

Post-stroke identification of EEG signals using recurrent neural networks and long short-term memory



Wanodya Sansiagi ^{a,1}, Esmeralda Contessa Djamel ^{a,2,*}, Daswara Djajasasmita ^{b,3}, Arlisa Wulandari ^{b,4}

^a Department of Informatics, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

^b Faculty of Medicine, Universitas Jenderal Achmad Yani, Cimahi, Indonesia

¹ wanodyasansiagi@gmail.com ; ² esmeralda.contessa@lecture.unjani.ac.id; ³ daswaradj@yahoo.com; ⁴ arlisawulandari@yahoo.com

* corresponding author

ARTICLE INFO

Article history

Received June 22, 2020

Revised April 26, 2021

Accepted May 7, 2021

Available online July 31, 2021

Keywords

Post-stroke

EEG signal

Wavelet

Recurrent neural networks (RNN)

Long Short-term memory

ABSTRACT

Stroke often causes disability, so patients need rehabilitation for recovery. Therefore, it is necessary to measure its effectiveness. An Electroencephalogram (EEG) can capture the improvement of activity in the brain in stroke rehabilitation. Therefore, the focus is on the identification of several post-rehabilitation conditions. This paper proposed identifying post-stroke EEG signals using Recurrent Neural Networks (RNN) to process sequential data. Memory control in the use of RNN adopted Long Short-Term Memory. Identification was provided out on two classes based on patient condition, particularly "No Stroke" and "Stroke". EEG signals are filtered using Wavelet to get the waves that characterize a stroke. The four waves and the average amplitude are features of the identification model. The experiment also varied the weight correction, i.e., Adaptive Moment Optimization (Adam) and Stochastic Gradient Descent (SGD). This research showed the highest accuracy using Wavelet without amplitude features of 94.80% for new data with Adam optimization model. Meanwhile, the feature configuration tested effect shows that the use of the amplitude feature slightly reduces the accuracy to 91.38%. The results also show that the effect of the optimization model, namely Adam has a higher accuracy of 94.8% compared to SGD, only 74.14%. The number of hidden layers showed that three hidden layers could slightly increase the accuracy from 93.10% to 94.8%. Therefore, wavelets as extraction are more significant than other configurations, which slightly differ in performance. Adam's model achieved convergence in earlier times, but the speed of each iteration is slower than the SGD model. Experiments also showed that the optimization model, number of epochs, configuration, and duration of the EEG signal provide the best accuracy settings.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



1. Introduction

A stroke occurs when the brain experiences disturbances such as blood supply. One in six people in the world will experience a stroke for life. Stroke can cause permanent damage, including partial paralysis and impaired speech, comprehension, and memory [1]. The situation of patients who have experienced a stroke is usually called post-stroke. In post-stroke patients, rehabilitation is needed to restore the function of body parts or minimize disability caused by stroke. Rehabilitation provided includes physical, cognitive, and mental, accompanying medication. However, monitoring and evaluation need to be done to determine the next treatment step.

One method for observing the general state of post-stroke patients is the National Institutes of Health Stroke Scale (NIHSS). NIHSS has 11 assessment indicators. There are vigilance, personal data, instructions to open and close eyes, moving the eyeball according to directions, focus on seeing an object, moving hands, moving legs, motor examination, testing body sensors such as needle insertion, testing ability language and imagination, the ability to read and repeat pronunciation, and the ability of the five senses [2]. NIHSS provides five categories of post-stroke patients based on the assessment of 11 indicators, particularly "No Stroke" for recovery, "Minor Stroke" for mild stroke conditions, "Moderate Stroke" for middle stroke conditions, "Moderate to Severe Stroke" concerning stroke conditions intermediate to heavy, and "Severe Stroke" for severe stroke conditions. Previous studies carried out electroencephalogram (EEG) signal analysis to calculate the NIHSS method [3].

EEG can be used in post-stroke patients to monitor their progress conditions. EEG signals can be obtained through the scalp by capturing the electrical potential of the brain. EEG signals have complexity in processing considering the low amplitude so that it is undoubtedly buried in noise and has an uncertain shape. However, the advantages of EEG are that it is cheap in operation and can be applied in real-time. The EEG signal that the Neurologist reads related to post-stroke patients is by observing the rhythm, changes in amplitude, and wave density on the EEG signal. In previous studies using EEG signals to detect emotions [4], detect epilepsy [5], hand rehabilitation in post-stroke patients with BCI display [6], and extraction significant variables for the recovery of post-stroke EEG signals [7].

EEG signal recording in the time-domain. Meanwhile, stroke analysis needs a rhythm pattern called Alpha, Beta, Theta, and Delta [8]. One method that is fitting for non-stationary signals like EEG signals to frequency extraction is Wavelet. This method has been used in the emotional classification of stroke patients [9], for classification in epilepsy patients [10], and diagnosis of depression through EEG signals [11].

Recurrent Neural Networks (RNN) is a machine learning method that can overcome memory optimization and overfitting limitations. RNN can connect sequential data. Besides RNN, the method used to process time-series data is fuzzy relations that can process linguistic values [12]. RNN is a learning model that continually preserves past information in sequential data, which is analogous to when the human brain makes decisions by remembering what has been learned [13]. Past studies used RNN for emotion recognition [13], motor imagery [14], and neuropsychological identification [15]. Other studies used Deep Neural Network to study motor imagery patterns in stroke patients [16].

This study identified post-stroke patients based on EEG signals using RNN and LSTM. Identification was provided out in two classes, namely "No Stroke" and "Stroke". Data set were recorded at Al-Islam Bandung Hospital from 25 patients with a post-stroke history with research ethics. As a comparison, 25 no-stroke people were recorded. Data used from previous studies and testing significant features of EEG signals [7]. Wavelet extraction is accomplished to separate each element, mainly Alpha, Mu, Theta, Beta, and Delta waves. The results of the extraction are processed using RNN and LSTM.

2. Method

This research is using the phase of Wavelet extraction and Recurrent Neural Networks for identification.

