Jurnal
Teknologi

Full Paper

COMPARISON OF CORONA DISCHARGE IDENTIFICATION IN 20 kV CUBICLES BASED ON VOLTAGE AND NOISE USING ED, HMM, AND FCM

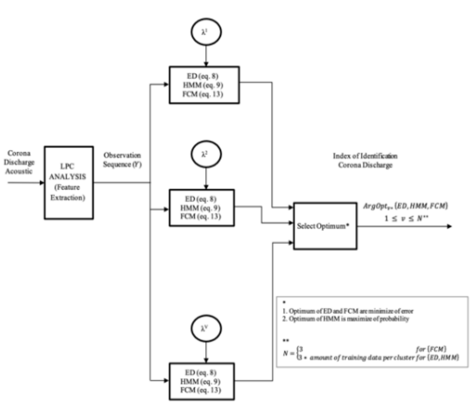
Miftahul Fikri[a,b], Christiono[b], Iwa Garniwa Mulyana K.[b,c], Zulkurnain Abdul-Malek[a]*, Muhammad Luthfiansyah Romadhoni[d], Mona Riza Mohd Esa[a], Eko Supriyanto[a]

[a]High Voltage and High Current Institute, Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia
[b]Faculty of Electrical and Renewable Energy, Institut Teknologi PLN, Jakarta Indonesia
[c]Department of Electrical Engineering, Universitas Indonesia, Depok, Indonesia
[d]UIP3B Kalimantan, PT PLN (Persero), Kalimantan Indonesia

Article history
Received
9 March 2023
Received in revised form
27 January 2024
Accepted
22 February 2024
Published Online
20 August 2024

*Corresponding author
miftahul@itpln.ac.id

Graphical abstract



Gambar 2 Flowchart Identifikasi Suara Corona Discharge Berdasarkan Voltage dan Noise

Abstract

Phenomena such as corona discharge (CD) still occurs in many electrical systems in Indonesia. As a first step for early detection of insulation failure. Identification of CD acoustic in this study namely clustering based on voltage and based on noise. So that the CD acoustic classification is set into 3 clusters. In addition, this study also classifies CD acoustic based on noise with three clusters, namely pure CD, CD with noise, and pure noise. Clustering was performed using the linear predictive coding (LPC) method as feature extraction, then a comparison of pattern matching results of feature extraction was performed using Euclidean distance (ED), hidden Markov model (HMM) and fuzzy cluster mean (FCM). The temperature in the cubical is between 27.5 ℃ - 35.3 ℃ and humidity ranges from 70% - 95%. The results of clustering accuracy on the average base voltage using the ED, HMM and FCM methods were obtained respectively 100%, 100% 93.93% for training data and 80.74%, 84.44%, 80.55% for testing data. While the results of the average base noise clustering accuracy using the ED, HMM and FCM methods were obtained respectively 100%, 100%, 94.69% for training data and 100%, 100%, 100% for testing data.

Keywords: Corona discharge, linear predictive coding, Euclidean distance, hidden Markov model, fuzzy c-means

Abstrak

Fenomena corona discharge (CD) masih kerap berlaku dalam sistem elektrik di Indonesia. Sebagai langkah pertama dalam pengesanan awal kegagalan penebat. Pengenalpastian akustik CD dalam penyelidikan ini adalah pengelompokan berdasarkan voltan dan berdasarkan hingar. Jadi klasifikasi akustik CD disusun kepada 3 kelompok. Selain itu, kajian ini juga mengklasifikasikan CD akustik berdasarkan hingar dengan tiga kelompok iaitu CD tulen, CD bunyi bising, dan hingar tulen. Pengekstrakan dilakukan menggunakan kaedah pengekodan ramalan linear (LPC) sebagai pengekstrakan ciri, kemudian perbandingan padanan pola hasil pengekstrakan ciri dijalankan menggunakan jarak Euclidean (ED), model Hidden Markov (HMM) dan min kelompok kabur (FCM). . Suhu di dalam kiub adalah antara 27.5 ℃ - 35.3 ℃ dan kelembapan antara 70% - 95%. Keputusan ketepatan pengelompokan pada purata tegasan asas menggunakan kaedah ED, HMM dan FCM diperolehi masing-masing 100%, 100% 93.93% untuk data latihan dan 80.74%, 84.44%, 80.55% untuk data ujian. Manakala, keputusan purata ketepatan pengelompokan hingar asas menggunakan kaedah ED, HMM dan FCM diperolehi masing-masing pada 100%, 100%, 94.69% untuk data latihan dan 100%, 100%, 100% untuk data ujian.

Kata kunci: Corona discharge, linear predictive coding, Euclidean distance, model hidden Markov, fuzzy c-means.

1.0 INTRODUCTION

The total installed capacity and number of PLN generating units in Indonesia reached 44,174.79 MW, of which 30,970.93 MW (64.24%) were in Java. The installed capacity and number of distribution substation transformers is 61,556 MVA and 527,544 units [1]. Based on such a large capacity, it requires nearly three cubicles more than the number of installed distribution transformer units [2]. As an illustration, cubicle damage has occurred due to the corona discharge of more than 300 units yearly at the Distribution Substation of PT. PLN (Persero) Distribution Jaya and Tangerang. This problem can be overcome by immediately carrying out repairs, but information about the condition of the cubicle can only be known by visiting the location of the cubicle [3]. Figure 1 is an example of 20 kV cubicle damage due to CD.

Figure 1 Cubicle Condition Due to Corona Discharge

Based on the many cases of damage that occurred, an experimental simulation was carried out. In this experiment, two electrodes were arranged, and conditions that could trigger corona discharger symptoms were manipulated [4]. This can happen because the electrode which is affected by the electric field will result in an ionization process that can cause positive ions or electrons in the medium voltage cubicle insulation of 20 kV. These positive ions or electrons will gather at one point in the form of a flood of electrons, which can then cause acoustic emission or become a source of acoustic. If the flood of electrons at this point increases, the greater the acoustic generated by the electrodes [1]. So, the parameter for measuring the symptoms of corona discharge can be measured by the size of the acoustic level. In this study, identification of the acoustic of the corona discharge was carried out based on the voltage level and the influence of noise from the surrounding environment [5].

Most of the identification of corona discharge is carried out in a laboratory environment and a noise-free environment. Research [6] discusses the detection of corona discharge acoustic using a microphone on a laboratory scale but cannot distinguish noise automatically. Then, in research [7], acoustic emission waves are used to avoid noise. The weakness of this acoustic emission method lies in the placement of the sensor in the high-voltage cable [8], and the sensor cannot work or measure perfectly because the induction is too high in the high-voltage cable. Meanwhile, the microphone sensor is a sensor for detecting acoustic waves in the area being detected or measured [9]. However, in reality, the corona discharge measurements in the field experience lower detection sensitivity due to external noise interference or noise [10]. So, in this study, other than stress-based tests were also carried out, observations were also carried out close to field conditions, namely the presence of noise from cubicle equipment [11].

