JurnalResti5Nov2022

by Kristiawan Nugroho

Submission date: 05-Nov-2022 06:58PM (UTC+0700)
Submission ID: 1945274674
File name: Kristiawan_RESTI04Nov2022.docx (173.69K)
Word count: 2991
Character count: 16204

Accredited Ranking SINTA 2
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

Published online on: **http://jurnal.iaii.or.id**

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 6 No. x (2022) x - x ISSN Media Electronic: 2580-0760

Multi-Accent Speaker Detection Using Normalize Feature MFCC Neural Network Method

Kristiawan Nugroho1, Edi Winarno2, Isworo Nugroho3
[1]Master in Information Technology, Faculty of Information Technology and Industry, Universitas STIKUBANK
[2]Technical Information, Faculty of Information Technology and Industry, Universitas STIKUBANK
[1]kristiawan@edu.unisbank.ac.id, [2]edywin@edu.unisbank.ac.id, [3]isworo@edu.unisbank.ac.id

Abstract

Speech recognition is a field of research that continues to this day. Various methods have been developed to detect the human voice with greater precision and accuracy. Research on human speech recognition that is quite challenging is accent recognition. Detecting various types of human accents with different accents and ethnicities with high accuracy is a research that is quite difficult to do. According to the results of the research on the data preprocessing stage, feature extraction and the selection of the right classification method play a very important role in determining the accuracy results. This study uses a preprocessing approach with normalizing features combined with feature extraction techniques MFCC and Neural Network (NN) which is a classification method that works based on the workings of the human brain. The results obtained using the normalize feature with MFCC and Neural Network for multi-accent voice recognition, the accuracy performance reaches 82.68%, precision is 83% and recall is 82.88%..

Keywords: Detection,Multi Accent,MFCC,Neural Network

1. Introduction

Artificial intelligence technology in the field of speech recognition is growing rapidly nowadays. Various companies have produced smart tools for voice recognition such as Alexa, Siri and Google Assistant. These various products have become part of people's lives like personal assistants who help humans in every activity of their lives such as translating languages, playing entertainment in the form of music to recommending paths that are free from traffic jams as well as recommendations for places to eat, travel and the nearest gas station.

The development of speech recognition technology began in 1940 by a company by the name of the American Telephone and Telegraph Company (AT & T) by building a tool to recognize human speech. The research has progressed to date with the discovery of various methods and the results of the real contribution of speech recognition technology in the health, education, automotive and military fields. Voice recognition technology continues to help various areas of human life so that they can carry out various life activities well.

Speech recognition is a field of research that continues to grow and is one of the most interesting research themes to study. Various methods that have been frequently used to achieve maximum accuracy include Support Vector Machine (SVM), Random Forest (RF), Gaussian Mixture Model (GMM), Hidden Markov Model (HMM) and Neural Network (NN). SVM is a machine learning method that is also used in voice classification, such as the research conducted by Zade[1] which combines SVM with the feature extraction method of MFCC (Mel-Frequency Cepstral Coefficient) and LPC (Linear Predictive Coding) in the Azerbaijani DataSet classification. In another study, Pradana[2] also used SVM which was also combined with MFCC to detect Arabic speech resulting in an accuracy rate of 61.16%.

Another study on speech recognition was also conducted using the Random Forest method in research conducted by Rao[3] using the VoxCeleb dataset, which achieved a voice recognition accuracy rate of 84.53%. Nivetha [4] also uses Random Forest and MFCC LPC to detect sounds in Tamil. Another method that is often used in speech recognition is GMM such as research conducted by Chauhan [5], Nayana [6], and

Accepted: xx-xx-2022 | Received in revised: xx-04-2022 | Published: xx-04-2022

Rajan [7] which has achieved a fairly good level of accuracy. HMM is also a method used in speech recognition such as research by Chamidy[8], Nada[9] and Chen[10]. One of the challenging topics in speaker recognition is recognizing speakers with different accents. Recognizing speakers with different accents is a complex task that is not easy to do [11], so an appropriate method is needed to recognize the speaker's accent so as to produce a high level of accuracy. Various studies have been carried out to achieve a better level of speech recognition accuracy, including using the MFCC (Mel-Frequency Cepstral Coefficient) method such as the research conducted by Maurya [12] with the MFCC and GMM methods, Widyowati [13] which combines MFCC with Convolutional Neural Network (CNN) and Nugroho[14] which use MFCC and Deep Neural Network (DNN). In these various studies, MFCC has been proven to help improve speech recognition accuracy optimally.