2.1. Wavelet Extraction

The wavelet method is generally used as a pre-processing method for signal processing [17], noise reduction [18][19], image processing [20], texture [21] which gives excellent results [22]. Wavelet consists of the decomposition that has a frequency component and reconstruction to the time domain. Therefore, EEG signals that are non-stationary can be analyzed using Wavelets. Decomposition has two processes that are convolution and down-sampling. Discrete Wavelet transformations of signal $x(n)$ are given by decomposition in (1).

$$C(\sigma, \tau) = \sum_n x(n)\psi_{\sigma,\tau}(n) \quad (1)$$

And reconstruction as (2).

$$\hat{x}(n) = \sum_{\sigma} \sum_{\tau} C(\sigma, \tau) \psi_{\sigma, \tau}(n) \quad (2)$$

In (3), consist of σ and τ are a scale and shift factor which is basic Wavelet function.

$$\psi_{\sigma, \tau}(n) = \frac{1}{\sqrt{|\sigma|}} \psi\left(\frac{n-\tau}{\sigma}\right) \quad (3)$$

Each decomposition step produces an approximation signal as half a low-frequency band and details like a high-frequency filter. The approximation or low coefficient and high detail coefficient are obtained using (4) and (5), where $f(n)$ and $g(n)$ are the low pass and high pass filters, respectively. Various forms of Wavelet functions, including Symlet2 [4], contain four coefficients. Signal decomposition several steps to produce Alpha, Beta, Mu, Delta, and Theta waves.

$$y_{low}[k] = \sum_n x[n].g[2k - n] \quad (4)$$

$$y_{high}[k] = \sum_n x[n].h[2k - n] \quad (5)$$

2.2. Recurrent Neural Networks and Long Short-Term Memory

Deep learning is a branch of machine learning inspired by a collection of neurons like brain neurons. Several methods used Convolutional Neural Networks (CNN) for identification images [23]. CNN can also be used for time-series data types with a one-dimensional form [24]. The advantages of this method are to provide faster computing and provide more freedom in previous extractions [25]. Meanwhile, other methods for signal processing, such as EEG, can use the Recurrent Neural Networks (RNN) method [26]. RNN helps the connection of sequential data.

One of the widely used RNN configurations is Long Short-Term Memory (LSTM) [27][28]. However, there are other sequential data information management methods such as Backpropagation Through Time (BPTT) [13], Gated Recurrent Units (GRU), Conventional Gated Recurrent Neural Networks (C-RNNs), Inception Convolutional Gated Recurrent Neural Networks (IC-RNNs), and Convolutional Densely Connected Gated Recurrent Neural Network (C-DRNN) [15]. Although RNN has the advantage of facilitating the connection of each signal, it is necessary to be careful when using pre-processing signal extraction methods. Also, it requires more computing time than using 1D CNN.

This study uses RNN with LSTM architecture to process long data sequences [28]. LSTM is used to overcome the amount of data processed by the gate mechanism. LSTM architecture can be seen in Fig. 1, starting with x_0 as input processed in hidden LSTM produces h_0 .

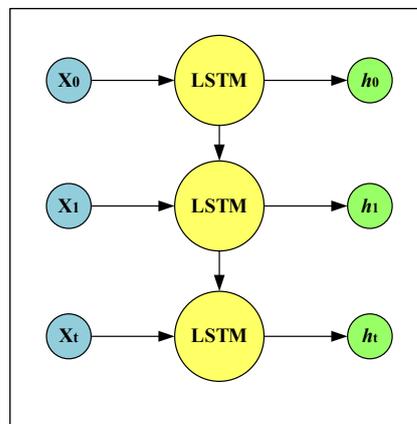


Fig. 1. LSTM architecture

Inside the hidden layer, LSTM has a standard unit called a memory block. This concept is another way to calculate the hidden state. A memory block called a cell can determine what to store in its memory storage temporarily [29]. LSTM memory input is taken from the initial state (h_{t-1}) and the current input (x_t). A set of cells whether to be stored or deleted in memory. LSTM has three gates: forget gate, input gate, and output gate by combining the previous state, current memory, and output shown in Fig. 2.

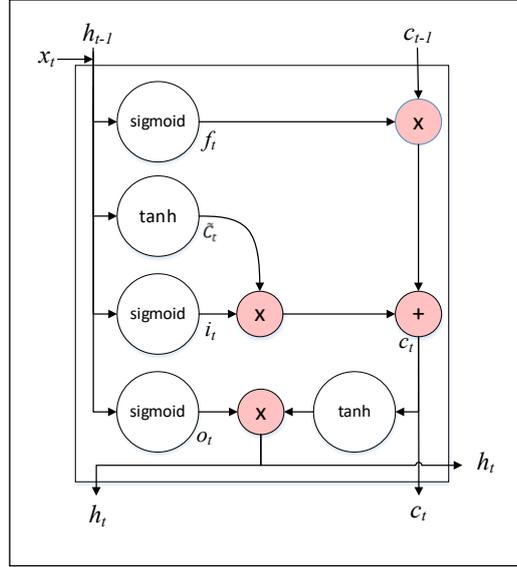


Fig. 2. LSTM cell of Recurrent Neural Networks architecture

The forget gate is the first gate (f) using the sigmoid layer to learn which data will be eliminated from the cell, as shown in (6). It used ReLU for the activation function with (7).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

$$ReLU(x) = \max(0, x) \quad (7)$$

The second is the input gate (i) so that the sigmoid layer (σ) will be updated, with tanh of and formulated as an updated vector. It can be seen at (8) and (9), where x_t is input for each current step.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (8)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (9)$$

Then cell from (6), (8), and (9) will be updated using (10) where C_t is the internal memory and C_{t-1} of the previous memory.

$$C_t = f_t * C_{t-1} + i_t * \tilde{c}_t \quad (10)$$

Finally, at the output gate (o), it will be calculated from the cell renewal and the sigmoid layer, which determines which cells will be received as the final result, such as (11) and (12).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (11)$$

$$h_t = o_t * \tanh(C_t) \quad (12)$$

2.3. Weight Correction

There are various ways to improve the weight of each training iteration. Some methods optimize to reach convergence and produce errors quickly or are marked with small cross-entropy values. This

method uses some ways, such as estimating the value of the subsequent iteration output, using not all output neurons, and gradually learning rate. This method is adapted to the Gradient Descent method: the Adaptive Moment Optimization (Adam) and Stochastic Gradient Descent (SGD) models.

- Stochastic Gradients Descent; Stochastic Gradient Descent (SGD) is a derivative learning method and the management of gradient values using a random sample in one iteration by taking one or more training data elements. In contrast to Gradient Descent which seeks local optimum using all training data because it can waste time. SGD correction uses (13). Variable θ is the weight function; η is the first learning rate, $x(i)$ and $y(i)$ parameter labels on the training data. In the comparison of traditional gradient descent, SGD uses minimum memory and reaches convergent values faster. However, updating the weight with a high variant of SGD is often carried out to make many fluctuations [30].

