Indonesian Journal of Electrical Engineering and Computer Science
Vol. 39, No. 2, August 2025, pp. 906~913
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v39.i2.pp906-913

Enhancing acoustic environment classification for hearing-impaired individuals using hybrid CNN and RFE

Sunilkumar M. Hattaraki[1], Shankarayya G. Kambalimath[2]
[1]Department of Electronics and Communication Engineering, B.L.D.E.A's V.P.Dr. P.G. Halakatti, College of Engineering and Technology, Visvesveraya Technological University, Belagavi, India
[2]Department of Electronics and Communication Engineering, Basaveshwar Engineering College, Visvesveraya Technological University, Bagalkote, India

Article Info

Article history:

Received Nov 19, 2024
Revised Apr 16, 2025
Accepted Jul 2, 2025

Keywords:

Acoustic environment classification
CNN
Ensemble learning
Hearing aids
PCA
Random forest

ABSTRACT

Individuals who are deaf or hard of hearing experience considerable difficulties in distinguishing sounds in various acoustic environments, which affects their communication ability and overall quality of life. Existing auditory assistive technologies currently face challenges with real-time classification and adaptation to changing noise conditions, underscoring the need for more reliable and accurate classification models. This research bridges the existing gap by creating a hybrid classification framework that integrates convolutional neural networks (CNN) and random forest ensemble (RFE) to enhance the accuracy of environmental sound classification. The study utilizes Mel-frequency cepstral coefficients (MFCCs) for feature extraction and principal component analysis (PCA) for dimensionality reduction, thus facilitating the efficient processing of real-world audio data. The proposed methodology improves classification accuracy across various environmental conditions. Experimental evaluations demonstrate superior performance, achieving a training accuracy of 94.93% and a testing accuracy of 93.41%, thereby exceeding conventional machine learning methods. By overcoming limitations in existing models, this research contributes to the development of adaptive hearing assistance systems with enhanced noise classification capabilities. The results have significant implications for the development of smart hearing aids, real-time noise classification, and auditory scene analysis. Ultimately, this research enhances assistive hearing technologies, promoting greater accessibility, communication, and inclusion for hearing-impaired individuals, thus contributing positively to society.

Corresponding Author:

Sunilkumar M. Hattaraki
Department of Electronics and Communication Engineering, B.L.D.E.A's V.P.Dr. P.G.Halakatti
College of Engineering and Technology, Visvesveraya Technological University
Belagavi-590018, Karnataka, India
Email: sunilmh039@mail.com

1. INTRODUCTION

It can be very difficult for people with hearing impairments to adjust to a variety of acoustic environments, including quiet places, busy restaurants, and noisy streets. Because traditional hearing aids cannot accurately classify ambient noise, they frequently do not offer the best support under a variety of circumstances. Because of this, users find it difficult to discern speech from background noise, which impairs communication effectiveness and increases cognitive load. In order to overcome this restriction, intelligent hearing aids that can dynamically adjust to shifting auditory conditions must be developed [1].

– Despite advancements in the classification of acoustic environments, current models have a number of drawbacks. Firstly, they lack real-time adaptability, as the majority of earlier models do not incorporate adaptive learning techniques to dynamically alter classification outputs [2].
– Limitations of feature extraction: traditional methods mostly use manually created features, which might not be able to adequately capture intricate acoustic patterns [3].
– Limited model generalization: since many studies only consider a small number of environments, the models' applicability in actual situations is diminished.

This study presents a hybrid classification model that combines random forest ensemble (RFE) and convolutional neural networks (CNNs) to address these issues. Mel-frequency cepstral coefficients (MFCCs) are used by the CNN component to extract robust features, and the RFE uses feature selection and decision tree-based classification to improve predictive accuracy.

