorrhage stroke is 97%, 100% and 99%, respectively. This outcome could serve as an insight to improve the healthcare of the community by providing better solutions using such intelligent system.

1. Introduction
Stroke is increasingly recognized as a serious, worldwide public health concern. Recent developments in stroke have heightened the need for post-treatment imaging. Lack of post-treatment imaging has existed as a most feared effect to the patient that undergone acute infarction treatment, bleeding, to a great extent hemorrhage stroke [1]. Diffusion-weighted imaging (DWI) is one of the most widely imaging tools of medical and have been extensively used for stroke imaging [2]. Stroke imaging showed by this modality has become a strong independent predictor. However, the diagnosis evaluation consideration is high to diagnose the availability and reliability of the imaging technique, time and expertise required to perform and interpret the diagnosis [3].

Recent studies suggest that machine learning is needed to detect and classify medical images. Revolution of machine learning is a continuing concern within it can tackle complex tasks [4]. Ensemble classifier or combining classifier is fast becoming a new direction in improving the
individual classifier in machine learning technique [5]. Central to the entire discipline of stroke imaging is the concept of stroke detection and classification. A primary concern of the classification is to generate more certain and accurate results.

Vecchio, A. et al. (2017), found that Bagged Tree classifier obtain high accuracy to classify the posture of the user is inferred from the interdistances between the set of devices worn by the user compared to other machine learning technique of Neural Network, weighted k-NN (k- Nearest-Neighbour with weight equal to the inverse of the squared distance), and subspace k-NN [6]. In analysis of Bagged Tree classifier, Chaeibakhsh, S., et al. (2016) found that the Bagged Tree classifier is a reliable classification algorithm in the presence of unstable data but is less sensitive to this instability when compared to the Decision Tree approach, the effect of the instability remains apparent [7].

The major objective of this study is to develop a classification method for brain stroke lesion using Bagged Tree classifier. The initial parts which are image pre-processing and segmentation is already discussed in [8]. The segmentation of stroke lesion proposed in [8] is used for the next step for the features extraction and classification. The features extraction is essential to extract useful information to feed into the Bagged Tree classifier. Then, the analysis is verified by using statistical calculation to obtain the accuracy. In this analysis, we hypothesized that details information can be estimated, to identify and classify stroke by the proposed machine learning techniques.

The structure of this paper consists of four sections. The first section is about a general overview of the study including some studies and idea of the study. Section two presents the approach and proposed methodology of the study. It consists of analysis framework, features extraction, classification and performance verification. The overall results of the approach study will be shown in section three. Section four is the overall discussion and conclusion of the study.

2. Research Methodology

2.1. Analysis Framework

The analysis framework begins with 30 sample of DWI segmented images from the Fuzzy C-Means (FCM) segmentation algorithm discussed in [8]. The features from the region of interest (ROI) of the brain stroke lesion is extracted to obtain some information for classification. Stroke is classified into three main types which are acute stroke, chronic stroke and hemorrhage stroke. Finally, the accuracy of the classification method is verified by using performance indices such as accuracy and area under curve (AUC).

![Figure 1. Analysis Framework](image)

2.2. Features Extraction

Features are the reliable information that helps to characterize the type of lesion in stroke patient by gathering some relevant information [9]. In this study, the important features are extracted to analyze and represent the characteristic of acute, chronic and hemorrhage strokes. The purpose of the feature extraction in this study is as the input for the classifier to determine the type of stroke by the information given.

This study analyzes the region pixels of the DWI images by using first order statistical analysis which are the mode, standard deviation, mean, median, and mean of the lesion boundary. Since this image depends on the signal of intensity, the stroke lesion is classified into two parts of the image which are hyperintense and hypointense lesions. Mean, median and mode are used to separate the image between hyperintense and hypointense. For hyperintense image, standard deviation is used.
while in hypointense mean of region boundary is used. The standard deviation and mean of boundary are used to differentiate each characteristic of stroke lesion.

