

FinalPaper-IJAI-17April2023-1

by Kristiawan Nugroho

Submission date: 17-Apr-2023 07:29AM (UTC+0700)

Submission ID: 2066395680

File name: Paper-Kristiawan-IJAI-Final-17042023.docx (479.57K)

Word count: 4447

Character count: 24731

Improving Indonesian Multiethnic Speaker Recognition Using Pitch Shifting Data Augmentation

Kristiawan Nugroho¹, Isworo Nugroho¹, De Rosal Ignatius Moses Setiadi², Omar Farooq³

¹Department of Information Technology and Industry, Universitas Stikubank, Semarang, Indonesia

²Department of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

³Department of Electronics Engineering, Z. H. College of Engg. and Technology, A. M. U. Aligarh, India

Article Info

Article history:

Received month dd, yyyy

Revised month dd, yyyy

Accepted month dd, yyyy

Keywords:

Speaker Recognition

Deep Learning

Data Augmentation

Pitch Shifting

ABSTRACT

Speaker Recognition to recognize multiethnic speakers is an interesting research topic. Various studies involving many ethnicities require the right approach to achieve optimal model performance. The Deep Learning approach has been used in Speaker Recognition research involving many classes to achieve high accuracy results with promising results. However, multi-class and imbalanced datasets are still obstacles encountered in various studies using the Deep Learning method which cause overfitting and decreased accuracy. Data Augmentation is an approach model used in overcoming the problem of small amounts of data and multiclass problems. This approach can improve the quality of research data according to the method applied. This study proposes a data augmentation method using Pitch Shifting with a Deep Neural Network called PSDA-DNN (Pitch Shifting Data Augmentation Deep Neural Network) to identify multiethnic Indonesian speakers. The results of the research that has been done prove that the PSDA-DNN approach is the best method in Multi-Ethnic Speaker Recognition where accuracy reaches 99.27% and the Precision, Recall, F1 score is 97.60%.

This is an open access article under the CC BY-SA license.



Corresponding Author:

Kristiawan Nugroho

Department of Information Technology and Industry

Universitas Stikubank Semarang

Jl. Tri Lomba Juang, Semarang, Indonesia

Email: kristiawan@edu.unisbank.ac.id

1. INTRODUCTION

Speaker Recognition is one of the challenging research fields. Various kinds of problems need to be solved to produce the discovery of new theories that can make a positive contribution to human life. Research results in the field of Speaker Recognition have proven to have led to various forms of new technology applied to voice authentication, Surveillance Speaker Recognition, Forensic Speaker Recognition, Security, and Multi Speaker Tracking. Tech giant companies such as Google, Microsoft, and Amazon have also taken advantage of this technology, including Google Voice technology, Apple Siri, and Amazon's Alexa.

The application of this technology is proven to be able to help human work, such as in the field of security and authentication smart homes to help people with disabilities to recognize sound patterns and images. Initially, research in the field of Speaker Recognition used classical methods in Machine Learning, GMM (Gaussian Mixture Model), such as studies carried out by Motlicek[1], and Veena[2]. In speaker

recognition[3], the Hidden Markov Model (HMM) strategy is also utilized, as demonstrated by Maghsoodi[3], Hussein[4] and the Support Vector Machine (SVM) approach on research that has been done by Chaunan[5].

However, along with the growth of data that is getting bigger and the complexity of the problems faced today, researchers are starting to use the Deep Learning method in Speaker Recognition research. Deep Learning (DL) is an approach developed from the Neural Network algorithm. This method continues to be developed by researchers to solve various problems in Machine Learning. DL has advantages in terms of the ability to handle computational processes that involve very large data. In addition, DL is also able to process data representations of various forms such as text, images, and sound, which makes it a good ability to process information in the form of multi modal so that this method outperforms the previous Machine Learning method, especially used in the field of Computer Vision.

Various DL methods are used in voice signal processing research using the DNN approach as on research that has been done by Guo[6] combined with I-Vector in research on port speech. This study improves the Equal Error Rate (EER) to 26.47%. In other research Mohan[7] using Self Organizing Map (SOM) and Latent Dirichlet Allocation (LDA) succeeded in increasing crop prediction accuracy by 7-23%. DNN is also used by Saleem[8] in research on Speech Separation, which produces promising performance. However, behind the advantages of DL, this method also has several weaknesses, including higher accuracy performance requiring large data sets, overfitting, and computational process efforts that require large resources. In addition to the need for large data volumes, research in the field of modern classification also faces the problem of multiple classes when it comes to processing an explosion of features and limited data.