The goal of this research is to create a stable and effective system for categorizing various acoustic environments. The suggested hybrid model provides a reliable answer by combining the predictive ability of an RFE with the ability to extract features using CNN. This model can dynamically adjust to various environmental conditions, improving speech intelligibility in difficult acoustic environments and thus the listening experience for those who are hard of hearing [4].

This study affects real-time hearing aids by allowing for automatic environmental adaptation. The research helps create intelligent hearing aids that can adjust sound output according to ambient noise, facilitating more effective communication by increasing the precision and dependability of environment classification. By making auditory systems more responsive and user-centered, this work may lessen the cognitive strain that people experience when switching between environments [5].

The rest of this paper is organized as follows: section 2 offers a thorough analysis of current methods for classifying environmental sounds. The suggested methodology, including feature extraction, model design, and dataset preparation, is described in section 3. Experimental results are shown in section 4, which contrasts the hybrid CNN-RFE model with cutting-edge methods. In contrast, section 5 wraps up the study and suggests areas for further research.

2. LITERATURE SURVEY

Zaheer *et al.* [6] provided a comprehensive review of artificial intelligence (AI)-based acoustic source identification (ASI) techniques. In their analysis, they examined the strengths and weaknesses of various AI-driven ASI processes and the methods proposed by researchers in the literature. Additionally, they conducted an in-depth survey of ASI applications across diverse fields, including machinery, underwater acoustics, environmental/event source recognition, healthcare, and more. The review also highlights significant research directions for future exploration in this area.

Abayomi-Alli *et al.* [7] implemented screening exclusion criteria and snowballing techniques, resulting in the selection of 56 articles. They identified several shortcomings in prior research, such as insufficient, weakly labeled, imbalanced, and noisy datasets, as well as inadequate sound feature representations and ineffective augmentation strategies that hinder classifier performance. Sound datasets, feature extraction techniques, data augmentation techniques, and their applications in sound classification are all briefly discussed in the article. In their conclusion, the authors provide answers to research questions, a synopsis of the systematic literature review (SLR), and suggestions for improving sound classification tasks.

Mutanu *et al.* [8] examined 124 studies spanning eight years, emphasizing important application areas in feature extraction, audio transformation, and bioacoustics research. Along with discussing the field's present difficulties, prospects, and future directions, the survey also examines the classification algorithms used in bioacoustics systems. Zhang *et al.* [9] examine new developments in emotion recognition systems, with an emphasis on architectures for classification that use inputs from text, audio, and vision, as well as fusion and feature engineering techniques. To enable reliable multi-modal analysis, the paper highlights creative pipeline interventions, from preprocessing raw signals to predicting emotion labels. By offering insights into the current state-of-the-art, highlighting unresolved issues, and investigating exciting avenues in emotion detection via cross-modal learning, this study seeks to stimulate additional research through theoretical discussions and real-world case studies.

Sangala *et al.* [10] investigates the creation of a voice assistant system intended for people with visual impairments. Through the use of sophisticated speech recognition and natural language processing, the system makes voice commands a smooth way to interact. Key issues like usability, accessibility, and contextual understanding are addressed. User reviews attest to its efficacy in boosting self-reliance and enhancing day-to-day living, indicating its potential as a useful assistive technology.

2.1. Identifying the gap
Current hearing aids use basic noise suppression and amplification techniques, which do not adequately adjust to complex, dynamic listening environments. Many existing models fail to accurately classify nuanced acoustic settings like cocktail party noise or reverberant spaces, leading to suboptimal performance. Additionally, most classification models rely on a single machine-learning approach, which limits the generalization ability across varied acoustic environments [11], [12].

2.2. Overcoming the gap
This research addresses the gap by integrating CNNs with RFE techniques to improve classification accuracy and robustness. The hybrid model takes advantage of CNN's capability to capture complex audio patterns, while the random forest provides effective generalization across varying data distributions. Furthermore, dimensionality reduction through principal component analysis (PCA) ensures computational efficiency, making the model suitable for real-time applications [13], [14].

