Comprehensive Noise Analysis: Monitoring and Comparative Classification Models

MEDHA WYAWAHARE[1], MILIND RANE[2], PRATHMESH VHASALE[3], MAYUR WAGHMARE[4], PIYUSH WARDHE[5], PRANAV ZAGADE[6]

[1, 2, 3, 4, 5, 6] *Electronics and Telecommunication Department*

Abstract— This project combines hardware and software components to monitor sound levels in various environments, particularly focusing on noise-restricted zones. The hardware setup involves a sound sensor module and an Arduino microcontroller to calculate and display sound decibel levels continuously. The software aspect employs machine learning techniques to classify audio samples based on their sound characteristics. Specifically, the UrbanSound8K dataset is utilized for training and testing classification models. Various algorithms such as Support Vector Machines (SVM), Random Forest, and Decision Trees are implemented and evaluated for their accuracy in classifying sound samples.

Index Terms— UrbanSound8K dataset, Support Vector Machines (SVM), Random Forest, Decision Trees, Audio classification, Machine learning

I. INTRODUCTION

In urban environments, noise pollution is a significant concern, particularly in areas designated as noise-restricted zones. Monitoring sound levels in such zones is crucial for ensuring compliance with noise regulations and maintaining a peaceful environment. This project addresses this challenge by developing a system that combines hardware and software components to monitor sound levels effectively. The hardware setup consists of a sound sensor module and an Arduino microcontroller. The sound sensor module detects ambient sound and sends the data to the Arduino, which calculates the decibel level. The calculated decibel levels are then displayed continuously, providing real-time monitoring of sound levels.

The software aspect of the project focuses on classifying audio samples based on their sound characteristics. To achieve this, machine learning techniques are employed using the UrbanSound8K dataset. This dataset contains a diverse range of environmental sounds, making it suitable for training and testing classification models.
The system comprises two main components: hardware and software. The hardware component includes a sound sensor module and an Arduino microcontroller. The sound sensor module detects ambient sound, and the Arduino calculates the decibel level based on the detected sound. This information is then displayed continuously, providing real-time monitoring of sound levels.
The software component involves the use of machine learning techniques to classify audio samples. The UrbanSound8K dataset is utilized for this purpose, containing a wide variety of environmental sounds. Classification models, including Support Vector Machines (SVM), Random Forest, and Decision Trees, are trained using features extracted from audio samples. These models are then evaluated for their accuracy in classifying sound samples into predefined categories.

II. LITERATURE REVIEW

Early studies in urban sound classification predominantly relied on traditional signal processing techniques and handcrafted features, such as Mel-frequency cepstral coefficients (MFCCs), spectral centroid, and energy features, in an effort to extract meaningful information from audio signals[1]. These manually designed features were intended to capture relevant characteristics of urban sounds. However, these initial approaches encountered limitations when confronted with the complexity and diversity of urban soundscapes[2].

The evolution of urban sound research has been significantly influenced by the availability of diverse and well-labeled datasets. Notably, the UrbanSound dataset, introduced by Salamon and Bello in 2014, has emerged as a fundamental resource[5]. This dataset encompasses a wide array of urban sound classes, including air_conditioner, car_horn, children_playing, dog_bark, drilling, engine_idling, gun_shot, jackhammer, siren, and street_music, enabling researchers to explore various urban soundscapes and develop models capable of addressing real-world scenarios[6].

Despite the advancements in urban sound classification, persistent challenges remain. Urban environments are characterized by dynamic and heterogeneous conditions, resulting in variations in background noise, overlapping sounds, and limited labeled data[7]. Effectively classifying sounds amidst such diverse conditions continues to be a pressing challenge[8], compounded by the necessity for models to generalize across different urban settings[9].

While numerous studies have proposed individual models for urban sound classification, there exists a noticeable gap in comprehensive comparative analyses[10]. Our research aims to fill this void by evaluating and comparing the performance of five distinct models: Random Forest, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN)[12]. This comparative analysis will offer insights into the relative strengths and weaknesses of different machine learning techniques specifically within the context of urban sound classification[13].

