

Spectral Analysis of Musical and Random Noise Signals with Audio Classification: A Comparative Study

Sandip Bordoloi¹, Mayuri Sarmah Mazinder², Rashmi Rekha Roy³

¹ Department of Electrical Engineering, Girijananda Chowdhury University, Azara, Guwahati, INDIA

² Department of Computer Science and Engineering, Girijananda Chowdhury University, Azara, Guwahati, INDIA

³ Department of Electronics and Communication Engineering, Assam Science and Technology University, Guwahati, INDIA

¹sandip_ee@gcuniversity.ac.in*, ²mayurisarmahmazinder@gmail.com, ³rashmirekharoy17@gmail.com

Abstract: *This study presents a comparative analysis of two types of audio signals: random noise and music audio. The random noise was recorded in a crowded public bus on an urban road, while the music audio was captured using the same device in a controlled environment, free from external disturbances. Statistical and spectral analyses, including histogram analysis, power spectrum, Fast Fourier Transform (FFT), and Auto-Correlation Function (ACF), were employed to evaluate the signals. The noise signal showed a bell-shaped histogram with samples concentrated around the mean, an ACF that decreased with lag, and the FFT spectrum, which shows that the energy is distributed uniformly across all frequencies. In contrast, the music signal displayed a sharp histogram, distinct voice and music components, rhythmic repetitions, and specific frequencies in its ACF and FFT spectra. Additionally, Support Vector Machine (SVM) classification was performed on both signals. The noise signal was accurately categorized into three noise levels with an accuracy of approximately 82%, while the music audio was categorized into voice, music, and noise components with 90% accuracy. The superior classification performance for music is attributed to the clear separation between voice and music components. This study demonstrates the potential of SVM in audio signal classification, providing a robust approach for distinguishing between noise and music audio.*

Keywords: Audio signal analysis; Random Noise; Music audio; FFT; ACF; Support Vector Machine (SVM).

Open Access. Copyright ©Authors(s). Distributed under [Creative Commons Attribution 4.0 International License \(CC-BY\)](#).
Article history- Received: 30 January 2025 and Accepted: 25 April 2025.

1. Introduction

Random Noise Signals (RNS) are often characterized by their chaotic and unpredictable nature. They lack any regularities in their amplitude and frequency variations over time. They typically exhibit a wide range of frequencies with varying intensities and there exist no phase relationships among different frequency components. Most random noises are man-made. These man-made noises are sourced from automobiles, machines, electric motors, ignitions, etc. Man-made noise is more effective in the frequency range of 1 MHz- 500 MHz, in urban, suburban, and Industrial areas [1]. This type of noise is under human control and can be eliminated by removing the noise source. However, in the rapid human-made development, era, removing such noise sources is quite impossible. A random noise process consists of a random sample (independent observations) from a normal distribution with constant mean and standard deviation. There are no trends because, as a result of

independence, the observations “have no memory” of the past behavior of the series [2].

A signal is often characterized by the signal-to-noise ratio (SNR). It is defined as the amount of power of the desired signal relative to background noise (undesired signal) [3]. SNR can be expressed by the power density spectrum. In the case of a random noise signal, the noise component is usually high, making the SNR value low. Therefore, random noise signals are often unpleasant in nature and create disturbances to the human hearing sensors i.e. the ears.

On the other hand, Music signals possess particular structural and audio features that set them apart from spoken language or other nonmusical signals [4]. The music signals have a structured pattern. They exhibit well-defined patterns in the form of melodies, rhythms, and harmonies. They are organized in time, with clear beginnings, endings,

and internal structure. Strong correlations between various frequency components are evident in them. Random noise signals are random, yet music signals can be classified into different categories, i.e. they are characterized by disorder and unpredictability. Music and noise signals are both the creations of humans, however, music is the creation of humans with creation and intentions, often made to arouse human feelings and emotions, while random noise signals are typically generated artificially or naturally, but without a specific artistic purpose.

This brief overview provides the contrasting characteristics of random noise signals and music audio signals, giving the path for a more in-depth and scientific comparison between these two signals.