3. METHOD
3.1. Model architecture
The proposed model integrates CNNs and a RFE, as illustrated in Figure 1. The CNN extracts features from raw audio data in the form of MFCCs, which are widely used in audio signal processing. These features are then reduced in dimensionality through PCA to ensure computational efficiency.

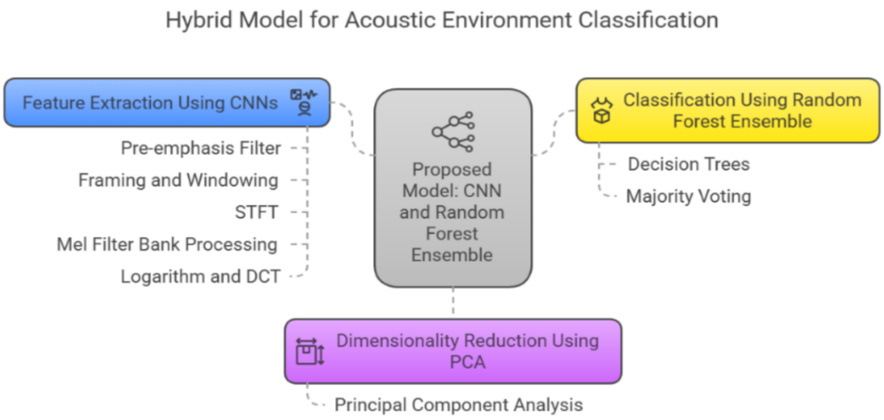

Figure 1. Proposed model architecture

3.2. Dataset and preprocessing
The dataset includes real-world acoustic environments, such as those in Table 1. Table 1 summarizes the different types of environments and the quantity of audio files available for testing and training. It describes a wide range of situations, such as those that are quiet, noisy cars, cocktail parties, restaurants, streets, train stations, airports, group settings, reverberant spaces, and phone conversations.

Table 1. Dataset distribution for environmental sound classification - training and testing files

S.No.	Environment type	Training (Audio files)	Testing (Audio files)
1	Quiet environment	2,000	400
2	Car noise environment	100	10
3	Cocktail environment	100	10
4	Restaurant environment	100	10
5	Street environment	100	10
6	Airport environment	100	10
7	Train station environment	100	10
8	Group setting environment	140	14
9	Reverberant spaces environment	50	05
10	Telephone conversations	100	10
	Total No. of audio files	2,890	489

In order to evaluate and develop audio processing algorithms or models, each type of environment is distinguished by the number of audio recordings that are available for training and testing. With 2,890 audio files available for training and 489 for testing, a comprehensive analysis and assessment across a variety of acoustic environments is made possible. Audio samples were collected, and MFCCs were extracted as input features. PCA was applied to reduce the feature dimensionality, which also improved training time without sacrificing accuracy [15]-[20].

3.3. Feature extraction using CNNs

The CNN extracts meaningful features from raw audio data using MFCCs. The MFCCs for an audio signal x(n) are computed as follows:

a. Pre-emphasis filter [21], [22]: the signal passes through a high-pass filter to balance the frequency spectrum:

$$y(n) = x(n) - \alpha x(n-1) \tag{1}$$

where α is typically set to 0.95.

b. Framing and windowing [23], [24]: the signal is divided into overlapping frames, each multiplied by a Hamming window to minimize spectral leakage:

$$w(n) = 0.54 - 0.46cos(\frac{2\pi n}{N-1}) \tag{2}$$

c. Short-time fourier transform (STFT): each frame's frequency representation is obtained by applying the discrete fourier transform (DFT) [25].

