PENGEMBANGAN MODEL KLASIFIKASI SUARA DALAM IDENTIFIKASI IRAMA AZAN MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK* (CNN)

SKRIPSI

Disusun Oleh:

RANGGA GALANG PRAMUDYA

09020621043

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN AMPEL

SURABAYA

2026

LEMBAR PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama	:	RANGGA GALANG PRAMUDYA
NIM	:	09020621043
Prodi	:	SISTEM INFORMASI
Angkatan	:	2021

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul " PENGEMBANGAN MODEL KLASIFIKASI SUARA DALAM IDENTIFIKASI IRAMA AZAN MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat maka bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 30 Juni 2026

Yang menyatakan,

Rangga Galang Pramudya
NIM. 09020621043

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh:

Nama : RANGGA GALANG PRAMUDYA

NIM : 09020621043

Judul : PENGEMBANGAN MODEL KLASIFIKASI SUARA DALAM IDENTIFIKASI IRAMA AZAN MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)

Ini telah diperiksa dan disetujui untuk diujikan.

Surabaya, 4 Mei 2026

Dosen Pembimbing 1

Dr. Eng. Anang Kunaefi, M. Kom
NIP. 197911132014031001

Dosen Pembimbing 2

Khalid, M. Kom.
NIP. 197906092014031002

LEMBAR PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi Rangga Galang Pramudya ini telah dipertahankan

Di depan tim penguji Skripsi

Di Surabaya, 4 Mei 2026

Mengesahkan,

Dewan Penguji

Menyetujui,

Dosen Penguji 1

Bayu Adhi Nugroho, Ph.D.
NIP. 197905182014031001

Dosen Penguji 2

Mujib Ridwan, S.Kom, M.T.
NIP. 198604272014031004

Dosen Penguji 3

Dr. Eng. Anang Kunaefi, M. Kom
NIP. 197911132014031001

Dosen Penguji 4

Khalid, M. Kom.
NIP. 197906092014031002

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya

Dr. Jamal Hamdani, M.Pd
NIP. 196507312000031002

UIN SUNAN AMPEL
S U R A B A Y A

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN
Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : RANGGA GALANG PRAMUDYA
NIM : 09020621043
Fakultas/Jurusan : SAINS DAN TEKNOLOGI / SISTEM INFORMASI
E-mail address : ranggagalang07@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :
■ Sekripsi ☐ Tesis ☐ Desertasi ☐ Lain-lain (.....)
yang berjudul :

PENGEMBANGAN MODEL KLASIFIKASI SUARA DALAM IDENTIFIKASI IRAMA AZAN MENGGUNAKAN CONVOLUTIONAL NEURAL NETWORK (CNN)

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Ekslusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara ***fulltext*** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 30 Juni 2026

Penulis

Rangga Galang Pramudya

ABSTRAK

PENGEMBANGAN MODEL KLASIFIKASI SUARA DALAM IDENTIFIKASI IRAMA AZAN MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK* (CNN)

Oleh:
Rangga Galang Pramudya

Azan merupakan bagian penting dalam praktik ibadah Islam yang dilantunkan menggunakan pola melodi khas yang dikenal sebagai *maqam*. Setiap *maqam* memiliki karakteristik tonal dan ritmis yang unik, sehingga menyulitkan pembelajar untuk membedakan berbagai gaya irama azan secara akurat. Keterbatasan sumber belajar yang terstruktur semakin meningkatkan kesulitan dalam menguasai pola-pola tersebut. Oleh karena itu, klasifikasi otomatis terhadap irama azan menggunakan pendekatan teknologi dapat mendukung proses pembelajaran. Penelitian ini mengusulkan pendekatan berbasis *deep learning* menggunakan *Convolutional Neural Network* (CNN) untuk klasifikasi irama azan. Dataset yang digunakan terdiri atas 250 sampel audio asli yang diperluas menjadi 2.500 sampel melalui teknik augmentasi data, meliputi *pitch shifting*, *speed perturbation*, penambahan *Gaussian noise*, dan penyesuaian volume. Beberapa metode ekstraksi fitur audio, yaitu *Mel-Frequency Cepstral Coefficients* (MFCC), *Mel Spectrogram*, dan *Chroma*, digunakan sebagai representasi masukan bagi model. Hasil eksperimen menunjukkan bahwa MFCC memberikan kinerja terbaik dengan akurasi sebesar 0,86 dan *F1-score* sebesar 0,8609, yang menunjukkan kemampuan generalisasi yang baik. Fitur *Chroma* juga menunjukkan hasil yang kompetitif, sedangkan *Mel Spectrogram* menghasilkan kinerja yang lebih rendah karena kompleksitas representasinya yang lebih tinggi. Selain itu, penggabungan beberapa fitur tidak meningkatkan performa klasifikasi dan pada beberapa kasus justru menurunkan akurasi. Temuan ini menegaskan pentingnya pemilihan representasi fitur yang tepat untuk memperoleh kinerja model yang optimal..

