ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

CORRELATION ANALYSIS APPROACH BETWEEN FEATURES AND MOTOR MOVEMENT STIMULUS FOR STROKE SEVERITY CLASSIFICATION OF EEG SIGNAL BASED ON TIME, FREQUENCY, AND SIGNAL DECOMPOSITION DOMAIN

Marcelinus Yosep Teguh Sulistyono[1,2], Evi Septiana Pane[4],
Eko Mulyanto Yuniarno[1,3], Mauridhi Hery Purnomo[1,3]

[1] Departement of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2] Departement of Information System, Universitas Dian Nuswantoro, Semarang, Indonesia
[3] Departement of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[4] Industrial Training and Education of Surabaya, Ministry of Industry RI, Surabaya, Indonesia

email: 07111960010007@student.its.ac.id[1], evi-septiana@kemenperin.go.id[2], ekomulyanto@ee.its.ac.id[3], hery@ee.its.ac.id[4]

Abstract

Stroke is a clinically defined illness characterized by immediate and focal neurological impairments arising from vascular injury (infarction, haemorrhage) to the central nervous system, which can lead to blockage and rupture of blood vessels, resulting in mortality and lifelong disability. The healing process necessitates a methodology for assessing criteria utilized in monitoring, evaluation, and medical rehabilitation. The measurement of appropriate parameters, assessed through the severity of stroke via the identification of components such as data retrieval, preprocessing, feature extraction, and feature selection, ensures that information is preserved, thereby facilitating the objectives of monitoring, evaluation, and medical rehabilitation for the recovery of stroke patients.

The challenge in rehabilitating stroke patients through monitoring, evaluation, and medical rehabilitation lies in the absence of accurate measurement correlating motor movement stimuli and features due to the signal's unstable, non-linear, and non-stationary characteristics. This necessitates an appropriate correlation method for stabilization, linearization, and stationarity within a reduced dimensional space.

In light of these challenges, it is essential to employ the correlation analysis method developed through EEG signal data recording, signal processing, feature extraction, feature selection via normality and significance tests, and correlation analysis to precisely identify the appropriate parameters for monitoring, evaluation, and medical rehabilitation through stroke severity classification patterns.

Our experimental findings indicate that the correlation analysis approach for assessing stroke severity classification patterns is evident in the Hajorth Complexity feature using shoulder motor movement stimuli and SVM classification, achieving a significant value of 98% accuracy. This finding corroborates the correlation analysis method between EEG signal features and motor movement stimuli in identifying the optimal parameters within a reduced dimensional space to assess stroke severity accurately.

Identifying suitable parameter measurements in stroke severity categorization enables physicians and physiotherapists to utilize these measurements for monitoring, evaluating, and rehabilitating stroke patients.

Keywords : Correlation Analysis, Feature, Motoric Movement Stimulus, EEG Signal, Stroke Severity

Received: 07-10-2024 | **Revised:** 05-11-2024 | **Accepted:** 09-11-2024
DOI: https://doi.org/10.23887/janapati.v13i3.85550

INTRODUCTION

Stroke is an acute and localized neurological condition resulting from arterial damage (infarction, haemorrhage) to the central nervous system, leading to mortality and permanent disability [1]. According to projections from the World Health Organization (WHO) spanning 2009 to 2019, stroke ranks as the second largest cause of death and the third major cause of disability globally, with disability-adjusted life-years lost (DALYs) indicating that 86% result in mortality and 89% lead to permanent disability [2]. Regular medical checks and maintaining a healthy lifestyle are essential to avert such issues. Doctors conduct examinations by eye observation, employing specific measurement techniques as tools for assessment. These visualization metrics yield varying analyses from each physician, which may lead to potential misdiagnosis. A brain examination tool that does not induce side effects, such as radioactivity and cellular death, is essential; thus, a safe instrument, specifically Electroencephalography (EEG), is required to ensure consistent diagnostic results. Electroencephalography (EEG) is effective for identifying electrical wave activity in the brain by electrodes affixed to the scalp, utilizing many channels as a data collection instrument [3]. Data collection from stroke patients via EEG, paired with a motor movement stimulus, seeks to assess the strengths and weaknesses of stroke-affected motion, hence serving as a metric for guiding subsequent medical interventions when translated into a signal [4].

EEG data obtained by scalp sensors can record up to 256 Hz or channels, with each channel comprising five frequency bands: alpha (8-13 Hz), beta (13-30 Hz), delta (0.1-4 Hz), gamma (30-100 Hz), and theta (4-8 Hz). The Gamma frequency sub-band exhibits the lowest amplitude and the highest frequency, whereas the Beta frequency sub-band has a frequency of > 14 Hz and a voltage of up to 25 mV. The Alpha frequency sub-band exhibits a voltage range of 10-150 mV, the Theta frequency sub-band is situated in the occipital or vertex temporal area, and the Delta frequency sub-band is characterized by a substantial amplitude and a low frequency, specifically below 3 Hz [5]. The overutilization of sensors will correlate with the number of channels and band components, leading to increased data complexity. The complexity of data will enhance the utilization of extracted features, hence identifying a diverse array of features that facilitate the selection of suitable features to minimize dimensional space usage.

Our project studies the link between EEG signal properties and the motion stimulus utilized in data collection, aiming to find patterns that will aid in post-stroke monitoring, evaluation, and rehabilitation. This study builds upon prior research that examines the selection of suitable parameters by comparing each feature associated with the subband and hand movement during data collection via stimuli. The study results revealed a correlation between the features and motions of the healthy and stroke-affected sides, aimed at identifying appropriate parameters for monitoring and rehabilitation through individual analysis [6]. This research exhibits limitations in feature use, as it employs only four features across two domains: the temporal and frequency domains. As the number of characteristics increases, the diversity in selecting parameters will also increase. A further limitation is that the accuracy rates remain approximately 53% for Elbow, 60% for Shoulder, and 61% for Grasp. Consequently, this classification pattern is less effective in assessing parameters, as it relies on individual analyses that cannot validate the accuracy of the parameters employed in monitoring, evaluation, and medical rehabilitation processes.

Furthermore, further studies have identified efficacy patterns in rehabilitation therapy, whether conducted at home or in a hospital setting, through movement stimuli associated with enhancements in morbidity and mortality by monitoring motor coordination skills in stroke rehabilitation. This work aims to establish a correlation between characteristics and movement stimuli utilized for rehabilitation during a pandemic [7]. This research exhibits limitations in feature utilization, as it employs only a single feature in the time domain, resulting in a classification pattern lacking comparative analysis. Stroke examination encompasses both movement and various aspects, including dimensions, frequency bands, features, and domains, which collectively contribute to EEG signal-based research. A limitation of this study is the elevated accuracy rate of 95% in the classification pattern, derived from a single feature associated with three motions, leaving no basis for comparison.

This study examines the influence of characteristics, domains, frequency bands, and motion stimuli on EEG signal patterns corresponding to different stroke severity levels, serving as a foundation for selecting further treatment for stroke patients. This research aims to identify the categorization pattern of stroke severity in the temporal domain that links characteristics with motor movements [8]. This research needs to improve in identifying categorization patterns due to an inadequate application of features and domains, as it just employs statistical features and time domains in data processing. A further limitation in assessing the accuracy value remains at around 82.5% in shoulder movements.