In several studies regarding speaker recognition multi ethnic, the problem of limited data is an initial challenge in conducting research such as the study conducted by Hanifa[9], which only had 62 recorded data of Malaysian speakers, and the study of Cole[10] to identify speakers in South East England with 227 speakers. To solve the problem of limited data, various approaches are used, among others, by using Data Augmentation (DA). DA is a method of expanding the quantity of data and has proven to be effective in conducting training on Neural Network. In speech signal processing, DA is proven to increase accuracy in audio classification using Deep Learning[11]. In the field of speech signal processing, several DA methods that are often used are Adding White Noise(AWN), Time Stretching(TS), Pitch Shifting(PS), Mixup and Speech Augment. Some of these DA approaches are proven to be able to increase the quantity of research data so that research using Deep Learning which requires relatively large amounts of data can be done well so as to achieve a high level of accuracy.

Currently, in Machine Learning, the problems faced by researchers in an effort to increase the accuracy of speaker recognition include the number of classes that must be classified in the imbalance dataset. One of the problems encountered in machine learning classification is unbalanced multiclass which will cause model inaccuracies in predicting data. This problem can cause prediction errors in machine learning algorithms, so it needs to be resolved immediately. In research conducted by Khan[12] and Mi[13] the data augmentation approach used the Generative Adversarial Network (GAN) method to solve multi-class problems which achieved an accuracy increase of 6.4%. In several studies related to the implementation of DA in speech signal processing, various methods used include Adding White Noise (AWN), such on research that has been done by Morales[14], which seeks to increase speaker recognition accuracy by adding noise effects.

Jacques[15] also used the approach to adding noise in the research conducted. Time Stretching (TS) is also used by several researchers in speech signal processing research, TS functions to change the speed of audio signal duration as on research that has been done by Sasaki[16] and Aquiar[17] to classify music genres with Convolutional Neural Network (CNN). However, the AWN and TS methods used still cannot produce high accuracy for speech recognition because they only achieve an accuracy level of around 70% to 80%. Another approach used in voice-based augmentation data is Pitch Shifting (PS). PS method is also used in several studies conducted by Morbale[18] in processing audio files, Rai[19] in processing musical instruments and, Ye[20] with CNN using UME and TIMIT datasets. In research conducted by Ye the Pitch Shifting method has achieved an accuracy rate above 90% with the highest accuracy of 98.72%.

This paper aims to improve the performance of the multi-ethnic speaker recognition model with a pitch shifting method based on deep neural networks and consists of several parts, namely the Introduction section in chapter 1, which describes the research problems and other research that has been carried out and the relationship with several other studies. Chapter 2 is a Literature Review that contains a section that contains the study of related topics. In Chapter 3, Proposed Methodology contains the proposed model in solving the problem. Chapter 4 is a Results and Discussion that contains Experiment results from the research that has been done. The final part is chapter 5, namely Conclusion, which contains several conclusions and solutions for the research carried out.

2. METHOD

This study proposed a PSDA-DNN (Pitch Shifting Deep Neural Network) which has the best performance when implemented in voice signal processing supported by MFCC as a feature extraction method

and a DNN that processes multiethnic speaker recognition classification. The search begins with processing the dataset of multiethnic speakers followed by the preprocessing process. Data Augmentation is a step that provides a solution to the limited dataset of multiethnic speakers. The dataset is then extracted using the MFCC approach and then the result of feature extraction is included in the 7 Layer DNN architecture. As the last step, measuring the performance of the proposed framework, among others, uses measures of accuracy, precision, recall, and F1. The proposed method can be seen in figure 1.

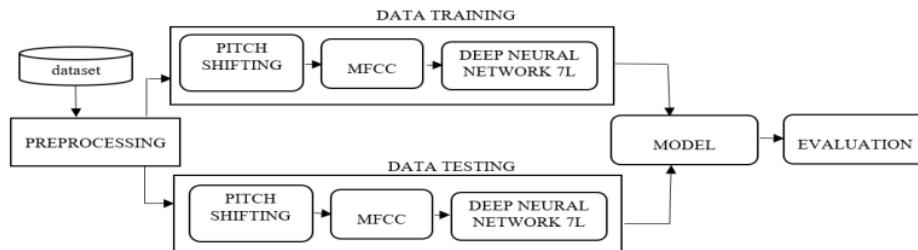


Figure. 1 Proposed Method

2.1. Datasets and Preprocessing

Research on multi-ethnic speaker recognition uses the sample for recognizing ethnic speakers dataset from Indonesia taken from Youtube 301 Language in Indonesia [21]. The dataset was compiled using 70 ethnic male speakers from hundreds of ethnic groups in Indonesia. Sound processing using Adobe Audition CS6, the voice of tribal speakers is carried out by a sampling process by taking 10 tribal speakers each with a duration of 1 second. The sampling process uses a standard sample rate (SR) of 44,100 Hz with 32 bit Depth, Mono 32-bit Floating Point. The next stage is the preprocessing process, which is an important process in data mining processing. Preprocessing is a step that needs to be done in Data Mining to get good data quality to reduce processing time and get the desired results. This study uses the Adobe Audition CS6 application in preprocessing data with Noise Reduction facilities to remove noise in the speaker's speech data.