Kata kunci: *Klasifikasi irama azan, Convolutional Neural Network (CNN), Mel-Frequency Cepstral Coefficients (MFCC), ekstraksi fitur audio, augmentasi data.*

ABSTRACT

DEVELOPMENT OF A VOICE CLASSIFICATION MODEL FOR ADHAN RHYTHM IDENTIFICATION USING CONVOLUTIONAL NEURAL NETWORKS (CNN)

Oleh:
Rangga Galang Pramudya

Adhan is an essential part of Islamic worship, delivered through distinctive melodic patterns known as *maqam*. Each *maqam* possesses unique tonal and rhythmic characteristics, making it challenging for learners to accurately distinguish among different adhan rhythms. Furthermore, the limited availability of structured learning resources increases the difficulty of mastering these melodic patterns. Therefore, an automatic adhan rhythm classification system can serve as a valuable tool to support the learning process. This study proposes a Convolutional Neural Network (CNN)-based approach for adhan rhythm classification. The dataset consists of 250 original audio recordings, which were expanded to 2,500 samples using data augmentation techniques, including pitch shifting, speed perturbation, Gaussian noise addition, and volume perturbation. Three audio feature extraction methods—Mel-Frequency Cepstral Coefficients (MFCC), Mel Spectrogram, and Chroma—as well as several feature fusion combinations, were evaluated as input representations for the CNN model. Experimental results demonstrate that MFCC achieved the best classification performance, with an accuracy of 86.00% and an F1-score of 0.8609, indicating its effectiveness in capturing the discriminative characteristics of adhan rhythms. Chroma also produced competitive results, whereas Mel Spectrogram yielded lower performance. Moreover, feature fusion approaches did not improve the classification performance and, in several cases, reduced the model's accuracy due to increased feature complexity and redundancy. These findings demonstrate that selecting an appropriate feature representation has a greater impact on classification performance than increasing feature complexity, making MFCC the most suitable feature extraction method for CNN-based adhan rhythm classification.

Kata kunci: *Adhan rhythm classification, Convolutional Neural Network (CNN), Mel-Frequency Cepstral Coefficients (MFCC), audio feature extraction, data augmentation.*

Daftar Isi

Daftar Isi ix
Daftar Tabel xii
Daftar Gambar xiii
BAB I PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Perumusan Masalah 3
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 4
1.5 Manfaat Penelitian 5
BAB II TINJAUAN PUSTAKA 6
2.1 Tinjauan Penelitian Terdahulu 6
2.2 Teori Dasar 11
2.2.1 Irama Azan 11
2.2.2 Voice Recognition 12
2.2.3 Augmentasi Data 13
2.2.4 Convolutional Neural Network (CNN) 20
2.2.5 MFCC 22
2.2.6 Chroma 23
2.2.7 Mel-Spectogram 26
2.2.8 Feature Fusion 28
2.2.8 Kfold Cross-Validation 29
2.2.9 Confusion Matrix 31
2.3 Integrasi Keilmuan 34
BAB III METODELOGI PENELITIAN 36

