Recognizing Five Major Dialects in Indonesia Based on MFCC and DRNN

B Tawaqal and S Suyanto

1,2 School of Computing, Telkom University, Bandung, Indonesia
Email: bilaltawaqal@student.telkomuniversity.ac.id

Abstract. Dialect is a variation of the language used by a group of people, sometimes in a particular region. It plays an essential role in automatic speech recognition (ASR). In general, an ASR gives high accuracy for a dialect-specific case, but it obtains a low accuracy for the multi-dialect application, such as for the Indonesian language that has hundreds of dialects. In this research, a system to recognize various dialects in Indonesia is developed. First, an utterance is preprocessed using both normalization and framing. Second, its features are then extracted using the Mel frequency cepstrum coefficients (MFCC), which is one of the feature extraction methods for the best acoustic signals. Finally, a deep recurrent neural network (DRNN) is used to learn and classify dialect characteristics. Evaluation of the dataset of five major dialects in Indonesia shows that the greater the Epoch and Bath Size, the greater the accuracy produced by the DRNN. However, accuracy is not directly proportional to the value of both parameters. The Epoch of 30 and Batch Size of 30 are the optimum parameters that yield the highest accuracy of 87.0% for the training set. Evaluation of the testing set shows that it gives an accuracy of 85.4% for the unseen dialects.

1. Introduction
Language is an organized communication tool in the form of units, such as words, groups of words, clauses, and sentences expressed verbally and in writing [1]. Both understanding and modeling of individual variations in spoken language are challenging in speech technology. People have various speaking styles based on dialect, accent, and socioeconomic background [2].

In Indonesia, there are many dialects from various regions. Dialects are varieties of speech in individual languages. The Oxford English Dictionary (OED) defines the dialect as one of the subordinate forms or varieties of language that arises from the local distinctiveness of vocabulary, pronunciation, and idioms. The variations can exist at all linguistics levels, namely vocabulary, idioms, grammar, and pronunciation [3]. Recognition of speech with multi-dialect remains challenging, even though the dialect-specific acoustic model works well in general [4].

Previous research in [5] proposes a model of dialect recognition using a combination of MFCC and pitch feature candidates to extract the features and a conventional machine learning approach to recognize two dialects in Indonesia: Javanese and Sundanese. The experimental results show that the use of the combination of MFCC and pitch feature and the I-Vector modeling technique gives a bit advantage. Unfortunately, the use of the conventional machine learning approach gives a low accuracy.

In this research, a modern machine learning-based model called DRNN is exploited to develop a new dialect model. The model is then evaluated to recognize five major dialects in Indonesia: Javanese, Sundanese, Banjar, Bugis, and Malay), which is more challenging than the previous work.
The feature extraction used here is the standard MFCC since it is not the focus of this research, and the appropriate features (or patterns) can be extracted using DRNN that is effective for both classification and feature engineering.

2. Literature Review

In [6], there are two methods for building a voice-command-based system: speech recognition and voice recognition. Speech recognition converts the analog speech signals into digital data matched with specific patterns stored in the database. Hence, the results are in the form of text following the given speech patterns. In contrast, voice recognition recognizes a speech by comparing the patterns of features with the previously stored speech. In other words, speech recognition recognizes what someone said, but voice recognition identifies a person through his voice.

In [7] and [8], speech recognition is a process to convert a given utterance into a text using an algorithm implemented in a computer program. It allows several devices to update and solve the utterance by converting them into a digital form and then matching the digital speech. The speech recognition area aims to make techniques and systems to receive input into system acceptance.

In general, a dialect-specific acoustic model works well [4]. However, it is hard to multi-dialect. Therefore, researches in dialect recognition are crucial. A dialect recognition system can be developed using either single conventional machine learning: such as support vector machines (SVM) [9], [10], artificial neural network (ANN) [11], Gaussian Mixture Model (GMM) [12], ensemble learning [10], [13], [14], or deep learning, such as recurrent neural network (RNN) [15], [16] long short-term memory (LSTM) [17], combined convolutional neural networks and bidirectional gated recurrent unit (CNN-BiGRU) [18], bidirectional LSTM (BLSTM) [19], combination of Hidden Markov Model and LSTM (HMM-LSTM) [20].

