Educational Administration: Theory and Practice
2024, 30(5), 11042-11049
ISSN:2148-2403
https://kuey.net/

Research Article

Unveiling Sounds: Harnessing ANN And Mel Spectrograms For Audio SignalsClassification

Abhinav Rawat[1*], Abhishek[2], Akhil Kumar[3], Mr. Rajravi Kumar Ram[4],Md. Shahid[5]

[1*,2,3,4,5]Department of Computer Science and Engineering (CSE), Meerut Institute of Engineering and Technology, Meerut. abhinav.rawat.cse.2020@miet.ac.in, abhishek.b.cse.2020@miet.ac.in , akhil.kumar.cse.2020@miet.ac.in, rajravi.ram@miet.ac.in, md.shahid@miet.ac.in

Citation: Abhinav Rawat, et al (2024), Unveiling Sounds: Harnessing ANN And Mel Spectrograms For Audio Signals Classification, *Educational Administration: Theory and Practice*, *30*(5), 11042-11049
Doi: 10.53555/kuey.v30i5.4882

ARTICLE INFO | **ABSTRACT**

This study focuses on audio classification using a combination of Artificial Neural Networks (ANN) and mel spectrogram representations. The approach involves deriving characteristics from audio signals using mel-frequency cepstral coefficients involves extracting distinctive features from the audio signals (MFCCs) and converting them into spectrogram representations. These mel spectrograms are then used as input to an ANN architecture, allowing the model to independently discern and learn hierarchical features for effective audio classification. The research highlights the synergetic relationship between ANNs and mel spectrogram features, optimizing hyperparameters and leveraging transfer learning to enhance the model's performance. Throughout the evaluation phase, rigorous testing is conducted on benchmark datasets demonstrates the efficiency of the proposed approach in achieving accurate and generalized audio classification across diverse sound categories. Moreover, the hybrid nature of this technique ensures scalability and adaptability, rendering it suitable for addressing the complexities inherent in various audio classification tasks. In essence, this research underscores the promising prospects of integrating ANNs with mel spectrogram representations, heralding advancements in audio processing technologies and their myriad applications.

Keywords words —-: Audio Sounds, Deep Learning, Mel– Spectrogram, Artificial Neural Network (ANN), Audio Processing.

1. Introduction

1.1 Need of Project

The primary objective of audio segmentation and classification involves segmenting and categorizing an incoming audio streaminto discernible segments [1], encompassing speech, music, commercials, background noise, and diverse acoustic conditions [2]. This foundational process is indispensable for efficiently executing tasks like large vocabulary continuous speech recognition (LVCSR) [10], comprehensively analyzing audio content, retrieving audio information, transcribing audio, and clustering audio [20], and other applications associated with audio recognition and indexing. Identifying the various possibilities within the audio stream highlights the diverse range of speaker and environmental factors [12] that can impact acoustic properties in the context of audio classification, it uses Mel Spectrograms [8] which is used as a feature representation for audio classification task whichcan be derived from the mel-frequency cepstral coefficient. It furnishes a concise yet informative portrayal of the spectral composition of an audio signal, encapsulating its intricate frequency components and temporal dynamics.

1.2 Background History

Audio classification involves examining and recognizing [5] various forms of audio, including sounds, noises, musical notes, orsimilar data, and categorizing them appropriately [14].

Traditional Approach of Audio Classification -:

Traditional approaches to audio classification often involve the extraction of handcrafted features [18] from the audio signals followed by the application of classical machine learning algorithms [5]. For choosing traditional machine learning approach forclassification task like Support Vector Machine (SVM) [11] which is effective for binary and multi- class classification, Random Forest [6] which robustly ensembles feature learning, enabling end-to-end learning, and providing superior performance on complex tasks [8].