3.1 Desain Penelitian 36
3.1.1 Perumusan Masalah 37
3.1.2 Studi Literatur 37
3.1.3 Pengumpulan Data 37
3.1.5 Augmentation Data 38
3.1.6 Feature Selection 38
3.1.7 Validasi Data 41
3.1.8 Training Model CNN 42
3.1.9 Pengujian 42
3.1.10 Analisis Model 43
BAB IV HASIL DAN PEMBAHASAN 45
4.1 Hasil 45
4.1.1 Preprocessing 45
4.1.1.1 Pitch Shift 45
4.1.1.2 Speed 47
4.1.1.3 Noise 49
4.1.1.4 Volume 50
4.1.2 Validasi Data 51
4.1.3 Ektraksi Fitur 52
4.1.3.1 MFCC (*Mel-Frequency Cepstral Coefficients*) 53
4.1.3.2 Chroma 54
4.1.3.3 Mel Spectogram 56
4.1.3.4 MFCC + Mel Spectogram 57
4.1.3.5 Chroma + Mel Spectogram 59
4.1.3.6 Chroma + MFCC 61
4.1.3.7 Chroma + Mel Spectogram + MFCC 63
4.1.4 Pemuatan Dataset dan Pelabelan 66
4.1.5 Padding 68
4.1.6 Encoding Label 69
4.1.7 Arsitektur Model Convulutional Neural Network (CNN) 69
4.1.8 Pengujian K-Fold Cross Validation 70

4.1.9 Pengujian Model pada Data Testing 76
4.2 Analisis Hasil 77
4.2.1 Perbandingan Performa berdasarkan Fitur 77
4.2.2 Analisis Learning Curve 80
4.2.3 Interpretasi 83
BAB V KESIMPULAN DAN SARAN 87
5.1 Kesimpulan 87
5.2 Saran 88
DAFTAR PUSTAKA 89
Lampiran 96

Daftar Tabel

Tabel 2. 1 Penelitian Terdahulu 6
Tabel 3. 1 Jadwal penilitan..... **Error! Bookmark not defined.**
Tabel 4. 1 Validasi Data..... 52
Tabel 4. 2 Hasil Pelatihan 73
Tabel 4. 3 Hasil Testing 77
Tabel 4. 4 Hasil Pelatihan dan Testing..... 78

Daftar Gambar

Gambar 2. 1 Arsitektur CNN 21
Gambar 2. 2 Logaritma Mel Spectogram 27
Gambar 2. 3 Arsitektur K fold Cross-validation 30
Gambar 2. 4 Confusion Matrix 32
Gambar 3. 1 Tahapan Penelitian 36
Gambar 4. 1 Spektogram MFCC 54
Gambar 4. 2 Spektogram Chroma 55
Gambar 4. 3 Spektogram Mel 57
Gambar 4. 4 Spektogram Mel + MFCC 59
Gambar 4. 5 Spektogram Mel + Chroma 61
Gambar 4. 6 Spektogram MFCC + Chroma 63
Gambar 4. 7 Spektogram MFCC + Chroma + Mel 66
Gambar 4. 8 Learning curve MFCC 75
Gambar 4. 9 Learning curve Mel spectogram 80
Gambar 4. 10 Learning curve MFCC 80
Gambar 4. 11 Learning curve Chroma 80
Gambar 4. 12 Learning curve Mel + Chroma 80
Gambar 4. 13 Learning curve Mel + MFCC 80
Gambar 4. 14 Learning curve Chroma + MFCC 80
Gambar 4. 15 Learning curve Mel + MFCC + Chroma 81

DAFTAR PUSTAKA

Al Bakri, T., Mallah, M., & Nuserat, N. (2019). Al Adhan: Documenting historical background, practice rules, and musicological features of the Muslim call for prayer in Hashemite Kingdom of Jordan. *Musicologica Brunensia*, *54*(1), 167–185. https://doi.org/10.5817/MB2019-1-12