In [9], a dialect recognition system is developed using the features of cepstral coefficients, which are then classified using support vector machines designed with the sequential minimal-optimization (SMO-SVM). Evaluation of two dialect datasets: the Kannada dataset of five dialects and the English dataset of nine dialects, show that the model gives a good performance. In [10], two models based on SVM and extreme gradient boosting (XGB) are exploited to recognize Kannada dialects. Experimental results show that the stops in shorter duration present the cues of dialectal linguistic. Combined properties of spectral contribute to identifying the Kannada dialects. In [11], the researchers show that ANN gives higher accuracy than SVM for the dataset of 5 spoken Kannada dialects, where spectral and prosodic features are the most prominent features.

In [12], for German dialects, the GMM and maximum-a-posteriori (MAP) give a recognition accuracy from 32.4% to 54.9% for a 3-dialects test and from 19.6% to 31.4% for classifying 9-dialects. Several dialects can be recognized more easily using spectral features. Meanwhile, other dialects need a different set of features or other complicated classifiers.

In [13] and [14], three spectral features of MFCCs, spectral flux, and entropy are used to classify dialects using a single SVM and an ensemble SVM (ESVM). Evaluation of the Kannada dialects shows that the derived feature vectors are better than the raw ones, and ESVM outperforms the SVM, where ESVM produces a recognition rate of 91.38%.

Meanwhile, the deep learning approach gives higher accuracies than conventional machine learning. RNN is reported to give a high accuracy for Chinese dialects recognition for both short and long utterance conditions [16]. Besides, it requires less training time. A combination of the new one-against-all (NOAA) binary classifier and LSTM gives a better performance than single LSTM for six major China dialects [17]. The CNN-BiGRU gives better features and increases recognition accuracy [18]. Furthermore, in [19], a hybrid connectionist temporal classification (CTC)/attention, where a BLSTM is exploited as an encoder, was shown to give a better result than the single CTC or single attention. In [20], the HMM-LSTM achieves higher accuracy than a deep neural network (DNN).

In this research, a DRNN-based dialect recognition model is developed to recognize five major dialects in Indonesia using the features of MFCC.

3. Proposed Model

The developed recognition system of five dialects in Indonesian is illustrated in Fig. 1. First, the system will receive speech input. The speech is then preprocessed using both normalization and
framing. The next step is feature extraction, where the speech is extracted and made into a template for the testing set. After that, the system learns the data and do the classification. If the training results have high accuracy, the system will carry out a test phase. At this stage, the system recognizes the dialect from the audio input.

**Figure 1.** Developed system.

### 3.1. Dataset

Based on data from the Central Statistics Agency in 2010, Javanese is the largest ethnic group, 40.22% of the total Indonesian population. The second-largest ethnic group is Sundanese, with a population of 15.50%. Furthermore, in the third, fourth, and fifth ranks are the Batak, other Sulawesi Origin, and the Madurese, which respectively have a population of 3.58%, 3.22%, and 3.03%. Therefore, the Javanese, Sundanese, Makassarese, and Balinese dialects (representing 60% of dialects in Indonesia) are selected to develop the dataset.

The dataset is developed by recording some speakers using a wireless microphone and Adobe Audition software. There are 500 speakers; each reads 100 sentences from the transcription in the text corpus. They have five dialects: Javanese, Sundanese, Banjar, Bugis, and Malay; each dialect has 100 speakers: 50 males and 50 females. They also have five age category (according to the Ministry of Health): Early Adolescence: 12-16 years (10 people in each dialect and gender), Late Adolescence: 17-25 years (10 people in each dialect and gender), Early Adults: 26 - 35 years (10 people in each dialect and gender), Late Adult: 36 - 45 years (10 people in each dialect and gender), and Seniors: more than 46 years (10 people in each dialect and gender).

### 3.2. Feature extraction

The speech feature is extracted using the MFCC, which is one of the best feature extraction methods for the acoustic signals [21], [22]. Speech analysis on Mel-frequency is based on human hearing perception because the human ear has been observed to function as a filter at specific frequencies. MFCC is a widely used method in the field of speech processing [23], [24]. This method is used to perform feature extraction, a process that converts speech signals into several parameters. This filter is used to capture the critical phonetic characteristics of a speech. The MFCC is represented on a Mel-frequency scale, which is a linear frequency below 1000Hz and logarithmic above 1000Hz.