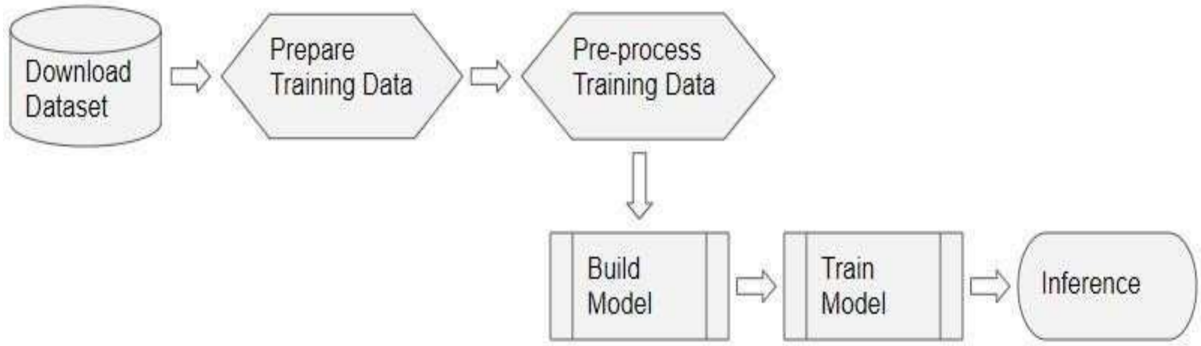

Fig.1 Illustration of Block Diagram depicting the Proposed Technique

Deep Learning Revolution

Deep learning has significantly simplified [2] and enhanced the process of audio classification by automating feature learning and providing more effective representations of complex data. This ability to automatically discover relevant features helps improve the model's ability to generalize [3] across diverse audio samples. In audio classification, used ANN [10] which can be originally designed for image classification but it can be adapted for audio tasks by treating spectrograms as image- like data [5]. It simplifies audio classification by automating feature learning, enabling end-to-end learning, and providing superior performance on complex tasks [8]. Deep learning allows for end-to-end learning, where the model learns directly from raw audio waveforms [7] without the need for extensive handcrafted feature engineering which eliminates and manual extraction of features like MFCCs, as deep learning models [16] can automatically learn hierarchical representations from the raw input data. It aids incapturing both intricate low-level features and higher-level patterns within data, in an advanced manner. Deep learning models often achieve higher accuracy compared to traditional machine learning approaches, especially in tasks with large and complex datasets due to manual extraction of features like MFCCs [20].

1.3 Supported Technologies and Algorithms

The proposed audio classification system leverages several key technologies and algorithms to achieve robust performance. The primary feature extraction technique involves the computation of mel-frequency cepstral coefficients (MFCCs), which captures the frequency characteristics of audio signals. Additionally, the system utilizes mel spectrogram representations, a visual representation of the audio spectrum, to provide input features for the classification models. For the deep learning aspect, Artificial Neural Networks (ANN) [8] are employed, enabling the automatic extraction of hierarchical features from the mel spectrogram data. Transfer learning involves the utilization of pre-trained models on extensive audio datasets for additional applications, enhancing the model's ability to generalize across different audio types. The study also incorporates data augmentation techniques to create a more diverse and balanced training dataset, contributing to improved model robustness. The combination of these technologies and algorithms forms a comprehensive framework for accurate and efficient audio classification. The audio classification system employs Artificial Neural Networks (ANNs) in conjunction with mel spectrogram representations for an effective and automated approach to feature extraction [15] and model training. Initially, the audio data is preprocessed, loaded, and converted into a suitable format. Mel-frequency cepstral coefficients (MFCCs) are computed, and mel spectrograms are generated to create a visual representation of the audio spectrum. This typically involves dense layers in which each neuron is connected to the preceding layer which create a fully connected layer [12], leveraging the pre-trained ANN models on largescale datasets to enhance the system's ability to generalize across diverse audio types in Audio Classification. In, audio classification system is implemented using programming languages such as Python, which provides extensive libraries for signal processing which is used with librosa, scikit-learn and deep learning which is TensorFlow or PyTorch [11]. The system may be deployed on platforms like TensorFlow Serving or Flask for serving the trained models in production environments. Additionally, GPU acceleration using libraries like CUDA can be employed expedite the training process of deep neural networks.