This research is a research that combines MFCC which has previously been preprocessed using the Normalize Feature combined with the Neural Network (NN) method so that it can achieve a high level of accuracy for recognizing different accent voices. NN is used because it has various advantages, including advantages in predicting nonlinear cases, having good performance in parallel processing and the ability to tolerate errors [15] so that this approach is suitable for use in this accent speech recognition research.

2. Research Methods

Research on the speaker accents recognition who use various accents is carried out in stages as shown in Figure 1 below:

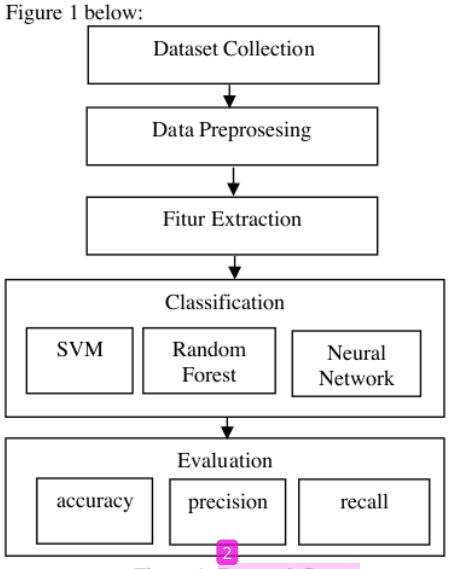

Figure 1. Research Stages

The research stages in Figure 1 are the stages in detecting speakers who have various kinds of accents with the following steps:

A. Dataset Collection

The dataset is the main data source used in a study. This research on the detection of speakers who use many accents takes a dataset from the UCI Dataset with the following website address https://archive.ics.uci.edu/ml/datasets/Speaker+Accent+Recognition# This dataset consists of 12 attributes with 1 label named language which contains 6 different types of speaker accents, including those from France, England, America and Germany.

B. Data Preprocessing

This stage is a very important stage because it relates to the preparation of valid data so that it can be used properly in the next stage. In this study, a preprocessing approach is used in the form of deleting data that has empty values and the Normalize Feature which is an integral part of the Neural Network method [6] in the Orange application with a display as shown in Figure 2 as follows:

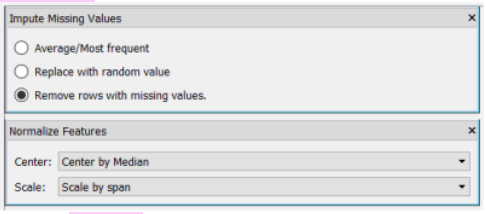

Figure 2. Preprocessing Stage

In the preprocessing process on figure 2, the multi-accent speaker data will be deleted which contains empty data and each feature will be normalized so that better data will be obtained and ready for feature extraction.

C. Feature Extraction

In this study, the MFCC method is used as one of the superior methods in extracting voice signal data. The MFCC method is often used because it has the advantage of being able to identify the characteristics of the voice signal properly [17] and is a method that is often used in speech recognition. In the MFCC method, the steps taken to extract the voice signal are:

1. Pre-emphasis

The initial stage of the MFCC method is carried out after the sound sampling stage, Pre-emphasis aims to obtain a smoother form of speech signal frequency spectral so that the quality of the voice signal will be better for processing in the next hold.

2 Frame Blocking

This stage is the process of dividing the voice signal in the form of shorter segments so that the time period changes.

DOI: https://doi.org/10.29207/resti.v6iX.xxx
Creative Commons Attribution 4.0 International License (CC BY 4.0)

3. Windowing
This function aims to manipulate the amplitude of the sound signal by using a mathematical formula.
4. Fast Fourier Transform (FFT)
FFT is an approach to perform DFT/Discrete Fourier Transform calculations quickly.
5. Mel Frequency Wrapping (MFW)
MFW is a step to determine the size of the frequency band in the voice signal that is needed before the next stage.
6. Discrete Continues Transform (DCT)
The DCT approach is needed in sound processing in changing from the frequency domain to the time domain [18] and performing spectrum compression.
7. Cepstral Lifting (CL)
In the MFCC method, CL is the last technique used to improve the quality of speech recognition signals.