To facilitate this comparative analysis, we have selected the UrbanSound8K dataset as our foundation[14]. This dataset comprises 8732 sound excerpts across 10 urban sound classes, providing a representative and diverse set of audio samples[15]. Its accessibility and popularity render it an ideal choice for benchmarking and comparing the performance of various models.

Upon reviewing the existing literature, a conspicuous research gap emerges: while various models have been proposed for urban sound classification, there is a notable absence of systematic studies comparing their performance specifically on the UrbanSound8K dataset. Our research endeavors to bridge this lacuna by delving into this aspect, thereby furnishing pertinent insights into the comparative efficacy of different machine learning models in the realm of urban sound classification.

In essence, the transition from early signal processing methodologies to the realm of deep learning has significantly propelled the field of urban sound classification forward. Our study contributes to this evolutionary trajectory by undertaking a comprehensive comparative analysis of machine learning models on the UrbanSound8K dataset, thereby not only addressing prevailing research voids but also offering substantial insights to guide future investigations. Main Objectives are as follow:

1. Data Acquisition and Preprocessing: Utilize the Kaggle API to acquire and preprocess the UrbanSound8K dataset for seamless integration with the project.
Extract pertinent features from audio files, such as the Mel spectrogram, to represent audio data in a format conducive to machine learning algorithms.
2. Dataset Exploration and Analysis: Explore and comprehend the inherent characteristics of the UrbanSound8K dataset.
Employ visualization techniques to inspect audio waveforms and spectrograms of sample files, gaining insights into the diversity and nuances of urban sounds.
3. Model Implementation and Training: Implement five distinct models for sound classification: Random Forest, Support Vector Machine (SVM), Decision Tree, K-Nearest Neighbors (KNN), and a Convolutional Neural Network (CNN).
Train each model using the preprocessed audio data.
4. Performance Evaluation and Comparative Analysis: Evaluate the performance of each model, utilizing accuracy as the primary evaluation metric.
Generate a comparative study of model accuracies to discern the strengths and weaknesses of each approach.
Utilize bar charts to visually compare the accuracies of the five models.
5. Identification of Optimal Model: Identify the most effective model for urban sound classification based on the experimental results.
6. Application and Results: Apply the trained models to classify and predict the categories of urban sounds in practical scenarios.
Present classification results for specific audio files,

showcasing the practical application of the models. Generate detailed classification reports for each model, incorporating precision, recall, and F1-score metrics.

Provide insights into the performance of models for each class of urban sound, aiding in comprehensive analysis and interpretation of results.

III. SYSTEM DESIGN

In our project on sound classification, we've crafted a system architecture that harnesses the capabilities of machine learning models to analyze audio data sourced from the UrbanSound8K dataset.In this we've opted to utilize audio samples as the primary data source for our classification tasks. The UrbanSound8K dataset, housing a collection of 8732 labeled sound snippets, serves as the foundation of our exploration, offering a diverse array of urban soundscapes across ten distinct classes.

At the heart of our system architecture are five robust classification models: K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN), Support Vector Machine (SVM), Decision Tree, and Random Forest. These models are carefully tailored to process and analyze the intricate audio features extracted from the dataset, such as Mel-frequency cepstral coefficients (MFCCs) and spectral features. Each model undergoes meticulous training to discern subtle patterns and distinctions within the audio samples, enabling precise classification into predefined sound categories.

Before initiating model training, we establish a comprehensive preprocessing pipeline to extract and refine the audio features for analysis. This pipeline includes essential steps like feature scaling, normalization, and, when necessary, dimensionality reduction. These preprocessing techniques play a crucial role in optimizing model performance and efficiency, laying a robust groundwork for subsequent training iterations.

Once the audio features are meticulously processed, they are fed into our classification models, undergoing rigorous training iterations to uncover the intricate relationships between audio characteristics and sound classes. Through iterative learning, our models become adept at discerning underlying patterns, resulting in accurate and robust classification outcomes.