Processing of measurement data based on categories on measured decibel values so that it can be read how severe the insulation failure is [12]. The linear predictive coding (LPC) method is a method that has long been known and is reliable for obtaining feature extraction from any acoustic, so that the recorded corona discharge acoustic is then calculated for feature extraction using the LPC method. After this feature extraction is obtained, pattern matching is carried out using the Euclidean distance (ED) method, hidden Markov models (HMM), and fuzzy c-means (FCM). The ED method is used because of its simplicity which is only based on calculating the distance using the Pythagorean theorem [13], [14]. The FCM method is used because it can conclude vague problems, is based on fuzzy measures, and is an unsupervised learning category [15], [16], [17], [18]. Meanwhile, HMM is chance-based and has been known for quite a long time for speech recognition [19], [20], [21], [22]. These three methods are used because they have different characteristics and calculation bases, so it is hoped that the best method will be obtained, as well as being unique and contributing to this research.

Based on the processing of the data obtained, it is expected that a corona discharge monitoring system will be formed in cubicles in real time. The up-to-date conditions that are monitored can be used by technicians in order to prevent breakdowns due to corona discharge [23]. As a first step, the focus of this research is to compare the identification of corona discharge acoustic based on voltage and noise using ED, HMM, and FCM. In order to obtain a more accurate method for identifying CD acoustic based on voltage and noise contamination from the noise environment.

Section 2 discusses the methods used to identify (clustering) corona discharge sounds, namely the LPC, ED, HMM, and FCM methods. Section 3 discusses the results obtained starting from prototype and design research, data and extraction, learning process and results obtained. Section 4 contains conclusions about the research conducted.

2.0 METHODOLOGY

In observing the corona discharge acoustic, this study refers to the IEEE Standard for High-Voltage Testing Techniques 4 2013-1995 IEEE [24] and IEC TS 62478:2016 High voltage test techniques - Measurement of partial discharges by electromagnetic and acoustic methods [25] even though there is no PD detector available to collect the data. As for the gap spacing, the vertical testing needle-rod electrodes are at 3 cm, and the microphone is 5 cm from the gap. The standard bandwidth is 40 – 200 kHz, but this research takes a bandwidth of 40 kHz.

The corona discharge acoustic phenomenon was observed based on voltage and noise, with each classification being carried out ten times. Based on preliminary observations, the most miniature breakdown was obtained at 34.3 kV, so base voltage observations were carried out with voltage classifications of 20–24 kV, 25-29 kV, and 30-33 kV, while based on noise observations were carried out classification with pure CD, CD with noise and noise pure. Room temperature ranges from 27.5°C to 35.3°C, and humidity ranges from 70% to 95%.

The data was collected at the High Voltage Technology and Equipment Laboratory of the PLN Institute of Technology (TPTT ITPLN Lab) for one year (February 2021 to February 2022). The data obtained was then processed using the MATLAB version 2016b program. The clustering flowchart of the corona discharge phenomenon based on acoustic has a scheme that can be seen in Figure 2.

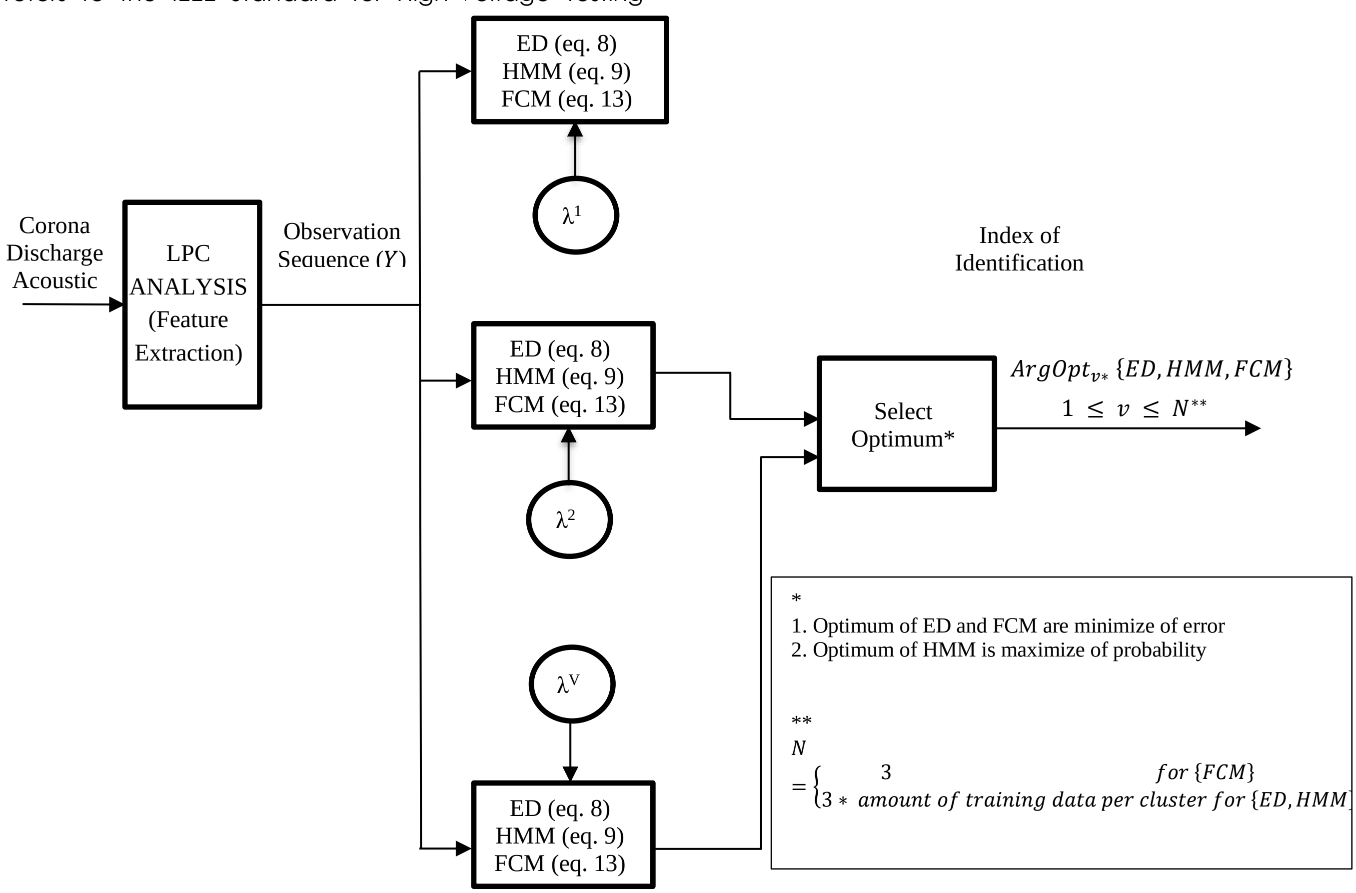


Figure 2 Flowchart of Corona Discharge Acoustic Identification Based on Voltage and Noise

2.1 Linear Predictive Coding

The LPC method is a method that has been known for a long time for speech recognition. According to [26], [27] the advantage of LPC is that it is simple and can be applied to various devices. There are two main components in LPC, namely Encoding and Decoding. Encoding functions examine and decompose the voice signal into several parts while decoding functions change the acoustic that has been recorded. The process of the LPC method is shown in Figure 3 [26].