D. Classification

In this study, the Neural Network (NN) method is used, which is a method that works like a neural network in the human brain. The architectural form of the NN method [19] can be seen in Figure 3 below:

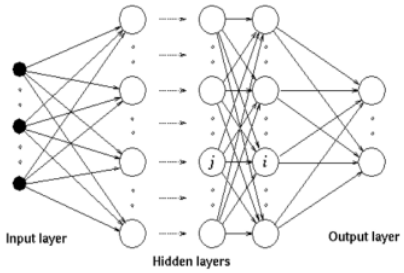

Figure 3. Neural Network architecture

In the NN architecture in Figure 3, it can be seen that there are 3 main parts of this method, namely the input layer, hidden layer and output layer which consists of several circular neurons. The input layer is the part to perform input in the form of features that will be processed in this algorithm, while the hidden layer consists of several network layers that are used to process data from the input layer and the results will be displayed in the output layer. The calculation of the value at the output layer (y) on the NN can be calculated using the following formula:

$$y = \sum (xiwi) + b \quad (1)$$

Where :

x = input value

w = weight

b = bias

This study also compares the results of research using NN with 2 other Machine Learning methods, namely SVM and Random Forest. These two methods are also used in this multi-accent voice detection research because they also have several advantages, SVM has the advantage of producing a good classification model even though it is trained to use only a small amount of data[20] while Random Forest is a robust method to overcome the problem of overfitting and data that is lacking. non-linear [21].

E. Evaluation

The evaluation stage is an important process in determining the performance of an algorithm. In research that uses Machine Learning, the evaluation method that is often used is by measuring the level of accuracy, precision and recall of the performance of each method. The calculation to determine the performance level of each evaluation is formulated as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (2)$$

$$\text{Precision} = \frac{TP}{FP+TP} * 100\% \quad (3)$$

$$\text{Recall} = \frac{TP}{FN+TP} * 100\% \quad (4)$$

Where :

TP = True Positive

TN=True Negative

FP=False Positive

FN=False Negative

3. Results and Discussions

Based on the results of research that has been carried out on multi-accent speaker recognition, after the multi-accent speaker dataset is preprocessed using the normalize feature technique, the next step is the feature extraction stage with MFCC resulting in 330 lines of extracted speech data. Then the classification process is carried out using the Orange application which is one of the processing applications in Data Mining.

The results of this multi-accent speech processing classification use a Neural Network (NN) which is then compared the results with 2 other methods, namely SVM and Random Forest. The results of measuring the performance of each model using the cross validation sampling technique with 10 folds were tested 5 times resulting in a performance comparison table as shown in table 1 as follows:

Table 1. Model Performance Test

DOI: https://doi.org/10.29207/resti.v6iX.xxx
Creative Commons Attribution 4.0 International License (CC BY 4.0)

If we observe in table 1 above, the Neural Network (NN) method is a method that produces the best level of performance compared to the other 2 methods. Comparison

Methods	Accuracy(%)	Precision(%)	Recall(%)
Test-1			
SVM	79.6	82.2	79.6
Random Forest	72.3	72.3	72.3
NN	**82.7**	**82.7**	**82.7**
Test-2			
SVM	79.6	82.4	79.6
Random Forest	74.5	74.0	74.5
NN	**83.6**	**84.0**	**83.6**
Test-3			
SVM	79.6	82.4	79.6
Random Forest	73.9	74.1	79.3
NN	**82.4**	**82.3**	**82.4**
Test-4			
SVM	79.6	82.4	79.6
Random Forest	74.8	75.4	74.8
NN	**83.0**	**83.4**	**83.0**
Test-5			
SVM	79.6	82.4	79.6
Random Forest	74.8	74.8	74.8
NN	**82.7**	**82.8**	**82.7**

of the results can be seen in the table of the average performance of the model below:

Table 2. Avarage Result of Model Test

Methods	Accuracy(%)	Precision(%)	Recall(%)
SVM	79.60	82.30	79.60
Random Forest	74,06	74,12	75,14
NN	**82,68**	**83,04**	**82,88**

Based on table 2 above, it can be concluded that the Neural Network method is the best method for its performance in recognizing multi-accent speakers compared to other methods. Neural Network achieves an accuracy rate of 82.68 %, precision 83.04% and recall of 82.88 outperforms other methods where the performance level of the method is less than 80%. A more detailed explanation of the performance of each method used in multi-accent speaker recognition can be described in a comparison chart so that it can be seen more clearly about the performance of each model when compared to one another. Graphics regarding the performance of these models can be seen in Figure 4 as follows:

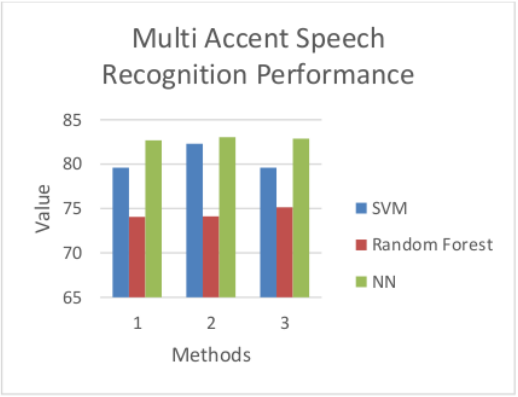

Figure 4. Model Performance Comparation

In Figure 4, it can be seen that the Neural Network method dominates the level of excellence in terms of accuracy, precision and recall. A more complete comparison of the level of accuracy can be seen in the image visualization and calculations on the Confusion Matrix as below:

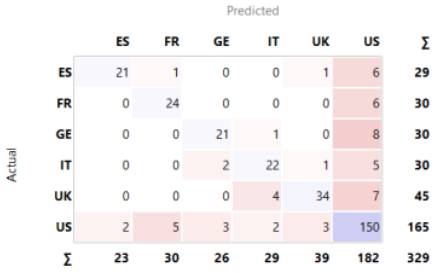

Figure 5. Confusion Matric Result

Based on Figure 5 above, the calculation of the accuracy of the Neural Network method on the confusion matrix is as follows:

$$\text{Accuracy} = \frac{(21+24+21+22+34+150)}{329} * 100\%$$

$$\text{Accuracy} = 82.68\%$$

The results of measuring the level of accuracy using NN which reached 82.68% showed that the NN method was superior to other methods in detecting multi-accented speech sounds.

DOI: https://doi.org/10.29207/resti.v6iX.xxx
Creative Commons Attribution 4.0 International License (CC BY 4.0)

4. Conclusion

Recognition of speakers with different accents is an interesting topic for research in speech recognition research. Various methods have been developed in recognizing speakers who use various accents. One method that is often used in speech recognition is the Neural Network (NN) which is a method that works based on the neural performance of the human brain.

In this research, a preprocessing approach using normalize feature and Neural Network is used to identify multi-accented speakers. The results show that the Neural Network method produces the best performance with an average accuracy of 82.68%, precision of 83.04% and recall of 82.88%, outperforming other methods such as Random Forest and Support Vector Machine.

Reference

[1] K. Aida-Zade, A. Xocayev, and S. Rustamov, "Speech recognition using Support Vector Machines," *Appl. Inf. Commun. Technol. AICT 2016 - Conf. Proc.*, vol. 1, 2017, doi: 10.1109/ICAICT.2016.7991664.

[2] W. A. Pradana, Adiwijaya, and U. N. Wisesty, "Implementation of support vector machine for classification of speech marked hijaiyah letters based on Mel frequency cepstrum coefficient feature extraction," *J. Phys. Conf. Ser.*, vol. 971, no. 1, 2018, doi: 10.1088/1742-6596/971/1/012050.

[3] M. S. Rao, G. B. Lakshmi, P. Gowri, and K. B. Chowdary, "Random Forest Based Automatic Speaker Recognition System," *Int. J. Anal. Exp. Model Anal.*, vol. 12, no. 4, pp. 526–535, 2020, [Online]. Available: http://www.ijaema.com/gallery/63-ijaema-april-3748.pdf

[4] N. S, D. R. A, and M. G. S, "Speech Recognition System for Isolated Tamil Words using Random Forest Algorithm," *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 2431–2435, 2020, doi: 10.35940/ijrte.a1467.059120.

[5] V. Chauhan, S. Dwivedi, P. Karale, and P. S. M. Potdar, "Speech to Text Converter Using Gaussian Mixture Model (GMM) of Electronics and Telecommunication Engineering," *Int. Res. J. Eng. Technol.*, pp. 125–129, 2016.

[6] P. K. Nayana, D. Mathew, and A. Thomas, "Comparison of Text Independent Speaker Identification Systems using GMM and i-Vector Methods," *Procedia Comput. Sci.*, vol. 115, pp. 47–54, 2017, doi: 10.1016/j.procs.2017.09.075.

[7] R. R. K and A. P. Joseph, "Domestic Language Accent Detector Using MFCC and GMM," *Int. J. Appl. Eng. Res.*, vol. 15, no. 8, p. 800, 2020, doi: 10.37622/ijaer/15.8.2020.800-803.