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^i; x^i) \tag{13}$$

- Adaptive Moment Optimization; Adam is a learning method that makes predictions by calculating the level of individual learning for each parameter which can adaptively minimize the possibility of error. Adam training converges faster than SGD [31] because it uses the estimation of the initial-moment gradient and v_t as an exponential slope as like (14). The second moment, s_t as the average of the quadratic exponents in (15) adapts the learning rate for every weight. The learning level is multiplied using the mean of the gradient. Then the result is calculated (16) and the weight update in (17).

$$v_t = \beta_1 * v_{t-1} - (1 - \beta_1) * g_t \tag{14}$$

$$s_t = \beta_2 * s_{t-1} - (1 - \beta_2) * g_t^2 \tag{15}$$

$$\Delta\omega_t = -\eta \frac{v_t}{\sqrt{s_t + \epsilon}} * g_t \tag{16}$$

$$\omega_{t+1} = \omega_t + \Delta\omega_t \tag{17}$$

2.4. Identification Model

At this stage, post-stroke patient EEG signal processing will be carried out for identification using the previously conducted Wavelet and RNN learning. As shown in Fig. 3, a computational model provides the output of stroke levels, namely "No Stroke" and "Stroke".

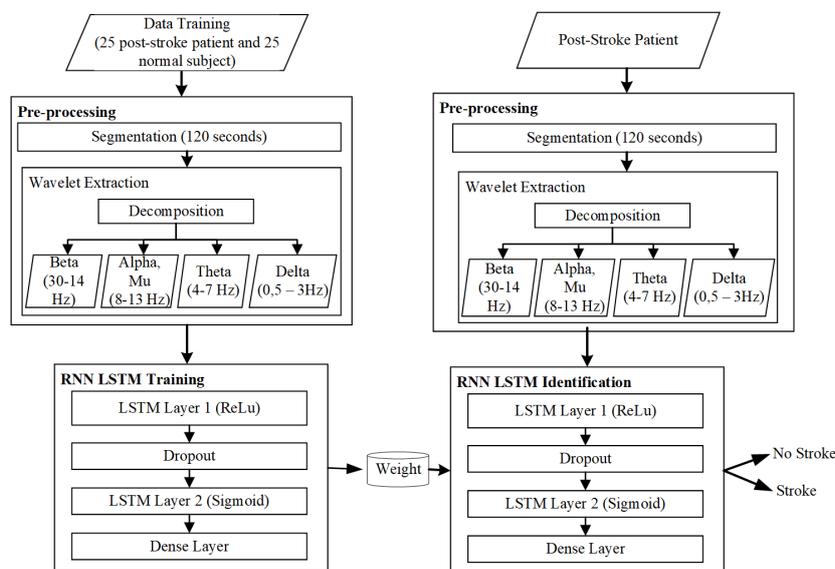


Fig. 3. Identification model of the post-stroke patient.

2.4.1. Data Acquisition

This study used EEG signals from 25 post-stroke patients at Al-Islam Hospital Bandung and 25 people without stroke as a comparison. Data on post-stroke patients were recorded after obtaining research ethics permit used in previous studies [7]. Classes used are 2, namely "No Stroke" and "Stroke". EEG signals are recorded from 14 channels consisting of AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8 with the Modified Combinatorial Nomenclature (MCN) system as shown in Fig. 4.

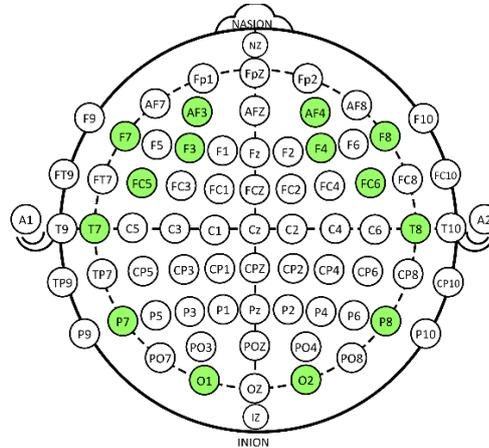


Fig. 4. Electrode configuration using the MCN system.

2.4.2. Wavelet Extraction

The EEG signal is decomposed in six steps using (4) and (5) with a sampling frequency of 128 Hz to obtain a wave, as shown in Fig. 5.

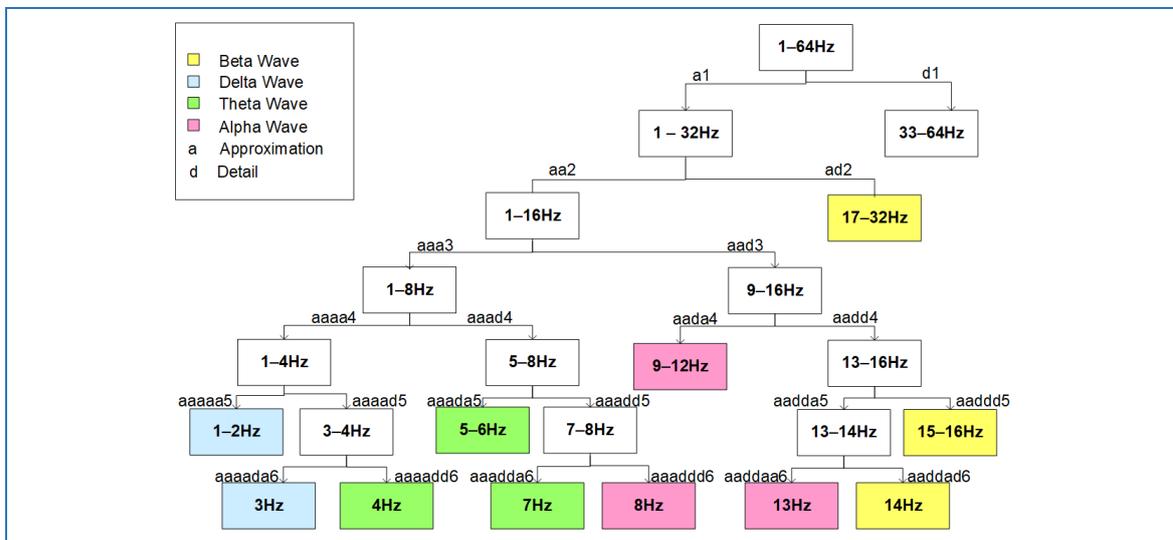


Fig. 5. Wavelet decomposition

Delta waves are obtained by six-step decomposition and reduce from 1,280 to 60 data, while Theta waves experienced a six-step decomposition process and reduced from 1,280 data to 80 data. In Alfa and Mu waves, the decomposition process is done in six steps and reduces 1,280 data to 120. However, Mu waves are just obtained from FC5 and FC6 channels. Beta waves experience six stages of decomposition and reduce from 1,280 data to 380 data.