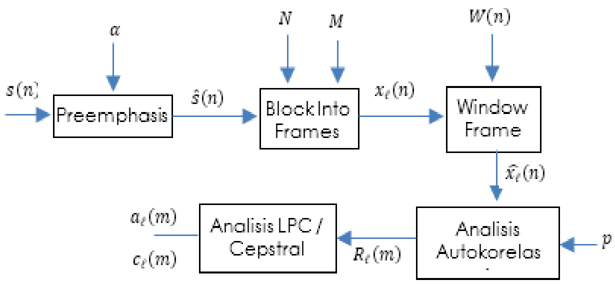


Figure 3 LPC Process as Feature Extraction of Corona Discharge Acoustic [13], [14], [15], [26], [27]

with

$$\tilde{s}(n) = s(n) - \tilde{a}\, s(n-1), \tag{1}$$

where s is data/ acoustic signal,
n is the number of data in the acoustic signal

$$0.9 \leq a \leq 1.$$

$$x_\ell(n) = \tilde{s}(M\ell + n) \tag{2}$$

where M denotes the number of frames, $\ell = 1,2,\dots,M$.

$$\tilde{x}_\ell(n) = x_\ell(n)\, w(n), \tag{3}$$

where $w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), 0 \leq n \leq N-1$

$$r_\ell(m) = \sum_{n=0}^{N-1-m} \tilde{x}_\ell(n)\, \tilde{x}_\ell(n+m) \tag{4}$$

where $m = 0,1,\dots,p$.

$$a_m = \alpha_m^{(p)} \tag{5}$$

with $1 \leq m \leq p$

$$E^{(0)} = r(0)$$

$$k_i = \frac{\left\{r(i) - \sum_{j=1}^{L-1} \alpha_j^{(i-1)} r(|i-j|)\right\}}{E^{(i-1)}}$$

$$\alpha_i^{(i)} = k_i$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)}$$

$$E^{(i)} = (1 - k_i^2)\, E^{(i-1)}$$

$$c_0 = \ln 1 \tag{6}$$

$$y_m = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \tag{7}$$

where $1 \leq m \leq p$.

The stages in Figure 2 above can be explained as follows [26], [27].

1. Pre-emphasis is a method that can remove acoustic noise from the original auditory input. Noise is a disturbance or interference from the original auditory object that occurs around it.
2. Frame blocking is the process of dividing acoustic signals into the form of acoustic segments.
3. Windowing is detecting and sorting out acoustic signal differences divided into several segments.
4. Autocorrelation analysis is a step to equalize the signal from each frame.
5. LPC analysis this is the part that converts the magnitude of the autocorrelation analysis results into the magnitude of the LPC parameter (logarithmic area ratio coefficient, reflection coefficient, and LPC coefficient).
6. Cepstral analysis converts LPC parameters into cepstral coefficients using the Fourier transform.

2.2 Euclidean Distance

Euclidean distance is used to measure the level of similarity of the feature extraction results on the data to be tested (all data) with all training data concerning the Pythagorean formula as follows:

$$Ed_{ij} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} (W_i - Y_j)^2} \tag{8}$$

where Ed_{ij} : Euclidean distance
W_i : Data training (LPC output)
Y_j : Data testing (LPC output)

The greater the value obtained, the smaller the degree of similarity between the training and test data. Conversely, the smaller the value received, the greater the level of similarity between the training data and the test data [13], [14], [28].

2.3 Hidden Markov Model

The standard hidden Markov model (NHMM) is a secret Markov model $\{X_t, Y_t\}_{t\in\mathbb{N}}$ that assumes the probability Y_t of observing X_t provided that it is known to have a normal distribution. $\{X_t\}_{t\in\mathbb{N}}$ The event's cause is deemed not to be followed and forms a Markov chain with state space $S_X = \{1,2,\cdots,m\}$. The main problem with the hidden Markov model is maximizing the likelihood function of the Y observation process which is defined as follows [26], [27], [19], [20].

$$L_T(\phi) = P(Y_1 = y_1, Y_2 = y_2, \dots, Y_T = y_T|\phi)$$

$$= \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} \left(\pi_{y_1 i_1} \pi_{y_2 i_2} \cdots \pi_{y_T i_T}\right) \times \left(\delta_{i_1} \gamma_{i_1 i_2} \gamma_{i_2 i_3} \cdots \gamma_{i_{T-1} i_T}\right)$$

$$= \sum_{i_1=1}^{m} \cdots \sum_{i_T=1}^{m} \delta_{i_1} \pi_{y_1 i_1} \prod_{t=2}^{T} \gamma_{i_{t-1} i_t} \pi_{y_t i_t}. \tag{9}$$

In this research, $y_1, y_2, . . . , y_T$ are a feature extraction of the corona discharge acoustic using the linear predictive coding (LPC) method. In order to maximize the likelihood function, parameter estimation is carried out, namely the mean, variance and transition matrix from the normal hidden Markov model. Optimization of the likelihood function in equation (9) is carried out using the Expectation Maximization (EM algorithm) iteratively, with each formulation as follows [20], [22].

$$\mu_i^{(k+1)} = \frac{\sum_{t=1}^{T} \alpha_t^{(k)}(i) \beta_t^{(k)}(i) y_t}{\sum_{t=1}^{T} \alpha_t^{(k)}(i) \beta_t^{(k)}(i)}, \tag{10}$$

$$\sigma_i^{2(k+1)} = \frac{\sum_{t=1}^{T} \alpha_t^{(k)}(i) \beta_t^{(k)}(i) \left(y_t - \mu_i^{(k+1)}\right)^2}{\sum_{t=1}^{T} \alpha_t^{(k)}(i) \beta_t^{(k)}(i)} \tag{11}$$

$$\gamma_{ij}^{(k+1)} = \frac{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i) \gamma_{ij}^{(k)} P^{(k)}(y_{t+1}|j) \beta_t^{(k)}(j)}{\sum_{t=1}^{T-1} \alpha_t^{(k)}(i) \beta_t^{(k)}(i)}. \tag{12}$$

with $\alpha_t(i) = P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_t = y_t, X_t = i|\phi)$ and $\beta_t(i|\phi) = P(Y_{t+1} = y_{t+1}, \ldots, Y_T = y_T | X_t = i, \phi)$ can be calculated recursively using forward and backward algorithms respectively. These results are then used for clustering the corona discharge phenomenon both for voltage and noise by taking into calculate the maximum likelihood.