[8] T. Chamidy, "Metode Mel Frequency Cepstral Coeffisients (MFCC) Pada klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada Penutur Indonesia," *Matics*, vol. 8, no. 1, p. 36, 2016, doi: 10.18860/mat.v8i1.3482.

[9] P. Huruf, Q. Nada, C. Ridhuandi, P. Santoso, and D. Apriyanto, "Speech Recognition dengan Hidden Markov Model untuk," *J. AL-AZHAR Indones. SERI SAINS DAN Teknol.*, vol. 5, no. 1, pp. 19–26, 2019.

[10] Y. Chen, "A hidden Markov optimization model for processing and recognition of English speech feature signals," *J. Intell. Syst.*, vol. 31, no. 1, pp. 716–725, 2022, doi: 10.1515/jisys-2022-0057.

[11] Y. Singh, A. Pillay, and E. Jembere, "Features of speech audio for accent recognition," *2020 Int. Conf. Artif. Intell. Big Data, Comput. Data Commun. Syst. icABCD 2020 - Proc.*, 2020, doi: 10.1109/icABCD49160.2020.9183893.

[12] A. Maurya, D. Kumar, and R. K. Agarwal, "Speaker Recognition for Hindi Speech Signal using MFCC-GMM Approach," *Procedia Comput. Sci.*, vol. 125, pp. 880–887, 2018, doi: 10.1016/j.procs.2017.12.112.

[13] D. S. Widyowaty, A. Sunyoto, and H. Al Fatta, "Accent Recognition Using Mel-Frequency Cepstral Coefficients and Convolutional Neural Network," *Proc. Int. Conf. Innov. Sci. Technol. (ICIST 2020)*, vol. 208, no. Icist 2020, pp. 43–46, 2021, [Online]. Available: https://doi.org/10.2991/aer.k.211129.010

[14] K. Nugroho, E. Noersasongko, D. R. Ignatius, and M. Setiadi, "Enhanced Indonesian Ethnic Speaker Recognition using Data Augmentation Deep Neural Network," *J. King Saud Univ. - Comput. Inf. Sci.*, no. xxxx, 2021, doi: 10.1016/j.jksuci.2021.04.002.

DOI: https://doi.org/10.29207/resti.v6iX.xxx
Creative Commons Attribution 4.0 International License (CC BY 4.0)

[15] M. Badrul, "Optimasi Neural Network dengan Algoritma Genetika untuk Prediksi Hasil Pemilukada," *Bina Insa. ICT J.*, vol. 3, no. 1, pp. 229–242, 2016.

[16] B. Li, F. Wu, S. N. Lim, S. Belongie, and K. Q. Weinberger, "On feature normalization and data augmentation," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 12378–12387, 2021, doi: 10.1109/CVPR46437.2021.01220.

[17] M. Susanti, B. Susilo, and D. Andreswari, "Aplikasi Speech-To-Text Dengan Metode Mel Frequency Cepstral Coefficient (Mfcc) Dan Hidden Markov Model (Hmm) Dalam Pencarian Kode," *J. Rekursif*, vol. 6, no. 1, pp. 48–58, 2018, [Online]. Available: https://ejournal.unib.ac.id/index.php/rekursif/article/view/6497%0Ahttps://ejournal.unib.ac.id/index.php/rekursif/article/download/6497/3102

[18] C. G. K. Leon, "Robust computer voice recognition using improved MFCC algorithm," *Proc. - 2009 Int. Conf. New Trends Inf. Serv. Sci. NISS 2009*, pp. 835–840, 2009, doi: 10.1109/NISS.2009.12.

[19] B. S. Santoso, J. P. Tanjung, U. P. Indonesia, B. Gandum, and A. N. Network, "Classification of Wheat Seeds Using Neural Network Backpropagation," *JITE (Journal Informatics Telecommun. Eng. Available*, vol. 4, no. January, pp. 188–197, 2021.

[20] M. Ichwan, I. A. Dewi, and Z. M. S, "Klasifikasi Support Vector Machine (SVM) Untuk Menentukan TingkatKemanisan Mangga Berdasarkan Fitur Warna," *MIND J.*, vol. 3, no. 2, pp. 16–23, 2019, doi: 10.26760/mindjournal.v3i2.16-23.