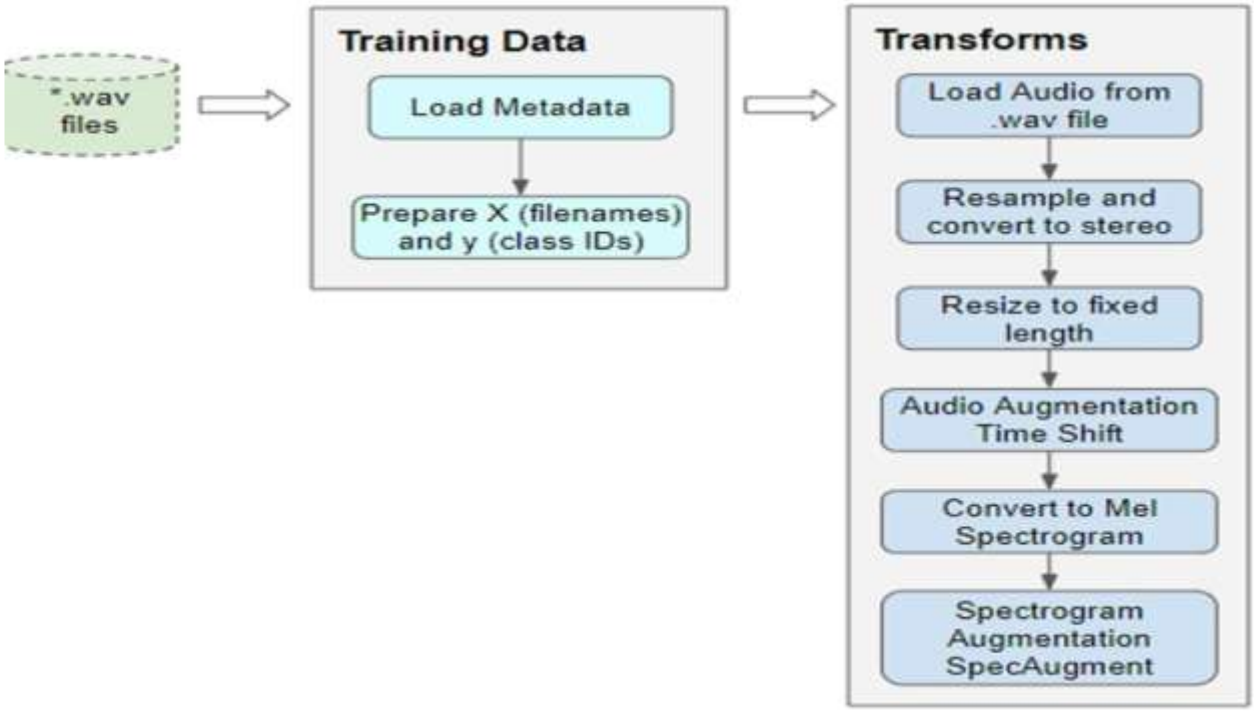

Fig.2 Training Data

2. Proposed Methodology

Audio Classification through machine learning involves the application of algorithms to analyze and categorize diverse audio data. It's finds applications is to classify the sound based on the dataset and provide output according to user requirement. This methodology combines Mel Spectrogram which makes Generate spectrogram images and construct an ANN framework [6] to analyze and interpret these images [8][21].

- **Data Collection** -: Gather a diverse dataset of audio samples [14] representing different classes or categories. that gauges the percentage of accurate predictions.
- **Feature Extraction** -: Extract relevant features, such as spectrogram [6] representations to capture essential informationabout the audio signal.
- **Labeling** -: Assign labels to each audio sample indicating its respective class [17], forming the labeled dataset for model training.
- **Audio Detection** -: Use Create Mel Spectrograms to capture fundamental characteristics of audio, as they are frequentlythe optimal representation for feeding audio data into deep learning models. In a CNN, each layer [8] employs filters toprogressively enhance the image depth, also known as the number of channels [31][32][33][34].
- **Training** -: Create a training loop to train the model which train the model over multiple epochs [5], handling a batch of data in each iteration. Monitor a straightforward accuracy metric that gauges the percentage of accurate predictions.
- **Deployment** -: Deploy the trained model for real-time audio classification in practical applications [9], such asintegration into mobile apps or web services[22].