2.4 Fuzzy c-Means

Fuzzy c-Means (FcM) is a method used for clustering data based on the parameters of the mean and degree of membership with the principle of fuzzy logic. The main problem of the FcM method is to minimize the error of the FCM Objective Function as follows [29], [30],[31], [32].

$$J_m(U, v) = \sum_{k=1}^{n} \sum_{i=1}^{c} (U_{ik})^m \|x_k - v_i\|^2 \tag{13}$$

The results of feature extraction of the corona discharge data $(y_1, y_2, . . . , y_T)$ use the linear predictive coding (LPC) method, then input them to the FcM method to minimize the error which in line with estimating the FCM parameters, namely the mean and degree of membership. Error minimization in equation (13) is carried out by iteration using the following formula [29], [33], [34].

$$(v_i^*) = \frac{\sum_{k=1}^{n} (u_{ik}^*)^m \, y_k}{\sum_{k=1}^{n} (u_{ik}^*)^m} \tag{14}$$

$$u_{ik}^* = \left(\sum_{j=1}^{c} a_{ijk}^*\right)^{-1} \tag{15}$$

With v cluster centers and u degrees of membership. These results are then used for clustering the corona discharge phenomenon both for voltage and noise by paying attention to the smallest error [35].

3.0 RESULT AND DISCUSION

3.1 Prototype and Design Research

The prototype and design that was carried out can be seen in Figure 4 and Figure 5 below.

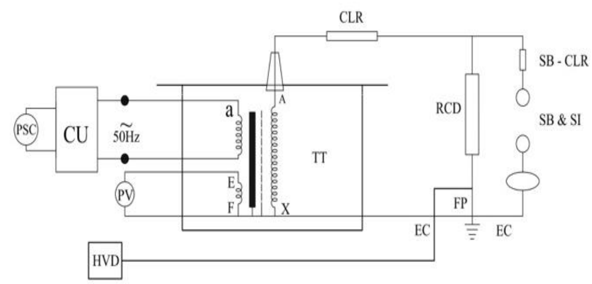


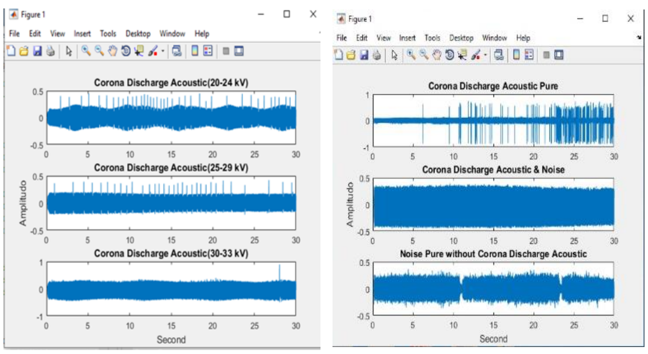
Figure 4 Prototype of Corona Discharge Acoustic Recording in Cubicle 20 kV

Figure 5 Research design of Corona Discharge Acoustic Recording in Cubicle 20 kV

Figure 4 is the prototype in this study, Figure 5 is the design of the corona discharge acoustic recording which was carried out at the Laboratorium High Voltage Institut Teknologi PLN which is generated by a test transformer with a 200 V scale ratio that can be step up to 100 kV. Based on the results of observations made as a preliminary analysis of the study, the breakdown voltage was 34.3 kV. So that in this study the recording of the corona discharge acoustic was carried out starting from a voltage of 20 kV and then increasing it gradually so that three corona discharge acoustic clusters were formed, namely 20-24 kV, 25-29 kV and 30-33 kV. In addition, noise-based CD acoustic data was also collected with three categories, namely pure CD, CD with noise and pure noise [36]. The observed temperature conditions in the cubicle during data collection ranged from 27.5 ℃ to 35.3 ℃ and humidity ranged from 70% to 95%. During the period the increase in voltage is accompanied by a process of recording the acoustic of the corona discharge that occurs between the two electrodes with a distance between the electrodes of 3 cm [37]. As for acoustic recording using a microphone device that is 5 cm from the sidelines of the electrodes, then the results of the recording are saved to a computer for processing [38].

3.2 Data and Extraction of Corona Discharge Acoustic

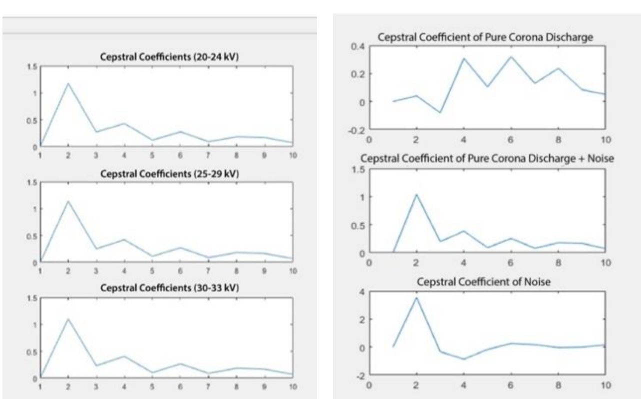
The research design in Figure 5 is used to record the corona discharge acoustic based on voltage with each cluster 20-24 kV, 25-29 kV, 30-33 kV and based on noise with each cluster pure CD, CD with noise and pure noise which can be seen in Figure 6 [39].



(a) (b)

Figure 6 Corona Discharge Acoustic (a) Based on Voltage (b) Based on Noise

The acoustic data in Figures 6(a) and 6(b) are then calculated using equations (1) – (5) so that LPC parameters are obtained, then for the feature extraction obtained to be more robust it is analyzed using the Fourier transform [15] using Equation (6) – (7) and produce LPC cepstral coefficients in Figures 7(a) and 7(b) respectively.



(a) (b)

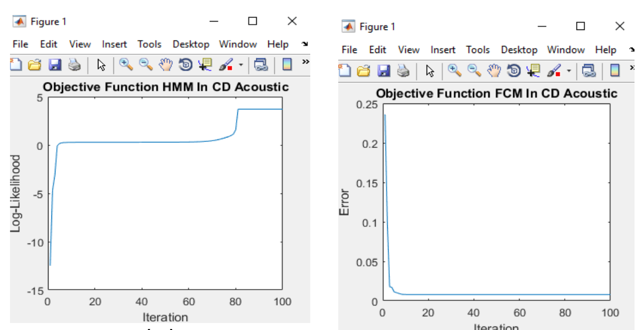
Figure 7 LPC Cepstral Coefficient of Corona Discharge Acoustic (a) Based on Voltage (b) Based on Noise

Figure 7(a) is similar because there is no outside interference (noise), so it is possible to obtain only moderately accurate results [40]. Meanwhile, Figure 7(b) differs significantly due to noise effects, so it is expected to provide exact clustering results. Feature extraction in cepstral LPC is then used as input for the Euclidean distance, HMM, and FCM methods for identifying/clustering corona discharge acoustic.