MFCC is based on discrete Fourier transforms (DFT) [25]. MFCC is considered good enough to present features and signals as humans recognize speech characteristics. Its process can be illustrated in a block diagram in Fig. 2.
Figure 2. Block diagram of MFCC [26].

- **Pre-emphasis**
  A speech pre-emphasis requires a recording or sampling. Filtering is used to smooth the spectral of the given speech. It is calculated in the time-domain speech depending on the input/output relation.

- **Framing**
  The spectral of the given speech is extracted with framing to reduce its non-stationary. In other words, framing the speech is expected to make it be stationary [27].

- **Windowing**
  Each frame is multiplied with a specific window function. It is used to make sure the continuity of the speech [28].

- **Fast Fourier Transform**
  One of the Fourier transform methods used to get a discrete signal is a fast Fourier transform (FTT), which is the fast-computational version of the discrete Fourier transform (DFT) algorithm. It is applied to each windowed frame.

- **Mel-Frequency Wrapping**
  Mel frequency wrapping is implemented using a filter bank, which is made to determine the energy size of a particular frequency-band (in the frequency domain). The filter bank uses a convolution representation. In [26], it can be implemented by multiplying the signal spectrum and its coefficient.

- **Log**
  The logarithm is produced from the DFT. The values of Mel filter-bank are then reduced by performing a natural log. The log feature is used to estimate entail and less sensitiveness in the input [29].

- **Discrete Cosines Transform**
  Discrete cosine transform (DCT), the same as the inverse Fourier transform, is the final step. It is the Mel spectrum's correlating value, which is used to produce a fair representation.
Cepstrum
Cepstrum is the inverse value of the spectrum [9]. It functions to get critical information from a given speech. In [26], the DCT is applied to transform the log Mel spectrum into the cepstrum.

3.3. Deep Recurrent Neural Network
In general, humans do not make a single decision every time [7], [14]. Humans always take into account the past in making a decision. This way of thinking is the basis of developing an RNN. Just like the analogy, RNN does not merely throw away information from the past in its learning process. This is what distinguishes RNN from ordinary ANN. In short, RNN is one of the neural network models for processing sequential data. The way that RNN can store information from the past is by looping in its architecture, which automatically keeps information from the past stored. The difference between RNN and DRNN is the number of hidden layers. As the name suggests, DRNN has more hidden layers than RNN.

A single direction architecture is used here since it is sufficient to classify five dialect classes in Indonesia. First, the data is transformed into a one-dimensional array. DRNN has a Dense layer to adjust the number of nodes to be created and softmax nodes to return a probability value. If summed up, the result is 1. Before training, there are three rules that must be added: 1) the loss function to measure how accurate the model is when doing training; 2) the optimizer to define how the model updates the model based on the data learned and the loss function; and 3) the metric to monitor the training and testing process, usually using accuracy.

The preprocessing results are several frames, which are then used to train the DRNN. The hidden state information is given to the next time step of the current layer and the existing layer of the present time step.

At time step $t$, it is assumed a minibatch $X_t \in \mathbb{R}^{n \times d}$, where $n$ is the number of examples, and $d$ is the number of inputs. The hidden state in the hidden layer $\ell$ ($\ell = 1, \ldots, T$) is $H^{(\ell)}_t \in \mathbb{R}^{n \times h}$ (where $h$ is the number of hidden units), the output layer variable is $O_t \in \mathbb{R}^{n \times q}$ (where $q$ is the number of outputs), and a hidden layer activation function $f_\ell$ for layer $\ell$. We compute the hidden state of layer 1 as before, using $X_1$ as input. For all subsequent layers, the hidden state of the previous layer is used in its place.

![Figure 3. DRNN Structure.](image-url)
Finally, the output of the output-layer is only based on the hidden state of the hidden layer \( L \). We use the output function \( g \) to address

\[
O_t = g\left( H_t^{(L)} \right)
\]

(2)

Just like the multilayer perceptrons, the number of hidden layers \( L \) and the number of hidden units \( h \) in DRNN are hyperparameters [30].