Signals can be analyzed by different tools. This comparative study has employed statistical analysis tools- mean, standard deviation, time-domain and frequency-domain analysis, probability density function (PDF), power spectral density (PSD), and the Fast Fourier Transform (FFT). After using these basic tools for characterizing the two signals in focus, the study uses the Support Vector Machines (SVM) tool for audio classification. Here SVM will self-classify the audio signals and classify the signals into different classes based on the features of the audio signals. The study here will focus more on the application of SVM to a random noise signal and a music audio signal.

The analysis of random noise signals and the generation of random numbers using Support Vector Machines (SVM) can be informed by various methodologies outlined in the literature. Support Vector Machines (SVM) have been extensively utilized to classify audio signals, particularly in the domain of music genre classification. SVM is favored for its ability to manage data with large dimensions and its effectiveness in binary classification tasks, which can be extended to multi-class problems through techniques like one-vs-one or one-vs-all. The paper discusses using SVM for interpolating noisy signals, highlighting Primal's performance and dual signal models. It highlights the status of parameter tuning and the ability of SVM to yield sparse solutions in the occurrence of random noise. [5]. In music genre classification, other algorithms have been compared with SVM like the Naive Bayes and k-Nearest Neighbors, demonstrating its robustness and accuracy when evaluated using metrics like precision, recall, and F1-score [6]. The classification process typically involves two main stages: feature extraction and classification. Features such as short-time energy, zero-crossing rate, bandwidth, spectral centroid, and Mel-frequency cepstral coefficients (MFCC) are commonly extracted from audio signals

to serve as inputs for the SVM model [7]. Various kernel functions, including linear, sigmoid, radial basis function (RBF), and polynomial, can be used within SVM to enhance its performance across different datasets [8]. Moreover, improvements to the SVM algorithm, such as normalization of vocal signals and advanced feature selection, have been proposed to increase its efficacy in complex noise environments, further enhancing its application in audio classification tasks [9]. These studies collectively highlight the versatility and effectiveness of SVM in audio signal classification, making it a valuable tool for practitioners in the field. The study presents a music vocal spectrum classification model that utilizes an improved SVM algorithm, which is selected after analyzing the differences between linear and non-linear methods in complex noise environments. The proposed Support Vector Machine (SVM) in [10] is utilized for the effective classification of acoustic signals by leveraging spectral features. The study highlights the significance of audio signal feature extraction and shows that the SVM works more accurately than other machine learning techniques. By employing various settings for feature extraction, the methodology enhances the categorization of audio signals, showcasing SVM's capability in addressing challenges in acoustic signal classification. The proposed machine learning-based architecture effectively classifies audio signals by leveraging both Chebychev moments and Mel-Frequency Cepstrum Coefficients (MFCC), enhancing the feature extraction process through the unique properties of Chebychev moments, particularly their symmetry characteristics [9].

This work undermines the application of statistical methods, along with a few tools for audio analysis, and SVM for creating a classification for random noise signals and musical audio signals.

2. Theoretical Overview of the Work

The experimental work aims to investigate the statistical, and spectral characteristics of a random noise signal and a musical audio signal. The study further classifies the two types of audio signals by using the SVM technique. Both audio signals are recorded using the same device. The random noise signal was captured within a crowded bus while traveling a distance of around 200 meters in a busy urban area, encompassing diverse sounds. The random sound comprises of a mixture of audio signals coming from the bus engine, internal vibration, traffic, people's conversation, as well as external environmental sounds. To provide a distinct comparison with the random noise signal, the music audio was recorded in a quiet environment, inside a room by playing a musical audio and recorded using

the same recording device. Both the audio signals were then processed using an R program implemented in the RStudio tool [14]. The R program uses the statistical, and machine learning library packages with graphic display packages to visualize the program's output.

2.1 Statistical and Spectral Analyses

The statistical analysis evaluates the statistical analysis of the two signal types using the mean and standard deviation. The time-domain and frequency-domain analyses were used to perform the spectrum analysis, probability density function (PDF), power spectral density (PSD), and the Fast Fourier Transform (FFT). A comparison of the two signals under study is shown in Table 1. In contrast to the random noise signal, the music audio (MA) signal exhibits distinct characteristics. The number of samples in the MA signal is 1141632, compared to the RNS's 943488. The duration and sampling rate for both signals are detailed in Table 1. Both signals are 16-bit. As depicted in Figure 1, the spectra of both the RNS and MA signals are presented.