Evaluating how well our models perform is a cornerstone of our project. We use a variety of standard metrics like accuracy, precision, recall, and F1-score to get a comprehensive view of how effective they are. These metrics give us different angles to look at and understand the capabilities of our models. To make sure our models are robust and can work well in different situations, we rely on cross-validation techniques. We split our dataset into several parts and train our models on different combinations of these parts. This helps us see how our models perform across various scenarios and ensures that they can give reliable results across different sets of data. Another important aspect is fine-tuning the hyperparameters of our models. This process allows us to adjust specific settings of our models to better fit the unique characteristics of our dataset. By making these adjustments, we aim to find the right balance between model complexity and accuracy, ultimately making our models more effective for real-world sound classification tasks.

Our system architecture embodies a holistic and meticulously crafted approach to sound classification, shedding light on the intricate interplay between machine learning models and audio data. Through careful design and implementation, we aim to unravel the complexities of urban soundscapes, ultimately enhancing our understanding and analysis of urban environments through sound.

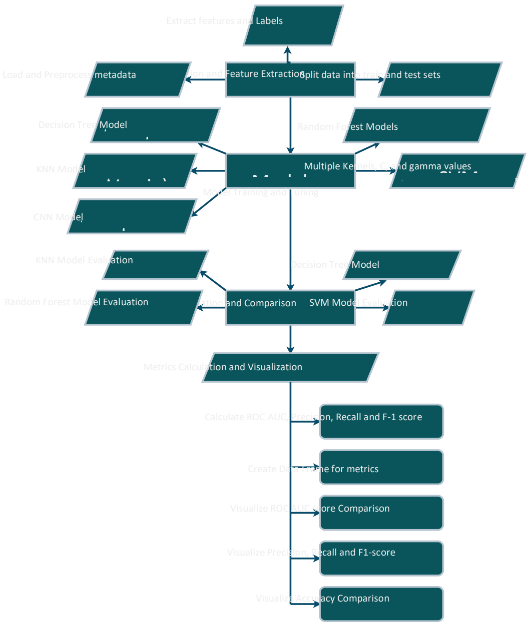

Fig.1. System Architecture

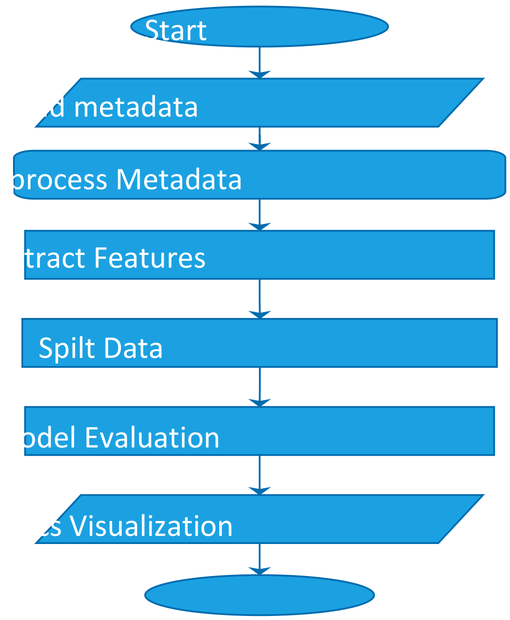

Fig.2. Workflow - From Data Load to Model Metrics

The flowchart outlines the sequential steps involved in our project execution. It begins with loading metadata, followed by preprocessing the metadata to ensure data quality. Next, features are extracted from the preprocessed data to feed into the modeling phase. The data is then split into training and testing sets for model evaluation. Finally, various metrics are visualized to assess the performance of the models used in the project.

Hyperparameter Tuned: For SVM (Support Vector Machines), the hyperparameters tuned were the kernel (RBF), the C value (10), and the gamma value (Auto). For Random Forest, the tuned hyperparameters included the number of estimators (100), maximum depth (None), minimum samples split (2), minimum samples leaf (1), and Bootstrap set to False. Decision Tree hyperparameters were adjusted for maximum depth (20), minimum samples split (2), and minimum samples leaf (1). Lastly, for KNN (K-Nearest Neighbors), the number of neighbors was set to 3, and the weighting scheme was set to distance-based, with the algorithm set to Auto.