4. Results and Discussion

The developed system is tested by changing the value of the Epoch and Batch Size. In each test, the kernel is restarted so that the learning rate is also redefined. Experimental results in Table 1 show the effect of epoch to the system accuracy in the training process. The best epoch is 30, which gives the highest accuracy of 89.0%. This result is higher than both Epochs of 10 and 20, which give an accuracy of 83.0% and 85.0%, respectively. It is also better than the epoch of 40 and 50, which just gives an accuracy of 84.0% and 86.0%, respectively. It means that the epoch is relatively sensitive to the accuracy of DRNN.

| Experiment | Epoch | Batch Size | Accuracy (%) |
|------------|-------|------------|--------------|
| I          | 10    | 27         | 83.0         |
| II         | 20    | 27         | 85.0         |
| III        | 30    | 27         | **89.0**     |
| IV         | 40    | 27         | 84.0         |
| V          | 50    | 27         | 86.0         |

Table 1. Effect of epoch on the accuracy in the training process.

The results in Table 2 show that the optimum Batch Size is 30 to produce the highest accuracy of 87.0%. This result is higher than both Batch Sizes of 10 and 20, which gives an accuracy of 84.0% and 83.0%, respectively. It is also better than both Batch Sizes of 40 and 50, which gives an accuracy of 84.0% and 82.0%, respectively. It means that Batch Size is also relatively sensitive to the accuracy of DRNN.

| Experiment | Epoch | Batch Size | Accuracy (%) |
|------------|-------|------------|--------------|
| I          | 30    | 10         | 84.0         |
| II         | 30    | 20         | 83.0         |
| III        | 30    | 30         | **87.0**     |
| IV         | 30    | 40         | 84.0         |
| V          | 30    | 50         | 82.0         |

Table 2. Effect of batch size on the accuracy in the training process.

Therefore, the epoch of 30 is used for the next experiment. Furthermore, using the optimum Epoch and Batch Size, the developed system can predict the unrecognized speech in the testing set. It can predict dialects from various regions in Indonesia, not only predict a specific dialect. It also has a high degree of accuracy. Evaluation of the testing set shows that it gives an average accuracy of 85.4%.

Since a syllable is strongly related to the dialect [31], a syllable-based feature extracted using an automatic speech segmentation as proposed in [32] and an automatic syllabification [33]–[35] will be exploited to enhance this dialect recognition model in our future works. An optimization method based
on evolutionary algorithms [36] or swarm intelligences [37], [38] will also be investigated to optimize the DRNN parameters.

5. Conclusion
The system to recognize five major dialects in Indonesia using MFCC and DRNN is successfully developed. Some experiments show that the greater the Epoch and Bath Size, the greater the accuracy is obtained. However, there is a limit where the accuracy is not directly proportional to the value of both parameters. Evaluation of the testing set shows that it gives an average accuracy of 85.4%.