[21] A. Sarica, A. Cerasa, and A. Quattrone, "Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer ' s Disease : A Systematic Review," vol. 9, no. October, pp. 1–12, 2017, doi: 10.3389/fnagi.2017.00329.

DOI: https://doi.org/10.29207/resti.v6iX.xxx
Creative Commons Attribution 4.0 International License (CC BY 4.0)

JurnalResti5Nov2022

ORIGINALITY REPORT

21%	14%	16%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	repository.mercubuana.ac.id Internet Source	4%
2	Kristiawan Nugroho, Edy Winarno. "Spoofing Detection of Fake Speech Using Deep Neural Network Algorithm", 2022 International Seminar on Application for Technology of Information and Communication (iSemantic), 2022 Publication	3%
3	jurnal.iaii.or.id Internet Source	3%
4	Jinsi Jose, Deepa V Jose, Karna Srinivasa Rao, Justin Janz. "Impact of Machine Learning Algorithms in Intrusion Detection Systems for Internet of Things", 2021 International Conference on Advances in Computing and Communications (ICACC), 2021 Publication	1%
5	krishi.icar.gov.in Internet Source	1%

6	Kristiawan Nugroho, Edi Noersasongko, Purwanto, Muljono, De Rosal Ignatius Moses Setiadi. "Enhanced Indonesian Ethnic Speaker Recognition using Data Augmentation Deep Neural Network", Journal of King Saud University - Computer and Information Sciences, 2022 Publication	1 %
7	repository.usd.ac.id Internet Source	1 %
8	patents.google.com Internet Source	1 %
9	ijece.iaescore.com Internet Source	1 %
10	"Deep Learning Approaches for Spoken and Natural Language Processing", Springer Science and Business Media LLC, 2021 Publication	<1 %
11	Imam Husni Al Amin, Falah Hikamudin Arby, Edy Winarno, Budi Hartono, Wiwien Hadikurniawati. "Real-time Social Distance Detection using YOLO-v5 with Bird-eye View Perspective to Suppress the Spread of COVID-19", 2022 2nd International Conference on Information Technology and Education (ICIT&E), 2022 Publication	<1 %

12	Trisna Ari Roshinta, Hartatik, Elya Kumala Fauziyah, Ivan Fausta Dinata, Nurul Firdaus, Fiddin Yusfida A'la. "A Comparison of Text Classification Methods: Towards Fake News Detection for Indonesian Websites", 2022 1st International Conference on Smart Technology, Applied Informatics, and Engineering (APICS), 2022 Publication	<1 %
13	dokumen.pub Internet Source	<1 %
14	proceeding.researchsynergypress.com Internet Source	<1 %
15	Sarita Limbu, Cyril Zakka, Sivanesan Dakshanamurthy. "Predicting Environmental Chemical Toxicity using a New Hybrid Deep Machine Learning Method", American Chemical Society (ACS), 2021 Publication	<1 %
16	ejournal.st3telkom.ac.id Internet Source	<1 %
17	nebula.wsimg.com Internet Source	<1 %
18	www.science.gov Internet Source	<1 %

19	Xinman Zhang, Dongxu Cheng, Pukun Jia, Yixuan Dai, Xuebin Xu. "An Efficient Android-Based Multimodal Biometric Authentication System with Face and Voice", IEEE Access, 2020 Publication	<1 %
20	"Artificial Intelligence and Speech Technology", Springer Science and Business Media LLC, 2022 Publication	<1 %
21	Communications in Computer and Information Science, 2011. Publication	<1 %
22	Mohamad Irfan, Imam Zainal Mutaqin, Rio Guntur Utomo. "Implementation of Dynamic Time Warping algorithm on an Android based application to write and pronounce Hijaiyah letters", 2016 4th International Conference on Cyber and IT Service Management, 2016 Publication	<1 %
23	Mohammad Ali Humayun, Hayati Yassin, Pg Emeroylariffion Abas. "Native language identification for Indian-speakers by an ensemble of phoneme-specific, and text-independent convolutions", Speech Communication, 2022 Publication	<1 %

24	Teguh Puji Laksono, Ahmad Fathan Hidayatullah, Chanifah Indah Ratnasari. "Speech to Text of Patient Complaints for Bahasa Indonesia", 2018 International Conference on Asian Language Processing (IALP), 2018 Publication	<1%
25	bbrc.in Internet Source	<1%
26	proceedings.stis.ac.id Internet Source	<1%
27	www.ijraset.com Internet Source	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		