Arslan, M., Guzel, M., Demirci, M., & Ozdemir, S. (2019). SMOTE and Gaussian Noise Based Sensor Data Augmentation. *UBMK 2019 - Proceedings, 4th International Conference on Computer Science and Engineering*, 458–462. https://doi.org/10.1109/UBMK.2019.8907003

Atrey, P. K., Hossain, M. A., Saddik, A. El, & Kankanhalli, M. S. (2010). *Multimodal fusion for multimedia analysis : a survey*. https://doi.org/10.1007/s00530-010-0182-0

Azmi, K., Defit, S., & Sumijan, S. (2023). Implementasi Convolutional Neural Network (CNN) Untuk Klasifikasi Batik Tanah Liat Sumatera Barat. *Jurnal Unitek*, *16*(1), 28–40. https://doi.org/10.52072/unitek.v16i1.504

Beinecke, J., & Heider, D. (2021). Gaussian noise up-sampling is better suited than SMOTE and ADASYN for clinical decision making. *BioData Mining*, *14*(1), 1–11. https://doi.org/10.1186/s13040-021-00283-6

Chu, H. C., Zhang, Y. L., & Chiang, H. C. (2023). A CNN Sound Classification Mechanism Using Data Augmentation. *Sensors*, *23*(15). https://doi.org/10.3390/s23156972

Deng, X., Liu, Q., Deng, Y., & Mahadevan, S. (2016). An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, *340–341*, 250–261. https://doi.org/10.1016/j.ins.2016.01.033

Dou, H. X., Lu, X. S., Wang, C., Shen, H. Z., Zhuo, Y. W., & Deng, L. J. (2022). PatchMask: A Data Augmentation Strategy with Gaussian Noise in Hyperspectral Images. *Remote Sensing*, *14*(24), 1–22. https://doi.org/10.3390/rs14246308

Garg, U., Agarwal, S., Gupta, S., Dutt, R., & Singh, D. (n.d.). *Prediction of Emotions from the Audio Speech Signals using MFCC , MEL and Chroma*. 87–91. https://doi.org/10.1109/CICN.2020.17

Geng, M., Xie, X., Liu, S., Yu, J., Hu, S., Liu, X., & Meng, H. (2020). Investigation of data augmentation techniques for disordered speech recognition. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2020-Octob*, 696–700. https://doi.org/10.21437/Interspeech.2020-1161

Gourisaria, M. K., Agrawal, R., Sahni, M., & Singh, P. K. (2024). Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques. *Discover Internet of Things*, *4*(1). https://doi.org/10.1007/s43926-023-00049-y

Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., & Wilson, K. (2017). CNN architectures for large-scale audio classification. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 131–135. https://doi.org/10.1109/ICASSP.2017.7952132

Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. *IEEE Access*, *10*, 19083–19095. https://doi.org/10.1109/ACCESS.2022.3151048

Hill, M. (1997). *a __*. *6393*(97).

Huang, X., Shirahama, K., Irshad, M. T., Nisar, M. A., Piet, A., & Grzegorzek, M.

(2023). *Gaussian Noise Data Augmentation*.

Huh, M., Ray, R., & Karnei, C. (2023). *A Comparison of Speech Data Augmentation Methods Using S3PRL Toolkit*. http://arxiv.org/abs/2303.00510

Jiang, G., & Wang, W. (2017). Error estimation based on variance analysis of k-fold cross-validation. *Pattern Recognition*, *69*, 94–106. https://doi.org/10.1016/j.patcog.2017.03.025

Joysingh, S. J., Vijayalakshmi, P., & Nagarajan, T. (2025). Significance of chirp MFCC as a feature in speech and audio applications. *Computer Speech and Language*, *89*, 1–7. https://doi.org/10.1016/j.csl.2024.101713

Kaneko, T., Tanaka, K., Kameoka, H., & Seki, S. (2022). Istftnet: Fast and Lightweight Mel-Spectrogram Vocoder Incorporating Inverse Short-Time Fourier Transform. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2022-May*, 6207–6211. https://doi.org/10.1109/ICASSP43922.2022.9746713