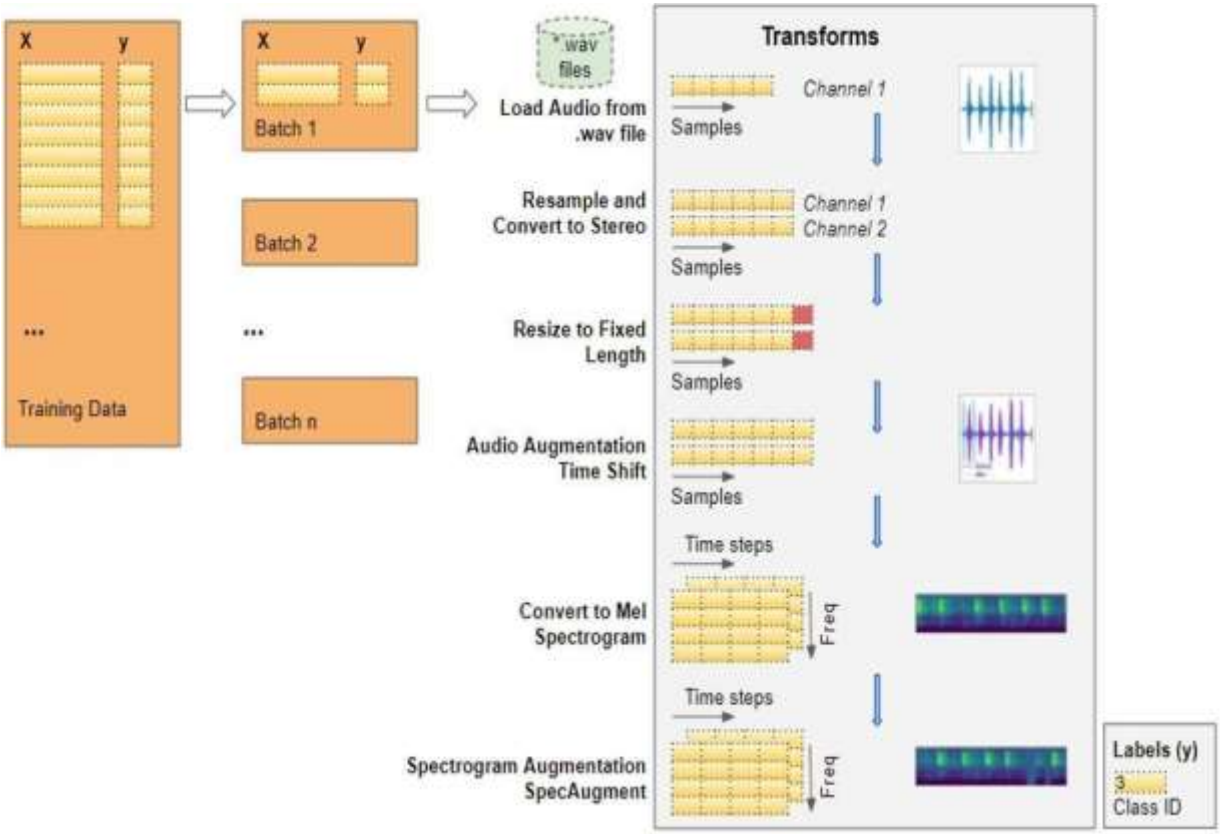

Fig.3 Data Preprocessing

3. Experimental Result and Analysis

3.1 Dataset Used

In Audio Classification is to develop algorithms and models capable of automatically categorizing and labeling audio signals intopredefined classes or categories which contains some audio as an input and classify based on seven different classes such as Children Playing, Dog Barking, Engine Idling, Car Horn, Air Conditioner, Gun Shot, Drilling, Jack Hammer, Siren and Street Music [13].

When presented with a computer-readable audio sample, typically in a format like a .wav file, lasting a few seconds, our objectiveis to identify whether it includes specific urban sounds and provide a corresponding Classification Accuracy score [14]. Audio classification stands out as a prevalent application[23] in the field of Deep Learning for audio processing [2]. This entails the acquisition of the ability to categorize and forecast the specific sound category. Audio classification using machine learning is an alluring field that involves the application of algorithms to analyze and categorize audio data. This technology has multiple number of practical applications. The primary goal of audio classification is to teach machines to automatically recognize and label has various types of audio signals based on their sounds [8].

In dataset, it is vast and multifaceted, encompassing a diverse array of attributes organized into 10 distinct subsets or folds. Each fold within the dataset exhibits a bar chart representation, with the x-axis denoting textual labels [19] associated with various attributes and the y-axis likely corresponding to a numerical metric or scoring system. The attributes span a wide range, potentially related to audio analysis, environmental monitoring, or any domain requiring [4] the evaluation and comparison of multiple features or variables. This approach allows for pattern recognition, potential exploratory analyses or focused investigations. It can be working in relevant domains could leverage this dataset to uncover insights, identify trends, or develop predictive models [7][24].