2.2. Data Augmentation

Data Augmentation (DA) is one method that is often used in increasing the quantity of dataset needed research. DA is an approach that aims to increase the size of the data quantity and is a powerful technique used in the field of data mining and data processing for regression and classification purposes. PS is a DA method in sound signal processing by raising or lowering the original voice pitch in audio without affecting the long duration of the recorded sound. PS is used in this study because it has the advantage that the overall spectral envelope does not change so that it can achieve high-quality output. The process in the PS approach can be seen in Figure 2. In this study, PS results will be compared with 2 other DA methods, namely Adding White Noise (AWN) and Time Stretching (TS) which are widely implemented in various research fields such as sound recording, music production, music learning, and foreign languages. The original speaker's voice signal is processed using the Pitch Shifting approach. The results of processing the voice signal before and after using the PS method can be seen in figure 2 as follows:

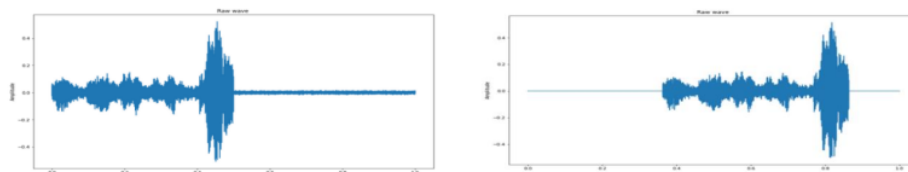


Figure 2. Original and Pitch Shifting Speech Signal

2.3. Feature Extraction and Deep Neural Network Seven Layer

2.3.1 Mel-Frequency Cepstral Coefficient

MFCC is a feature extraction method used in this study. MFCC is a robust approach used in speech recognition. The speech signal of the ethnic speaker consisting of 700 wav voice from 70 Ethnic was extracted using the python application with MFCC settings of frame 900lengths 25, frame shifts 10, window type hamming, preemphasis coefficient 0.97, number of cepstral coefficient 13 and number of lifter 22. In the MFCC approach, the voice signal will be processed through the following steps :

35 Preemphasis

Is a process that will be carried out after the sound sampling process, this process serves to reduce noise at the sound source. The preemphasis process in the time domain can be formulated as:

$$y(n) = x(n) - ax(n-1) \quad (1)$$

a indicates the constant which is at $0.9 < a < 1.0$

2. Frame Blocking

The speech signal will then enter the Frame Blocking process which functions to divide the sound into several parts of the frame.

3. Windowing

Windowing is 2 step to analyze a long signal by taking the right part to be processed in the next stage. If the window is defined as $w(n), 0 < n < N - 1$ where N is the number of samples in each frame, then the windowing signal can be formulated as:

$$x_1(n) = x(n) w(n), 0 \leq n \leq N - 1 \quad (2)$$

4. Fast Fourier Transform (FFT)

Fourier transform is used to convert the time series of signals in the form of a limited time domain into a frequency spectrum, while FFT is part of fast Discrete Fourier Transform (DFT) algorithm that converts frames into N samples starting from the time domain to the frequency domain. The result of this processing is referred to as Cepstrum which is formulated as follows:

$$X_1(n) = \sum_{k=0}^{N-1} x_k e^{-2\pi jkn/N} \quad (3)$$

Where $n = 0, 1, 2, \dots, N-1$ and $j = \sqrt{-1}$ while $X[n]$ is an n-frequency form resulting from the Fourier Transform mechanism.

5. Mel-Frequency wrapping

In this step the existing FFT signals are grouped in a triangular filter which aims to multiply the FFT value with the appropriate filter gain which then the results will be summed. The wrapping process into a signal in the frequency domain can be formulated by:

$$x_i = \log_{10}(\sum_{k=0}^{N-1} |x(k) H_i(k)|) \quad (4)$$

Where $i = 1, 2, 3, \dots, M$, M is the sum of the triangle filters and $H_i(k)$ is the value for the triangular filter for the acoustic frequency k.

6. Cepstrum

In order for the signal to be heard by humans, it is necessary to convert the signal into a time domain using a Discrete Cosine Transform (DCT). The final result of this process is referred to as Mel Frequency Cepstral Coefficients which is formulated as follows:

$$c_j = \sum_{i=1}^K x_i \cos(j(i-1)/2\pi/K) \quad (5)$$

C_j shows the MFCC coefficient. X_j is the strength of the Mel Frequency spectrum, $j = 1, 2, 3, \dots, K$ is the expected coefficient and M indicates the number of filters.