\[
\text{Mode} = \frac{\text{mod} e - \mu}{\sigma}
\]  
\[
\text{Standard Deviation} = \sqrt{\frac{s^2}{N}}
\]  
\[
\text{Mean} = \frac{\sum x}{N}
\]  
\[
\text{Median} = \frac{1}{2} (n+1)^{th} \text{ value}
\]  
\[
\text{Compactness} = \frac{\text{Perimeter}^2}{\text{Area}}
\]

2.3. Classification
In the classification stage, Bagged Tree classifier is proposed. A Bagged classifier is a technique where it ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then combine their individual forecast to form a final forecast. The data that are used to classify the type of stroke are extract from the brain stroke lesions. It is randomly selected and then is classified with the given calculation or cases. Below is the step of classifying the type of stroke by considering the type of stroke lesion and a binary response variable.

Step 1: A random data of each type of stroke lesion is selected.
Step 2: Constructed all of the random data by using Bagged Tree.
Step 3: Assign each type of stroke lesion to each terminal node and observe each terminal node that is attached with a given case.

Figure 2. Feature Extraction Flow Chart
Figure 3 shows the Bagged Tree classifier diagram with 5 indicated nodes. The first and third nodes are assigned as a branch node while the second, fourth and fifth nodes are assigned as a leaf node. The first branch node is also known as the root of tree. It classified two cases which are chronic stroke and another branch at node 3. If the variable is less than 0.361024 it is classified as a chronic stroke but if the value is greater than 0.361024 it is assigned to a second branch. The second branch at node 3 classified two parts of cases which is acute stroke and hemorrhage stroke. Acute stroke has a variable in between 0.361024 and 0.509774 while hemorrhage stroke has a variable above than 0.590774.

Table 1 shows the parameters of the Bagged Tree classifier. It shows that, node 1 is identified as the root of tree and the branch for node 2 and node 3 with its predictor value of \( x_1 < 0.361024 \) and \( x_1 \geq 0.361024 \), respectively. Node 2 is the leaf that assigned chronic stroke with variable range \( x_1 < 0.361024 \). Node 3 is the branch for node 4 and node 5 with its predictor value of \( x_1 < 0.590774 \). Node 4 is the leaf that assigned acute stroke, it is ranged \( 0.361024 \leq x_1 < 0.590774 \). Lastly, node 5 is assigned as hemorrhage stroke and its value is \( x_1 \geq 0.590774 \). Class membership for the classifier is shown in Table 2.

![Figure 3. Bagged Tree Classifier Diagram](image)

The step of the decision tree of the stroke classification can be summarized as:

- If \( x_1 < 0.367101 \) then node 2 else if \( x_1 \geq 0.367101 \) then node 3 else Chronic Stroke
- Class = Chronic Stroke
- If \( x_1 < 0.590774 \) then node 4 else if \( x_1 \geq 0.590774 \) then node 5 else Hemorrhage Stroke
- Class = Acute Stroke
- Class = Hemorrhage Stroke

| Node | Identity | Predictor | Variable Range        |
|------|----------|-----------|-----------------------|
| Node 1 | Branch   | \( x_1 < 0.361024 \) | Root of Tree          |
| Node 2 | Leaf     | Chronic Stroke | \( x_1 < 0.361024 \) |
| Node 3 | Branch   | \( x_1 < 0.590774 \) | \( 0.361024 \leq x_1 \) |
| Node 4 | Leaf     | Acute Stroke  | \( 0.361024 < x_1 < 0.590774 \) |
| Node 5 | Leaf     | Hemorrhage Stroke | \( 0.590774 \leq x_1 \) |
Table 2. Class Membership of Bagged Tree Classifier

| Node  | Acute Stroke | Chronic Stroke | Hemorrhage Stroke |
|-------|--------------|----------------|-------------------|
| Node 1 | 7            | 15             | 8                 |
| Node 2 | 0            | 15             | 0                 |
| Node 3 | 7            | 0              | 8                 |
| Node 4 | 7            | 0              | 0                 |
| Node 5 | 0            | 0              | 8                 |