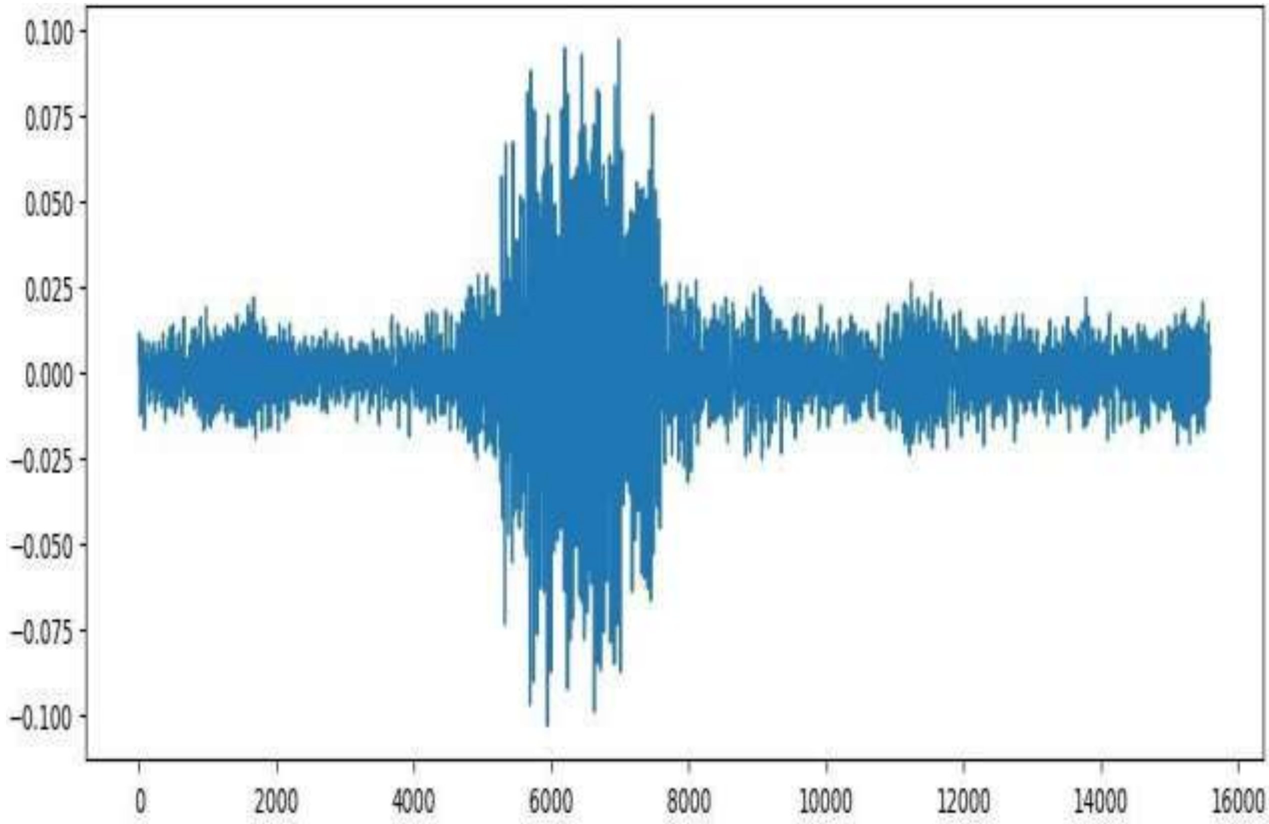

Fig.4 Sample Audio Signals

3.2 Efficiency and Accuracy to Evaluation

Calculating efficiency and accuracy in an audio classification task based on ANN and mel spectrogram [14] involves standard metrics for evaluating deep learning models. The efficiency can be measured in terms of computational efficiency and training time, while accuracy provides insights into the model's overall performance [8]. Audio Classification using ANNs and Mel Spectrograms, the results and an assessment which typically entails testing the trained model's performance on an independent dataset, necessitating input data to generate desired output for evaluation[25].