Appropriate parameters must be established through classification types, standardized Manual Muscle Testing (MMT) movements, and features with a reduced dimensional space encompassing multiple domains to facilitate the monitoring, evaluation, and medical rehabilitation of stroke patients. By correlating acceptable movement norms and

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

features from multiple domains, numerous patterns for classifying stroke severity will emerge. Appropriate metrics can be identified for monitoring, evaluating, and rehabilitating stroke patients through various classification patterns.

A correlation classification model between EEG signal features and motor movement stimuli is required, utilizing exact parameter measurements through signal processing, feature selection, and correlation analysis. Establishing a correlation classification model for signal processing analysis necessitates the selection of signals and the correlation of EEG signal features with motor movement stimuli. A relationship between inferential statistical analysis and correlation analysis is essential to minimize redundant resources in processing, thereby enhancing signal reduction and improving processing performance. The methodology employed involves inferential statistical analysis and correlation analysis, comparing EEG signal features with one another, motor movement stimuli with each other, and EEG signal features with motor movement stimuli. This approach aims to identify differences or similarities in patterns that are pertinent for developing stroke severity classification models, thereby yielding accurate correlation analysis results. To develop classification patterns by enhancing accuracy and complexity through the analysis of differences and similarities, inferential statistical methods and correlation analysis are required, offering computational efficiency and robustness in pre-processing.

This paper aims to identify classification patterns that reveal similarities and differences, enhancing accuracy. To achieve this, feature selection via feature analysis and correlation analysis is essential for determining optimal parameters, stabilizing signals, and ensuring computational efficiency and robustness. We propose a correlation analysis method between EEG signal features and motor movement stimuli, facilitating feature selection and the identification of variances or commonalities in variables and parameters for classifying stroke severity patterns. This study contributes to research in the following manner:

1. Facilitate the utilization of private datasets that have undergone ethical feasibility assessments by the Hospital.
2. Contribute to identifying optimal EEG signal characteristics and motor movement stimuli that significantly impact monitoring, assessment, and medical rehabilitation.
3. Enhance the classification significance in identifying appropriate parameters as materials for developing stroke severity classification patterns through a correlation analysis methodology.

METHOD

This research proposes a correlation analysis approach between EEG signal properties and motor movement stimuli, illustrated in the block diagram in Figure 1.

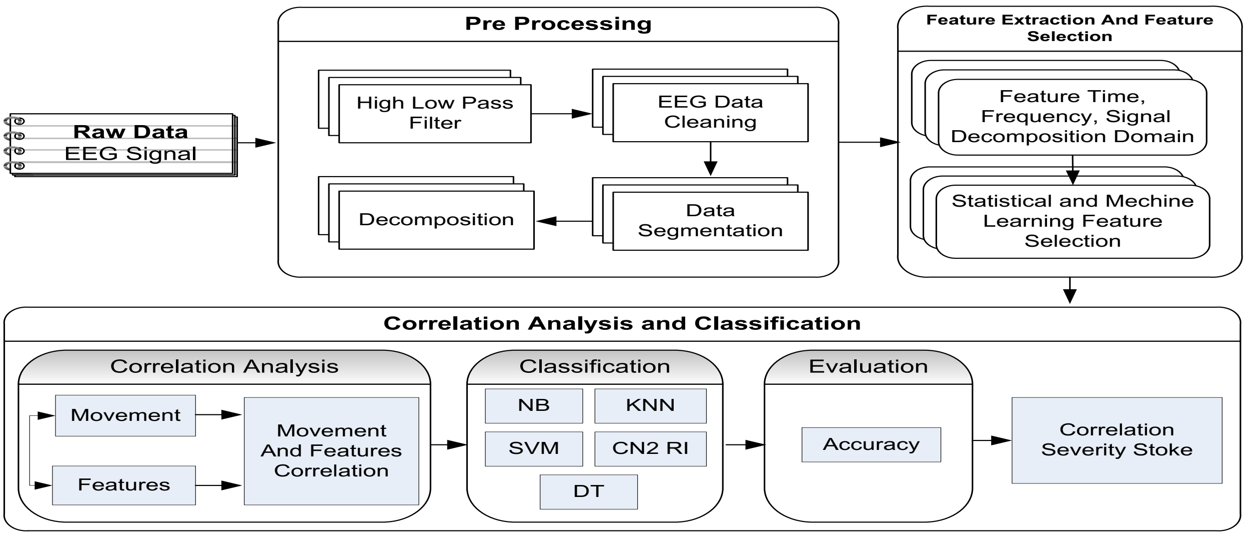


Figure 1. Flowchart of Correlation Analysis Method for Stroke Severity Classification

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

The initial phase involves raw EEG signal data, encompassing data collected from stroke victims. The second stage involves preprocessing, encompassing high- and low-pass filters, EEG data cleansing, data segmentation, and decomposition. The third stage entails feature extraction and selection, which includes time-domain feature extraction, frequency-domain, signal decomposition-domain (multi-domain), and the optimal feature selection process. The fourth stage involves correlation analysis for classification, encompassing correlation analysis, classification, and evaluation.

A. EEG Signal Raw Data

The chosen patient data comprise individuals with stroke from Simpang Lima Gumul Hospital in Kediri, East Java, with the etiology of stroke attributed to diminished cerebral blood flow resulting from ischemia (thrombotic obstruction or arterial embolism). Data collection involved EEG readings from 22 ischemic stroke patients, all of whom provided consent during the data-gathering process. Data collection must adhere to recording protocols, specifically established regulations that apply before, during, and after the recording process. Physicians have classified all data collected as mild, moderate, or severe conditions and have had ethical approval from the Hospital.

Table 1. Stroke Patient Data

Patient	Age	Sex	Affected Hand	Category
P01	52	Male	Left	Severe
P02	61	Male	Left	Severe
P03	66	Male	Left	Severe
P04	48	Male	Left	Severe
P05	60	Male	Left	Moderate
P06	58	Male	Left	Moderate
P07	60	Male	Right	Moderate
P08	56	Female	Left	Mild
P09	53	Female	Right	Mild
P10	50	Female	Right	Mild
P11	50	Female	Left	Mild
P12	58	Male	Left	Mild
P13	61	Male	Left	Severe
P14	48	Male	Left	Severe
P15	58	Male	Left	Severe
P16	50	Female	Right	Severe
P17	50	Female	Left	Moderate
P18	56	Female	Left	Moderate
P19	52	Male	Left	Moderate
P20	53	Female	Right	Mild
P21	60	Male	Right	Mild
P22	67	Male	Right	Mild

Table 1 elucidates the correlation between the affected hand and the stroke incident impacting the left or right hand. This indicates that the impact of the assault results in weakness or an attack on either the right or left hand.