Kethireddy, R., Kadiri, S. R., Alku, P., & Gangashetty, S. V. (2020). Mel-weighted single frequency filtering spectrogram for dialect identification. *IEEE Access*, *8*, 174871–174879. https://doi.org/10.1109/ACCESS.2020.3020506

Khan, W., Daud, A., Khan, K., Nasir, J. A., Basheri, M., Aljohani, N., & Alotaibi, F. S. (2019). Part of Speech Tagging in Urdu: Comparison of Machine and Deep Learning Approaches. *IEEE Access*, *7*, 38918–38936. https://doi.org/10.1109/ACCESS.2019.2897327

Kharitonov, E., Riviere, M., Synnaeve, G., Wolf, L., Mazare, P. E., Douze, M., & Dupoux, E. (2021). Data Augmenting Contrastive Learning of Speech Representations in the Time Domain. *2021 IEEE Spoken Language Technology Workshop, SLT 2021 - Proceedings*, 215–222. https://doi.org/10.1109/SLT48900.2021.9383605

Kons, Z., & Toledo-Ronen, O. (2013). Audio event classification using deep neural

networks. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, August 2013*, 1482–1486. https://doi.org/10.13064/ksss.2015.7.4.027

Kronvall, T., Juhlin, M., Swärd, J., Adalbjörnsson, S. I., & Jakobsson, A. (2017). Sparse modeling of chroma features. *Signal Processing*, *130*, 105–117. https://doi.org/10.1016/j.sigpro.2016.06.020

Likitha, M. S., Gupta, S. R. R., Hasitha, K., & Raju, A. U. (2017). Speech based human emotion recognition using MFCC. *Proceedings of the 2017 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2017*, *2018-Janua*, 2257–2260. https://doi.org/10.1109/WiSPNET.2017.8300161

Liu, Z., Wang, S., Inoue, S., Bai, Q., & Li, H. (2024). *Autoregressive Diffusion Transformer for Text-to-Speech Synthesis*. 1–24. http://arxiv.org/abs/2406.05551

Lopes, R. G., Yin, D., Poole, B., Gilmer, J., & Cubuk, E. D. (2019). *Improving Robustness Without Sacrificing Accuracy with Patch Gaussian Augmentation*. 1–18. http://arxiv.org/abs/1906.02611

Mariya Celin, T. A., Vijayalakshmi, P., & Nagarajan, T. (2023). Data Augmentation Techniques for Transfer Learning-Based Continuous Dysarthric Speech Recognition. *Circuits, Systems, and Signal Processing*, *42*(1), 601–622. https://doi.org/10.1007/s00034-022-02156-7

Markoulidakis, I., & Markoulidakis, G. (2024). Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis. *Technologies*, *12*(7). https://doi.org/10.3390/technologies12070113

Mohd Hanifa, R., Isa, K., & Mohamad, S. (2021). A review on speaker recognition: Technology and challenges. *Computers and Electrical Engineering*, *90*(April 2020), 107005. https://doi.org/10.1016/j.compeleceng.2021.107005

Müller, M., & Ewert, S. (2011). Chroma toolbox: Matlab implementations for

extracting variants of chroma-based audio features. *Proceedings of the 12th International Society for Music Information Retrieval Conference, ISMIR 2011*, *Ismir*, 215–220.

Müller, M., Kurth, F., & Clausen, M. (2005). Audio matching via chroma-based statistical features. *ISMIR 2005 - 6th International Conference on Music Information Retrieval*, *May*, 288–295.

Nugroho, K., Noersasongko, E., Purwanto, Muljono, & Setiadi, D. R. I. M. (2022). Enhanced Indonesian Ethnic Speaker Recognition using Data Augmentation Deep Neural Network. *Journal of King Saud University - Computer and Information Sciences*, *34*(7), 4375–4384. https://doi.org/10.1016/j.jksuci.2021.04.002

Paul S, B. S., Glittas, A. X., & Gopalakrishnan, L. (2021). A low latency modular-level deeply integrated MFCC feature extraction architecture for speech recognition. *Integration*, *76*(September 2020), 69–75. https://doi.org/10.1016/j.vlsi.2020.09.002

Pinson, H., Lenaerts, J., & Ginis, V. (2023). Linear CNNs Discover the Statistical Structure of the Dataset Using Only the Most Dominant Frequencies. *Proceedings of Machine Learning Research*, *202*, 27876–27906.