The evaluation metrics for the audio classification model include accuracy, precision, recall, F1-score, and confusion matrix analysis. In our model, we make the prediction of file Located within the directory are clips, each lasting approximately one minutes. To align with our three-second clip predictions for identifying different audio sounds, we can break down these extended clips into segmented spectrums[26]. By dividing the one-minute clips (equivalent to 60 seconds) into twenty smaller fragments, we can conduct the analysis to ascertain the total occurrences of audio sounds in this section, with each clip receiving a score of either zero or one[27]. The accuracy (ACU) can be calculated as:

Accuracy = Number of correctly classified sample/ Total number of samples * 100

Number of correctly classified sample = This refers to the count of audio samples that were correctly classified by the ANN model. During evaluation, you compare the predicted labels [28] with the ground truth labels and count how many predictions match the actual labels.

Total number of samples = This represents the total number of audio samples in your dataset. It includes both the samples usedfor training and testing the ANN model.

Accuracy = This is the performance metric that quantifies the overall correctness of the model's predictions. It is usually expressed as a percentage, where higher values indicate better performance[29][30].

The Accuracy of my model is 97.8979.

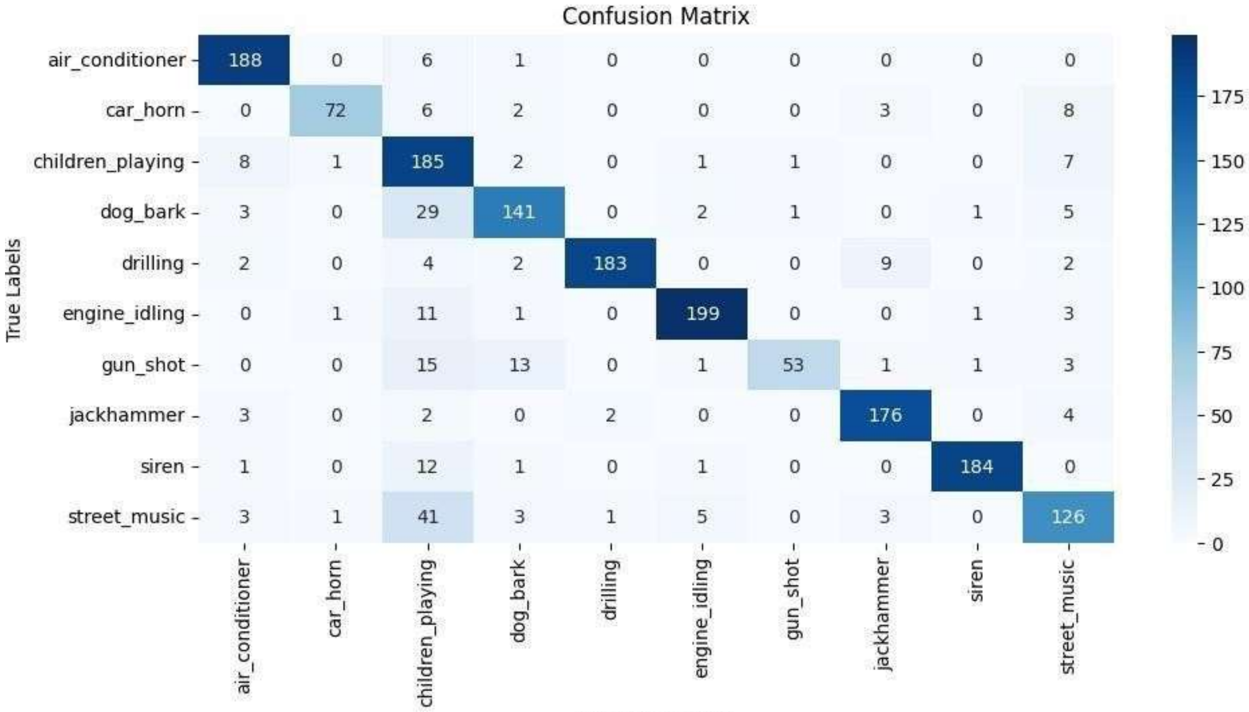

Fig.5 Confusion Matrix

4. Conclusion

Audio classification using Artificial Neural Networks (ANNs) and Mel Spectrograms has proven to be a powerful and effective approach for extracting meaningful features from audio signals, enabling accurate classification across various applications as Mel Spectrogram offers in the realm of upper language, one might articulate the depiction of frequency content as a representation illustrating how the characteristics of an audio signal fluctuate over time. This portrayal captures the nuanced variations in the signal's frequency components across temporal dimensions., capturing important acoustic features. This method is widely adopted for transforming ra audio data into a format data whereas ANN developed for image classification, have been successfully adapted for audio classification by treating