Meanwhile, the amplitude change variable is taken every 10 seconds and segmented every 16 points to take the average value to produce 80 data from 1,280 data. The result of Wavelet extraction will be a feature vector for the RNN method, as shown in Table 1.

2.4.3. Recurrent Neural Networks

The identification used RNN with the LSTM architecture. The model uses six types of features: Alpha, Mu, Beta, Theta, Delta, and changes in amplitude. Input vectors can be seen in Table 1, which starts from each feature taken in all channels beginning from $x_1 - x_{10,080}$ for inputs that enter the RNN.

Table 1. Feature of Recurrent Neural Networks

Feature Vector	Point number each channel	Channel	Point number
Delta Wave	60	14	840
Theta Wave	80	14	1,120
Alpha Wave	120	12	1,440
Mu Wave	120	2	240
Beta Wave	380	14	5,320
Amplitude	80	14	1,120
Total			10,080

3. Results and Discussion

This experiment used EEG signal data as many as 50 datasets of 25 post-stroke patients and 25 non-stroke patients. The data was used 80% for learning with RNN and 20% for identification. The research began with Wavelet extraction to obtain frequency-based waves such as Alpha, Mu, Beta, Theta, and Delta waves and changes in amplitude. Then the extraction results are input the RNN model for identification.

3.1. Wavelet Extraction

The frequency recorded by the EEG signal is 0-64 Hz, while the extraction stage processes 0.5-30 Hz. The first stage of the experiment is to extract Wavelet using the Symlet2 coefficient, as shown in Fig. 6(a). The results of Wavelet extraction, as shown in Fig. 6(b), will be used as input vectors for input into the RNN method.

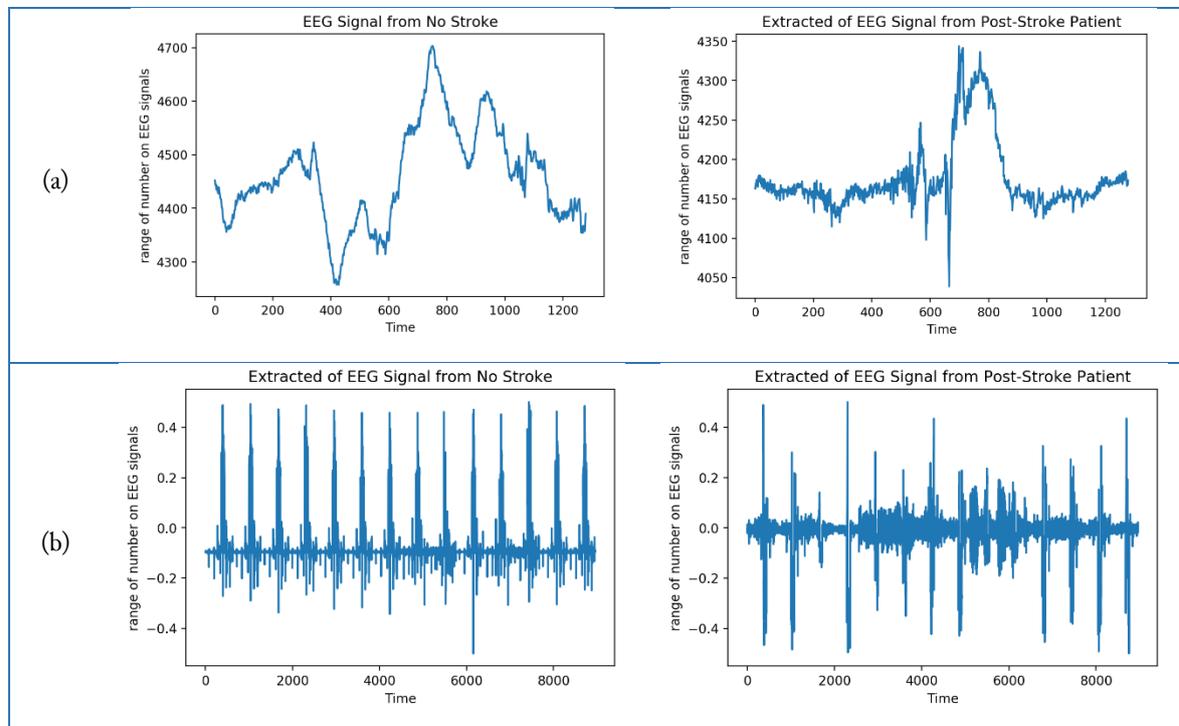


Fig. 6. Original EEG signal (a) and Extracted EEG signal (b)

3.2. Testing the Model Optimization, Wavelet Extraction, and Amplitude Feature

This experiment used two optimization models, remarkably Adam and SGD models. Both models improved weight, reduce error values, minimize output deviations, and increase speed during learning. Adam's model has fast convergence properties but is just not stable due to swift error reduction. Meanwhile, the SGD model samples data randomly and uses only a single sample for each accuracy. In testing using amplitude and wavelet extraction, the Adam optimization model obtained training data accuracy of 99.61% and new data of 91.38%. Meanwhile, the SGD optimization model got the training data accuracy of 67.57%, and the new data was 62.07%. It shows in Fig. 7, after 100 epochs. Compared to the Adam model, the small accuracy of SGD shows that the loss value is 0.6502 for the training data and 0.6788 for the new data. At the same time, the Adam model provides a smaller loss of 0.5090, as shown in Fig. 8. Therefore, this test shows if the SGD model has become saturated, so it does not help initialize weights that affect accuracy.