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{\frac{-j2\pi kn}{N}} \tag{3}$$

d. Mel filter bank processing: the power spectrum is passed through a set of triangular filters spaced on the Mel scale, defined as:

$$f_{mel} = 2595log_{10}(1 + \frac{f}{700}) \tag{4}$$

e. Logarithm and discrete cosine transform (DCT): the logarithm of the Mel-filtered energy is computed, followed by a DCT to obtain MFCCs:

$$C_m = \sum_{n=0}^{N-1} L(n)\, cos\left[m\left(n - \frac{1}{2}\right)\frac{\pi}{N}\right] \tag{5}$$

where L(n) represents the log-energy outputs of the Mel filter banks.

3.4. Dimensionality reduction using PCA

To improve computational performance, PCA minimizes the dimensionality of the generated MFCC features. The transformation is given by,

$$Z = W^T X \tag{6}$$

where W is the matrix of principal components, and X represents the original feature matrix. PCA ensures that only the most relevant features are retained for classification.

3.5. Classification using RFE

Once the features have been retrieved and refined, they are put into the random forest classifier, which is made up of several decision trees. Each decision tree is trained on a random portion of the dataset, and the final classification is determined via majority voting.

$$P(y = c) = \frac{1}{T}\sum_{t=1}^{T} I(h_t(X) = c) \tag{7}$$

where T is the total number of trees, $h_t(X)$ represents the prediction from tree t, and I is an indicator function. This ensemble method enhances classification accuracy and generalization performance.

4. RESULTS AND DISCUSSION

The model's performance aligns well with the research objectives of enhancing classification accuracy and robustness across diverse environments. By using a combination of CNNs and RFE, the system exhibited better generalization than standalone models. The confusion matrix as shown in Figure 2 illustrate the model's ability to accurately classify the test samples with minimal misclassifications.