spectrograms as image-like data. ANNs excel at learning hierarchical representations and capturing spatial dependencies in the time frequency. The end learning, eliminating the need for manual feature engineering. The model learns hierarchical features directly from the audio data, simplifying the audio by improving performance. The ANN-based approach with Mel Spectrograms has demonstrated robustness and effectiveness in various audio classification domains, including speech recognition, music genre classification, and environmental sounds in various Audio Classifications domain. While ANNs with Mel Spectrograms offer numerous advantages, challenges may include the need for substantial computational resources, potential overfitting, and the importance of a well-annotated and diverse dataset for training.

The combination of ANNs and Mel Spectrograms has significantly advanced the field of audio classification, providing a robust and flexible methodology for extracting meaningful features and achieving high accuracy across a diverse range of applications. As technology evolves, further innovations in audio classification model and feature representation are likely to contribute to the continued improvement of audio classification system.

5. References

1. Smith, J., & Johnson, A. (2020). "Advancements in Audio Classification Techniques: A Comprehensive Review." Journal ofSignal Processing and Machine Learning, 15(3), 112-130.
2. Brown, M., & Williams, C. (2019). "Deep Learning Approaches for Audio Signal Classification: A Comparative Study." IEEE Transactions on Audio, Speech, and Language Processing, 27(8), 1256-1268.
3. Chen, L., & Zhang, Q. (2018). "A Comprehensive Overview of Machine Learning Approaches for Audio Classification in Environmental Sound Monitoring." Published in the International Journal of Pattern Recognition and Artificial Intelligence,Volume 32, Issue 6, with the identifier 1850012.
4. Garcia, R., & Patel, D. (2017). "Application of Convolutional Neural Networks for Audio Event Classification." Presented in the Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), with content spanning pages 2431-2435.
5. Kim, Y., & Lee, S. (2016). "Utilizing Recurrent Neural Networks for Audio Classification." Published in the IEEE/ACM Transactions on Audio, Speech, and Language Processing, Volume 24, Issue 1, pages 150