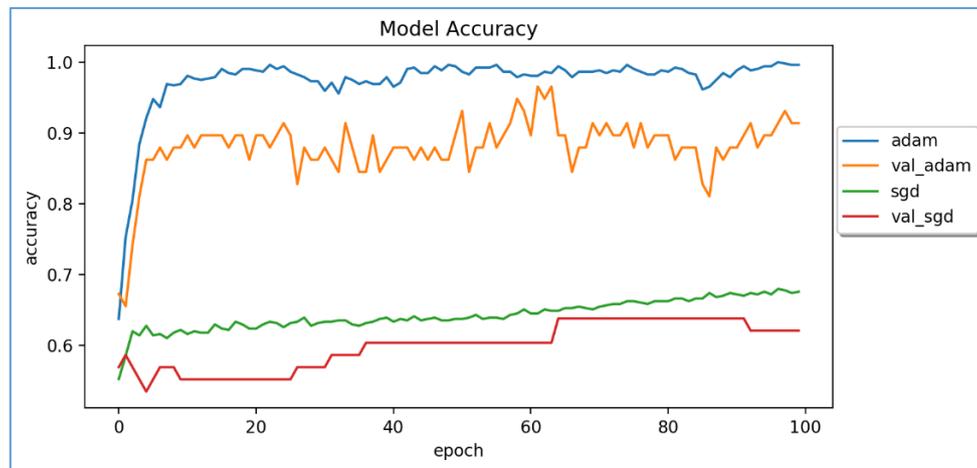


Fig. 7. Accuracy with amplitude feature

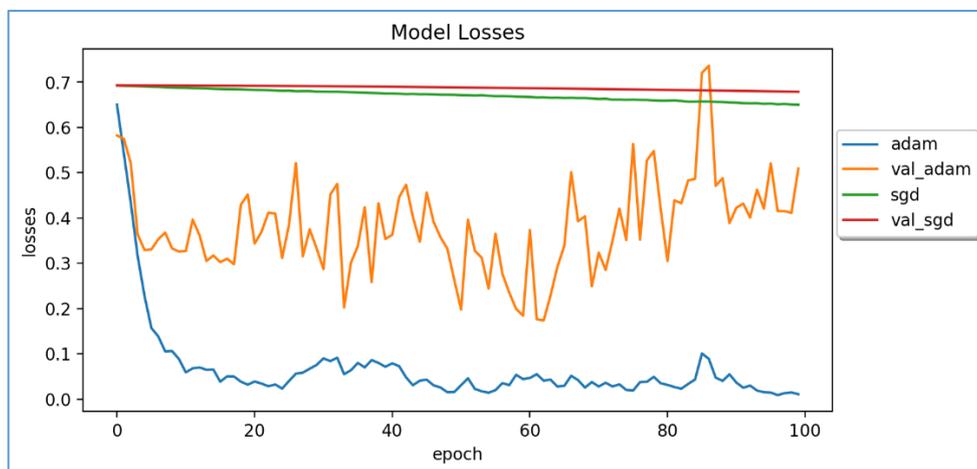


Fig. 8. Losses with amplitude feature

Meanwhile, the test without amplitude yields an accuracy of 93% for the new data with the Adam optimization model. Meanwhile, the SGD optimization model produces 70.08% for the training data and 58.62% for the new data. This result is shown in Fig. 9(a). Compared with using the amplitude feature, accuracy without using the amplitude feature is not better. However, the exactness is higher with the Adam model, no more than 3%, so there are still random factors. While the SGD model's correctness, using the amplitude feature is better than 3%, as shown in Fig. 9(a). While the loss value from using the amplitude feature is relatively the same, both were using the Adam or SGD model, as in Fig. 9(c).