2.3.2 Deep Neural Network Seven Layer (DNN7L)

One of the finest Deep Learning techniques is the Deep Neural Network (DNN). DNN has the advantage of building more accurate models. In this work used DNN 7 layers with architectural network as in Table 1. The DNN architecture used in this study consists of seven layers consisting of dense or also called fully connected layers, indicating that the layer in which there are neurons connected to neurons in the previous layer. Layer 1 consists of 193 nodes which is an input layer that shows 193 features generated from the extracted features. Between layers is given a dropout function which is a technique used to solve overfitting problems and prediction problems in large neural networks. Layers two to seven use half of the number of nodes in the previous layer in order to reduce the complexity of calculations on each layer.

36

Table 1. DNN7L Architecture

Layer (type)	Output Shape	Layer (type)	Output Shape
Dense 1 (dense)	(None,193)	Dense 5 (dense)	(None,50)
Dense 2 (dense)	(None,400)	Dropout_4 (Dropout)	(None,50)
Dropout_13 (Dropout)	(None,400)	Dense 6 (dense)	(None,25)
Dense 3 (dense)	(None,200)	Dropout_5 (Dropout)	(None,25)
Dropout_2 (Dropout)	(None,200)	Dense 7 (dense)	(None,15)
Dense 4 (dense)	(None,100)	Dropout_6 (Dropout)	(None,15)
Dropout_3 (Dropout)	(None,100)	Dense 8 (dense)	(None,8)

2.4. Evaluation

The result of the process of Indonesian ethnic speakers recognition is an evaluation process by measuring the performance of the proposed model by evaluating the level of Accuracy, Precision, Recall, and Measure. Accuracy is the ratio of the number of cases that were predicted with correct answers compared to the total number of cases, while recall is the ratio of number of positive cases that were predicted correctly to the number of positive cases that were predicted. Accuracy, precision, recall and F1-score measure can be calculated using the Eq. (6), (7), (8) and (9), respectively.

$$\frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (6)$$

$$\frac{TP}{(TP+FP)} \times 100\% \quad (7)$$

$$\frac{TP}{(TP+FN)} \times 100\% \quad (8)$$

$$\frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (9)$$

Where TP, TN, FP, and FN stand for True Positive, True Negative, False Positive, and False Negative, respectively. Performance measurement in this study also uses recall which is the ratio of true positive cases that are predicted to be positive.

3. RESULTS AND DISCUSSION

Dataset on multi-ethnic speakers in Indonesia were tested using Data Augmentation approach, namely AWN, PS and TS then trained using DNN using a split ratio of 70:30, 80:20 and 90:10. The test results show that the PSDA-DNN (Pitch Shifting Data Augmentation Deep Neural Network) has better performance than AWNDA- DNN (Adding White Noise Data Augmentation Deep Neural Network), and TSDA-DNN (Time Stretching Data Augmentation Deep Neural Network) methods. The following are the results of the PSDA-DNN model performance testing process using a 70:30 split ratio as depicted in figure 3.

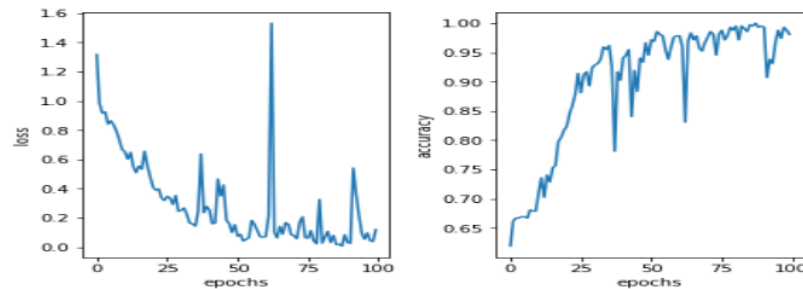


Figure 3. PSDA-DNN performance

Testing on the 70:30 split ratio resulted in an Accuracy level of 98.55%, Precision, Recall and F1 measure each of 94.37%. This performance result shows that PSDA-DNN will be more robust when used on larger data. Classification using various Machine Learning methods using many classes is not easy and cause various problems in learning[22]. An appropriate approach is needed in managing the dataset. The following is a comparison of the number of classes processed between the various types of research shown in table 2.