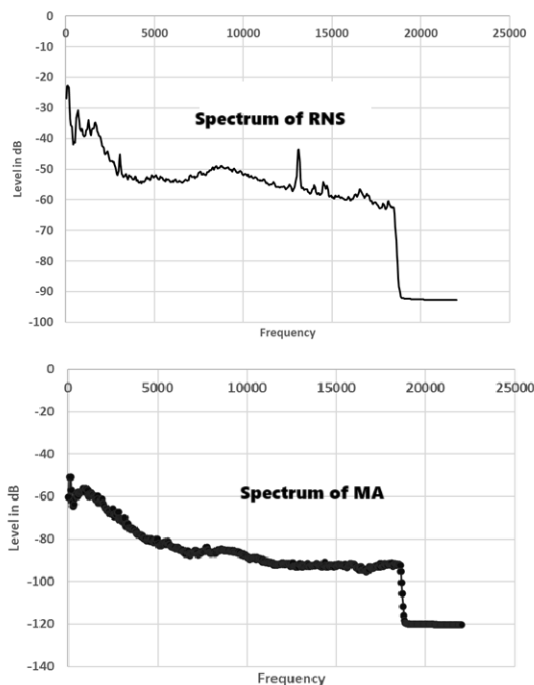


Figure 1: Spectrum of Random noise signal and Music audio signal (Frequency vs. Level in dB)

The spectrum of the RNS signal is maximum at -22.5 dB, while the spectrum of the MA signal is maximum at -50.3 dB. A higher dB (less negative) generally indicates a stronger signal, while a lower dB signal indicates a weaker signal. The difference between these two signals is 27.5 dB, indicating a large difference in power signal, with the -22.5 dB signal being significantly stronger than

the -50.3 dB signal. Therefore, it can be said from the spectrums that the RNS is a stronger signal compared to the MA signal. The next section shows the analysis of these two signals.

3. Experimental Analysis

3.1 Audio signals used for the study

To carry out this study, the audio signals, as described in section 2.1 have been used. The power spectrum in Figure 1 illustrates the type of signals used.

From the Power Spectrums a Random noise signal and a Music audio signal as shown in Figure 2, it is found that the random noise signal's (RNS) power range is quite high as compared to the music audio (MA). The amount of power (or energy) across various frequencies within a signal is expressed by a power spectrum. The x-axis typically represents frequency, while the y-axis signifies the power associated with each frequency. The range on the x-axis (frequency axis) indicates the span of frequencies present in the signal. The RNS is less structured and typically has a more uniform distribution of energy across frequencies. While it might contain a wide range of frequencies, there may be a less visible peak in the power spectrum at certain frequencies because the power associated with each frequency may be slightly comparable. In the case of MA, the power spectrum shows distinct peaks and valleys, indicating a concentration of power at specific frequencies. The high-frequency range of the MA is due to the music and human voice that contribute to the higher frequency content.

Table 1: Properties and statistical data of noise and music signal

Signal Type	No of samples	Duration of signal in Sec	Sampling Rate in Hz	Mean Signal	Std. Dev
RNS	943488	21.39	44100	-0.5174491	1723.234
MA	1141632	25.89	44100	-0.567538	85.62847

For analyzing the spectral analysis, the following tools were used and the analysis is given as follows.

Figure 3 shows the histogram of the RNS and MA respectively. The histogram of the RNS reflects its random nature, characterized by a bell-shaped distribution, widespread, and low density. This indicates that the noise samples are dispersed randomly around the mean value, with fewer instances occurring at the extremes and the majority of samples concentrated near the centre.