6. References
[1] L. Shen, “Context and Text,” Theory Pract. Lang. Stud., vol. 2, no. 12, pp. 2663–2669, 2012, doi: 10.4304/tpls.2.12.2663-2669.
[2] F. Biadsy, “Automatic dialect and accent recognition and its application to speech recognition,” Autom. Dialect Acc Ent Recogni. Its Appl. To Speech Recogn., 2011.
[3] H. Behravan, “Dialect and Accent Recognition,” University of Eastern Finland, 2012.
[4] S. Yoo, I. Song, and Y. Bengio, “A Highly Adaptive Acoustic Model for Accurate Multi-Dialect Speech Recognition,” ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc., vol. 2019-May, pp. 5716–5720, 2019, doi: 10.1109/ICASSP.2019.8683705.
[5] R. Rahmawati and D. P. Lestari, “Java and Sundanese Dialect Recognition from Indonesian Speech Using GMM and I-Vector,” 2017, doi: https://doi.org/10.1109/tssa.2017.8272892.
[6] T. Fukuda et al., “Data Augmentation Improves Recognition of Foreign Accented Speech,” in INTERSPEECH, 2018, no. September, pp. 2409–2413.
[7] S. K. Gaikwad, B. W. Gawali, and P. Yannawar, “A Review on Speech Recognition Technique,” Int. J. Comput. Appl., vol. 10, no. 3, pp. 16–24, 2010, doi: 10.5120/1462-1976.
[8] J. Home, “RNN Dialek Manado,” Medicus, vol. 5, no. 3, pp. 3–4, 2018.
[9] A. R. Choudhury, N. B. Chittaragi, and S. G. Koolagudi, “Dialect Recognition System Using Excitation Source Features,” 2018, doi: 10.1109/INDICON45594.2018.8987055.
[10] N. B. Chittaragi, P. Hegde, S. K. P. Mothukuri, and S. G. Koolagudi, “Spectral Feature Based Kannada Dialect Classification from Stop Consonants,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11941 LNCS, pp. 82–90, 2019, doi: 10.1007/978-3-030-34869-4_10.
[11] N. B. Chittaragi, A. Limaye, N. T. Chandana, B. Annappa, and S. G. Koolagudi, “Automatic text-independent Kannada dialect identification system,” Adv. Intell. Syst. Comput., vol. 863, pp. 79–87, 2019, doi: 10.1007/978-981-3-3338-5_8.
[12] J. Dobbriner and O. Jokisch, “Towards a dialect classification in german speech samples,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11658 LNAI, pp. 64–74, 2019, doi: 10.1007/978-3-030-26061-3_7.
[13] S. Ye, R. Zhao, and X. Fang, “An Ensemble Learning Method for Dialect Classification,” in IOP Conference Series: Materials Science and Engineering, 2019, vol. 569, no. 5, doi: 10.1088/1757-899X/569/5/052064.
[14] N. B. Chittaragi and S. G. Koolagudi, “Automatic dialect identification system for Kannada language using single and ensemble SVM algorithms,” Lang. Resour. Eval., vol. 54, no. 2, pp. 553–585, 2020, doi: 10.1007/s10579-019-09481-5.
[15] T. N. Trong, K. Jokinen, and V. Hautamäki, “Enabling spoken dialogue systems for low-resourced languages—End-to-end dialect recognition for north sami,” Lect. Notes Electr. Eng., vol. 579, pp. 221–235, 2019, doi: 10.1007/978-981-3-9443-0_19.
[16] Z. Ren, G. Yang, and S. Xu, “Two-stage training for Chinese dialect recognition,” in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2019, vol. 2019-Septe, pp. 4050–4054, doi: 10.21437/Interspeech.2019-1522.
[17] S. Ye, C. Li, R. Zhao, and W. Wu, “NOAA-LSTM: A New Method of Dialect Identification,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 11632 LNCS, pp. 16–26, 2019, doi: 10.1007/978-3-030-24274-9_2.
[18] Q. Zhang et al., “End-to-end Chinese dialects identification in short utterances using CNN-BiGRU,” in Proceedings of 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference, ITAIC 2019, 2019, pp. 340–344, doi: 10.1109/ITAIC.2019.8785614.

[19] J. Sun, G. Zhou, H. Yang, and M. Wang, “End-to-end Tibetan Ando dialect speech recognition based on hybrid CTC/attention architecture,” in 2019 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2019, 2019, pp. 628–632, doi: 10.1109/APSIPAASC47483.2019.9023130.

[20] W. Ying, L. Zhang, and H. Deng, “Sichuan dialect speech recognition with deep LSTM network,” Front. Comput. Sci., vol. 14, no. 2, pp. 378–387, 2020, doi: 10.1007/s11704-018-8030-z.

[21] B. Alkhatib, “Voice Identification Using MFCC and Vector Quantization,” vol. 17, no. 3, pp. 1019–1028, 2020.

[22] R. Ahmad and S. Suyanto, “The Impact of Low-Pass Filter in Speaker Identification,” in 2019 2nd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2019, 2019, pp. 133–136.

[23] S. Suyanto, A. Arifianto, A. Sirwan, and A. P. Rizaendra, “End-to-End Speech Recognition Models for a Low-Resourced Indonesian Language,” in 2020 8th International Conference on Information and Communication Technology (ICoICT), Jun. 2020, pp. 1–6, doi: https://doi.org/10.1109/ICoICT49345.2020.9166346.