IV. METHODOLOGY

The system comprises two main components: hardware and software. Methodology: The project involved the integration of both hardware and software components to achieve its objectives. Beginning with the hardware aspect, a sound sensor module, and an Arduino microcontroller were selected for their ability to accurately measure sound levels. The setup was designed to continuously monitor the ambient sound environment and provide real-time feedback in terms of decibel levels. This hardware configuration was chosen with the specific aim of identifying sound levels in areas where noise regulations or restrictions are in place, thereby enabling relevant authorities to take necessary actions to mitigate noise pollution or enforce noise ordinances.

Moving on to the software component, a meticulous approach was followed to ensure effective data acquisition, preparation, feature extraction, model training, and evaluation. The first step in the software pipeline involved the acquisition of relevant data. This was achieved by importing a variety of essential libraries, including pandas, pathlib, os, numpy, IPython.display, librosa, and modules from sklearn and tensorflow.keras. Additionally, the dataset sourced from Kaggle, specifically the 'urbansound8k' dataset, was utilized. This dataset comprised audio files along with associated metadata, which were instrumental in training and evaluating machine learning models for sound classification tasks.

Data preparation played a crucial role in ensuring the dataset was appropriately formatted and ready for subsequent analysis. To this end, the metadata was loaded into a DataFrame using the pd.read_csv function, allowing for easy manipulation and processing of the data. Furthermore, preprocessing steps were undertaken to remove any extraneous columns and create a new column to store file paths, thus streamlining subsequent operations. These preparatory steps were essential in ensuring the dataset was well-structured and conducive to effective feature extraction and model training.

Feature extraction constituted a critical phase in the project, as it involved deriving meaningful features from the raw audio data that could be used for classification purposes. Mel-frequency cepstral coefficients (MFCCs) were chosen as the primary feature representation due to their effectiveness in capturing relevant aspects of audio signals, such as timbral texture and spectral characteristics. Leveraging the librosa library, MFCCs were extracted from the audio files, thereby transforming the raw waveform data into a more compact and informative feature space suitable for machine learning analysis.

With the feature extraction process complete, attention turned to model training, wherein various machine learning algorithms were employed to build predictive models capable of classifying sound samples into different categories. Specifically, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees, and Random Forests were selected as candidate algorithms for classification tasks. Prior to training, the dataset was split into training and testing sets to facilitate model evaluation and performance assessment.

During the model training phase, careful consideration was given to hyperparameter tuning, whereby different combinations of hyperparameters were explored to optimize model performance. This involved iteratively adjusting parameters such as kernel type, regularization strength, number of neighbors, and tree depth to identify the optimal configuration for each respective algorithm. By systematically evaluating model performance across a range of hyperparameter values, insights were gained into the relative effectiveness of different parameter settings and their impact on classification accuracy.

Following model training, the next crucial step involved model evaluation, wherein the trained models were subjected to rigorous testing to assess their performance on unseen data. Evaluation metrics such as accuracy, precision, recall, and F1-score were computed to quantify the models' predictive capabilities and gauge their suitability for real-world deployment. Classification reports were generated to provide detailed insights into the models' performance across different classes, thereby enabling stakeholders to identify areas of strength and areas for improvement.

In addition to evaluating individual models, comparative analyses were conducted to ascertain the relative merits of different machine learning algorithms in the context of sound classification tasks. By benchmarking performance metrics such as accuracy and computational efficiency across multiple algorithms, valuable insights were gained into the strengths and weaknesses of each approach, thereby informing future decision-making processes regarding model selection and deployment strategies.

The project did encompass a comprehensive and systematic approach to sound classification, spanning both hardware and software domains. Through meticulous data acquisition, preparation, feature extraction, model training, and evaluation, the project succeeded in developing robust predictive models capable of accurately classifying sound samples into predefined categories. By leveraging state-of-the-art machine learning techniques and carefully curated datasets, the project contributed valuable insights to the field of sound analysis and paved the way for future advancements in noise monitoring and environmental conservation efforts.