3.3 Learning Process and Results of Corona Discharge Acoustic Identification

The feature extraction results obtained for both based on voltage and based on noise using the LPC method in Figures 7(a) and 7(b) are then used as input to identify both, i.e., training data compared to training data and training data correspond to testing data using Euclidean distance, HMM, and FCM.

The results of identifying CD acoustics based on voltage and noise using ED are only calculated by calculating the shortest distance between CD data using equation (8) so that the analysis process is carried out without any learning process. Meanwhile, with the HMM and FCM methods, identification is carried out by a learning process: looping to optimize the objective function. The HMM method is carried out by maximizing the likelihood function in equation (9) and FCM by minimizing the difference in distance in equation (13). The formula for looping used for HMM is equation (10)-(12), and FCM uses equation (14)-(15). The results obtained during the learning process for both HMM and FCM can be seen in Figure 8 below.



(a) (b)

Figure 8 Learning Process of Corona Discharge Acoustic Identification (a) HMM (b) FCM

Figure 8 shows that the learning processes for both HMM and FCM in acoustic CD have optimized their respective objective functions, this indicates that the method used for CD acoustic identification has been validated. The results of the CD base learning process acoustic identification above, based on voltage, can be seen in Figure 9, where 9(a) uses the Euclidean distance method, 9(b) uses the hidden Markov model method, 9(c) uses the Fuzzy c-mean method. The identification based on noise can be seen in Figure 10, where 10(a) uses the Euclidean distance method, 10(b) uses the hidden Markov model method, and 10(c) uses the Fuzzy c-mean method. This process is repeated with the treatment from 1 data training - 9 data testing up to 9 data training - 1 data testing in each corona discharge acoustic treatment cluster.

Identification of CD acoustic based on voltage using ED can be seen in Figure 9(a) where the rows correspond to all data, both training data and testing data, columns correspond to training data where columns 1-5 are training data for a voltage of 20-24 kV (cluster 1), columns 6-10 are training data for a voltage of 25-29 kV (cluster 2), columns 11-15 are training data for a voltage of 30-33 kV (cluster 3). Clustering results are obtained based on the smallest Euclidean distance value in Figure 9(a). So that the accuracy obtained can be seen in Table 1.

Table 1 Accuracy of CD acoustic Identification Based on Voltage Using ED

No.	Training		Testing	
	Amount of Data	Accuracy (%)	Amount of Data	Accuracy (%)
1	1	100	9	70.37
2	2	100	8	83.33
3	3	100	7	80.95
4	4	100	6	77.78
5	5	100	5	73.33
6	6	100	4	91.67
7	7	100	3	88,89
8	8	100	2	100
9	9	100	1	100
Average	**100 %**		**80.74 %**	

Table 1 shows the identification of CD acoustic based on voltage using ED with an average accuracy of 100% for training and 80.74% for testing.

Identification of CD acoustic based on voltage using HMM can be seen in Figure 9(b) where the rows correspond to all data, both training data and testing data, columns correspond to training data where columns 1-5 are training data for a voltage of 20-24 kV (cluster 1), columns 6-10 are coaching data for a voltage of 25-29 kV (cluster 2), columns 11-15 are training data for a voltage of 30-33 kV (cluster 3). Clustering results are obtained based on the most significant likelihood value in Figure 9(b). So that the accuracy obtained can be seen in Table 2.

Table 2 Accuracy of CD acoustic Identification Based on Voltage Using HMM

No.	Training		Testing	
	Amount of Data	Accuracy (%)	Amount of Data	Accuracy (%)
1	1	100	9	88.89
2	2	100	8	83.33
3	3	100	7	80.95
4	4	100	6	77.78
5	5	100	5	73.33
6	6	100	4	91.67
7	7	100	3	88,89
8	8	100	2	100
9	9	100	1	100
Average	**100 %**		**84.44 %**	

Table 2 shows the identification of the CD acoustic based on voltage using HMM, which obtained an average accuracy of 100% for training and 84.44% for testing.

The identification of CD acoustics based on voltage using FCM can be seen in Figure 9(c), where the rows correspond to all training and testing data, and columns 1-3 correspond to clusters 1-3. Clustering results are obtained based on the value of the smallest distance between the data and the cluster center formed by FCM, which can be seen in Figure 9(c). So that the accuracy obtained can be seen in Table 3.

Table 3 Accuracy of CD acoustic identification Based on Voltage Using FCM

No.	Training		Testing	
	Amount of Data	Accuracy(%)	Amount of Data	Accuracy (%)
1	1	-	9	-
2	2	83.33	8	83.33
3	3	100	7	80.95
4	4	100	6	77.78
5	5	100	5	73.33
6	6	94.44	4	75
7	7	90.47	3	100
8	8	83.33	2	66.67
9	9	100	1	100
Average	**93.93**		**80.55**	

Table 3 shows the identification of the CD acoustic based on voltage using FCM, which obtained an average accuracy of 93.93% for training and 80.55% for testing.

The identification of CD acoustics based on noise using ED can be seen in Figure 10(a), where the readings are similar to Figure 9(a). So that the accuracy obtained can be seen in Table 4.

Table 4 Accuracy of CD acoustic Identification Based on Noise Using ED

No.	Training		Testing	
	Amount of Data	Accuracy (%)	Amount of Data	Accuracy (%)
1	1	100	9	100
2	2	100	8	100
3	3	100	7	100
4	4	100	6	100
5	5	100	5	100
6	6	100	4	100
7	7	100	3	100
8	8	100	2	100
9	9	100	1	100
Average	**100 %**		**100 %**	

Table 4 shows the identification of CD acoustic based on noise using ED with an average accuracy of 100% for training and 100% for testing.