The data collection employs the EEG signal device OpenBCI (Open et al.) UltraCortex ‘Mark IV’ and the Cayton Board 8 Channel as a processing apparatus [9], as this device meets the study requirements for the motor cortex region. This device is portable, user-friendly, and affordable compared to EEG equipment for medical applications. According to a study by Jeremy Frey [10], this EEG device demonstrates a commendable accuracy in interpreting EEG signals, comparable to that of medical-grade EEG instruments. The EEG reading results from this instrument can be validated. The channel configuration of this tool can be modified, comprising six electrodes and channels F3, F4, C3, C4, O1, and O2.

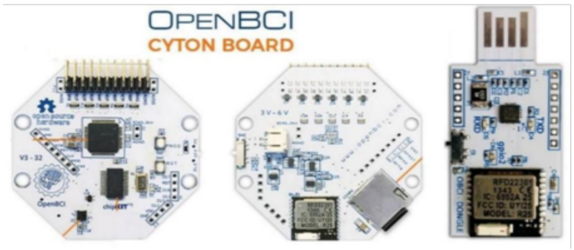


Figure 2. Headset UltraCortex “Mark IV”

Alongside the EEG signal device, data collection employs a motoric movement stimulus via the Manual Muscle Testing (MMT) approach, which includes hand-held, elbow, and shoulder movements. The objective of data collecting via EEG signal devices in conjunction with movement stimuli is to clinically analyze the patient's neuromotor performance, facilitating the analysis, evaluation, and assessment of motor performance in rehabilitation. Neuromotor impairment is characterized by compromised movement or motor function that impacts the neurological system. Integrating EEG and movement stimulus signaling methods is essential for elucidating neuromotor integrity or disease by observing brain activation and motor correlates [11].

Upon the availability of all necessary components, including patients, EEG equipment, and motions, data collection is executed based on the principle of recording data during the collection process as outlined below:

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

Figure 3. Movement Stimulus

1. The patient is allotted 60 seconds to execute three movements: shoulder, grip, and elbow, with a sampling frequency of 256 Hz. The recording begins with 5 seconds of relaxation, followed by 10 seconds of grip movement. Another 5 seconds of rest was succeeded by 10 seconds of elbow movement, followed by 5 seconds of relaxation, and the conclusion was 10 seconds of shoulder movement.
2. The data gathering technique involved recording data in OpenBCI software, which includes Time Series, FFT Plot, and Bandpower records.

Figure 4. OpenBCI Software Interface

The time series will archive EEG signals across six channels, with channel 1 denoting F3, channel 2 denoting F4, channel three denoting C3, channel four denoting C4, channel 5 denoting O1, and channel six denoting O2.

3. The data produced from the recording process is in a raw format compatible with Matlab (*.mat). The event marker and channel location are also imported to facilitate additional data processing. Event markers denote movement instructions per time unit (*.txt). The locations of EEG channels are referred to as channel locations.

B. Preprocessing

Raw data will be generated upon data acquisition, which necessitates preprocessing due to significant interference. This data does not accurately represent the original signal, as noise disrupts the separation of pertinent brain signals from extraneous activity. The preprocessing is conducted through the subsequent steps:

1. High Low Pass Filter
   Employing the Finite Impulse Response (FIR) Method with a Hamming Window to enhance the ripple quality in spectral sidelobes derived from the Fourier transform. The lowpass filter will be set at 1 Hz to eliminate the signal's muscular noise and low-frequency distortions. The highpass filter will be set at 40 Hz to eliminate Radio Frequency (RF) wave interference and high-frequency noise from the signal [9].
2. Data Cleaning
   Data cleaning is performed to eliminate noise from muscle action around the head, including swallowing, head movements, or excessive hand motions characterized by high-frequency activity or elevated amplitude in EEG time series data. The technique employed for data cleansing via Automatic Subspace Reconstruction (ASR). This approach autonomously does cleaning and effectively eliminates substantial quantities of artifact noise [10].
3. Data Segmentation
   Segmentation is the process of partitioning data based on predefined events, ensuring that the residual data exclusively comprises the intended event (movement command). To accommodate the motion capabilities of stroke patients, the data is truncated one second before the movement command and three seconds after the movement command. Each movement is captured by 0.5 seconds of data, yielding 20 data points in a single series of motions [11].
4. Decomposition
   Decomposition is the process of partitioning a band into multiple frequency bands. This study concentrates on four EEG signal sub-bands: alpha low (α low), alpha high (α high), beta low (β low), and beta high (β high). The decomposition procedure employs the Infinite Impulse Response (IIR) method. IIR is employed as the extraction method for α low, α high, β low, and β high waves because of its ability to generate wavelengths identical to the original signal. This research employs a Chebyshev type II convolution window in the IIR filter [14]. This

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

study will evaluate two of the four brain wave frequencies, specifically beta and alpha, for subsequent feature extraction. Alpha waves will be categorized into alpha low and alpha high, whereas beta waves will be divided into beta low and beta high. Motor activity in the brain is typically associated with alpha-low, alpha-high, beta-low, and beta-high waves. These four waves play a crucial role in motor motions requiring conscious execution.

C. Feature Extraction and Feature Selection

1. Feature Extractin

Feature extraction is the process of uncovering latent characteristic information by expressing an input signal through features that denote specific behaviors or patterns inherent in the signal [16] [17] [18] [19]. Feature extraction typically involves dimension reduction or data compression to minimize the resources needed for analyzing the input signal.

Table 2. Feature Formula

Domain	Feature	Features	Description
Time Domain	Variance	$\sigma^2 = \frac{1}{\mathrm{n}-1}\sum_{i=1}^{n}(I_n - \mu)^2$	Where σ is the EEG signal, n is the total number of EEG data, and μ is the average value.
Time Domain	Energy	$E = \pi \sum_{n=-\infty}^{\infty} \lvert x(n)\rvert^2$	Where E is the energy, π is the average value, and x(n) is the EEG signal.
Time Domain	Skewness	$Skewness = \frac{\sum_{i=1}^{N}\frac{(xi-\mu)^2}{N}}{(\sum_{i=1}^{N}\frac{(xi-\mu)^2}{N})^{2/2}}$	Where N is the total amount of EEG data, μ is the average value, and n is the EEG signal.
Time Domain	Zero Crossing	$Z(\mathrm{i}) = \frac{1}{2N}\lvert[X_i(n)] - [X_i(n-1)]\rvert$	Where Z(i) is the value of the zero crossing feature, x[n] is the amplitude value at the nth index data, N is the total number of bits in frame t.
Time Domain	Hjorth Activity	$\sigma_x^2 = \frac{\sum_{n=0}^{N-1}(x(n)-\bar{x})^2}{N}$	Where σ_0^2 is the variance of the signal $S_{norm}(n), \sigma_1^2$ is the first derivative looking for variance, n is the order
Time Domain	Hjorth Mobility	$M_x = \frac{\sigma_{x'}}{\sigma_x}$	Where Mx is the variance of the signal, σ_1^2 is the first derivative of finding the variance.
Time Domain	Hjorth Complexity	$\mathrm{FF} = \frac{M_{x'}}{M_x} = \frac{\sigma_{x''}/\sigma_{x'}}{\sigma_{x'}/\sigma_x}$	Where σ_0^2 is the variance of the signal $S_{norm}(n), \sigma_1^2$ is the first derivative looking for variance, n is the order
Signal Decomposition Domain	EMD	$Z(t) = \sum_{i=1}^{n} C_i\,(t) + r(t)$	Where Z(t) is the local extreme value, n is the number of iterations starting from i=1, Ci(t) is separating the residuals by equation, while r(t) is the residual data.
Frequency Domain	PSD	$PSD = \frac{[Xn]^2}{f} = \left(\frac{(amplitudo)^2}{Hz}\right)$	Where x[n] is the amplitude value at the n index data.
Frequency Domain	PP	$PP = \frac{\sum_f S(f)}{\sum_{f-f1}^{f2} S(f)}, f \in [8{,}12]$	Where PP is the power percentage account in the frequency band, and S(f) is the PSD account