2.4. Performance Verification

The performance verification takes part in the statistical calculation and is shown in the form of confusion matrix attributes using MATLAB APPS. Number of observation, true positive rate (TPR), false negative rate (FNR), false positive rate (FPR) is used to identify the performance of the classifier. Below is the statistical calculation for performance verification.

\[
TPR = \frac{TP}{TP + FN} \quad (6)
\]

\[
FNR = \frac{FN}{TP + FN} \quad (7)
\]

\[
FPR = \frac{FP}{FP + TN} \quad (8)
\]

where true positive (TP) is the number of samples correctly classified within their type of stroke in the positive sample, true negatives (TN) is patient without brain stroke lesion and correctly classified as a negative case, false positives (FP) is incorrectly classified negative cases of patient with brain stroke lesion, false negative (FN) is the number of samples incorrectly classified positive cases of patient without brain stroke lesions. Lastly, the performance of the classifier will be verified using accuracy as shown in equation 9.

\[
Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)
\]

Receiver operating characteristic (ROC) curve is used to show the TPR and FPR of each type of stroke. The receiver operating characteristic is a metric used to check the quality of classifiers by ensuring the stroke lesion is classified into its type of stroke [10]. For each type of stroke, ROC applies threshold values across the interval [0,1] to outputs. For each threshold, two values are calculated, the True Positive Ratio (the number of outputs greater or equal to the threshold, divided by the number of one targets), and the False Positive Ratio (the number of outputs less than the threshold, divided by the number of zero targets). The AUC number is a measure of the overall quality of the stroke. Larger AUC values indicate better classifier performance.

3. Results

3.1. Feature and Extraction

The brain stroke lesion is classified by using the first statistical method calculation in section 2.2. Figure 4 show the result of the feature extraction in scatter diagram for each type of stroke data.
plot indicates acute stroke, red plot indicates chronic stroke and yellow plot indicates hemorrhage stroke.

**Figure 4.** (a) Mean versus Median Scatter Plot Diagram (b) Mean boundary versus Compactness Scatter Diagram (c) Standard deviation mode scatter plot diagram

The scatter plot for each figure shows overall type of stroke is well classified in their group or within their range. Since acute and chronic strokes are present in bright lesion, some of the features are misclassified.

### 3.2. Classification

The performance of the confusion matrices of the number of observation of each type of brain stroke lesions is shown in Figure 5. Green box indicate each type of stroke is correctly classified to the assign classes while the red box indicates each type of stroke is misclassified to the assign classes. The red box indicates that a sample of hemorrhage stroke is misclassified to the assigned acute stroke class. The green color boxes indicate that the type of stroke is correctly classified.
Figure 5. The number of observation of each type of stroke lesion

| True Class          | Acute Stroke | Chronic Stroke | Hemorrhage Stroke |
|---------------------|--------------|----------------|-------------------|
| Acute Stroke        | 7            |                |                   |
| Chronic Stroke      |              | 15             |                   |
| Hemorrhage Stroke   | 1            |                | 7                 |

Predicted Class

The performance of the confusion matrices is based on the features from the feature extraction. One of the hemorrhage strokes is misclassified into acute stroke. The reason to this problem is maybe due to one of the hemorrhage stroke lesion that share the same features as acute stroke lesion.

3.3. Performances Matrics

Figure 6 shows the true and predicted class for the type of brain stroke lesions. The result for acute stroke and chronic stroke indicate 100% true positive rate as both of this stroke is correct classified. Meanwhile for hemorrhage stroke 13% is misclassified as acute stroke. The correct classification for hemorrhage stroke only reaches 88%.

| True Class          | Acute Stroke | Chronic Stroke | Hemorrhage Stroke | True Positive Rate | False Negative Rate |
|---------------------|--------------|----------------|-------------------|--------------------|---------------------|
| Acute Stroke        | 100%         |                |                   | 100%               |                     |
| Chronic Stroke      |              | 100%           |                   | 100%               |                     |
| Hemorrhage Stroke   | 13%          | 88%            | 88%               | 13%                |                     |

