

DAFTAR PUSTAKA

- Abbaschian, B. J., Sierra-sosa, D., & Elmaghraby, A. (2021). *Deep Learning Techniques for Speech Emotion Recognition , from Databases to Models*.
- Abd, S. M. (2026). Deep Hybrid CNN-LSTM Architecture with Mel-MFCC-Chroma Feature Fusion and Attention Mechanism for Enhanced Egyptian Arabic Speech Emotion Recognition. *International Journal on Science and Technology (IJSAT)*, 17(1), 1–11.
- Albawi, S., & Mohammed, T. A. (2017). *Understanding of a Convolutional Neural Network*.
- Alías, F., Socoró, J. C., & Sevillano, X. (2016). *applied sciences A Review of Physical and Perceptual Feature Extraction Techniques for Speech , Music and Environmental Sounds*. <https://doi.org/10.3390/app6050143>
- Alluhaidan, A. S., Saidani, O., Jahangir, R., & Nauman, M. A. (2023). *applied sciences Speech Emotion Recognition through Hybrid Features and Convolutional Neural Network*.
- Ambaye, G. A. (2020). *Time and Frequency Domain Analysis of Signals : A Review*. 9(12), 271–276.
- Anglin, K. (2024). *Addressing Threats to Validity in Supervised Machine Learning : A Framework and Best Practices for Education Researchers*. 10(1), 1–21. <https://doi.org/10.1177/23328584241303495>
- Atmaja, B. T., & Sasou, A. (2022). *Effect of different splitting criteria on the performance of speech emotion recognition*.
- Bagli, E., & Visani, G. (2020). *METRICS FOR MULTI-CLASS CLASSIFICATION: AN OVERVIEW*. 1–17.
- Bahdanau, D., Cho, K., & Bengio, Y. (2015). *N EURAL M ACHINE T RANSLATION*. 1–15.
- Barbierato, E., & Gatti, A. (2024). *The Challenges of Machine Learning : A Critical Review*.
- Barhoumi, C., & Benayed, Y. (2025). *Real-time speech emotion recognition using deep learning and data augmentation*.
- Budi, T., Khoirotul, Y., & Dutono, T. (2025). *Speech Emotion Recognition of Indonesian Movies by Using Convolutional Neural Network*. 9(November), 2569–2575.
- Bustamin, A., Rizky, A. M., Warni, E., & Sari, I. (2024). *IndoWaveSentiment : Indonesian audio dataset for emotion classification*. 57. <https://doi.org/10.1016/j.dib.2024.111138>
- Chaves-villota, A., Jimenez-martín, A., & Jojoa-acosta, M. (2026). Deep feature representations and fusion strategies for speech emotion recognition from acoustic and linguistic modalities : A systematic review. *Computer Speech &*