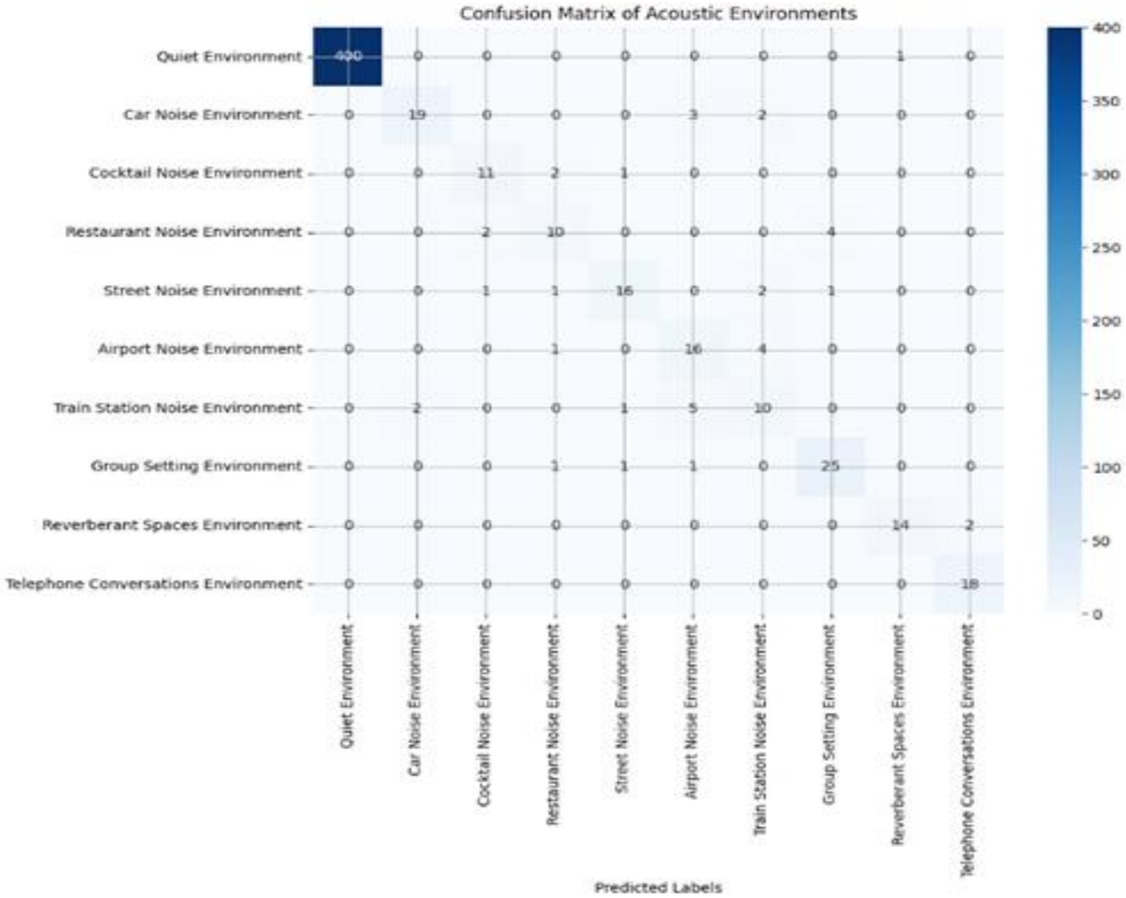

Figure 2. Confusion matrix

4.1. Linking results to objectives

The high accuracy achieved in both training and testing phases supports the objective of developing a robust classifier. Furthermore, the precision and recall scores for challenging environments such as cocktail noise and train station noise demonstrate the model's effectiveness in real-world scenarios. Table 2 highlights the performance of the hybrid CNN and RFE model in classifying diverse acoustic environments. With a high training accuracy of 94.93%, the model effectively learns from the dataset, while a strong test accuracy of 93.41% demonstrates its ability to generalize to new, unseen data. These results reflect the model's robustness and accuracy, making it a reliable tool for improving auditory systems for hearing-impaired individuals in various real-world environments.

Figure 3 displays the accuracy curve, which shows the hybrid CNN and RFE model's training and validation performance over several epochs. Accuracy in both training and validation increases gradually, eventually surpassing 90%, indicating the model's strong generalization to new data. The model is not overfitting, as indicated by the small difference between the two curves.

Table 2. Training and testing accuracy

Metric	Value
Training accuracy	94.93%
Test accuracy	93.41%

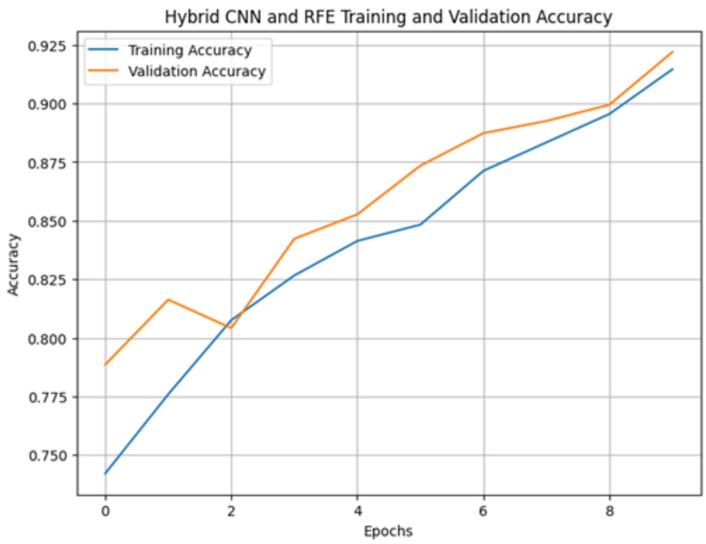

Figure 3. Hybrid CNN and RFE training and validation accuracy

As seen in Figure 4, effective learning is confirmed by the loss curve, which consistently decreases training and validation loss over epochs. The validation loss and training loss are still very similar, indicating a well-regularized model with good generalization power. The decrease in loss points to better performance in feature extraction and classification.