Ramadhan, H. (2023). *Klasifikasi irama azan berbasis mel frequency cepstral coefficients menggunakan multilayer perceptron*. 1–54. http://digilib.uinsa.ac.id/64838/2/Hafid Ramadhan_H96219046 ok.pdf

Schultz, A., Rekeczky, C., Szatmari, I., Roska, T., & Chua, L. O. (1998). Spatio-temporal CNN algorithm for object segmentation and object recognition. *Proceedings of the IEEE International Workshop on Cellular Neural Networks and Their Applications*, *4*(9208), 347–352. https://doi.org/10.1109/cnna.1998.685400

Serrano, S., Scarpa, M. L., & Serghini, O. (2024). Vggish for Music/Speech

Classification in Radio Broadcasting. *Proceedings - European Council for Modelling and Simulation, ECMS*, *38*(1), 550–557. https://doi.org/10.7148/2024-0550

Suganuma, M., Kobayashi, M., Shirakawa, S., & Nagao, T. (2020). Evolution of deep convolutional neural networks using cartesian genetic programming. *Evolutionary Computation*, *28*(1), 141–163. https://doi.org/10.1162/evco_a_00253

Sumit Srivastava, K. J. A. J. (2024). Analysis of Human Voice for Speaker Recognition: Concepts and Advancement. *Journal of Electrical Systems*, *20*(1s), 582–599. https://doi.org/10.52783/jes.806

Terashima, R., Yamamoto, R., Song, E., Shirahata, Y., Yoon, H. W., Kim, J. M., & Tachibana, K. (2022). Cross-Speaker Emotion Transfer for Low-Resource Text-to-Speech Using Non-Parallel Voice Conversion with Pitch-Shift Data Augmentation. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2022-Septe*(1), 3018–3022. https://doi.org/10.21437/Interspeech.2022-11278

Wang, X., Chu, Z., Han, B., Wang, J., Zhang, G., & Jiang, X. (2020). A Novel Data Augmentation Method for Intelligent Fault Diagnosis under Speed Fluctuation Condition. *IEEE Access*, *8*, 143383–143396. https://doi.org/10.1109/ACCESS.2020.3014340

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, *48*(9), 2839–2846. https://doi.org/10.1016/j.patcog.2015.03.009

Wong, T. T., & Yang, N. Y. (2017). Dependency Analysis of Accuracy Estimates in k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, *29*(11), 2417–2427. https://doi.org/10.1109/TKDE.2017.2740926

Wong, T., & Yeh, P. (n.d.). *10.1109@Tkde.2019.2912815*. *1*, 1.

Xu, J., Zhang, Y., & Miao, D. (2020). Three-way confusion matrix for classification: A measure driven view. *Information Sciences*, *507*, 772–794. https://doi.org/10.1016/j.ins.2019.06.064

Zhang, L., Zhang, J., Lei, B., Mukherjee, S., Pan, X., Zhao, B., Ding, C., Li, Y., & Xu, D. (2023). Accelerating Dataset Distillation via Model Augmentation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2023-June*, 11950–11959. https://doi.org/10.1109/CVPR52729.2023.01150

Zhang, T., Feng, G., Liang, J., & An, T. (2021). Acoustic scene classification based on Mel spectrogram decomposition and model merging. *Applied Acoustics*, *182*, 108258. https://doi.org/10.1016/j.apacoust.2021.108258

Zhao, C., Yuan, X., Long, J., Jin, L., & Guan, B. (2025). *Chinese stock market Pr ep rin t n ot pe er re v Pr ep rin t n ot pe er re v ed*. 1–20.