Table 2. Comparison with another datasets

Methods	Datasets	Σ Class	Acc (%)
Support Vector Machine/SVM[9]	Ethnicity of Malaysian Dataset	4	57.7
Convolutional Neural Network/CNN [23]	Urdu Speakers	4	87.5
Deep Belief Network/DBN [24]	Accented Spoken English Corpus	6	90.2

Deep Neural Network [25]	TITML-IDN, OpenSLR	4	98.9
Proposed Method (PSDA-DNN)	Indonesian Multiethnics Speakers	42	99.2

Based on the comparison results presented in table 2, even though it has more classes, the proposed model, namely PSDA-DNN, produces better model performance than other Machine Learning methods with an accuracy of up to 99.2%. The achievement of a high level of performance at PSDA-DNN was due to a good preprocessing process on the speech signal dataset which was then carried out by data augmentation process with Pitch Shifting and proper classification using the Deep Neural Network 7 layer approach. The PSDA-DNN method also produces the most effective performance in comparison to various other techniques using the Indonesian Multiethnics Speakers dataset. The results of the comparison of these methods can be seen in table 3 as follows:

Table 3. Comparison with another methods

Methods	Accuracy
K-Nearest Neighbor	92 %
Random-Forest	81 %
Deep Neural Network	98.4 %
PSDA-DNN (Ours)	99.2 %

The results of this study will be compared with several methods with classical machine learning and deep learning approaches. Table 3 shows that the PSDA-DNN method has better performance compared to other methods. In several comparisons of the performance of the methods as previously presented, it can be concluded that the proposed method has the best performance.

4. CONCLUSION

Study in the area of speaker identification is an interesting topic and challenges researchers around the world to work hard to make new scientific contributions, including research on Indonesian multiethnic speaker recognition. The Deep Learning method is an approach that has been chosen, especially for processing large amounts of data, including sound signal processing. However, the problem of multiple classes and data imbalance causes low accuracy because the model performance is not optimal. The Pitch Shifting Approach in Data Augmentation is a solution in increasing the quantity of data and as a solution for multiple class classification problems. Obtaining a high model accuracy performance of 99.27% through the proposed model, namely PS-DNN is one of the solutions to overcome the problem of multiple classes and imbalanced datasets in Machine Learning. This study proposes the PSDA-DNN approach which is a multi-ethnic speaker recognition method that uses the Pitch Shifting Data Augmentation technique which is supported by the MFCC and Deep Neural Network methods in processing speech signals. The research results show that PSDA-DNN has better performance compared to other approaches such as Adding White Noise (AWN) and Time Stretching (TS) which are also DNN-based in processing speech signals. The PSDA-DNN approach produces an average level of Accuracy of 99.27%, Precision, Recall, and F1 Measure of 97.60%.

REFERENCES

- [1] P. Motlicek, S. Dey, S. Madikeri, and L. Burget, "Employment of Subspace Gaussian Mixture Models in speaker recognition," *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 2015-Augus, pp. 4445–4449, 2015, doi: 10.1109/ICASSP.2015.7178811.
- [2] K. V. Veena and D. Mathew, "Speaker identification and verification of noisy speech using multitaper MFCC and Gaussian Mixture models," *Proc. 2015 IEEE Int. Conf. Power, Instrumentation, Control Comput. PCCC 2015*, no. Lc, 2016, doi: 10.1109/PICC.2015.7455806.
- [3] N. Maghsoodi, H. Sameti, H. Zeinali, and T. Stafylakis, "Speaker recognition with random digit strings using uncertainty normalized HMM-Based i-Vectors," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 11, pp. 1815–1825, 2019, doi: 10.1109/TASLP.2019.2928143.
- [4] J. S. Hussein, A. A. Salman, and T. R. Saeed, "Arabic speaker recognition using HMM," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 23, no. 2, pp. 1212–1218, Aug. 2021, doi: 10.11591/ijeecs.v23.i2.pp1212-1218.
- [5] N. Chauhan, T. Isshiki, and D. Li, "Speaker recognition using LPC, MFCC, ZCR features with ANN and SVM classifier for large input database," *2019 IEEE 4th Int. Conf. Comput. Commun. Syst. ICCCS 2019*, pp. 130–133, 2019, doi: 10.1109/ICCCS.2019.9000000.