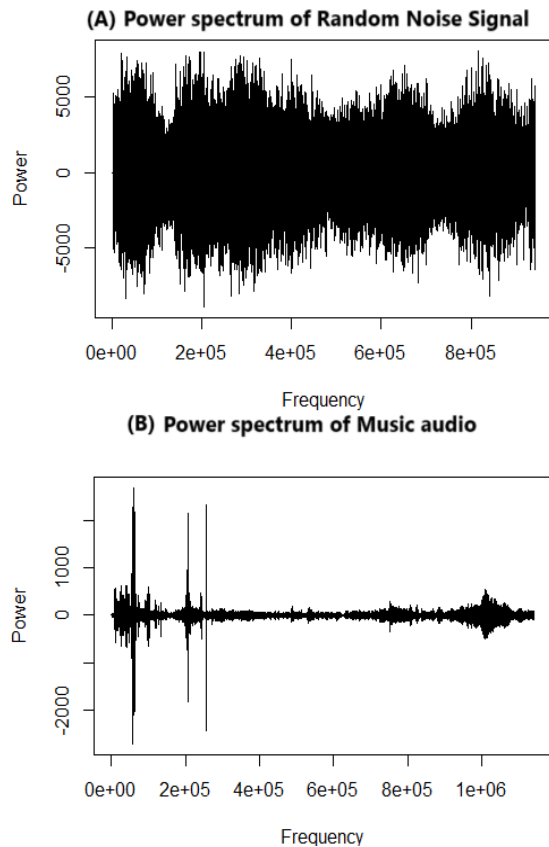


Figure 2: Power Spectrum of (a) Random noise signal and (b) Music audio signal

The histogram of the MA, on the other hand, has multiple peaks and valleys, suggesting a more complex distribution of the music samples. This is expected, as music signals typically contain a wide range of frequencies and amplitudes, leading to a less uniform distribution compared to random noise. The MA signal histogram is more concentrated, with the majority of the samples clustered around certain values. This indicates that the music signal has a more limited range of amplitudes and frequencies compared to the noise. The density of the music signal histogram is higher, particularly around the peaks. This proposes that the music signal has more energy intensity at certain frequency bands or amplitude ranges.

The auto-correlation function (ACF) of both RNS and MA signals is displayed in Figure 4. From Figure 4, it is noted that in the case of RNS, the ACF decays as the lag increases. This indicates that the signal has a short memory, meaning that past values have little influence on future values. In the MA signal, the ACF decreases rapidly and then increases. This is characteristic of music signals, which, in contrast to ACF, lacks any visible periodic components, frequently having repetitive structures and patterns that remain over longer time periods.

The FFT spectrum of the RNS appears relatively flat across the entire frequency range. This indicates that the noise contains energy distributed equally across all frequencies. In contrast, the FFT spectrum of the music signal is characterized by distinct peaks and valleys, revealing the occurrence of specific frequencies and their relative amplitudes.

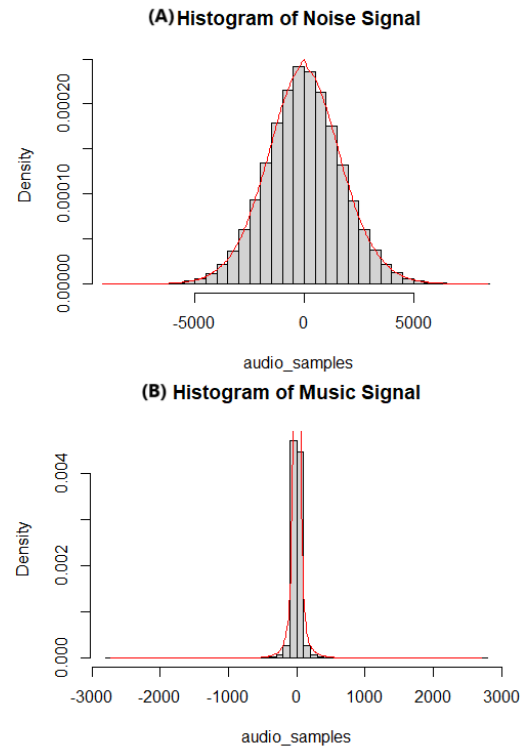


Figure 3: Histogram of noise and music signal

The FFT magnitude spectrum of RNS and MA signal is shown in Figure 5.