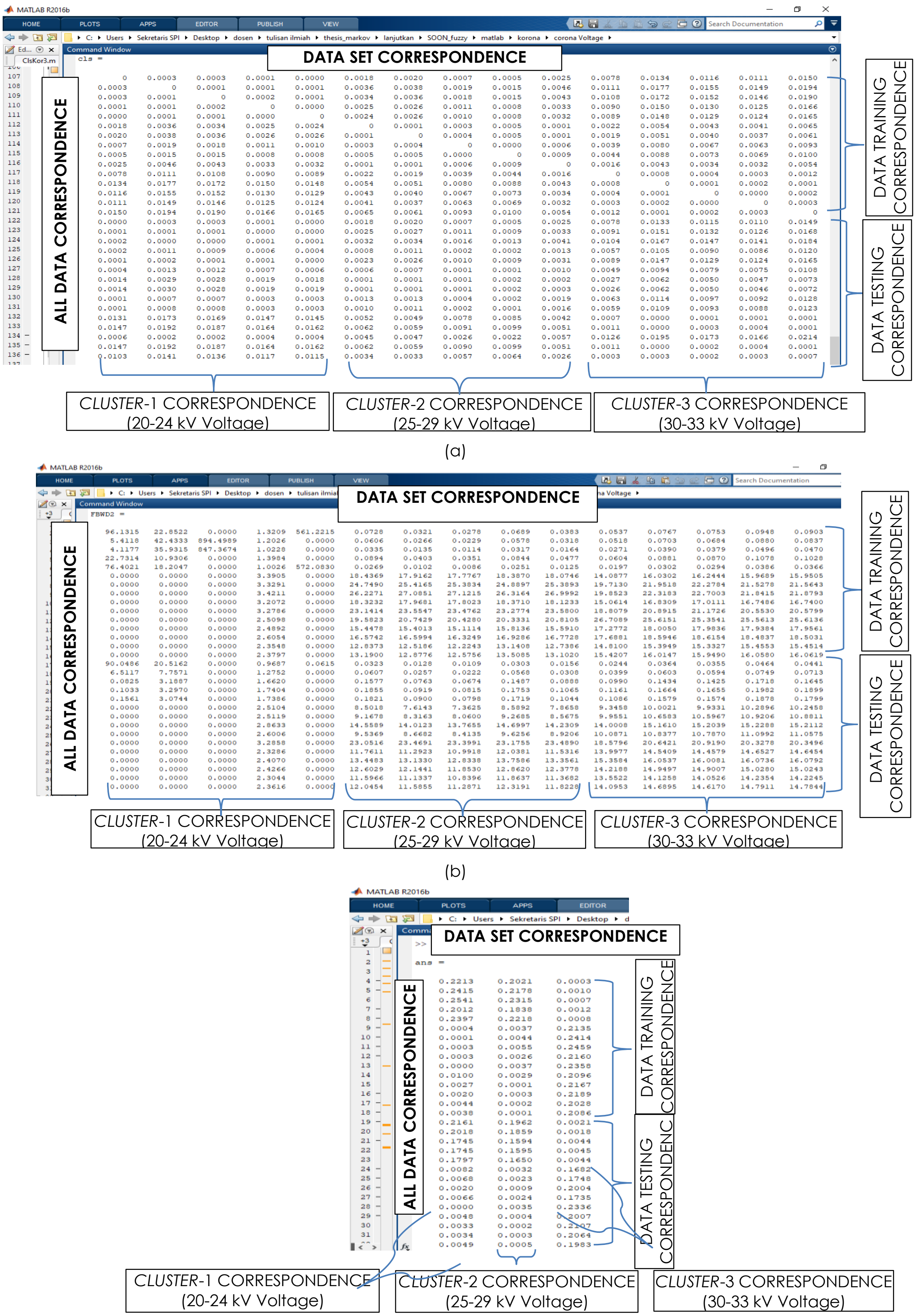


Figure 9 Identification of CD Acoustic Based on Voltage using (a) Euclidean Distance (b) hidden Markov models (c) Fuzzy c-Mean

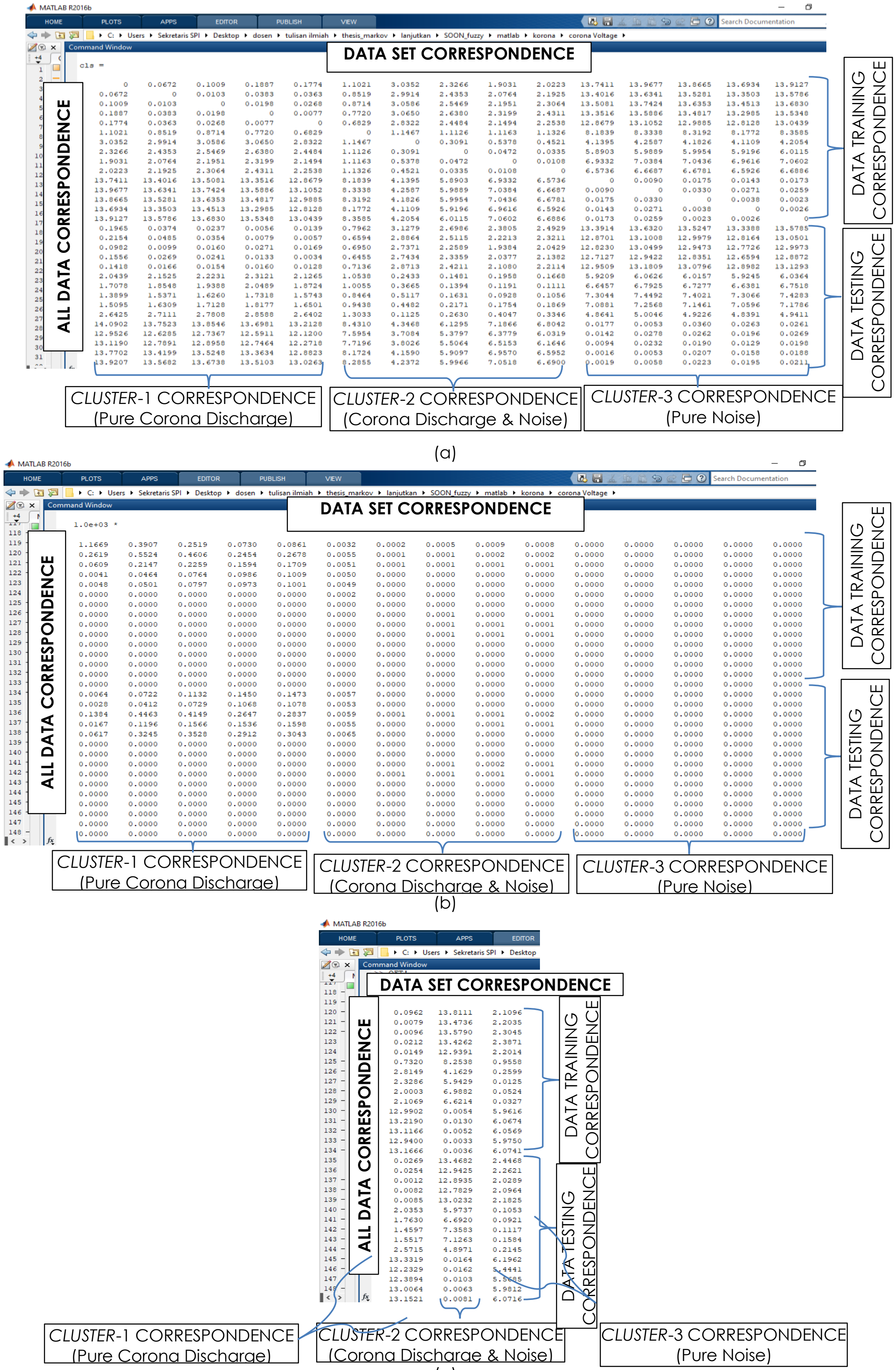


Figure 10 Identification of CD Acoustic Based on Noise using (a) Euclidean Distance (b) hidden Markov models (c) Fuzzy c-Mean

Identification of CD acoustic based on noise using HMM can be seen in Figure 10(b) where the readings are similar to Figure 9(b). So that the accuracy obtained can be seen in Table 5.