2. Feature Selection

a. Normality Test

A normality test is conducted to ascertain if the sample data is usually distributed, evaluating the distribution of the obtained data through statistical analysis.

$$X^2 = \sum \frac{(O_i - E_i)}{E_i} \quad (1)$$

Where X^2 represents the X^2 value, Oi denotes the observed value, Ei signifies the predicted value derived from the standard table multiplied by N, and N indicates the total count of data points.

A normality test is conducted on features derived from the time domain, frequency domain, and signal decomposition domain (multi-domain). Feature selection is conducted after identifying pertinent and optimal features using the normalcy test. A normality test is conducted to ascertain if the sample data is usually distributed, evaluating the distribution of the obtained data through statistical analysis.

b. Feature Selection

Feature selection is conducted using a significance test using Analysis of Variance, which assesses the extent of the average differences among groups by comparing their means.

Table 3. Feature Selection

Name	Formula	Description
SSW	$SSW = \sum_{j=1}^{k}(X - \bar{X}_j)^2$	$\bar{x}_j$ is the average value in the j group,
SSB	$SSB = \sum_{j=1}^{k}(\bar{X}_j - \bar{X})^2$	$\bar{x}_T$ is the overall average value
df_w	$df_w = k - 1$	x_{ij} data ke-i dalam kelompok ke j
df_b	$df_b = n - 1$	
MSW	$MSW = \frac{SSW}{df_w}$	
MSB	$MSB = \frac{SSB}{df_b}$	

The significance test is conducted by categorizing each variable into a specified target, followed by the computation of the Mean Squares value for each feature. The Mean Squares value determines the ratio of squares to the degrees of freedom for each feature, thereby enabling the ratio calculation for each Mean Squares value.To find the best feature, an F-test was conducted with the following equation:

$$F = \frac{MSB}{MSW} \quad (2)$$

In equation 2, MSB is the Mean Square Between, and MSW is the Mean Square Within.

D. Correlation Analysis and Classification

1. Correlation Analysis

Correlation analysis examines the extent of the relationship between variables as indicated by the correlation coefficient value. The correlation between these variables might be both positive and negative. In correlation analysis, there is no designation of independent variable (X) and dependent variable (Y).

This study determined the coefficient using the Pearson correlation method (r) [19] between the average values of mild, moderate, and severe situations, utilizing the C3 and C4 channels of EEG, involving 22 stroke patients.

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} \quad (3)$$

X is the variable value, Y is the variable value, r is the correlation coefficient, and n is the information quantity.

In this work, we use two variables, namely features and motor movement stimulus, and use a combination of inferential statistical analysis with correlation analysis, so the equation is developed using the development of inferential statistical analysis algorithms with correlation, which we compiled under the name SigKorelasi algorithm. In Equation 3, the application of only two variables will be compared.

The SigCorrelation method clarifies that we only add one variable as the counter of the new variable if there is more than one variable. From 1 variable X, if another variable, it will be X and Y as the input variables. The variable input is then repeated depending on the weight count limit.

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

Algorithm	: SigCorrelation
Input	: X-input, Y-input is the dataset, F-input is the statistical F value, n-value initial weight, j-value final weight
Output	: SigCorrelation
Step	:

1 : Initialise SigCorrelation
2 : Set Height Limit j = sum of weights
3 : Set F = statistic-value
4 : for n = 1 to j
5 :   If F < 0.05
6 :   A ← [n](XY) - (X[n]) (Y[n])
7 :   B ← √[[n](X^2 - X[n] ^2)] [[n](Y^2 - Y[n] ^2)]
8 :   SigCorrelation ← A/B
9 :   end if
7 : end for
8 : return SigCorrelation

2. Classification
This research employs machine learning methods to train and predict extracted attributes. Through feature extraction, suitable features were identified using five classification methods: Decision Tree (DT), K-Nearest Neighbor (KNN), Naive Bayes (NB), Support Vector Machine (SVM), and CN2 Rule Induction (CN2 RI). These methods evaluate the correlation between EEG signal features and motor movement stimuli to assess the performance of a stroke severity model, thereby facilitating effective classification. This will demonstrate the effectiveness of parameter adjustment in classifying the retrieved features, allocated as 80% for training and 20% for testing.

3. Evaluation
The assessment employs a comparative approach to classify results into severe, moderate, and mild. The three classes are evaluated using classification technique parameters in a matrix table that delineates the classification model's performance on a set of test data with known true values.

$$Accuracy\ (\%) = \frac{Number\ of\ corect\ predict\ sample}{Number\ of\ total\ sample} \quad (4)$$

The accuracy % is determined by the ratio of correctly predicted samples to the total number of samples.

RESULT AND DISCUSSION

A. Result

1. Results of EEG Signal Recording Dataset
Upon completion of the EEG signal recording procedure on stroke patients, the subsequent data set will be produced :

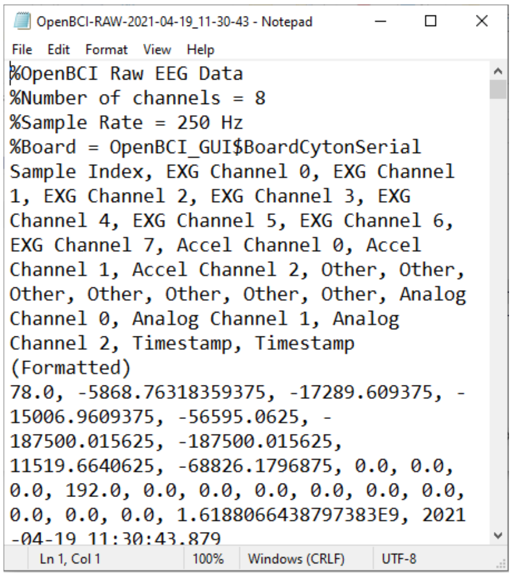

OpenBCI-RAW-2021-04-19_11-30-43 - Notepad
File Edit Format View Help
%OpenBCI Raw EEG Data
%Number of channels = 8
%Sample Rate = 250 Hz
%Board = OpenBCI_GUI$BoardCytonSerial
Sample Index, EXG Channel 0, EXG Channel 1, EXG Channel 2, EXG Channel 3, EXG Channel 4, EXG Channel 5, EXG Channel 6, EXG Channel 7, Accel Channel 0, Accel Channel 1, Accel Channel 2, Other, Other, Other, Other, Other, Other, Other, Analog Channel 0, Analog Channel 1, Analog Channel 2, Timestamp, Timestamp (Formatted)
78.0, -5868.76318359375, -17289.609375, -15006.9609375, -56595.0625, -187500.015625, -187500.015625, 11519.6640625, -68826.1796875, 0.0, 0.0, 0.0, 192.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.6188066438797383E9, 2021-04-19 11:30:43.879
Ln 1, Col 1 100% Windows (CRLF) UTF-8