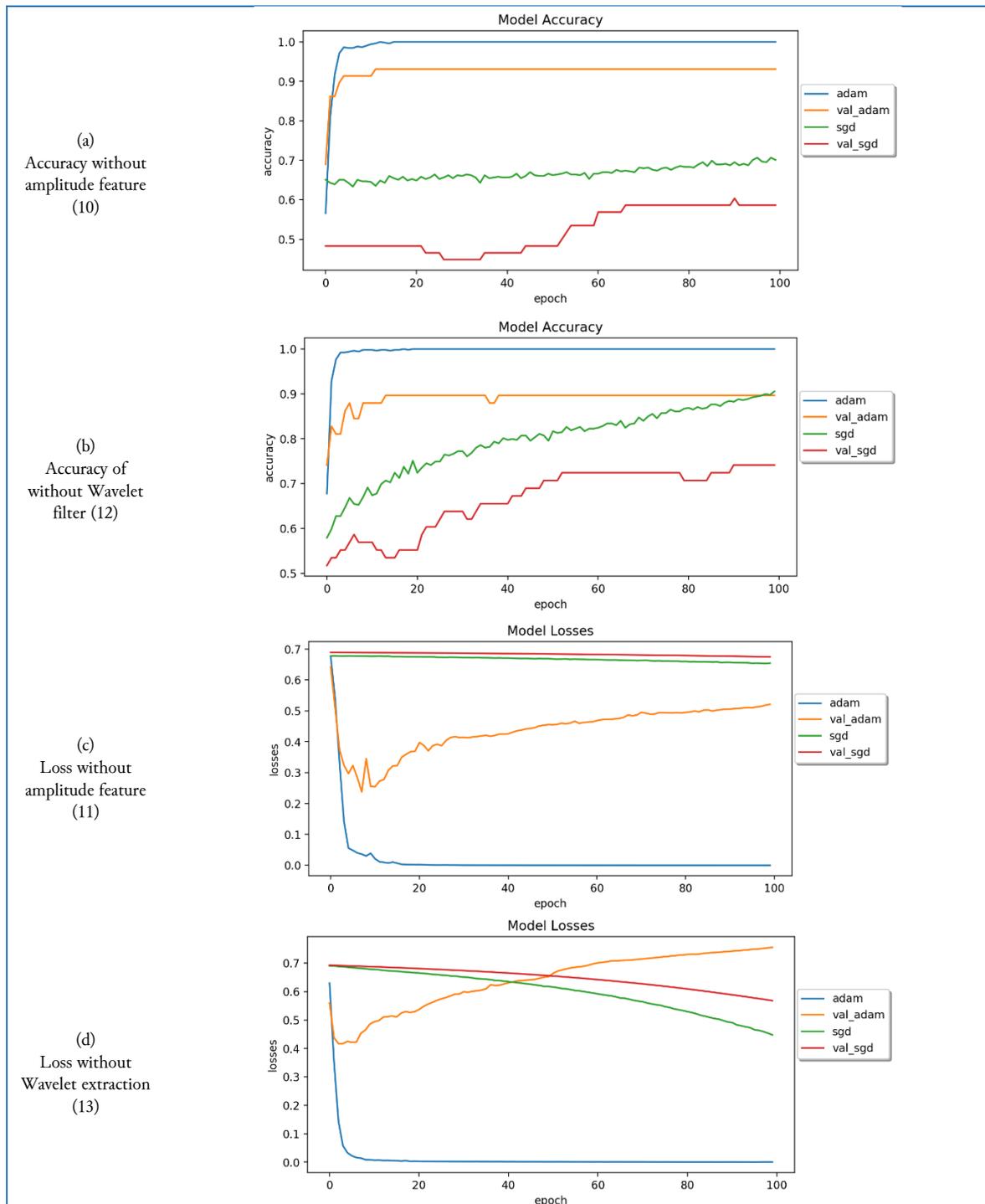


Fig. 9. Accuracy and loss without amplitude feature and wavelet extraction

This study tested the performance of Wavelet extraction. It shows that the accuracy of the Adam model using Wavelet increased from 89.66% to 93.10%. While the SGD optimization model produces accuracy, it grew from 62.07% to 74.11% for the new data, as shown in Fig. 9(b). From previous studies, the correctness of the system using Wavelet extraction can be improved. The opposite phenomenon in the use of SGD is likely, too large a range of random numbers in training, so it tends to be unstable. This result is shown by the value of losses from a system without Wavelet extractions extensive, as in Fig. 9(d).

The RNN model, which was tested with two models with 100 epoch iterations, had significant differences. The Adam optimization model used amplitude or does not have the accuracy of training

data ranges from 99% - 100%, and new data ranges from 89% to 93%. Meanwhile, the SGD optimization model gave accuracy ranges from 67% - 90% of training data and 58% - 74% of new data. The computing time between the two optimizations with 100 epochs and training was carried out using the GPU with RAM 6 GB. Adam can see that convergence in the epoch is faster, but the speed of each epoch is slower than SGD. A comparison of all test results can be seen in Table 2 with Adam and SGD optimization and amplitude features without amplitude features.

Table 2. Accuracy of various configurations using RNN and LSTM

Model	Method and Feature	Training Data		New Data		Computational time (second)
		Accuracy (%)	Loss	Accuracy (%)	Loss	
Adam	Wavelet with amplitude&wavelet	99.61	0.010600	91.38	0.5090	74.32
	Wavelet without amplitude, with wavelet	100.00	0.000009	93.10	0.5187	34.54
	Without Wavelet	100.00	0.000070	89.66	0.7551	126.00
SGD	Wavelet with amplitude&wavelet	67.57	0.650200	62.07	0.6788	49.75
	Wavelet without amplitude, with wavelet	70.08	0.654700	58.62	0.6749	24.00
	Without Wavelet	90.54	0.447500	74.14	0.5680	134.40

3.3. Testing the Number of Epochs

This study used the number of iterations of 500 epochs with the SGD optimization model. This experiment considers that in the 100 epoch, the SGD model provides a graph of increasing accuracy of new data in Table 3. Accuracy for training data addition of epochs does not experience increased accuracy. At the same time, the Adam model has converged.

Table 3. SGD model Wavelet without amplitude.

Epoch	Training Data		New Data	
	Accuracy (%)	Loss	Accuracy (%)	Loss
100	66.09	0.6814	62.07	0.6832
200	67.61	0.6602	63.79	0.6683
300	73.91	0.6134	70.69	0.6384
400	80.00	0.5075	74.14	0.5802
500	91.80	0.3316	81.03	0.4699

Fig. 10 shows the best accuracy of training data by 91.80%, and the accuracy of new data is 81.03%. In comparison, losses can be seen in Fig. 11, of 0.3316 for the training data and 0.4699 for the new data.

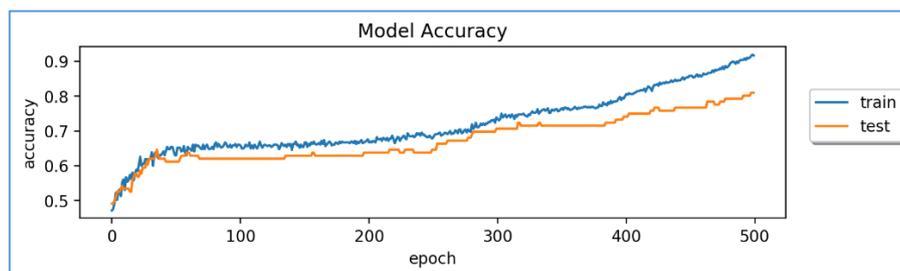


Fig. 10. Accuracy of 500 epoch using SGD model Wavelet without amplitude

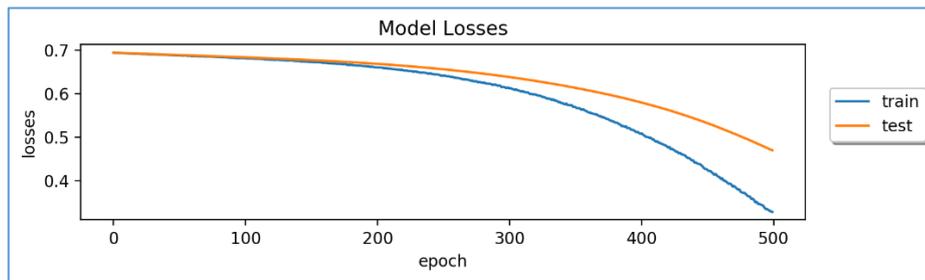


Fig. 11. Losses of 500 epoch using SGD model Wavelet without amplitude

3.4. Amount of Hidden Layers

This study was initially using two hidden layers from the LSTM method. From the previous test, the best accuracy was obtained using wavelet extraction and without amplitude. Then the test was conducted to determine the effect of the hidden layer in Table 4. The best accuracy result was 94.80%, with three hidden layers.

Table 4. Amount of Hidden Layers

Hidden Layers	Accuracy of New Data (%)
Two hidden layers	93.10
Three hidden layers	94.80
Four hidden layers	89.60
Five hidden layers	90.50

3.5. Discussion

Identification of post-stroke patients based on EEG signals was carried out using Wavelet and RNN. The features that provide the best accuracy are Alpha, Beta, Theta, and Delta waves at 93.1%. The RNN and LSTM methods have provided better accuracy than previous studies using Kohonen Self Organizing, only 74% [7].