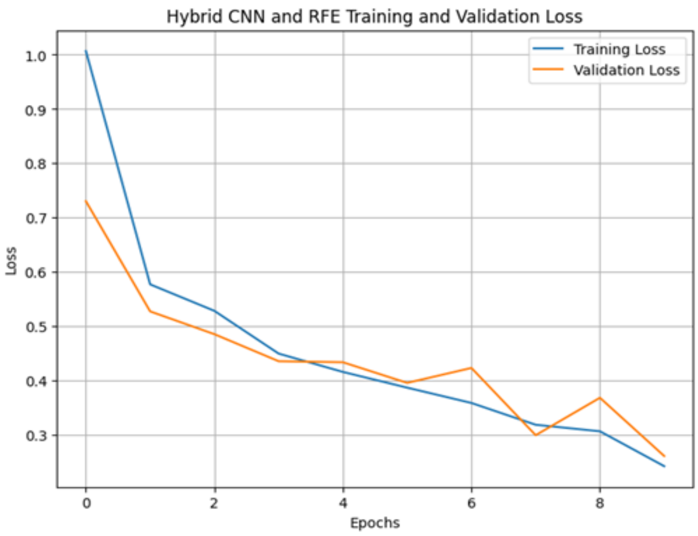

Figure 4. Hybrid CNN and RFE training and validation loss

5. CONCLUSION AND FUTURE SCOPE

The proposed hybrid CNN and RFE model successfully addresses the gap in acoustic environment classification for hearing-impaired individuals. By leveraging the combined strengths of deep learning and ensemble learning, the model achieves an impressive training accuracy of 94.93% and a test accuracy of 93.41%, demonstrating superior accuracy, robustness, and adaptability across diverse environments. This research contributes to the advancement of assistive hearing technologies and holds broader applications in smart devices and real-time environmental classification. Future work could explore optimizing the model for low-power devices, integrating more diverse acoustic environments, and enhancing real-time performance for greater applicability in wearable technology and edge computing solutions.

ACKNOWLEDGMENTS

We would like to express my gratitude to the Research Centre, Department of Electronics and Communication Engineering at Basaveshwar Engineering College, Bagalkote, Karnataka, India.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

AUTHOR CONTRIBUTIONS STATEMENT

Sunilkumar M. Hattaraki, collected the data, analyzed the data, implemented the proposed work and drafted the complete manuscript. Shankarayya G. Kambalimath defined the problem statement and provided critical reviews. All authors read and approved the final manuscript.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sunilkumar M. Hattaraki					✓			✓	✓	✓	✓			
Shankarayya G. Kambalimath	✓	✓			✓				✓	✓	✓			

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

REFERENCES

[1] B. D. Auerbach and H. J. Gritton, "Hearing in complex environments: auditory gain control, attention, and hearing loss," *Frontiers in Neuroscience*, vol. 16, Feb. 2022, doi: 10.3389/fnins.2022.799787.

[2] J. Abeßer, "A review of deep learning based methods for acoustic scene classification," *Applied Sciences (Switzerland)*, vol. 10, no. 6, p. 2020, Mar. 2020, doi: 10.3390/app10062020.

[3] D. Bonet-Solà and R. M. Alsina-Pagès, "A comparative survey of feature extraction and machine learning methods in diverse acoustic environments," *Sensors (Switzerland)*, vol. 21, no. 4, pp. 1–21, Feb. 2021, doi: 10.3390/s21041274.

[4] S. Sachdeva and M. Mulimani, "Acoustic scene classification using fusion of features and random forest classifier," in *Proceedings of the 2022 9th International Conference on Computing for Sustainable Global Development, INDIACom 2022*, Mar. 2022, pp. 654–658, doi: 10.23919/INDIACom54597.2022.9763271.