Figure 5. Dataset of Recording Results

The resultant dataset comprises text formatted as a time-series dataset, with a sampling frequency of 256 Hz, utilizing six channels: F3, F4, C3, C4, O1, and O2. Each patient's dataset has four text files: baseline, shoulder, grip, and elbow.
Once all the raw data is acquired, it is further segmented into the requisite channels. This research exclusively focuses on motorbikes, utilizing only channels C3 and C4, which are converted into digital signals and prepared for preprocessing.

2. Preprocessing Results
Preprocessing limits the high-low-pass filter so that the frequency is not too high or too low. Then, the data is cleaned from noise and interference due to the crackle during recording. Furthermore, the data is segmented as desired so that the pattern of the EEG signal waveform is visible for easy research. The next step is decomposition, which divides the signal into subbands so that each event is divided into predetermined subbands.
Table 4 below shows the results of preprocessing processing based on 22 patients who have processed the data.

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

Table 4. Preprocessing Results

Patient	Hjorth Complexity			
	Shoulder			
	Alpha Low	**Alpha High**	**Beta Low**	**Beta High**
P01	0.2137	0.2725	0.3748	0.5869
P02	0.2141	0.2619	0.3729	0.5863
P03	0.2104	0.2694	0.3777	0.5636
P04	0.208	0.2645	0.3757	0.5988
P05	0.2075	0.2716	0.3634	0.5720
P06	0.2108	0.2687	0.3635	0.5925
P07	0.2095	0.2572	0.3648	0.5921
P08	0.2118	0.2574	0.3626	0.5977
P09	0.212	0.2590	0.3687	0.5773
P10	0.2112	0.2599	0.3711	0.5926
P11	0.2091	0.269	0.3732	0.5783
P12	0.2158	0.2582	0.3761	0.5774
P13	0.2079	0.2539	0.3653	0.5803
P14	0.2122	0.2640	0.3610	0.5734
P15	0.2097	0.2632	0.3670	0.5799
P16	0.2094	0.2665	0.3690	0.5979
P17	0.2098	0.2615	0.3673	0.5877
P18	0.2097	0.2683	0.358	0.5461
P19	0.2151	0.2606	0.3651	0.5625
P20	0.2079	0.2755	0.363	0.5889
P20	0.2110	0.2579	0.3647	0.5767
P21	0.2079	0.2755	0.363	0.5889
P22	0.2137	0.2725	0.3748	0.5869

The processed data above is one example of the Hjorth Complexity feature in the time domain, with shoulder movements in the alpha low, alpha high, beta low, and beta high subbands. So, each feature has three movements, namely shoulder, grasp, and elbow, each stored in four subbands: alpha low, alpha high, beta low, and beta high. If each feature has 12 columns of data, then there are 120 columns if there are ten features.

3. Correlation Analysis Resuls

This study aims to ascertain the correlation between EEG signal features and motor movement stimuli, serving as a classification pattern for stroke severity to guide subsequent monitoring, evaluation, and medical rehabilitation, utilizing brain wave characteristics in stroke severity groups through correlation analysis methods. The features extracted from the time domain include Variance, Skewness, Zero-crossing, Energy, and Hajorth Parameters (activity, mobility, and complexity); from the frequency domain, they include Power Percentage (PP) and Power Spectral Density (PSD); and from the decomposition domain, Empirical Mode Decomposition (EMD) is utilized. This research employs an approach comprising three test scenarios: (a) Normality Test Results, (b) Feature Selection Results via Significant Test, (c) Correlation Analysis Results, and (d) Classification and Evaluation.

a. Analysis of Normalization Test Outcomes

A normality test evaluates whether a sample or data distribution from a population adheres to the assumption of normality, which is essential for statistical analysis. This study examined the EEG signal characteristics and hand motor movement stimuli to determine if the sample data had a normal distribution, as indicated by the results below :

Table 5. Feature and Movement Correlation Normality Test Results

Feature	Test Normality		
	Elbow	**Grasp**	**Shoulder**
Variance	0.001	0.006	0
Skewness	0	0	0.009
Zero Crossing	0	0.661	0
Hjorth Activity	0.001	0.01	0
Hjorth Mobility	**0.439**	**0.052**	**0.996**
Hjorth Complexity	**0.34**	**0.096**	**0.922**
Energy	0	0.002	0
PP	0	0	0.048
PSD	**7.495**	**6.495**	**9.945**
EMD	0	0.084	0

The analysis of the normalization test results for the correlation between EEG signal features and motor movement stimuli in Table 5 indicates that the classical assumption required for regression analysis is the normal distribution of residual values, necessitating a normality test on these residuals. The test result for each characteristic regarding a normally distributed residual value indicates a significant average value level with a p-value greater than 0.05. The results in Table 5 indicate that the EEG signal features and motor movement stimuli that exhibit a normal distribution include Hjorth Mobility, Hjorth Complexity, and PSD features (highlighted in bold red). The outcomes of this normalcy assessment serve as the foundation for identifying the characteristics that will be

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

utilized in important testing and feature selection.

b. Feature Selection Outcomes via Significance Testing

The outcomes of the normalcy test are next analyzed using the Analysis of Variance, categorized into stroke severity groups according to the most relevant EEG signal characteristics and motor movement stimuli. This technique will create the average value for each stroke severity group to identify the most suitable EEG signal characteristics and motor movement stimuli for monitoring, evaluation, and medical rehabilitation.

According to the examination of the substantial test findings in Table 6 for each feature, the F-Table value, derived from the number of independent variables and respondents, is 3.13. The F value (F-Count) is usually distributed when the F-Count exceeds the F-Table value, as indicated by the formula. According to the F (F-Count) value results in the z table, instances exist when F-Count exceeds 3.13, indicating that all features are regularly distributed, specifically the Hjorth Mobility Feature, Hjorth Complexity, and PSD during elbow, grip, and shoulder movements.

The study of the feature selection findings in the Table 6 indicates that each feature possesses a normally distributed residual value, as evidenced by a significant average p-value<0.05 for Hjorth Mobility, Hjorth Complexity, and PSD concerning elbow, grip, and shoulder movements.

c. Correlation Analysis Results

The correlation analysis results, employing the Pearson correlation method, indicate the degree of association between the EEG signal characteristic and the motor movement stimulus, represented by the correlation coefficient value. The link between these variables may manifest as features and motions or as movements categorized into mild, moderate, and severe, yielding both good and negative outcomes. This correlation analysis examines the relationship between characteristics and motor motions, with the results presented in Table 7.