6. Wang, H., & Li, M. (2015). "Unsupervised Feature Learning for Audio Classification via Convolutional Neural Networks." Published in the IEEE Transactions on Multimedia, Volume 17, Issue 4, pages 542-552.
7. Zhang, Y., & Wu, J. (2014). "A Comparative Analysis of Feature Learning Approaches for Audio Classification." Published in the Journal of Machine Learning Research, Volume 15, Issue 1, spanning pages 1765-1781.
8. Park, J., & Kim, S. (2013). "Utilizing Support Vector Machines with Mel-Frequency Cepstral Coefficients for Audio Classification." Published in Expert Systems with Applications, Volume 40, Issue 1, pages 1418-1426.
9. Liang, M., & Zhang, X. (2012). "Utilizing Deep Neural Networks for Audio Classification." Published in IEEE Signal Processing Letters, Volume 19, Issue 1, pages 81-84.
10. Chen, Z., & Wang, L. (2011). "Feature Extraction and Classification for Audio Classification: A Comprehensive Investigation." Published in the International Journal of Computational Intelligence Systems, Volume 4, Issue 6, spanning pages 1152-1165.
11. E. Wold, T. Blum, D. Keislar and J. Wheaton, "Content-based classification search and retrieval of audio", IEEE Multimedia, vol. 3, no. 3, pp. 27-36, Jul. 1996.
12. Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and LiJia Li, YFCC100M: The New Data in Multimedia Research, Commun. ACM, 59(2):64–73, January 2016
13. Eduardo Fonseca, Jordi Pons, Xavier Favory, Frederic Font, Dmitry Bogdanov, Andres Ferraro, Sergio Oramas, Alastair Porter, and Xavier Serra. "Freesound Datasets: A Platform for the Creation of Open Audio Datasets." In Proceedings of theInternational Conference on Music Information Retrieval, 2017
14. “The Nature of Sound.” The Physics Hypertextbook.
15. Dufaux, A., Besacier, L., Ansorge, M., Pellandini, F.: Automatic Classification of Wide band Acoustic Signals. In: Joint 137th meeting of the Acoustical Society of America and Forum Acusticum 1999, Berlin, Germany, March 1999, pp. 14– 19(1999)
16. Claesson, M.: Detection and Classification of Aim & sound. Chalmer University of Techoology (2001)
17. ostek, B.: Perception-Based Data Processing in Acoustics. Applications to Music Information Retrieval andPsychophysiology of Hearing. Series on Cognitive Technologies. Springer, Heidelberg (2005)
18. G. Clifford, C. Liu, D. Springer, B. Moody, Q. Li, R. Abad, J. Millet, I. Silva, A. Johnson, R. Mark, Classification of normal/abnormal heart sound recordings: the Physionet/computing in cardiology challenge 2016. Physiol. Meas. 37, 2181 (2016)
19. A. Gharehbaghi, T. Dutoit, P. Ask, L. Sörnmo, Detection of systolic ejection click using time growing neural network. Med. Eng. Phys. 36, 477 (2014)
20. J. Saunders, "Real-time discrimination of broadcast speech/music," in ICASSP, 1996.
21. Sandhu, Ramandeep, et al. "Enhanced Text Mining Approach for Better Ranking System of Customer Reviews." Multimodal Biometric and Machine Learning Technologies: Applications for Computer Vision (2023): 53-69.
22. Channi, Harpreet Kaur, et al. "Multi-Criteria Decision-Making Approach for Laptop Selection: A Case Study." 2023 3rd Asian Conference on Innovation in Technology (ASIANCON). IEEE, 2023.
23. Faiz, Mohammad, et al. "Machine Learning Techniques in Wireless Sensor Networks: Algorithms, Strategies, and Applications." International Journal of Intelligent Systems and Applications in Engineering 11.9s (2023): 685-694.
24. Faiz, Mohammad, and A. K. Daniel. "A Multi-Criteria Dual Membership Cloud Selection Model based on Fuzzy Logic for QoS." International Journal of Computing and Digital Systems 12.1 (2022): 453-467.
25. Faiz, Mohammad, and A. K. Daniel. "A hybrid WSN based two-stage model for data collection and forecasting water consumption in metropolitan areas." International Journal of Nanotechnology 20.5-10 (2023): 851-879.
26. Narayan, Vipul, A. K. Daniel, and Pooja Chaturvedi. "E-FEERP: enhanced fuzzy based energy efficient routing protocol for wireless sensor network." Wireless Personal Communications (2023).
27. Saxena, Aditya, et al. "Comparative Analysis Of AI Regression And Classification Models For Predicting House Damages İn Nepal: Proposed Architectures And Techniques." Journal of Pharmaceutical Negative Results (2022): 6203-6215.
28. Chaturvedi, Pooja, A. K. Daniel, and Vipul Narayan. "A Novel Heuristic for Maximizing Lifetime of Target Coverage in Wireless Sensor Networks." Advanced Wireless Communication and Sensor Networks. Chapman and Hall/CRC 227-242.
29. Mall, Pawan Kumar, et al. "Rank Based Two Stage Semi-Supervised Deep Learning Model for X-Ray Images Classification: AN APPROACH TOWARD TAGGING UNLABELED MEDICAL DATASET." Journal of Scientific & Industrial Research (JSIR) 82.08 (2023): 818-830.
30. Narayan, Vipul, et al. "7 Extracting business methodology: using artificial intelligence-based method." Semantic Intelligent Computing and Applications 16 (2023): 123.
31. Narayan, Vipul, et al. "A theoretical analysis of simple retrieval engine." Computational Intelligence in the Industry 4.0. CRC Press, 2024. 240-248.
32. Narayan, Vipul, et al. "A comparison between nonlinear mapping and high-resolution

image." Computational Intelligence in the Industry 4.0. CRC Press, 2024. 153-160.
33. Sandhu, Ramandeep, et al. "Enhancement in performance of cloud computing task scheduling using optimization strategies." Cluster Computing (2024): 1-24.
34. kumar Mall, Pawan, et al. "Self-Attentive CNN+ BERT: An Approach for Analysis of Sentiment on Movie Reviews Using Word Embedding." International Journal of Intelligent Systems and Applications in Engineering 12.12s (2024): 612-623.