[5] K. C. De Sousa, V. Manchaiah, D. R. Moore, M. A. Graham, and D. W. Swanepoel, "Effectiveness of an over-the-counter self-fitting hearing aid compared with an audiologist-fitted hearing aid: a randomized clinical trial," *JAMA Otolaryngology - Head and Neck Surgery*, vol. 149, no. 6, pp. 522–530, Jun. 2023, doi: 10.1001/jamaoto.2023.0376.

[6] R. Zaheer, I. Ahmad, D. Habibi, K. Y. Islam, and Q. V. Phung, "A survey on artificial intelligence-based acoustic source identification," *IEEE Access*, vol. 11, pp. 60078–60108, 2023, doi: 10.1109/ACCESS.2023.3283982.

[7] O. O. Abayomi-Alli, R. Damaševičius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: a systematic review," *Electronics (Switzerland)*, vol. 11, no. 22, p. 3795, Nov. 2022, doi: 10.3390/electronics11223795.

[8] L. Mutanu, J. Gohil, K. Gupta, P. Wagio, and G. Kotonya, "A review of automated bioacoustics and general acoustics classification research," *Sensors*, vol. 22, no. 21, p. 8361, Oct. 2022, doi: 10.3390/s22218361.

[9] S. Zhang, Y. Yang, C. Chen, X. Zhang, Q. Leng, and X. Zhao, "Deep learning-based multimodal emotion recognition from audio, visual, and text modalities: a systematic review of recent advancements and future prospects," *Expert Systems with Applications*, vol. 237, p. 121692, Mar. 2024, doi: 10.1016/j.eswa.2023.121692.

[10] T. Sangala, H. Kose, S. Chalkhure, S. Umare, and R. Chilbule, "Voice assistant for blind person," *International Journal of Advanced Research in Science, Communication and Technology*, pp. 578–583, May 2024, doi: 10.48175/ijarsct-18093.

[11] P. U. Diehl *et al.*, "Restoring speech intelligibility for hearing aid users with deep learning," *Scientific Reports*, vol. 13, no. 1, p. 2719, Feb. 2023, doi: 10.1038/s41598-023-29871-8.

[12] N. Z. Tasnim, A. Ni, E. Lobarinas, and N. Kehtarnavaz, "A review of machine learning approaches for the personalization of amplification in hearing aids," *Sensors*, vol. 24, no. 5, p. 1546, Feb. 2024, doi: 10.3390/s24051546.

[13] M. Ashraf *et al.*, "A hybrid CNN and RNN variant model for music classification," *Applied Sciences (Switzerland)*, vol. 13, no. 3, p. 1476, Jan. 2023, doi: 10.3390/app13031476.

[14] M. Turab, T. Kumar, M. Bendechache, and T. Saber, "Investigating multi-feature selection and ensembling for audio classification," *International Journal of Artificial Intelligence & Applications*, vol. 13, no. 3, pp. 69–84, May 2022, doi: 10.5121/ijaia.2022.13306.