The correlation analysis results in Table 7 assess the link between each EEG signal aspect and the movement stimulus, yielding a correlation coefficient with the following objectives:

The correlation analysis results indicate a robust relationship between the EEG signal features and the movement stimuli, with correlation coefficients ranging from 0.76 to 0.99 for the Hjorth Mobility feature, Power Spectral Density (PSD) with elbow, grasp, and shoulder movements, as well as the Hjorth Complexity feature in elbow and grasp movements. A high link, indicated by a coefficient value between 0.51 and 0.75, is shown in the Hjorth Complexity characteristic concerning shoulder movements.

Table 6. Significant Test Results and Feature Selection

		Significant Test and Feature Selection					
		Elbow		**Grasp**		**Shoulder**	
Domain	**Feature**	**F**	**Sig.**	**F**	**Sig.**	**F**	**Sig.**
Time Domain	Hjorth Mobility	176340.8	0.02	154744.9	0.02	91140.04	0.02
Time Domain	Hjorth Complexity	181836.6	0.02	180637.9	0.02	104731.1	0.02
Frequency Domain	PSD	40.27	0.02	30.323	0.02	69.667	0.02

Table 7. Correlation Analysis Results of Features and Motor Movement Stimulus

		Correlation Analysis of Features and Motor Movement Stimulus					
		Elbow		**Grasp**		**Shoulder**	
Domain	**Feature**	**F**	**Sig.**	**F**	**Sig.**	**F**	**Sig.**
Time Domain	Hjorth Mobility	0.779	0.05	0.841	0.05	0.708	0.05
Time Domain	Hjorth Complexity	0.774	0.05	0.872	0.05	0.664	0.01
Frequency Domain	PSD	0.776	0.05	0.893	0.05	0.889	0.05

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024

1) The correlation analysis results indicate that all coefficient values are positive, signifying a unidirectional link between the EEG signal characteristic and the movement stimulus.
2) The correlation analysis results indicate a substantial link between each EEG signal characteristic and each movement stimulus.

d. Classification and Evaluation Results

The comprehensive assessment of the proposed method for correlating EEG signal features with motor movement stimuli, evaluated through five classification algorithms—DT, KNN, SVM, NB, and CN2 RI—yields accuracy values that align with the conducted correlation analysis. Table 8 illustrates that the highest average accuracy value, 98%, is achieved by the correlation between EEG signal features and motor movement stimuli, as determined by the DT, KNN, and SVM classification methods. The association between the EEG signal features and the motor movement stimulus achieved an accuracy of 98% for Hjorth Mobility, Hjorth Complexity, and PSD, specifically for elbow and shoulder movements.

Figure 6 illustrates that the mean accuracy of the five classifications based on the EEG signal features that underwent feature selection—specifically, Hjorth Mobility, Hjorth Complexity, and PSD—along with the notable Elbow, Grasp, and Shoulder motor movement stimuli, reached an impressive accuracy of 98% utilizing DT, KNN, SVM, and CN2 RI classifications, 97% with KNN and CN2 RI, and 94% with DT and NB.

Table 8. Correlation Analysis Evaluation Results

Class	Movement	Feature		
		Hjorth Mobility	Hjorth Complexity	PSD
DT	Elbow	**0.98**	**0.98**	**0.98**
	Grasp	0.923	0.923	**0.98**
	Shoulder	0.884	0.884	**0.98**
KNN	Elbow	0.903	0.903	**0.98**
	Grasp	0.961	0.961	**0.98**
	Shoulder	**0.98**	**0.98**	**0.98**
SVM	Elbow	**0.98**	**0.98**	**0.98**
	Grasp	**0.98**	**0.98**	**0.98**
	Shoulder	0.769	0.769	**0.98**
NB	Elbow	0.923	0.923	**0.98**
	Grasp	0.923	0.923	**0.98**
	Shoulder	0.865	0.865	**0.98**
CN2 RI	Elbow	**0.98**	**0.98**	**0.98**
	Grasp	0.961	0.961	**0.98**
	Shoulder	0.961	0.961	**0.98**

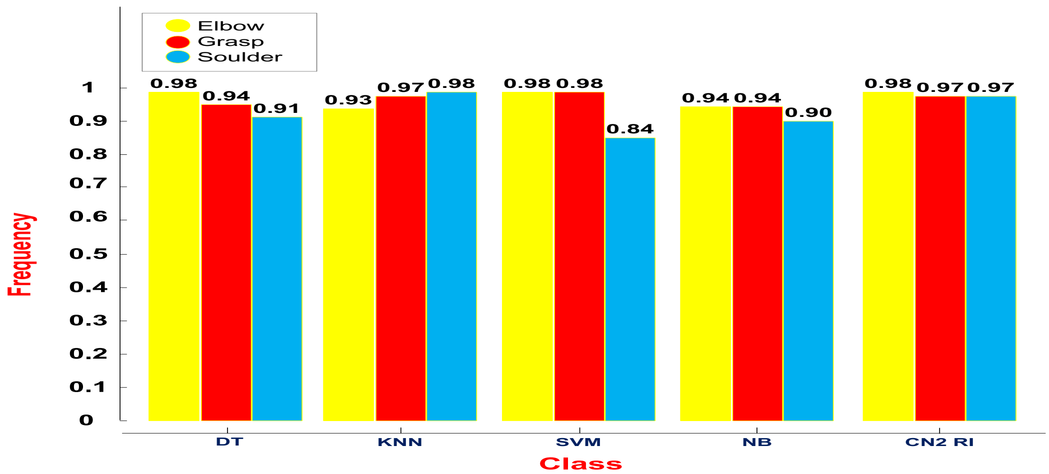
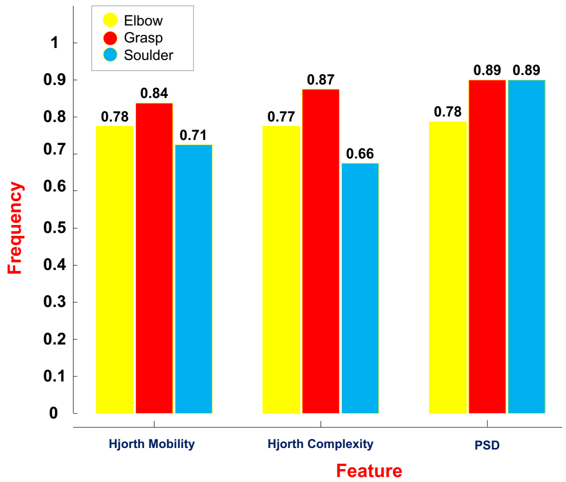


Figure 6. Average classification performance based on each motor movement stimulus

ISSN 2089-8673 (Print) | ISSN 2548-4265 (Online)
Volume 13, Issue 3, December 2024



Figur 7. Average classification performance of relationship strength based on each feature

Figure 7 illustrates that the average accuracy of the five classifications relating to the strength of the correlation between the EEG signal features and the motor movement stimulus achieved a remarkable accuracy of 89% for the PSD feature with Grasp and Shoulder movements, 87% for the Hjorth Complexity feature with Grasp movements, and 84% for the Hjorth Mobility feature with Shoulder movements.