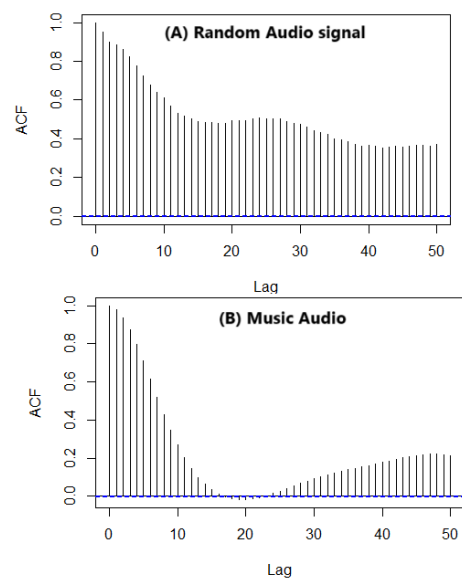


Figure 4: ACF of noise and music signal

Deep learning models can recognize intricate patterns in audio signals and achieve high accuracy when trained on large datasets [12]. One effective machine learning method that is frequently used for classification tasks, including audio classification, is the Support Vector Machine (SVM). SVM is used for audio classification as SVM performs well when good feature representations are available [12]. It performs well with High-dimensional, well-defined features (such as MFCC) and is able to perform binary classification tasks (speech vs. noise). SVM is better than the k-Nearest Neighbors (KNN) algorithm because KNN makes predictions based on nearest neighbors, which can lead to poor performance on noisy or high-dimensional data. While the Decision Trees tend to overfit when working with high-dimensional feature spaces and the Convolutional Neural Network (CNN) algorithms require extensive training, SVM functions well over decision trees and CNN, as SVM is based on optimal hyperplane that maximizes the margin between classes, leading to better generalization on unseen data and can train faster, especially with optimized kernels.

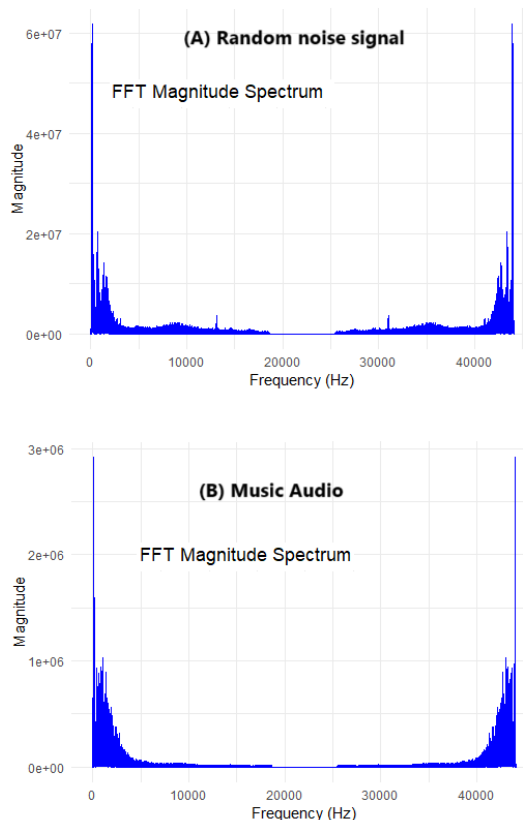


Figure 5: FFT of noise and music signal.

The generalised used for carrying out the SVM operation for audio signals are as follows-

- i. **Feature Extraction:** Due to its complexity, audio signals are transformed into numerical features that represent relevant characteristics using the Mel-Frequency Cepstral Coefficients (MFCCs) feature.
- ii. **Data Preprocessing:** The performance is enhanced by normalizing the extracted features from the SVM algorithm.
- iii. **SVM Training:** An SVM model is trained on a labeled dataset of audio samples. The training process involves finding the optimal hyperplane that maximizes the margin between different classes.
- iv. **Classification:** Once trained, the SVM model can classify new, unseen audio samples. The model assigns a class label to each audio sample based on its features.

The process of the SVM classification specific to this research is as follows-

- i. Audio samples are transformed into a Wave object, which represents the audio in a format suitable for processing. MFCCs (Mel-Frequency Cepstral Coefficients) are then calculated from the audio. MFCCs are numbers that summarize the sound properties. MFCC was computed and compared with other features like Spectral Contrast and Chroma Features. The MFCC showed the highest accuracy of 89.47%, compared to the Spectral contrast feature (78.32%) and the Chroma feature (81%) in an average of RNS and MA signals. This shows that MFCC is best used for MFCC classification.
- ii. The audio frames are divided into three groups, and each group is assigned a label (class 1, class 2, or class 3). A table (a data frame) is generated by combining the MFCCs with the labels that correspond with them. A machine-learning model will be trained and tested. The data is split into training and testing sets in the ratio of 80:20. Based on the training and testing dataset obtained from both the RNS and MA signal, the specifications of the dataset are stated in Table 2.