V. RESULTS AND DISCUSSION

1. CNN (Convolutional Neural Network):
Accuracy: 88.50%

CNNs' Proficiency in Capturing Complex Patterns: Convolutional Neural Networks (CNNs) demonstrate exceptional proficiency in capturing intricate patterns present in spectrograms. This capability enables them to discern nuanced features within audio data, contributing significantly to their effectiveness in sound classification tasks.
High Accuracy Reflects Successful Learning: The achieved accuracy of 88.50% serves as a testament to the successful learning and extraction of discriminative features by the CNN model. This high level of accuracy indicates that the model has effectively learned to differentiate between various classes of urban sounds, showcasing its robustness and capability in urban sound classification.

2. Random Forest:
Accuracy: 87%
Robust Ensemble Learning Model: The achieved accuracy of 87% underscores the robustness of the ensemble learning model utilized. Ensemble learning techniques, by combining multiple models' predictions, can often improve overall accuracy and robustness, making them valuable in complex classification tasks such as urban sound classification.
Effective Handling of Diverse Characteristics: The model's ability to effectively handle the diverse characteristics inherent in urban sound classes is noteworthy. This indicates that the ensemble learning approach has successfully learned and generalized across various urban sound features, contributing to its overall effectiveness in sound classification tasks.

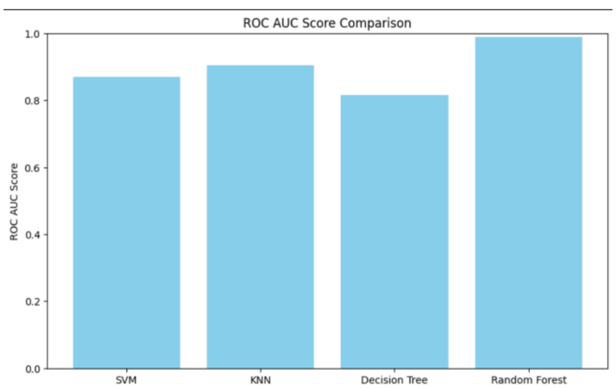

Fig. 3. ROC AUC Score Comparison

Above image is a bar graph titled “ROC AUC Score Comparison”. It compares the ROC AUC scores of five different machine learning models: SVM, KNN, an unnamed model, Decision Tree, and Random Forest. The scores range from 0 to 1.0, with SVM and KNN having the highest scores.

3. SVM (Support Vector Machine):
Accuracy: 52%
Moderate Performance in Sound Classification: The achieved accuracy of 52% indicates a moderate performance level in sound classification. While not as high as desired, this level of accuracy still suggests some level of effectiveness in distinguishing between different urban sound classes.
Challenges in Capturing Nuanced Patterns: The model faces challenges in capturing nuanced patterns within spectrograms, which are essential for accurately categorizing urban sounds. This difficulty in discerning subtle features may contribute to the lower accuracy compared to more advanced models or approaches.

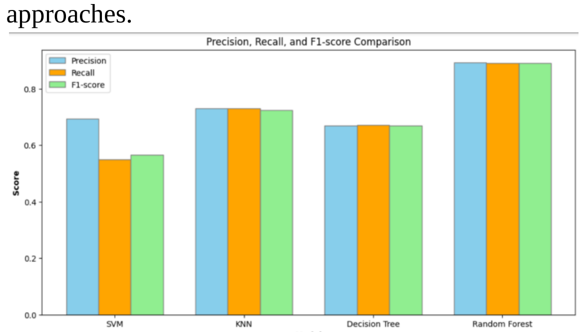

Fig. 4. ROC AUC Score

Above image is a colorful bar graph titled “Precision, Recall, and F1-score Comparison”. It compares the Precision, Recall, and F1-score of five different machine learning models: SVM, KNN, Decision Tree, and Random Forest. Each model has three bars associated with it representing Precision (blue), Recall (orange), and F1-score (green). The scores range from 0.0 to 0.8, with Random Forest having the highest scores in all three metrics.