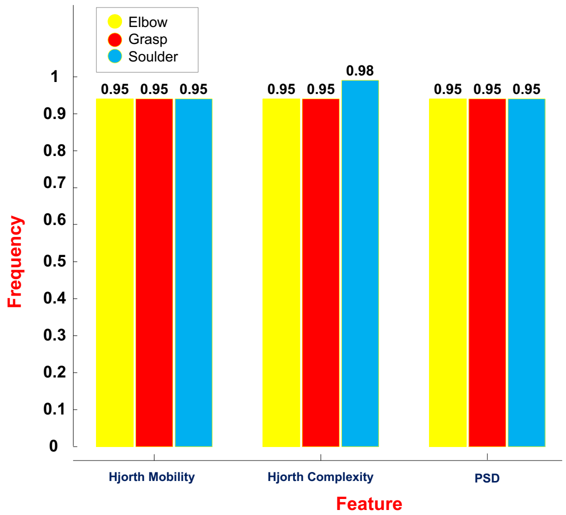


Figure 8. Average classification performance of significant value based on each feature

Figure 8 illustrates that the mean accuracy of the five classifications exhibiting a significant correlation between the EEG signal features and motor movement stimuli attains a notable 98% for the Hjorth Complexity feature associated with Shoulder movement. In contrast, the Hjorth Mobility and PSD features yield an accuracy of 95% for Elbow and Grasp movements, respectively.

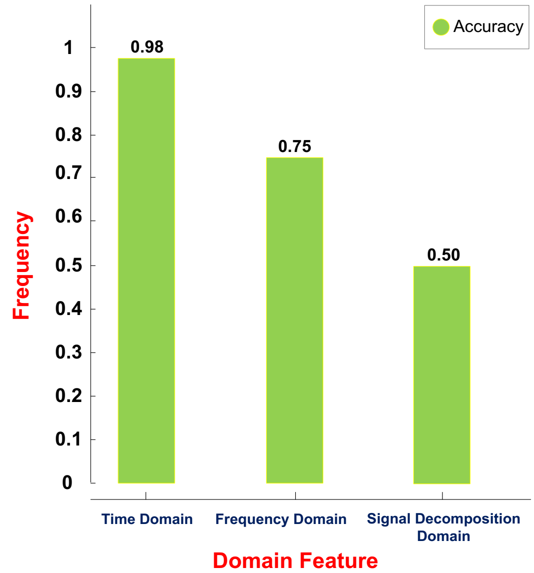


Figure 9. Average classification performance of significant values based on each feature domain

Figure 9 illustrates that the mean accuracy of the five classifications, which possess a high significance value for the domain, feature, and motion, attains peak accuracy levels of 98% in the time domain, 75% in the frequency domain, and 50% in the signal decomposition domain.

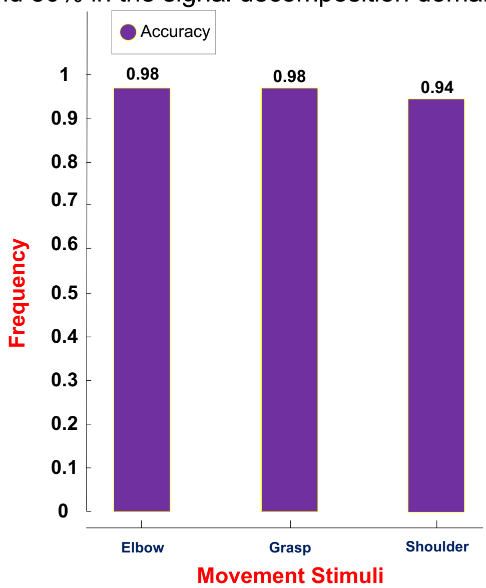


Figure 10. Average classification performance based on each motor movement stimulus

Figure 10 illustrates that the mean accuracy of five classifications utilizing three features with substantial significance to the motor movement stimulus is 98% for Elbow and Grasp movements and 94% for Shoulder movements.

The analysis above indicates that the correlation between EEG signal characteristics and motor movement stimuli is as follows:

Table 9. Overall average accuracy value of correlation

Class	Feature	Movement	Sig.
SVM	Hjorth Complexity	Shoulder	98%
DT	Hjorth Mobility	Elbow	95%
KNN	Hjorth Mobility	Elbow	95%
CN2RI	Hjorth Mobility	Elbow	95%
NB	PSD	Grasp	95%

Table 9 illustrates that the average classification accuracy using SVM achieves the highest performance in the time domain, scoring 98%. This is based on the Hjorth Complexity feature with shoulder movement. DT, KNN, CN2 RI, and NB also demonstrate optimal classification performance in both the time and frequency domains, scoring 95% with elbow and grasp movements.

B. Discussion

Dialogue This study employed a correlation analysis method to assess the relationship between EEG signal features and motor movement stimuli, aiming to establish a classification pattern of stroke severity by calculating the average values among groups. This approach facilitates the identification of relevant features and movements for monitoring, evaluating, and rehabilitating stroke patients. This research represents the inaugural correlation analysis, which requires the prior selection of appropriate characteristics through normalcy and significance testing for each parameter. Thus, with the appropriate selection of features, the correlation between features and motor motions yields accurate measurement parameters.

All prior studies must be comparable to ensure the research is consistent and pertinent; hence, earlier research should, at a minimum, involve the same disease type, condition, characteristic, and movement stimulus to facilitate proportionate comparison. The initial prior study provided insights into feature connection with motor motions, yielding accuracy values of 53% for Elbow, 60% for Shoulder, and 61% for Grasp [8]. Similarly, the preceding study exclusively emphasized characteristics in the time domain, with Shoulder motions achieving an accuracy rate of 95% [10]. The third prior study, which exclusively presents features in the time domain, reports an accuracy of 82.5% for shoulder movement [9].

In contrast to the three preceding studies that employed individual and cluster analysis methods, our research utilizes correlation analysis, encompassing three scenarios: normality test, significance test, and correlation analysis. Consequently, these scenarios yield an accuracy value of 98% for the Hjorth Complexity feature associated with shoulder movement and 95% for the Hjorth Mobility and PSD features related to elbow and grasp movements. This affirms that our research exhibits superior accuracy compared to prior studies; however, it has limitations, specifically that we must thoroughly investigate all aspects of correlation. Consequently, further research is warranted. This study solely correlates features with motor movements. Future investigations will aim to correlate features, frequency bands, and movements, thereby increasing the complexity of the problem.

CONCLUSION

The correlation analysis method between features and motoric movement stimuli for stroke severity classification, utilizing time domain, frequency, and signal decomposition, achieved a notable accuracy of 98% with the Hjorth Complexity feature in response to shoulder movement stimuli. The stroke severity classification pattern, derived from the correlation study between EEG signal features and motor movement stimuli, is identified in the SVM classification type, achieving an accuracy of 98%.