The connection between sequential features is very appropriate for EEG signals, which are the time domain, thus providing better accuracy. However, the addition of the amplitude features of post-stroke patients characteristic of the EEG signal is less precise than the use of RNN. This result considers the feature sequence to have the same type, namely the wavelet extraction waveform. So the addition of features of different types can interfere with connectivity that decreases accuracy. Unlike previous studies, amplitude features and symmetric wave differences between channels accompanying wave features provide the best accuracy.

The study also found that the addition of hidden layers does not always increase accuracy, so that obtained three hidden layers give the best accuracy of 94.8%. Therefore, the addition of hidden layers only aggravates the work of RNN in learning but does not improve accuracy. Therefore, the addition of hidden layers only aggravates the work of RNN in learning but does not improve accuracy. This study can improve the classification of EEG signals with motor imagery features using Wavelet and LSTM for post-stroke patients, which provides an accuracy of 93.3%. However, this study uses nine hidden layers [14].

Judging from the optimization model for weight improvement, the Adam model gives better accuracy than the SGD model. However, SGD models are more susceptible to adding features of different types. So that EEG signals without Wavelet extraction and amplitude features provided 74.1% of accuracy. Besides, the SGD model requires more iterations to achieve convergence, but each iteration requires a shorter time than the Adam model. Therefore, in general, the SGD model learning computing time is shorter.

4. Conclusion

This study shows that EEG signals using Wavelet and RNN can be used to identify post-stroke patients, which gives accuracy up to 94.8%. The model was developed on EEG signals from 50 people consisting of 25 people post-stroke patients and 25 non-stroke patients. EEG signal data is segmented every 10 seconds from 120 seconds of recording. The sequential data of the EEG signal is matched using RNN and LSTM to provide excellent accuracy. However, this method is slightly vulnerable to variations in feature types, so adding amplitude features does not offer better correctness. Although, this feature has a very significant effect on the SGD model compared to the Adam model. The configuration of RNN architecture in the identification of post-stroke patients using EEG signals provides that it is necessary to pay attention. It is such the number of hidden layers, and especially the weight optimization model. The results show that the Adam model provides better accuracy and is more robotic by adding different features than the SGD model. However, SGD's computation time is faster than Adam's, although it requires a longer iteration.

Acknowledgment

Thanks to Nerve Clinic and Rehabilitation, Al Islam Hospital Bandung, for recording EEG signals from post-stroke patients.

Declarations

Author contribution. The first and second authors create a computational model, implementing the ideas and approaches presented by the third and fourth authors. The third author supervises the data used. The third and fourth authors verified the simulation results of each part of the model from a medical point of view. All authors discussed the results and contributed to the final manuscript.

Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

- [1] World Stroke Organization, "Facts and Figures about Stroke," *Kenes Media*, 2015. Available at: <https://www.world-stroke.org/>
- [2] S. E. Kasner, "Clinical Interpretation and Use of Stroke Scales," *The Lancet Neurology*, vol. 5, no. 7, pp. 603–612, 2006. doi: [10.1016/S1474-4422\(06\)70495-1](https://doi.org/10.1016/S1474-4422(06)70495-1)
- [3] W. R. R. Omar, R. Jailani, M. N. Taib, N. A. A. Razak, N. H. A. Wahab, and W. N. Nafisah, "An analysis of male stroke patients' brain signal according to NIHSS score," *IEEE Conference on Systems, Process and Control (ICSPC 2014)*, pp. 183–187, 2014. doi: [10.1109/SPC.2014.7086254](https://doi.org/10.1109/SPC.2014.7086254)
- [4] E. C. Djamal and P. Lodaya, "EEG Based Emotion Monitoring Using Wavelet and Learning Vector Quantization," *2017 4Th International Conference on Electrical Engineering, Computer Science and Informatics (Eecsi)*, pp. 94–99, 2017. doi: [10.1109/EECSI.2017.8239090](https://doi.org/10.1109/EECSI.2017.8239090)
- [5] O. Faust, U. R. Acharya, H. Adeli, and A. Adeli, "Wavelet-based EEG Processing for Computer-aided Seizure Detection and Epilepsy Diagnosis," *Seizure: European Journal of Epilepsy*, vol. 26, 2015, [Online]. doi: [10.1016/j.seizure.2015.01.012](https://doi.org/10.1016/j.seizure.2015.01.012).
- [6] F. Cincotti *et al.*, "EEG-Based Brain-Computer Interface to Support Post-Stroke Motor Rehabilitation of The Upper Limb," *34th Annual International Conference of the IEEE EMBS*, pp. 4112–4115, 2012. doi: [10.1109/EMBC.2012.6346871](https://doi.org/10.1109/EMBC.2012.6346871)
- [7] E. C. Djamal, D. P. Gustiawan, and D. Djajasmita, "Significant Variables Extraction of Post-Stroke EEG Signal using Wavelet and SOM Kohonen," *TELKOMNIKA*, vol. 17, no. 3, pp. 1149–1158, 2019, doi: [10.12928/TELKOMNIKA.v17i3.11776](https://doi.org/10.12928/TELKOMNIKA.v17i3.11776).