Table 5 Accuracy of CD acoustic Identification Based on Noise Using HMM

No.	Training		Testing	
	Amount of Data	Accuracy (%)	Amount of Data	Accuracy (%)
1	1	100	9	100
2	2	100	8	100
3	3	100	7	100
4	4	100	6	100
5	5	100	5	100
6	6	100	4	100
7	7	100	3	100
8	8	100	2	100
9	9	100	1	100
Average	**100 %**		**100 %**	

Table 5 shows the CD acoustic identification based on noise using HMM, which obtained an average accuracy of 100% for training and 100% for testing.

Identification of CD acoustic based on noise using FCM can be seen in Figure 10(c) where the readings are similar to Figure 9(c). So that the accuracy obtained can be seen in Table 6.

Tabel 6 Accuracy of CD acoustic Identification Based on Noise Using FCM

No.	Training		Testing	
	Amount of Data	Accuracy (%)	Amount of Data	Accuracy (%)
1	1	-	9	-
2	2	100	8	100
3	3	88.89	7	100
4	4	91.67	6	100
5	5	93.33	5	100
6	6	94.44	4	100
7	7	95.24	3	100
8	8	95.83	2	100
9	9	96.29	1	100
Average	**94.69**		**100**	

Table 6 shows the identification of CD acoustics based on noise using FCM, which obtained an average accuracy of 94.69% for training and 100% for testing. The average training accuracy is smaller than the intermediate testing because the training data is more spread out than the testing data. Furthermore, in Table 6, when the training data is 1 data, the training process cannot be carried out according to [32] as stated in the formula in equations (13)-(15).

The results of CD acoustic identification based on voltage with each method can be seen in Table 1-3. The CD acoustic identification based on noise with each technique can be seen in Table 4-6. The results are summarized in Table 7.

Table 7 Recapitulation of CD acoustic identification Based on Voltage and Noise Using ED, HMM, FCM

No.	Method	Accuracy Mean (%)			
		Base on Voltage		Base on Noise	
		Training	Testing	Training	Testing
1	ED	100	80.74	100	100
2	HMM	100	84.44	100	100
3	FCM	93.93	80.55	94.69	100

Based on Table 7 for CD acoustic identification based on voltage, HMM obtained the best average accuracy of 100% for training and 84.44% for testing, followed by Euclidean with 100% accuracy for training and 80.74% for testing, and FCM with 93.93% accuracy for training and 80.55% for testing. While identifying CD base noise, HMM and ED obtained the best average accuracy of 100% for training and testing. Finally, FCM had an accuracy of 94.69% for training and 100% for testing. FCM has the least accuracy because this method is included in the unsupervised learning category, which has the advantage when the cluster category indicators are unknown. So, when the cluster indicators are unknown, the FCM method will still be reliable, while the HMM and ED methods cannot be used.

4.0 CONCLUSION

In this research, the corona discharge (CD) acoustic recording was carried out directly at the High Voltage Equipment and Technology Laboratory of the PLN Institute of Technology. The recording of the CD acoustic started with a voltage of 20 kV and then increased gradually before breakdown voltage occurs 34.3 kV so that three clusters of CD acoustic were formed, namely 20-24 kV, 25-29 kV and 30-33 kV. In addition, noise-based CD acoustic data was also collected with three categories, namely pure CD, CD with noise and pure noise. The observed temperature conditions in the cubicle during data collection ranged from 27.5 ℃ to 35.3 ℃ and humidity ranged from 70% to 95%. During the voltage regulation period, it is accompanied by a recording process of the CD acoustic that occurs between the two electrodes with a distance between the electrodes of 3 cm. Recording is done using a microphone device that is 5 cm from the sidelines of the electrodes. The results of the recording are then processed using the LPC, Euclidean distance, HMM, and FCM methods. The results of clustering accuracy on the average base voltage using the ED, HMM and FCM methods were obtained respectively 100%, 100% 93.93% for training data and 80.74%, 84.44%, 80.55% for testing data. While the results of the average base noise clustering accuracy using the ED, HMM and FCM methods were obtained respectively 100%, 100%, 94.69% for training data and 100%, 100%, 100% for testing data. Identification of the CD acoustic based on voltage acoustic obtained the highest average accuracy using the HMM method, namely 100% for training data and 84.4% for testing data. Meanwhile, identification of the CD acoustic based on noise was obtained using the HMM and ED methods, namely 100% for both training and testing data.

Acknowledgement

The first author would like to thank Institut Teknologi PLN for the financial support for his postgraduate research study at Universiti Teknologi Malaysia and the authors would like to thank Prof. Zulkurnain Abdul-Malek at Universiti Teknologi Malaysia for Supervision and Prof. Iwa Garniwa at Universitas Indonesia for the initial research idea. Their expertise and insights were invaluable in shaping our research and helping us to overcome challenges.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