Compared to prior studies on correlation analysis performance, the influence of features on motor movement stimuli yielded an optimal accuracy of 61% for Grasp movement and 95% and 82.5% for Shoulder movement. The results indicate that the proposed correlation analysis method, employing three approaches—normality test, significance test, and correlation analysis—effectively identifies patterns in stroke severity classification related to feature correlation with movement, facilitating monitoring, evaluation, and medical rehabilitation of stroke patients.

Future research will involve a novel technique for analyzing correlations between features, motion stimuli, and frequency subbands regarding stroke severity of EEG signals, utilizing time domain, frequency, and signal decomposition methodologies.

ACKNOWLEDGMENT

This research received financial support from Dian Nuswantoro University, Semarang, in the form of a PhD scholarship. This research also involved collaboration with the Multimedia Computing Laboratory at the Faculty of Electrical Technology and Intelligent Informatics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, where the authors conducted their research. Simpang Lima Gumul Hospital in Kediri, East Java, also participated in the collaborative

initiative to collect p rimary research data on stroke victims.

REFERENCES

[1] Johanna M. Ospel, Leon Rinkel, Aravind Ganesh, Andrew Demchuk, Manraj Heran, Eric Sauvageau, Manish Joshi, Diogo Haussen, Mahesh Jayaraman, Shelagh Coutts, Amy Yu, Volker Puetz, Dana Iancu, Oh Young Bang, Jason Tarpley, Staffan Holmin, Michael Kelly, Michael Tymianski, Michael Hill, Mayank Goyal. How Do Quantitative Tissue Imaging Outcomes in Acute Ischemic Stroke Relate to Clinical Outcomes?. *Journal of Stroke*. 2024; 26(2): 252–259.

[2] Valery L. Feigin, Michael Brainin, Bo Norrving, Sheila Martins, Ralph L. Sacco, Werner Hacke, Marc Fisher, Jeyaraj Pandian, Patrice Lindsay. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *International Journal of Stroke.* 2022; 17(1): 18-29.

[3] N. K. Al-Qazzaz, A. A. Aldoori, S. H. B. M. Ali, S. A. Ahmad, A. K. Mohammed, M. I. Mohyee, EEG Signal Complexity Measurements to Enhance BCI-Based Stroke Patients' Rehabilitation. *Sensors.* 2023; 23(8): 1-24.

[4] P. Ofner, A. Schwarz, J. Pereira, D. Wyss, R. Wildburger, and G. R. Müller-Putz. Attempted Arm and Hand Movements can be Decoded from Low-Frequency EEG from Persons with Spinal Cord Injury. *Scientific Reports.* 2019; 9(1): 1-15.

[5] M. Soufineyestani, D. Dowling, A. Khan. Electroencephalography (EEG) technology applications and available devices. *Applied Sciences (Switzerland).* 2020; 10(21): 1-23.

[6] H. Setiawan, W. R. Islamiyah, A. D. Wibawa, M. H. Purnomo. *Identifying EEG Parameters to Monitor Stroke Rehabilitation using Individual Analysis*. International Seminar on Intelligent Technology and Its Application, ISITIA. 2019; 337–342.

[7] M. M. Hartanti, A. D. Wibawa, M. H. Purnomo. *Homecare and Hospital Stroke Therapy Comparison Using EEG Analysis*. International Seminar Wireless Technologies and Intelligent Systems for Better Human Lives, IES. 2021; 179–184.

[8] M. H. P. Rosita Devi Kusumastuti, Adhi Dharma Wibawa. *Stroke Severity Classification based on EEG Statistical Features*. International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS). 2021; 138–142.

[9] W.R.W. Omar, M.N. Taib, R. Jailani, N. Fuad, R.M. Isa, A.H. Jahidin, Z. Sharif. *Acute Ischemic Stroke Brainwave Classification Using Relative Power Ratio Cluster Analysis*. Procedia - Social and Behavioral Sciences. 2013; 546–552.

[10] J. Frey. *Comparison of an open-hardware electroencephalography amplifier with medical grade device in brain-computer interface applications*. Proceedings of the 3rd International Conference on Physiological Computing Systems, PhyCS. 2016;105–114.

[11] C. Brambilla, I. Pirovano, R. M. Mira, G. Rizzo, A. Scano, A. Mastropietro. Combined use of emg and eeg techniques for neuromotor assessment in rehabilitative applications: A systematic review. *Sensors.* 2021; 21(21): 1–25.

[12] M. Y. T. Sulistyono, D. Ernawati, W. S. Sari, S. Hadiati Nugraini. *Artifact-EOG Denoising Using FIR-Filtering in EEG Channel Selection for Monitoring and Rehabilitation of Stroke Patients*. International Seminar on Application for Technology of Information and Communication, iSemantic. 2022; 82–88.

[13] Mita Amara, Esmeralda C Djamal, Asri Maspupah. *Identifikasi Sinyal EEG dari Pasien Pasca-Stroke Menggunakan Backpropagation dan Algoritma Genetika Daswara Djajasasmita*. Seminar Nasional Aplikasi Teknologi Informasi (SNATi). 2019; 1907–5022.

[14] A. F. Nurfirdausi, R. A. Apsari, S. K. Wijaya, P. Prajitno, N. Ibrahim. Wavelet Decomposition and Feedforward Neural Network for Classification of Acute Ischemic Stroke based on Electroencephalography. *International Journal of Technology*, 2022; 13(8): 1745–1754.

[15] N. R. Sims, H. Muyderman. Mitochondria, oxidative metabolism and cell death in stroke. *Biochimica et Biophysica Acta - Molecular Basis of Disease*. 2010; 1802(1): 80–91.

[16] A. W. de Weerd, R. J. Veldhuizen, M. M. Veering, D. C. J. Poortvliet, E. J. Jonkman. Recovery from cerebral ischaemia. EEG, cerebral blood flow and clinical symptomatology in the first three years after a stroke. *Electroencephalography and Clinical Neurophysiology*. 1998; 70(3): 197–204.

[17] J. J. M. F. Van Der Putten, J. C. Hobart, J. A. Freeman, A. J. Thompson.

Measuring change in disability after inpatient rehabilitation: Comparison of the responsiveness of the Barthel Index and the Functional Independence Measure. *Journal of Neurology Neurosurgery and Psychiatry*, 1999; 66(4): 480–484.

[18] H. S. Jørgensen, H. Nakayama, H. O. Raaschou, J. Vive-Larsen, M. Støier, T. S. Olsen. Outcome and time course of recovery in stroke. Part I: Outcome. The Copenhagen stroke study. *Archives of Physical Medicine and Rehabilitation*. 1995; 76(5): 399–405.

[19] H. F. Posada-Quintero, N. Reljin, J. B. Bolkhovsky, A. D. Orjuela-Cañón, K. H. Chon. Brain Activity Correlates With Cognitive Performance Deterioration During Sleep Deprivation. *Frontiers in Neuroscience.* 2019; 13(1001): 1–9.