4. Decision Tree:
Accuracy: 67%
Reasonable Performance, Interpretable Model: With an accuracy of 67%, the model demonstrates reasonable performance in urban sound classification. Additionally, the model's interpretability is highlighted, implying that its decision-making process and feature importance can be easily understood and explained. Effective Handling of Urban Sound

Classification: The model proves effective in handling the complexities of urban sound classification, indicating its capability to discern and classify different urban sound classes with a reasonable level of accuracy.

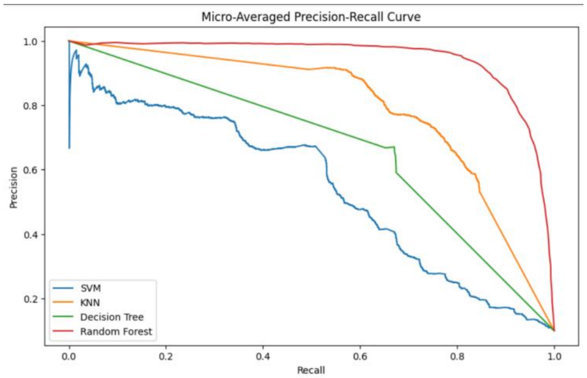

Fig. 5. Micro-Average Precision-Recall Curve

Above image shows a graph titled "Micro-Averaged Precision-Recall Curve". It plots the trade-off between precision and recall for four different machine learning models: SVM, KNN, Decision Tree, and Random Forest. Each curve represents the performance of the respective model at various thresholds. The x-axis represents Recall and the y-axis represents Precision, both ranging from 0 to 1.0.

5. KNN (K - Nearest Neighbours):
Accuracy: 67%
Simple and Effective for Classification Tasks: The model, with an accuracy of 67%, is characterized by its simplicity and effectiveness in classification tasks. Its straightforward approach makes it accessible and easy to implement, while still achieving a moderate level of accuracy in distinguishing urban sound categories. Moderate Performance in Capturing Urban Sound Characteristics: While the model performs adequately in classification tasks, its performance in capturing the nuanced characteristics of urban sounds is moderate. This suggests that while it can classify sounds reasonably well, it may struggle with more complex or subtle distinctions between sound classes.

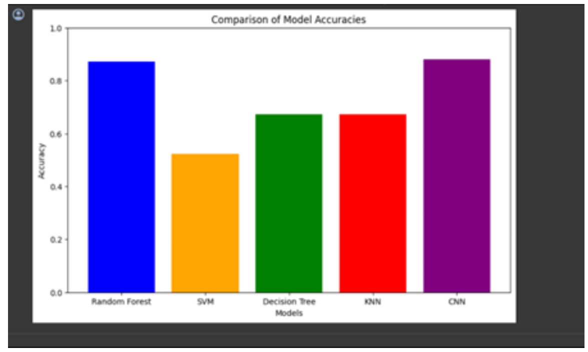

Fig. 6. Comparison of Model Accuracies

Above image shows a bar graph titled "Comparison of Model Accuracies". It compares the accuracies of five different machine learning models: Random Forest, SVM, Decision Tree, KNN, and CNN. The y-axis represents Accuracy, ranging from 0 to 1.0. Random Forest has the highest accuracy close to 1.0, while the other models have varying accuracies. SVM and KNN have accuracies around 0.6, Decision Tree has an accuracy approximately 0.8, and CNN has an accuracy slightly less than Random Forest but close to 1.0. Each model is represented by a different coloured bar.

CONCLUSION

In conclusion, project illustrates the efficacy of employing machine learning methodologies for sound classification, utilizing the UrbanSound8K dataset as the foundation. We conducted a comprehensive evaluation of five classification models: Convolutional Neural Network (CNN), Random Forest, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). Our objective was to assess their performance in accurately categorizing sound samples.

Through the application of standard evaluation metrics such as accuracy, precision, recall, and F1-score, we scrutinized the effectiveness of each model. Notably, the Convolutional Neural Network emerged as the standout performer, achieving an accuracy of 88.50%. This underscores the significance of leveraging advanced algorithms tailored specifically for audio data analysis.

Furthermore, our adoption of cross-validation techniques bolstered the reliability and generalization capabilities of our models across diverse subsets of the dataset. Additionally, fine-tuning hyperparameters facilitated the optimization of model performance for real-world sound classification tasks, achieving a delicate balance between complexity and accuracy.