- [8] W. R. W. Omar *et al.*, "Acute Ischemic Stroke Brainwave Classification Using Relative Power Ratio Cluster Analysis," *The 9th International Conference on Cognitive Science*, pp. 546–552, 2013. doi: [10.1016/j.sbspro.2013.10.271](https://doi.org/10.1016/j.sbspro.2013.10.271)
- [9] S. Z. Bong, K. Wan, M. Murugappan, N. M. Ibrahim, Y. Rajamanickam, and K. Mohamad, "Implementation of Wavelet Packet Transform and Non Linear Analysis for Emotion Classification in Stroke Patient using Brain Signals," *Biomedical Signal Processing and Control*, vol. 36, pp. 102–112, 2017. doi: [10.1016/j.bspc.2017.03.016](https://doi.org/10.1016/j.bspc.2017.03.016).
- [10] M. Kaleem, A. Guergachi, and S. Krishnan, "Biomedical Signal Processing and Control Patient-Specific Seizure Detection in Long-Term EEG using Wavelet Decomposition," *Biomedical Signal Processing and Control*, vol. 46, pp. 157–165, 2018. doi: [10.1016/j.bspc.2018.07.006](https://doi.org/10.1016/j.bspc.2018.07.006)
- [11] M. Sharma, P. V. Achuth, D. Deb, S. D. Puthankattil, and U. Rajendra Acharya, "An Automated Diagnosis of Depression Using Three-Channel Bandwidth-Duration Localized Wavelet Filter Bank with EEG Signals," *Cognitive Systems Research*, 2018. doi: [10.1016/j.cogsys.2018.07.010](https://doi.org/10.1016/j.cogsys.2018.07.010).
- [12] Y. L. Huang *et al.*, "An Improved Forecasting Model Based on The Weighted Fuzzy Relationship Matrix Combined With a PSO Adaptation for Enrollments," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 7 A, pp. 4027–4046, 2011. Available at: [Google Scholar](https://scholar.google.com/).
- [13] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion Recognition based on EEG using LSTM Recurrent Neural Network," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8, no. 10, pp. 355–358, 2017. doi: [10.14569/IJACSA.2017.081046](https://doi.org/10.14569/IJACSA.2017.081046)
- [14] M. Li, M. Zhang, X. Luo, and J. Yang, "Combined Long Short-Term Memory Based Network Employing Wavelet Coefficients for MI-EEG Recognition," *2016 IEEE International Conference on Mechatronics and Automation, IEEE ICMA 2016*, pp. 1971–1976, 2016. doi: [10.1109/ICMA.2016.7558868](https://doi.org/10.1109/ICMA.2016.7558868)
- [15] S. Roy, I. Kiral-kornek, and S. Harrer, "ChronoNet : A Deep Recurrent Neural Network for Abnormal EEG Identification," *International Joint Conferences on Artificial Intelligence (IJCAI)*, pp. 1–10, 2018. doi: [10.1007/978-3-030-21642-9_8](https://doi.org/10.1007/978-3-030-21642-9_8)
- [16] D. Cheng, Y. Liu, and L. Zhang, "Exploring Motor Imagery Eeg Patterns for Stroke Patients with Deep Neural Networks," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2561–2565, 2018. doi: [10.1109/ICASSP.2018.8461525](https://doi.org/10.1109/ICASSP.2018.8461525)
- [17] M. M. Shaker, "EEG Waves Classifier using Wavelet Transform and Fourier Transform," *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, vol. 1, no. 3, pp. 169–174, 2007. Available at: [Google Scholar](https://scholar.google.com/).
- [18] F. M. Bayer, A. J. Kozakevicius, and R. J. Cintra, "An Iterative Wavelet Threshold for Signal Denoising," *Signal Processing*, vol. 162, pp. 10–20, 2019, doi: [10.1016/j.sigpro.2019.04.005](https://doi.org/10.1016/j.sigpro.2019.04.005).
- [19] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, and M. A. Awadallah, "Hybridizing β -hill Climbing with Wavelet Transform for Denoising ECG Signals," *Information Sciences*, vol. 429, pp. 229–246, 2018, doi: [10.1016/j.ins.2017.11.026](https://doi.org/10.1016/j.ins.2017.11.026).
- [20] N. Gurudath and H. Bryan Riley, "Drowsy driving detection by EEG analysis using Wavelet Transform and K-means clustering," *Procedia Computer Science*, vol. 34, pp. 400–409, 2014, doi: [10.1016/j.procs.2014.07.045](https://doi.org/10.1016/j.procs.2014.07.045).
- [21] C. Li, Y. Huang, X. Yang, and H. Chen, "Marginal Distribution Covariance Model in The Multiple Wavelet Domain for Texture Representation," *Pattern Recognition*, vol. 92, pp. 246–257, 2019, doi: [10.1016/j.patcog.2019.04.003](https://doi.org/10.1016/j.patcog.2019.04.003).
- [22] P. V. Kasambe and S. S. Rathod, "VLSI Wavelet Based Denoising of PPG Signal," *Procedia Computer Science*, vol. 49, no. 1, pp. 282–288, 2015, doi: [10.1016/j.procs.2015.04.254](https://doi.org/10.1016/j.procs.2015.04.254).
- [23] H. Fadhilah, E. C. Djamal, and R. Ilyas, "Non-Halal Ingredients Detection of Food Packaging Image Using Convolutional Neural Networks," 2018. doi: [10.1109/SAIN.2018.8673376](https://doi.org/10.1109/SAIN.2018.8673376)
- [24] E. P. Giri, M. I. Fanany, A. M. Aryrnurthy, and S. K. Wijaya, "Ischemic Stroke Identification Based on EEG and EOG using 1D Convolutional Neural Network and Batch Normalization," in *2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, 2016, pp. 484–491. doi: [10.1109/ICACSIS.2016.7872780](https://doi.org/10.1109/ICACSIS.2016.7872780)

-
- [25] L. Torop, J. Morton, and J. B. West, "Unsupervised Feature Learning for Audio Classification using Convolutional Deep Belief Networks," *Proceedings of the 22th Annual Conference on Advances in Neural Information Processing Systems (NIPS)*, pp. 1096–1104, 2008. Available at: [Google Scholar](#).
- [26] N. F. Güler, E. D. Uğur, and I. Güler, "Recurrent Neural Networks Employing Lyapunov Exponents for EEG Signals Classification," *Expert Systems with Applications*, vol. 29, no. 3, pp. 506–514, 2005, doi: [10.1016/j.eswa.2005.04.011](#).
- [27] F. J. Ordóñez and D. Roggen, "Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition," *Sensors (Switzerland)*, vol. 16, no. 1, 2016. doi: [10.3390/s16010115](#)
- [28] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling," *INTERSPEECH*, 2014. Available at: [Google Scholar](#).
- [29] F. A. Gers and F. Cummins, "Learning to Forget : Continual Prediction with LSTM," *9th International Conference on Artificial Neural Networks*, vol. 12, no. 10, pp. 2451–2471, 2000. doi: [10.1162/089976600300015015](#)
- [30] S. Ruder, "An Overview of Gradient Descent Optimization Algorithms," *arXiv:1609.04747*, pp. 1–14, 2016. Available at: [Google Scholar](#).
- [31] D. P. Kingma and J. L. Ba, "Adam : A Method for Stochastic Optimization," *3rd International Conference for Learning Representations*, pp. 1–15, 2015. Available at: [Google Scholar](#).