- 10.1109/CCOMS.2019.8821751.
- [6] J. Guo *et al.*, "Deep neural network based i-vector mapping for speaker verification using short utterances," *Speech Commun.*, vol. 105, pp. 92–102, 2018, doi: 10.1016/j.specom.2018.10.004.
- [7] P. Mohan, "Deep Learning Based Weighted SOM to Forecast Weather and Crop Prediction for Agriculture Application," vol. 11, no. 4, 2018, doi: 10.22266/ijies2018.0831.17.
- [8] N. Saleem and M. Irfan Khattak, "Deep neural networks based binary classification for single channel speaker independent multi-talker speech separation," *Appl. Acoust.*, vol. 167, p. 107385, 2020, doi: 10.1016/j.apacoust.2020.107385.
- [9] R. M. Hanifa, K. Isa, and S. Mohamad, "Speaker ethnic identification for continuous speech in Malay language using pitch and MFCC," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 19, no. 1, pp. 207–214, 2020, doi: 10.11591/ijeeecs.v19.i1.pp207-214.
- [10] A. Cole, "Identifications of Speaker Ethnicity in South-East England: Multicultural London English as a Divisible Perceptual Variety," *Proc. Lr. 2020 Work. "Citizen Linguist. Lang. Resour. Dev.*, pp. 49–57, 2020, [Online]. Available: <https://www.aclweb.org/anthology/2020.cllrd-1.7>
- [11] L. Nanni, G. Maguolo, and M. Paci, "Data augmentation approaches for improving animal audio classification," *Ecol. Inform.*, vol. 57, no. December 2019, p. 101084, 2020, doi: 10.1016/j.ecoinf.2020.101084.
- [12] M. Heenaye-Mamode Khan *et al.*, "Multi-Class Skin Problem Classification Using Deep Generative Adversarial Network (DGAN)," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/1797471.
- [13] Q. Mi, Y. Hao, M. Wu, and L. Ou, "An Enhanced Data Augmentation Approach to Support Multi-Class Code Readability Classification," *Proc. Int. Conf. Softw. Eng. Knowl. Eng. SEKE*, pp. 48–53, 2022, doi: 10.18293/SEKE2022-130.
- [14] N. Morales, Nuance, and Y. Gao, "Adding Noise to Improve Noise Robustness in Speech Recognition," no. January 2007, 2014.
- [15] C. Jacques and A. Roebel, "Data augmentation for drum transcription with convolutional neural networks," *Eur. Signal Process. Conf.*, vol. 2019-Sept, 2019, doi: 10.23919/EUSIPCO.2019.8902980.
- [16] T. Sasaki *et al.*, "Time stretching: Illusory lengthening of filled auditory durations," *Attention, Perception, Psychophys.*, vol. 72, no. 5, pp. 1404–1421, 2010, doi: 10.3758/APP.72.5.1404.
- [17] L. Rafael Aguiar, M. G. Yandre Costa, and N. Carlos Silla, "Exploring Data Augmentation to Improve Music Genre Classification with ConvNets," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2018-July, pp. 1–8, 2018, doi: 10.1109/IJCNN.2018.8489166.
- [18] P. R. Morbale and M. Navale, "Design and Implementation of Real Time Audio Pitch Shifting on FPGA," no. 02, pp. 81–88, 2015.
- [19] A. Rai and B. D. Barkana, "Analysis of three pitch-shifting algorithms for different musical instruments," *2019 IEEE Long Isl. Syst. Appl. Technol. Conf. LISAT 2019*, pp. 1–6, 2019, doi: 10.1109/LISAT.2019.8817334.
- [20] Y. Ye, L. Lao, D. Yan, and R. Wang, "Identification of Weakly Pitch-Shifted Voice Based on Convolutional Neural Network," *Int. J. Digit. Multimed. Broadcast.*, vol. 2020, 2020, doi: 10.1155/2020/8927031.
- [21] Indonesia Ideas, "301 Languages In Indonesia - Bahasa Logat Dialek Daerah #2," 2019.
- [22] M. H. Xue, Yanbing, "Active Learning of Multi-class Classification Models from Ordered Class Sets," *Proc Conf AAAI Artif Intell*, 2019.
- [23] A. Ashar, M. S. Bhatti, and U. Mushtaq, "Speaker Identification Using a Hybrid CNN-MFCC Approach," *2020 Int. Conf. Emerg. Trends Smart Technol. ICETST 2020*, 2020, doi: 10.1109/ICETST49965.2020.9080730.
- [24] R. Upadhyay and S. Lui, "Foreign English Accent Classification Using Deep Belief Networks," *Proc. - 12th IEEE Int. Conf. Semant. Comput. ICSC 2018*, vol. 2018-Janua, pp. 290–293, 2018, doi: 10.1109/ICSC.2018.00053.
- [25] K. Azizah, M. Adriani, and W. Jatmiko, "Hierarchical Transfer Learning for Multilingual, Multi-Speaker, and Style Transfer DNN-Based TTS on Low-Resource Languages," *IEEE Access*, vol. 8, pp. 179798–179812, 2020, doi: 10.1109/access.2020.3027619.

ACKNOWLEDGEMENTS




We would like thank to the Universitas Stikubank for granted funding through scientific publication incentives.

BIOGRAPHIES OF AUTHORS






Kristiawan Nugroho    works as a lecturer at the faculty of information technology and industry, Universitas Stikubank. He obtained a bachelor's degree in 2001 in the information systems department, Faculty of Computer Science, Universitas Dian Nuswantoro Semarang, then in 2007 He graduated from Universitas Dian Nuswantoro with a master's degree in informatics engineering. He also obtained Doctoral degree in computer science with a concentration in Machine Learning and Artificial Intelligence in 2022 at Dian Nuswantoro University Semarang. He has conducted various researches in machine learning, speech recognition and sentiment analysis. He can be contacted via email kristiawan@edu.unisbank.ac.id.