[15] S. M. Hattaraki, S. G. Kambalimath, S. Hanamareddy, S. Bhusanur, V. Bilur, and S. Maranur, "Automatic detection and filtering of listening conditions in hearing aids using convolutional neural networks," in *2024 International Conference on Innovation and Novelty in Engineering and Technology, INNOVA 2024 - Proceedings*, Dec. 2024, pp. 1–6, doi: 10.1109/INNOVA63080.2024.10847027.
[16] S. M. Hattaraki, S. G. Kambalimath, P. N. Karjol, P. H. Guggari, N. S. Patil, and B. Naludi, "Objective assessment of speech signal filtering algorithms for hearing aids," in *2024 International Conference on Innovation and Novelty in Engineering and Technology, INNOVA 2024 - Proceedings*, Dec. 2024, pp. 1–6, doi: 10.1109/INNOVA63080.2024.10847038.
[17] S. M. Hattaraki, S. G. Kambalimath, B. P. Savukar, S. Bagali, U. D. Dixit, and A. S. Jadhav, "Detection and classification of various listening environments for hearing-impaired individuals using CRNN," in *2024 International Conference on Innovation and Novelty in Engineering and Technology, INNOVA 2024 - Proceedings*, Dec. 2024, pp. 1–4, doi: 10.1109/INNOVA63080.2024.10847013.
[18] S. M. Hattaraki, S. G. Kambalimath, L. Khedagi, K. Halemani, A. Bhairagond, and P. Hiremath, "Evaluation of speech enhancement algorithms for hearing aids," in *Proceedings of NKCon 2024 - 3rd Edition of IEEE NKSS's Flagship International Conference: Digital Transformation: Unleashing the Power of Information*, Sep. 2024, pp. 1–5, doi: 10.1109/NKCon62728.2024.10774613.
[19] S. M. Hattaraki and S. G. Kambalimath, "Detection and classification of diverse listening conditions for hearing-impaired individuals using RNN model and FIR filter," *Journal of Basic Science and Engineering*, vol. 21, no. 1, pp. 592–612, 2024.
[20] S. M. Hattaraki and S. G. Kambalimath, "Enhancing speech intelligibility in hearing aids using spectral subtraction," *Gongcheng Kexue Yu Jishu/Advanced Engineering Science*, vol. 56, no. 7, pp. 4793–4801, 2024.
[21] Z. Zhou *et al.*, "Impact of analog and digital pre-emphasis on the signal-to-noise ratio of bandwidth-limited optical transceivers," *IEEE Photonics Journal*, vol. 12, no. 2, pp. 1–12, Apr. 2020, doi: 10.1109/JPHOT.2020.2966617.
[22] A. Wright and V. Valimaki, "Perceptual loss function for neural modeling of audio systems," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, May 2020, vol. 2020-May, pp. 251–255, doi: 10.1109/ICASSP40776.2020.9052944.
[23] A. Hannah and G. K. Agordzo, "A design of a low-pass FIR filter using hamming window functions in Matlab," *Computer Engineering and Intelligent Systems*, Feb. 2020, doi: 10.7176/ceis/11-2-04.
[24] D. J. Jwo, W. Y. Chang, and I. H. Wu, "Windowing techniques, the welch method for improvement of power spectrum estimation," *Computers, Materials and Continua*, vol. 67, no. 3, pp. 3983–4003, 2021, doi: 10.32604/cmc.2021.014752.
[25] M. Li, Y. Liu, S. Zhi, T. Wang, and F. Chu, "Short-time fourier transform using odd symmetric window function," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 1, no. 1, pp. 37–45, Dec. 2022, doi: 10.37965/jdmd.v2i2.39.

BIOGRAPHIES OF AUTHORS

Sunilkumar M. Hattaraki is currently pursuing his Ph.D. in electronics and communication engineering, focusing on signal processing, speech processing, and artificial intelligence and machine learning. He completed his M.Tech. in digital electronics and communication from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, in 2011, and his B.E. in electronics and communication engineering from the same university in 2008. With 14 years of academic and research experience, he is presently serving as an assistant professor in the Department of Electronics and Communication Engineering at BLDEA's V.P. Dr. P.G. Halakatti College of Engineering and Technology, Vijayapura. He has published more than 15 research papers in reputed international conferences and journals. He can be contacted at email: sunilmh039@mail.com.

Dr. Shankarayya G. Kambalimath is an associate professor in the Department of Electronics and Communication Engineering at Basaveshwar Engineering College, Bagalkot. He holds a Ph.D., M.E., and B.E., and has 29 years of teaching experience along with 6 months in the industry. He is actively involved in various administrative roles, including TEQIP-III coordinator and placement officer. His research interests lie in VLSI and signal processing, and he has guided numerous UG and PG projects. He has published extensively in reputed conferences and journals and has organized and attended numerous professional development workshops and conferences. He can be contacted at email: kambalimath15@gmail.com and sgkec@becbgk.edu.