- Language*, 96(August 2025), 101873.
<https://doi.org/10.1016/j.csl.2025.101873>
- Cheng, H., Xie, Z., Wu, L., & Yu, Z. (2019). *Data prediction model in wireless sensor networks based on bidirectional LSTM*.
- Cherukuru, P., & Mustafa, M. B. (2024). *CNN-based noise reduction for multi-channel speech enhancement system with discrete wavelet transform (DWT) preprocessing*. <https://doi.org/10.7717/peerj-cs.1901>
- E, A. K., & Kulkarni, S. B. (2025). *Machine Learning Methods for Speech Emotion Recognition*. 03(September), 3793–3798.
- Farid, M. N., Rahman, A. F., & Wicaksono, H. (2023). *Analisis Pengaruh Kombinasi Fitur Spektral terhadap Tingkat Akurasi Speech Emotion Recognition*. 5(2), 120–129. <https://doi.org/10.37034/jsisfotek.v5i1.234>
- Finch, E. (2024). *Speech Emotion Recognition with Dynamic CNN and Bi-LSTM*. 3(6).
- Firdaus, R., Yasmin, F., Haider, S. A., Jahan, S., & Poulouva, P. (2025). *Emotion Detection Using Machine Learning: A Study on Human Computer Interaction*. 125–134.
- Gjoreski, M., Gjoreski, H., & Kulakov, A. (2014). *Machine Learning Approach for Emotion Recognition in Speech*. 38, 377–384.
- Gumelar, A. B., Yuniarno, E. M., Anggraeni, W., & Sugiarto, I. (2020). *Kombinasi Fitur Multispektrum Hilbert dan Cochleagram untuk Identifikasi Emosi Wicara (Spectrum Features Combination of Hilbert and Cochleagram for Speech Emotions Identification)*. 9(2), 180–189.
- He, L., & Dellwo, V. (2016). A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform. *Interspeech*, 034(2).
- Hinton, G. E. (2010). *Rectified Linear Units Improve Restricted Boltzmann Machines*. 3.
- Hochreiter, S., & Schmidhuber, J. (1997). *LONG SHORT-TERM MEMORY*. 9(8), 1–32.
- Hu, J., Lv, S., & Jie, R. (2024). *A Cardiac Audio Classification Method Based on Multidimensional Feature Expression Multidimensional Feature Expression*. 0–14.
- Hussein, E. E., Zerouali, B., Bailek, N., Derdour, A., Ghoneim, S. S. M., Santos, C. A. G., & Hashim, M. A. (2025). Harnessing Explainable AI for Sustainable Agriculture: SHAP-Based Feature Selection in Multi-Model Evaluation of Irrigation Water Quality Indices. *Water (Switzerland)*, 17(1). <https://doi.org/10.3390/w17010059>
- Ibrahim, H., Loo, C. H. U. K., & Alnajjar, F. (2021). *Speech Emotion Recognition by Late Fusion for Bidirectional Reservoir Computing With Random*

Projection. 122855–122871.

Iwana, B. K. (2022). *ON MINI-BATCH TRAINING WITH VARYING LENGTH TIME SERIES*.

Kamble, V. V, Deshmukh, R. R., & Karwankar, A. R. (2015). *Emotion Recognition for Instantaneous*. 2, 335–346. <https://doi.org/10.1007/978-3-319-12012-6>

Khamaisi, K., Keller, N., Krummenacher, S., Huber, V., & Rodrigues, B. (2025). *From Noise to Knowledge : A Comparative Study of Acoustic Anomaly Detection Models in Pumped-storage Hydropower Plants* (Vol. 1, Nomor 1). arXiv.

Kingma, D. P., & Ba, J. L. (2017). *A : a m s o*. 1–15.

Krichen, M. (2023). *Convolutional Neural Networks : A Survey*. 1–41.

Kshirsagar, S., Pendyala, A., & Falk, T. H. (2020). *Task-specific speech enhancement and data augmentation for improved multimodal emotion recognition under noisy conditions*.

Kumah, A., Nwogu, C. N., Issah, A., Obot, E., Kanamitie, D. T., Sifa, J. S., & Aidoo, L. A. (2024). *Quality and Safety Learning Corner Cause-and-Effect (Fishbone) Diagram : A Tool for Generating and Organizing Quality Improvement Ideas*. 7(2), 85–87. <https://doi.org/10.36401/JQSH-23-42>. This

Latif, S., Rana, R., Khalifa, S., Jurdak, R., & Qadir, J. (2021). *Deep Representation Learning in Speech Processing : Challenges , Recent Advances , and Future Trends*.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-Based Learning Applied to Document Recognition. *IEEE*.

Lee, S., & Kwon, H. (2020). *applied sciences A Preprocessing Strategy for Denoising of Speech Data Based on Speech Segment Detection*. 1–24. <https://doi.org/10.3390/app10207385>

Lieskovská, E., Jakubec, M., Jarina, R., & Chmul, M. (2021). *A Review on Speech Emotion Recognition Using Deep Learning and Attention Mechanism*.