Table 2: Specifications of RNS and MA dataset

Signal Type	RNS	MA
Train data dimension	1709 × 14	2069 × 14
Test Data dimension	428 × 14	508 × 14
Train data size in kB	189.8	229.1
Test Data size in kB	49.6	59.5

- iii. SVM is used to classify data. The SVM uses a radial basis function (RBF) kernel, which can handle complex patterns in the data. SVM hyperparameters are tuned (Cost and Gamma) using 10-fold cross-validation.
- iv. Each SVM kernel (default, polynomial, radial, sigmoid) has been tested with different values of Cost and Gamma to find the best combination that minimizes the classification error (performance score). The Radial Kernel (RBF) is the best choice for this dataset, as it achieves the lowest error (0.0662). The Sigmoid Kernel performs the worst and should be avoided (error: 0.2505). Polynomial Kernel performs decently (error: 0.1035), but not as good as RBF. The Default Kernel is not optimal, as its error is higher than RBF and Polynomial (error: 0.1716)
- v. The trained SVM model is used to forecast the labels (class1, class2, or class3) for the test data. The model's predictions are compared to the actual labels in the test data.
- vi. The confusion matrix is created to compare the predicted labels with the actual labels.

The RNS signal is classified into three classes viz- noise level 1, noise level 2 and noise level 3, as the sample contains the entire noise. After running the code, the SVM classifier gives the confusion matrix, along with Precision, Recall and F1 score in the following form as shown in Table 3.

Table 3: Confusion matrix, along with Precision, Recall and F1 score for RNS using SVM

	Actual					
Predicted	NL1	NL2	NL3	Precision	Recall	F1
NL1	112	13	7	0.8356	0.7393	0.7845
NL2	16	122	27	0.7718	0.8778	0.8214
NL3	5	11	115	0.8421	0.8484	0.8452

Where NL is the noise label. This means that 112 labels are correctly labelled as NL1, while 13 are NL2, but predicted as NL1, while 7 are NL3 but predicted as NL1. In this way, the entire confusion matrix can be interpreted. This model, for the given RNS, gives an accuracy of 81.54%.

The MA signal is classified into three labels i.e. voice, instruments, and noise [13]. The SVM classification's confusion matrix along with Precision, Recall and F1 score is shown in Table 4.

Table 4: Confusion matrix, along with Precision, Recall and F1 score for MA signal using SVM

	Actual					
Predicted	Instruments	Noise	Voice	Precision	Recall	F1
Instruments	125	9	17	0.8389	0.8278	0.8333
Noise	11	180	1	0.9473	0.9375	0.9424
Voice	13	1	161	0.8994	0.9200	0.9096

The model shows that the instrument label was predicted 125 times, while 9 instances of noise were predicted as instruments. 17 instances of actual voice were predicted to be instruments. Likewise, for the other labels Noise and Voice, the predicted values can be explained. This model gives an accuracy of 89.96%.

Regarding the SVM, the model works better in the instance of the MA signal, as the accuracy percent is higher by nearly 8%, i.e., approximately 90% in the MA signal compared to nearly 82% in the RNS signal. Thus, in SVM, the same model, when used with an MA signal gives nearly 90%, while the RNS gives an approximate accuracy of 82%, marking a contrast between the signals under study.

The SVM is compared with KNN and a Decision tree for performance evaluation. The comparison has been done based on Precision, Recall and F1 score as shown in Table 5.

Table 5: Comparison of SVM, KNN and Decision tree on RNS and MA signal

Classifier	Classification	Precision	Recall	F1
Random Noise Signal				
KNN	NL1	0.7945	0.7435	0.7682
	NL2	0.7852	0.8239	0.8239
	NL3	0.8421	0.8629	0.8517
Decision Tree	NL1	0.6369	0.6690	0.6526
	NL2	0.6912	0.6560	0.6732
	NL3	0.8045	0.8106	0.8106
SVM	NL1	0.8356	0.7393	0.7845
	NL2	0.7718	0.8778	0.8214
	NL3	0.8421	0.8484	0.8452
Music Audio				
KNN	Inst	0.8255	0.8601	0.8424
	Noise	0.9526	0.9187	0.9354
	Voice	0.9106	0.9157	0.9131
Decision Tree	Inst	0.7248	0.6625	0.6923
	Noise	0.8421	0.8695	0.8556
	Voice	0.7262	0.7602	0.7428
SVM	Inst	0.8389	0.8278	0.8333
	Noise	0.9473	0.9375	0.9424
	Voice	0.8994	0.9200	0.9096