Isworo Nugroho    works as a lecturer at the faculty of information technology and industry, Universitas Stikubank. He obtained a bachelor's degree in 2001 in the management department, Faculty of Economics, Stikubank University, then in 2003 he obtained a Master's degree in computer science, Gajah Mada University. He has conducted various researches in text processing, data mining and statistical science. He can be contacted via email isworo@edu.unisbank.ac.id.



De Rosal Ignatius Moses Setiadi    a Bachelor of Science in Informatics Engineering from Universitas Soegijaprana, Indonesia, and a Master of Science in Informatics Engineering from Universitas Dian Nuswantoro Semarang, both in 2012. He is presently a lecturer and researcher at the Faculty of Computer Science at Universitas Dian Nuswantoro in Semarang, Indonesia. He has written more than 138 peer-reviewed journal and conference publications that Scopus has indexed. His areas of interest in study include machine learning, cryptography, image steganography, and watermarking. His email address is moses@dsn.dinus.ac.id.



Prof Omar Farooq    Omar Farooq joined the Department of Electronics Engineering, AMU Aligarh as Lecturer in 1992 and is currently working as a Professor. He was awarded Commonwealth Scholar from 1999-2002 towards PhD at Loughborough University, UK, and a one-year postdoctoral fellowship with the UKIERI in 2007-2008. With a focus on speech recognition, signal processing is his broad area of study interest. He has approximately 250 publications published by him or with him in reputable academic journals and conference proceedings, and he has helped 9 scholars complete their PhDs. He is a Senior Member, Institute of Electrical and Electronics Engineers, (IEEE, USA). He can be contacted via email omar.farooq@amu.ac.in.

ORIGINALITY REPORT

20%

SIMILARITY INDEX

12%

INTERNET SOURCES

17%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

- | | | |
|--|---|-----------|
| <div style="background-color: red; color: white; padding: 5px; text-align: center; width: 40px; height: 40px; line-height: 40px;">1</div> | <p>Dyah Anggraeni, W. S. Mada Sanjaya, Madinatul Munawwaroh, M. Yusuf Solih Nurasyidiek, Ikhsan Purnama Santika. "Control of robot arm based on speech recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and K-Nearest Neighbors (KNN) method", 2017 International Conference on Advanced Mechatronics, Intelligent Manufacture, and Industrial Automation (ICAMIMIA), 2017</p> <p>Publication</p> | <p>2%</p> |
| <hr/> | | |
| <div style="background-color: purple; color: white; padding: 5px; text-align: center; width: 40px; height: 40px; line-height: 40px;">2</div> | <p>www.researchgate.net</p> <p>Internet Source</p> | <p>1%</p> |
| <hr/> | | |
| <div style="background-color: purple; color: white; padding: 5px; text-align: center; width: 40px; height: 40px; line-height: 40px;">3</div> | <p>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, 2021</p> <p>Publication</p> | <p>1%</p> |
-

4

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

1 %

5

digilib.uinsgd.ac.id

Internet Source

1 %

6

Guobo Xie, Changhai Wu, Guosheng Gu, Bin Huang. "HAUBRW: Hybrid algorithm and unbalanced bi-random walk for predicting lncRNA-disease associations", Genomics, 2020

Publication

1 %

7

ijeecs.iaescore.com

Internet Source

1 %

8

Herlambang Setiadi, Muhammad Abdillah, Yusrizal Afif, Rezi Delfianti. "Adaptive virtual inertia controller based on machine learning for superconducting magnetic energy storage for dynamic response enhanced", International Journal of Electrical and Computer Engineering (IJECE), 2023

Publication

1 %

9

Edy Winarno, Imam Husni Al Amin, Wiwien Hadikurniawati. "Asymmetrical Half-join Method on Dual Vision Face Recognition",

1 %

10

www.cscjournals.org

Internet Source

<1 %

11

www.iosrjournals.org

Internet Source

<1 %

12

I Putu Agus Eka Darma Udayana, Made Sudarma, I Ketut Gede Darma Putra, I Made Sukarsa. "Effect on signal magnitude thresholding on detecting student engagement through EEG in various screen size environment", Bulletin of Electrical Engineering and Informatics, 2023

Publication

<1 %

13

Biswajeet Pradhan, Ratiranjan Jena, Debojit Talukdar, Manoranjan Mohanty et al. "A New Method to Evaluate Gold Mineralisation-Potential Mapping Using Deep Learning and an Explainable Artificial Intelligence (XAI) Model", Remote Sensing, 2022

Publication

<1 %

14

D Anggraeni, W S M Sanjaya, M Y S Nurasyidiek, M Munawwaroh. "The Implementation of Speech Recognition using Mel-Frequency Cepstrum Coefficients (MFCC) and Support Vector Machine (SVM) method based on Python to Control Robot Arm", IOP