Liu, G., Cai, S., & Wang, C. (2023). Speech emotion recognition based on emotion perception. *EURASIP Journal on Audio, Speech, and Music Processing*, 4. <https://doi.org/10.1186/s13636-023-00289-4>

Lu, C., Zong, Y., Lian, H., Zhao, Y., Schuller, W., & Zheng, W. (2024). *IMPROVING SPEAKER-INDEPENDENT SPEECH EMOTION RECOGNITION USING DYNAMIC JOINT DISTRIBUTION ADAPTATION*.

Lu, X. (2022). *Deep Learning Based Emotion Recognition and Visualization of Figural Representation Deep Learning and Emotion Recognition*. 12(January), 1–12. <https://doi.org/10.3389/fpsyg.2021.818833>

Madanian, S., Chen, T., Adeleye, O., Michael, J., Poellabauer, C., Parry, D., & Schneider, S. L. (2023). *Intelligent Systems with Applications Speech*

- emotion recognition using machine learning — A systematic review. *Intelligent Systems with Applications*, 20(September 2022), 200266. <https://doi.org/10.1016/j.iswa.2023.200266>
- Majiid, M., Setiawan, K., Pamungkas, P., Annas, T., & Setiawan, L. (2025). *Advancing Indonesian Audio Emotion Classification: A Comparative Study Using IndoWaveSentiment*. JURNAL EMACS (Engineering, MAThematics and Computer Science). <https://journal.binus.ac.id/index.php/EMACS/article/view/13415/5468>
- Muda, L., Begam, M., & Elamvazuthi, I. (2010). *Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques*. 2(3), 138–143.
- Mustaqeem, & Kwon, S. (2020). A CNN-Assisted Enhanced Audio Signal Processing. *Sensors*.
- Olawale, I. O., Iworiso, J., & Amaunam, I. A. (2023). *Predicting Stock Investments Based on Sentiment and Historical Price Data*. 4(1), 1–32. <https://doi.org/10.12691/ijdeaor-4-1-1>
- Ong, K. A. H. L., Lee, C. P. O. O., Lim, H. S., & Lim, K. M. (2023). SCQT-MaxViT: Speech Emotion Recognition with Constant-Q Transform and Multi-Axis Vision Transformer. *IEEE Access*, PP, 1. <https://doi.org/10.1109/ACCESS.2023.3288526>
- Orzikulova, A., Lee, S., Kwak, J., & Shin, J. (2024). *Federated Learning with Incomplete Sensing Modalities Jaemin Shin*.
- Park, D. S., Chan, W., Zhang, Y., Chiu, C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). *SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition*.
- Pedregosa, F., Grisel, O., Andreas, M., Weiss, R., Passos, A., & Brucher, M. (2011). *Scikit-learn: Machine Learning in Python*. 12, 2825–2830.
- Peng, Z., He, W., Li, Y., Du, Y., & Dang, J. (2023). *Multi-Level Attention-Based Categorical Emotion Recognition Using Modulation-Filtered Cochleagram*.
- Powers, D. M. . (2011). *EVALUATION: FROM PRECISION, RECALL AND F-MEASURE TO ROC, INFORMEDNESS, MARKEDNESS & CORRELATION*. 37–63.
- Rascon, C. (2023). *Characterization of Deep Learning-Based Speech-Enhancement Techniques in Online Audio Processing Applications*.
- Sailor, H. B. (2025). *MERaLiON-SER: Robust Speech Emotion Recognition Model for English and SEA Languages*. 1–11.
- Saito, T., & Rehmsmeier, M. (2015). *The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets*. 1–21. <https://doi.org/10.1371/journal.pone.0118432>
- Sajjad, M., & Kwon, S. (2020). *Clustering-Based Speech Emotion Recognition by*