Figure 7 show the TPR and FNR of each type of stroke lesion

Figure 7 show the ROC plot for each type of stroke. Figure 7(a) shows the result of acute stroke is 4% incorrectly to the positive type of acute stroke. Acute (Figure 7(b)) and chronic (Figure 7(c)) stroke assign 100% on the true positive rate axis showing that both type of stroke is correctly classified within their type of stroke. The TPR of hemorrhage in Figure (Figure 7(c)) indicates that about 88% is correctly positive to hemorrhage stroke. Chronic and hemorrhage stroke both obtain 0% FPR. The results for acute, chronic and hemorrhage stroke for AUC is 97%, 100% and 99%, respectively. Overall accuracy of the classifier proposed is 96.7%.
Figure 7. (a) Acute stroke ROC curve (b) Chronic stroke ROC curve (c) Hemorrhage stroke ROC curve

4. Conclusion
In this study, feature extraction is applied by using first-order statistical calculation. By then, the information of the features is used as inputs to classify the brain stroke lesion in each type of stroke according to its class. From the results, only chronic stroke is correctly classified into its class. This is because the chronic stroke lesion was dark image compare to acute stroke and hemorrhage stroke that were categorized as the same bright image. The overall accuracy for this classification method is 96.7% with the AUC of acute stroke, chronic stroke and hemorrhage stroke for is 97%, 100% and 99%, respectively.

Acknowledgment
The authors would like to thank to the Universiti Teknikal Malaysia Melaka (UTeM), Rehabilitation Engineering & Assistive Technology (REAT) Research Group under Center for Robotics Industrial Automation (CeRIA), Advanced Digital Signal Processing (ADSP) Lab, Center for Telecommunication Research and Innovation (CeTRI), Faculty of Electronic and Computer
Engineering (FKEKK) and Ministry of Higher Education (MOHE), Malaysia for sponsoring this work under project FRGS/2018/FKEKK-CeRIA/F00363 and the use of the existing facilities to complete this project.

References
[1] Clinical Relevance. European Journal of Radiology.
[2] Thomalla, G., Boutitie, F., Fiebach, J.B., Simonsen, C.Z., Pedraza, S., Lemmens, R., Nighoghossian, N., Roy, P., Muir, K.W., Ebinger, M. and Ford, I., 2017. Clinical characteristics of unknown symptom onset stroke patients with and without diffusion-weighted imaging and fluid-attenuated inversion recovery mismatch. International Journal of Stroke.
[3] Akbik, F., Hirsch, J.A., Chandra, R.V., Frei, D., Patel, A.B., Rabinov, J.D., Rost, N., Schwamm, L.H. and Leslie-Mazwi, T.M., 2016. Telestroke—the promise and the challenge. Part one: growth and current practice. Journal of neurointerventional surgery, pp.neurintsurg-2016.
[4] Vitola, J., Vejar, M.A., Burgos, D.A.T. and Pozo, F., 2016. Data-Driven Methodologies for Structural Damage Detection Based on Machine Learning Applications. In Pattern Recognition-Analysis and Applications. InTech.
[5] Santucci, E., Didaci, L., Fumera, G. and Roli, F., 2017. A parameter randomization approach for constructing classifier ensembles. Pattern Recognition, 69, pp.1-13.
[6] Vecchio, A., Mulas, F. and Cola, G., 2017. Posture Recognition Using the Interdistances Between Wearable Devices. IEEE Sensors Letters, 1(4), pp.1-4.
[7] Chaeibakhsh, S., Phillips, E., Buchanan, A. and Wade, E., 2016, August. Upper extremity post-stroke motion quality estimation with decision trees and bagging forests. In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the (pp. 4585-4588). IEEE.
[8] Saad, N.M., Noor, N.S.M., Abdullah, A.R., Muda, S., Muda, A.F. and Rahman, N.A., 2017. Automated Stroke Lesion Detection and Diagnosis System. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1).
[9] Rathi, V.P. and Palani, S., 2012. Brain tumor MRI image classification with feature selection and extraction using linear discriminant analysis. arXiv preprint arXiv:1208.2128.
[10] Krupinski, E.A., 2017. Receiver Operating Characteristic (ROC) Analysis. Frontline Learning Research, 5(2), pp. 31 – 42.