<1 %

Conference Series: Materials Science and Engineering, 2018

Publication

15

Natnael Tilahun Sinshaw, Beakal Gizachew Assefa, Sudhir Kumar Mohapatra. "Transfer Learning and Data Augmentation Based CNN Model for Potato Late Blight Disease Detection", 2021 International Conference on Information and Communication Technology for Development for Africa (ICT4DA), 2021

Publication

<1 %

16

ipfs.io

Internet Source

<1 %

17

www.amu.ac.in

Internet Source

<1 %

18

ismej.iitd.ac.in

Internet Source

<1 %

19

id.123dok.com

Internet Source

<1 %

20

serisc.org

Internet Source

<1 %

21

www.ijsr.net

Internet Source

<1 %

22

"Advances in Soft Computing", Springer Science and Business Media LLC, 2020

Publication

<1 %

23

Bartłomiej Nita, Krzysztof Nowosielski, Zdzisław Kes, Olena Sidor et al. "Machine learning in the enrolment management process: a case study of using GANs in postgraduate students' structure prediction", Procedia Computer Science, 2022

Publication

<1 %

24

Edy Winarno, Agus Harjoko, Aniati Murni Arymurthy, Edi Winarko. "Face Recognition Based on Symmetrical Half-Join Method using Stereo Vision Camera", International Journal of Electrical and Computer Engineering (IJECE), 2016

Publication

<1 %

25

Muhammad Hatta Rahmatul Kholiq, Wiranto Wiranto, Sari Widya Sihwi. "News classification using light gradient boosted machine algorithm", Indonesian Journal of Electrical Engineering and Computer Science, 2022

Publication

<1 %

26

Samsuryadi Samsuryadi, Rudi Kurniawan, Fatma Susilawati Mohamad. "Automated handwriting analysis based on pattern recognition: a survey", Indonesian Journal of Electrical Engineering and Computer Science, 2021

Publication

<1 %

27

Internet Source

<1 %

28

Vincent Didiek Wiet Aryanto, Yohan Wismantoro, Karis Widyatmoko.

"Implementing Eco-Innovation by Utilizing the Internet to Enhance Firm's Marketing Performance", International Journal of E-Business Research, 2018

Publication

<1 %

29

github.com

Internet Source

<1 %

30

repo.lib.tokushima-u.ac.jp

Internet Source

<1 %

31

Ali Bou Nassif, Ismail Shahin, Imtinan Attali, Mohammad Azzeh, Khaled Shaalan.. "Speech Recognition Using Deep Neural Networks: a Systematic Review", IEEE Access, 2019

Publication

<1 %

32

Junhao Zheng, Xiaoming Chen, Xiaobo Liu, Ming Zhang, Binhui Liu, Yi Huang. "An Improved Method for Reconstructing Antenna Radiation Pattern in a Loaded Reverberation Chamber", IEEE Transactions on Instrumentation and Measurement, 2022

Publication

<1 %

33

dyuthi.cusat.ac.in

Internet Source

<1 %

34	jurnal.ugm.ac.id Internet Source	<1 %
35	shareok.org Internet Source	<1 %
36	Hitesh Tekchandani, Shrish Verma, Narendra Londhe. "Performance improvement of mediastinal lymph node severity detection using GAN and Inception network", Computer Methods and Programs in Biomedicine, 2020 Publication	<1 %
37	archive.org Internet Source	<1 %
38	dspace.lib.cranfield.ac.uk Internet Source	<1 %
39	ijece.iaescore.com Internet Source	<1 %
40	ijesrt.com Internet Source	<1 %
41	journal.waocp.org Internet Source	<1 %
42	link.springer.com Internet Source	<1 %
43	www.ijitee.org Internet Source	<1 %

45

Hari Soetanto, Sri Hartati, Retyanto Wardoyo, Samekto Wibowo. "Hypertension Drug Suitability Evaluation Based On Patient Condition with Improved Profile Matching", Indonesian Journal of Electrical Engineering and Computer Science, 2018

Publication

<1 %

46

Pooja Singh, Sushil Kumar Mishra. "Alzheimer's Detection And Categorization using a Deep-Learning Approach", 2022 Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT), 2022

Publication

<1 %

47

Xinyu Zhang, Yang Xu, Sicong Zhang, Xiaojian Li. "A Highly Stealthy Adaptive Decay Attack Against Speaker Recognition", IEEE Access, 2022

Publication

<1 %

48

"Recent Advances on Soft Computing and Data Mining", Springer Science and Business Media LLC, 2017

Publication

<1 %

49

Mohd Wajid, Meenu Kumari, Mohammed Usman, Omar Farooq. "Estimation of Mixing

<1 %

Coefficients using the Joint Probability Distribution of Mixed Signals", IETE Journal of Research, 2019

Publication

Exclude quotes On

Exclude matches Off

Exclude bibliography On