- Incorporating Learned Features and Deep BiLSTM*. 8, 1–15.
- Seema, P. (2020). *Speech Signal Analysis and Speaker Recognition by Signal Processing*. 8(11), 7–10.
- Sekkate, S., Khalil, M., Adib, A., & Jebara, S. Ben. (2019). *An Investigation of a Feature-Level Fusion for Noisy Speech Emotion Recognition*. 1–15. <https://doi.org/10.3390/computers8040091>
- Septiyanto, M., Susanto, E. B., & Sugianti, D. (2025). *IMPROVING INDONESIAN SPEECH EMOTION CLASSIFICATION USING MFCC AND BILSTM WITH AUDIO AUGMENTATION*. 8(3), 244–251. <https://doi.org/10.33387/jiko.v8i3.10820>
- Shixin, P., Kai, C., Tian, T., & Jingying, C. (2024). An autoencoder-based feature level fusion for speech emotion recognition. *Digital Communications and Networks*, 10(5), 1341–1351. <https://doi.org/10.1016/j.dcan.2022.10.018>
- Sun, L. (2022). *Speech emotion recognition based on genetic algorithm – decision tree fusion of deep and acoustic features*. 44(March 2021), 462–475. <https://doi.org/10.4218/etrij.2020-0458>
- Tao, H., Shan, S., Hu, Z., Zhu, C., & Ge, H. (2023). *Strong Generalized Speech Emotion Recognition Based on*.
- Tiwari, S. (2025). *Emotion Detection from Speech Using CNN-BiLSTM with Feature Rich Audio Inputs*. 1, 80–89.
- Truong, N., Ngoc, D., Dang, M., Duy, N., & Thi, T. (2023). Hybrid data augmentation and deep attention-based dilated convolutional-recurrent neural networks for speech emotion recognition. *Expert Systems With Applications*, 230(April), 120608. <https://doi.org/10.1016/j.eswa.2023.120608>
- Tyagi, S., & Szénási, S. (2024). Semantic speech analysis using machine learning and deep learning techniques: a comprehensive review. *Multimedia Tools and Applications*, 83(29), 73427–73456. <https://doi.org/10.1007/s11042-023-17769-6>
- Vakalopoulou, M., Christodoulidis, S., Burgos, N., Colliot, O., & Lepetit, V. (2023). *Deep Learning: Basics and Convolutional Neural Networks (CNNs)* (Vol. 197). <https://doi.org/10.1007/978-1-0716-3195-9>
- Vaswani, A. (2017). *Attention Is All You Need*. *Nips*.
- Vazhenina, D., & Markov, K. (2020). *End-to-End Noisy Speech Recognition Using Fourier and Hilbert Spectrum Features*. <https://doi.org/10.3390/electronics9071157>
- Waleed Alsabhan. (2023). Human – Computer Interaction with a Real-Time Speech Emotion Recognition with Ensembling Techniques 1D. *Sensors* 2023, 23, 1386, 1–21.
- Wang, Y. (2024). *Speech Emotion Recognition Using Multi-Scale Global – Local Representation Learning with Feature Pyramid Network*.

- Wang, Y., Gu, Y., Yin, Y., Han, Y., Zhang, H., Wang, S., Li, C., & Quan, D. (2021). *Multimodal transformer augmented fusion for speech emotion recognition*.
- Yoon, S. (2020). *applied sciences A Simple Distortion-Free Method to Handle Variable Length Sequences for Recurrent Neural Networks in Text Dependent Speaker Verification*.
- Zennou, H. (2026). *Real-Time Speech Emotion Recognition with a CNN-BiLSTM-Attention Deep Learning Model*. 16(3), 35047–35055.
- Zhang, C., Yang, Z., He, X., & Deng, L. (2020). *Multimodal Intelligence : Representation Learning , Information Fusion , and Applications*. 1–15.
- Zhao, Y., & Shu, X. (2023). Speech emotion analysis using convolutional neural network (CNN) and gamma classifier - based error correcting output codes (ECOC). *Scientific Reports*, 1–18. <https://doi.org/10.1038/s41598-023-47118-4>