[24] A. Prayitno and S. Suyanto, “Segment Repetition Based on High Amplitude to Enhance a Speech Emotion Recognition,” Procedia Comput. Sci., vol. 157, pp. 420–426, 2019, doi: https://doi.org/10.1016/j.procs.2019.08.234.

[25] M. Y. Faisal and S. Suyanto, “SpecAugment Impact on Automatic Speaker Verification System,” in 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Dec. 2019, pp. 305–308, doi: https://doi.org/10.1109/ISRITI48646.2019.9034603.

[26] Y. Afrillia, H. Mawengkang, M. Raml, F. Fadlisyah, and R. P. Fhonna, “Performance Measurement of Mel Frequency Ceptral Coefficient (MFCC) Method in Learning System of Al-Qur’an Based in Nagham Pattern Recognition,” J. Phys. Conf. Ser., vol. 930, no. 1, 2017, doi: 10.1088/1742-6596/930/1/012036.

[27] J. Li, L. Deng, R. Haeb-Umbach, and Y. Gong, “Fundamentals of speech recognition,” Robust Automatic Speech Recognition. pp. 9–40, 2016, doi: 10.1016/b978-0-12-802398-3.00002-7.

[28] A. Pahwa, “Speech Feature Extraction for Gender Recognition,” Int. J. Image, Graph. Signal Process., vol. 8, no. 9, pp. 17–25, 2016, doi: 10.5815/ijigsp.2016.09.03.

[29] M. A. For and Q. Rule, “MFCC-VQ Approach For QalqalahTajweed Rule Checking . pp 275 -293,” vol. 27, no. 4, pp. 275–293, 2014.

[30] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, “Dive into Deep Learning,” p. 639, 2018.

[31] R. Janakiraman, J. C. Kumar, and H. A. Murthy, “Robust syllable segmentation and its application to syllable-centric continuous speech recognition,” in National Conference on Communications (NCC), Jan. 2010, pp. 1–5, doi: 10.1109/NCC.2010.5430189.

[32] S. Suyanto and A. E. Putra, “Automatic Segmentation of Indonesian Speech into Syllables using Fuzzy Smoothed Energy Contour with Local Normalization, Splitting, and Assimilation,” J. ICT Res. Appl., vol. 8, no. 2, pp. 97–112, 2014, doi: http://dx.doi.org/10.5614%2Fitbj.ict.res.appl.2014.8.2.2.

[33] S. Suyanto, “Phonological similarity-based backoff smoothing to boost a bigram syllable boundary detection,” Int. J. Speech Technol., vol. 23, no. 1, pp. 191–204, 2020, doi: https://doi.org/10.1007/s10772-020-09677-z.

[34] S. Suyanto, “Flipping onsets to enhance syllabification,” Int. J. Speech Technol., vol. 22, no. 4, pp. 1031–1038, 2019, doi: https://doi.org/10.1007/s10772-019-09649-y.

[35] E. A. Parande and S. Suyanto, “Indonesian graphemic syllabification using a nearest neighbour classifier and recovery procedure,” Int. J. Speech Technol., vol. 22, no. 1, pp. 13–20, 2019, doi: https://doi.org/10.1007/s10772-018-09569-3.
[36] M. H. Aliefa and S. Suyanto, “Variable-Length Chromosome for Optimizing the Structure of Recurrent Neural Network,” in 2020 International Conference on Data Science and Its Applications (ICoDSA), Aug. 2020, doi: https://doi.org/10.1109/ICoDSA50139.2020.9213012.

[37] F. Ahyar and S. Suyanto, “Firefly Algorithm-based Hyperparameters Setting of DRNN for Weather Prediction,” in 2020 International Conference on Data Science and Its Applications (ICoDSA), Aug. 2020, doi: https://doi.org/10.1109/ICoDSA50139.2020.9212921.

[38] B. Z. Aufa, S. Suyanto, and A. Arifianto, “Hyperparameter Setting of LSTM-based Language Model using Grey Wolf Optimizer,” in 2020 International Conference on Data Science and Its Applications (ICoDSA), Aug. 2020, pp. 1–5, doi: https://doi.org/10.1109/ICoDSA50139.2020.9213031.