References

[1] I. Jonan. 2018. Rencana Usaha Penyediaan Tenaga Listrik 2018-2028. 2018th edition. Jakarta: Menteri Energi dan Sumber Daya Mineral.
[2] I. E. Portugues, dkk. 2009. RF-based Partial Discharge Early Warning System for Air-insulated Substations. *IEEE Trans. Power Deliv*. 24(1): 20-29.
[3] S. Rusdi. 2019. Analisa Terbentuknya Korona Pada Saluran Kubicle Tegangan 20Kv Serta Pengaruhnya Terhadap Rugi-Rugi Daya. *Lensa*. 2(48): 14-21.
[4] Alsumaidaee, Y., Yaw, C., Koh, S., Tiong, S., Chen, C., Ali, K. 2022. Review of Medium-Voltage Switchgear Fault Detection in a Condition-Based Monitoring System by using Deep Learning. *Energies.* 15(18): 6762.
[5] X. Wang, N. Taylor, dan H. Edin. (2016). Effect of Humidity on Partial Discharge in a Metal-Dielectric Air Gap on Machine Insulation at Trapezoidal Testing Voltages. *J. Electrostat*. 83: 88-96.
[6] IY. A. M. Alsumaidaee, C. T. Yaw, S. P. Koh, S. K. Tiong, C. P. Chen, and K. Ali. 2022. Review of Medium-voltage Switchgear Fault Detection in a Condition-based Monitoring System by Using Deep Learning. *Energies (Basel)*. 15(18): 1-34.
[7] G. Ciaburro and G. Iannace. 2022. Machine-Learning-based Methods for Acoustic Emission Testing: A Review. *Applied Sciences.* 12(20): 10476.
[8] T. Zhang *et al.* 2016. A Fiber-Optic Sensor for Acoustic Emission Detection in a High Voltage Cable System. *Sensors.* 16(12).
[9] M. X. Zhu, dkk. 2017. Localization of Multiple Partial Discharge Sources in Air-insulated Substation using Probability-based Algorithm. *IEEE Trans. Dielectr. Electr. Insul.* 24(1): 157-166.
[10] P. Bas-Calopa, J.-R. Riba, and M. Moreno-Eguilaz. 2022. Measurement of Corona Discharges under Variable Geometry, Frequency and Pressure Environment. *Sensors.* 22(5): 1856.
[11] N. A. Al-geelani, M. A. M. Piah, dan Z. Abdul-Malek. 2010. Identification of Acoustic Signals of Corona Discharges under Different Contamination Levels using Wavelet Transform. *Electr. Eng*. 100(2): 1059-1067.
[12] N. Taylor. 2010. Dielectric Response And Partial Discharge Measurements on Stator Insulation at Varied Low Frequency.
[13] M. Fikri, C. and I. G. Mulyana K. 2022. Clustering Fenomena Corona Discharge Berdasarkan Suara menggunakan Metode LPC dan Euclidean Distance. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*. 10(4): 689-701.
[14] N. Pasra, M. Fikri, K. T. Mauriraya, T. Rijanto and I. G. P. A. Buditjahjanto. 2023. Deteksi Suara Corona Discharge berdasarkan. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*. 11: 72-85.
[15] M. Fikri *et al.* 2023. Clustering Suara Corona Discharge berdasarkan Tegangan menggunakan Metode Fuzzy C-Mean. *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*. 11(3): 609. Doi: 10.26760/elkomika.v11i3.609.
[16] M. Fikri, Z. Rustam, Kurniawan Atmadja, and Nurhadi Hadi. 2022. The Proof of Quantum and Fuzzy Measures as Generalization of Measure that Does Not Generalize Each Other. *J Med Sci.* 10: 548-555.
[17] M. Fikri *et al*. 2023. Optimization Objective Function Corona Discharge Acoustic Using Fuzzy c-Means (FcM). *ELKHA: Journal Teknik Elektro*. 15(2): 84.
[18] M. Fikri and Z. Abdul-Malek. 2023. Partial Discharge Diagnosis and Remaining Useful Lifetime in XLPE Extruded Power Cables under DC Voltage: A Review. *Electrical Engineering*. 105(6): 4195-4212. Doi: 10.1007/s00202-023-01935-y.
[19] M. Fikri, B. Setiyawaty and I. G. P. Purnaba. 2016. Pendugaan Parameter dan Kekonvergenan Penduga Parameter Model Poisson Hidden Markov. *MILANG Journal of Mathematics and Its Applications.* 15(1).
[20] M. Fikri, Samsurizal, Christiono and K. Tresya. 2020. Pemodelan Cuaca Menggunakan Model Hidden Markov Untuk Pemanfaatan Energi Surya. *Kilat.* 9(2): 217-224.
[21] M. Fikri, Z. Abdul-Malek, M. R. M. Esa, and E. Supriyanto. 2023. Recursive Parameter Estimation and Its Convergence for Multivariate Normal Hidden Markov Inhomogeneous Models. *Malaysian Journal of Fundamental and Applied Sciences*. 19(5): 840-854. Doi: 10.11113/mjfas.v19n5.3041.
[22] M. Fikri, C. Chirstiono, and K. Garniwa Mulyana. 2022. Optimasi Likelihood Function Suara Corona Discharge Menggunakan Model Normal Hidden Markov Sebagai Langkah Awal Deteksi Dini Kegagalan Isolasi. *Prosiding Seminar Nasional Energi*. 3:1-8.
[23] N. R. Pal, K. Pal, J. M. Keller and J. C. Bezdek. 2005. A Possibilistic Fuzzy c-means Clustering Algorithm. *IEEE Transactions on Fuzzy Systems*. 13(4): 517-530.
[24] IEEE. 2013. IEEE Standard for High-Voltage Testing Techniques Sponsored by the Power System Instrumentation and Measurements Committee IEEE Power and Energy Society. *IEEE Power and Energy Society*. 1-231.
[25] IEC. 2016. High Voltage Test Techniques - Measurement of Partial Discharge by Electromagnetic and Accoustic Methods. *IEC TS 62478;2016*. 1-38.
[26] L. Rabiner dan B.-H. Juang. 1993. *Fundamentals of Speech Recognition.* Prentice Hall.
[27] L. R. Rabiner. 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Re. *Proc. IEEE*. 77(2): 257-288,
[28] M. Thakker. 2013. Microphone Sensor for Sound Analysis and Breakdown Detection Using LabVIEW. *ResearchGate.*
[29] L. Gröll and J. Jäkel. 2005. A New Convergence Proof of Fuzzy c-means. *IEEE Transactions on Fuzzy Systems*. 13(5): 717-720.
[30] H. Illias, Teo Soon Yuan, A. H. A. Bakar, H. Mokhlis, G. Chen, dan P. L. Lewin. 2012. Partial Discharge Patterns in High Voltage Insulation. *PECon 2012 - 2012 IEEE International Conference on Power and Energy*. 750-755.
[31] N. R. Pal, K. Pal, J. M. Keller and J. C. Bezdek. 2005. A Possibilistic Fuzzy c-means Clustering Algorithm. *IEEE Transactions on Fuzzy Systems.* 13(4): 517-530.

[32] N. R. Pal and J. C. Bezdek. 1995. On Cluster Validity for the Fuzzy c-Means Model. *IEEE Transactions on Fuzzy Systems.* 3(3): 370-379.

[33] R. J. Hathaway and J. C. Bezdek. 1988. Recent Convergence Results for the Fuzzy C-means Clustering Algorithms. *Journal of Classification*. 5(2): 237-247.

[34] J. C. Bezdek. 1984. FCM: The Fuzzy c-Means Clustering Algorithm. *Computer & Geosciences*. 10(2-3): 191-203.

[35] M. Borghei dan M. Ghassemi. 2021. A Deep Learning Approach for Discrimination of Single- And Multi-source Corona Discharges. *IEEE Trans. Plasma Sci*. 49(9): 2936-2945.

[36] G. C. Montanari, R. Ghosh, L. Cirioni, G. Galvagno, dan S. Mastroeni. (2022). Partial Discharge Monitoring of Medium Voltage Switchgears : Self-condition Assessment using an Embedded Bushing Sensor. *IEEE Trans. Power Deliv*. 37(1): 85-92.

[37] M. Karimi, M. Majidi, H. Mirsaeedi, M. M. Arefi, dan S. Member. 2019. IEEE Transactions on Industrial Electronics A Novel Application of Deep Belief Networks in Learning Partial Discharge Patterns for Classifying Corona, Surface and Internal Discharges. *IEEE Trans. Ind. Electron*. 1.

[38] Y. Liu, J. Xv, Y. Liu, H. Yuan, dan Y. Cui. 2020. A Method for the Indirect Detection of Audible Noise from High-Voltage Direct Current Transmission Lines. *IEEE Trans. Instrum. Meas.* 69(7): 4358-4369.

[39] J. H. Pengfei Xu, Sören Hedtke, Bo Zhang, Martin Pfeiffer, Christian M. Franck. 2021. Research Collection. *IEEE Trans. Power Deliv*. 36(1): 1-8.

[40] S. Bolat dan Ö. Kalenderli. 2017. Estimation of Corona Location from Audible Noise. *Electr. Eng*.