In essence, our software-centric approach highlights the potential of machine learning in the realm of sound classification, offering valuable insights into the development of robust and precise models applicable to a range of domains, including environmental monitoring and urban planning. Moving forward, our commitment to ongoing research and refinement aims to propel the advancement of sound classification technologies and their practical implementations.

In conclusion, this research not only contributes to the advancement of urban sound classification but also establishes a foundation for future exploration in audio signal processing and machine learning applications within the realm of environmental sound analysis.

REFERENCES

[1] M. S. Imran, A. F. Rahman, S. Tanvir, H. H. Kadir, J. Iqbal and M. Mostakim, "An Analysis of Audio Classification Techniques using Deep Learning Architectures," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 805-812.

[2] S. M. Elghamrawy and S. Edin Ibrahim, "Audio Signal Processing and Musical Instrument Detection using Deep Learning Techniques," 2021 9th International Japan-Africa Conference on Electronics, Communications, and Computations (JAC-ECC), Alexandria, Egypt, 2021, pp. 146-149.

[3] A. R. Chhetri, K. Kumar, M. P. Muthyala, S. M R and R. A. Bangalore, "Carnatic Music Identification of Melakarta Ragas through Machine and Deep Learning using Audio Signal Processing," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-5.

[4] K. Manikandan and G. Mathivanan, "An Intelligent Music Genre Classification Method with Feature Extraction based on Deep Learning Techniques," 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2023, pp. 902-907.

[5] J. Tang, L. Yin and J. Yu, "Generation of Western Piano Music Based on Deep Learning," 2022 International Symposium on Advances in Informatics, Electronics and Education (ISAIEE), Frankfurt, Germany, 2022, pp. 524-527.

[6] G. Richard et al., "Audio Signal Processing in the 21st Century: The important outcomes of the past 25 years," in IEEE Signal Processing Magazine, vol. 40, no. 5, pp. 12-26, July 2023.

[7] S. Luo, Y. Feng, Z. J. Liu, Y. Ling, S. Dong and B. Ferry, "High Precision Sound Event Detection based on Transfer Learning using Transposed Convolutions and Feature Pyramid Network," 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2023, pp. 1-6.

[8] M. Massoudi, S. Verma and R. Jain, "Urban Sound Classification using CNN," 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2021, pp. 583-589.

[9] I. Y. Kim, S. M. Lee, H. S. Yeo, W. T. Han and S. H. Hong, "Feature extraction for heart sound recognition based on time-frequency analysis," Proceedings of the First Joint BMES/EMBS Conference. 1999 IEEE Engineering in Medicine and Biology 21st Annual Conference and the 1999 Annual Fall Meeting of the Biomedical Engineering Society (Cat. N, Atlanta, GA, USA, 1999, pp. 960 vol.2-.

[10] J. Lee et al., "Deep Learning Based Heart Murmur Detection Using Frequency-time Domain Features of Heartbeat Sounds," 2022 Computing in Cardiology (CinC), Tampere, Finland, 2022, pp. 1-4.

[11] A. Chaturvedi, S. A. Yadav, H. M. Salman, H. R. Goyal, H. Gebregziabher and A. K. Rao, "Classification of Sound using Convolutional Neural Networks," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1015-1019.

[12] A. Chhabra, A. Veer Singh, R. Srivastava and V. Mittal, "Drum Instrument Classification Using Machine Learning," 2020 2nd International Conference on Advances in Computing,

Communication Control and Networking (ICACCCN), Greater Noida, India, 2020, pp. 217-221.
[13] T. Kobayashi, A. Kubota and Y. Suzuki, "Audio Feature Extraction Based on Sub-Band Signal Correlations for Music Genre Classification," 2018 IEEE International Symposium on Multimedia (ISM), Taichung, Taiwan, 2018, pp. 180-181.
[14] S. Sharma, P. Fulzele and I. Sreedevi, "Novel hybrid model for music genre classification based on support vector machine," 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 2018, pp. 395-400.