From Table 5, the comparison of classification models for audio signals demonstrates that SVM and KNN significantly give better results

than Decision Trees. For RNS classification, SVM achieves the highest average F1-score (0.8170), closely followed by KNN (0.8146), while the Decision Tree lags behind (0.7121). In MA classification, KNN marginally surpasses SVM (0.8970 vs. 0.8951) in overall performance, with both classifiers exhibiting strong precision and recall, particularly in distinguishing voice and instrumental sounds. Decision Tree has the lowest F1-scores in both RSA and MA signal classification. These findings show that SVM is the best overall classifier for audio classification, while KNN is preferable for music-related tasks. Decision Trees is not well-suited for complex audio classification problems.

4. Conclusion

This study compares two types of audio signals: random noise and music. The noise signal was recorded in a crowded public bus, while the music signal was recorded under controlled conditions. Both signals were analyzed using spectral and statistical methods, including histogram analysis, power spectrum, FFT, and ACF.

The noise signal exhibited a bell-shaped histogram with most samples clustered around the mean, while its ACF decayed with increasing lag. FFT analysis confirmed that noise energy is uniformly distributed across frequencies. In contrast, the music signal showed a sharper histogram due to distinct voice and instrumental components. Its ACF and FFT spectrum revealed rhythmic patterns and specific frequencies, distinguishing it from noise.

A Support Vector Machine (SVM) classifier was employed for further analysis. It classified noise into three levels with approximately 82% accuracy and categorized speech, music, and noise in the music signal with approximately 90% accuracy. The higher accuracy for music classification is attributed to the clear distinction between voice and instrumental elements. This study demonstrates that SVM is an effective classifier for audio signal analysis.

References

- [1] A. F. Siegel and M. R. Wagner, "Time series," in *Practical Business Statistics*, Elsevier, 2022, pp. 445–482. doi: 10.1016/b978-0-12-820025-4.00014-2.
- [2] M. Müller and A. Klapuri, "Music Signal Processing," in *Academic press library in signal processing*, Elsevier, 2014, pp. 713–756. doi: 10.1016/b978-0-12-396501-1.00027-3.
- [3] R. Awati, R. Sheldon and J. Burke, "What is signal-to-noise ratio and how is it measured?," *Search Networking, TechTarget*, Jun. 3, 2025. Available: <https://www.techtarget.com/searchnetworking/definition/signal-to-noise-ratio> [Accessed: Jun. 4, 2025]
- [4] M. Muller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 6, pp. 1088–1110, Oct. 2011, doi: 10.1109/jstsp.2011.2112333.
- [5] J. L. Rojo-Alvarez, C. Figuera-Pozuelo, C. E. Martinez-Cruz, G. Camps-Valls, F. Alonso-Atienza, and M. Martinez-Ramon, "Nonuniform Interpolation of Noisy Signals Using Support Vector Machines," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4116–4126, Aug. 2007, doi: 10.1109/tsp.2007.896029.
- [6] J. R. Reddy, K. K. P. S. Reddy, M. M. Ali, and U. Sesadri, "Music Genre Classification Using Machine Learning," *International Journal for Multidisciplinary Research*, vol. 6, no. 1, Jan. 2024, doi: 10.36948/ijfmr.2024.v06i01.8857.
- [7] C. Wang and X. Geng, "Simulation of music style classification model based on support vector machine algorithm," in *Sixth International Conference on Intelligent Computing, Communication, and Devices (ICCD 2023)*, S. Patnaik, Ed., Hong Kong, China: SPIE, Jun. 2023, p. 86. doi: 10.1117/12.2682982.
- [8] K. Kumar, "Audio classification using ML methods," *arXiv (Cornell University)*, May. 2023, doi: 10.48550/arxiv.2305.19304.
- [9] M. Xiang, "Music Vocal Spectrum Classification Model based on Improved SVM Algorithm," *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, pp. 1244–1247, Nov. 2022, doi: 10.1109/icaiss55157.2022.10010883.
- [10] S. Kavitha and J. Manikandan, "Improved Methodology of SVM to Classify Acoustic Signal by Spectral Centroid," *Journal of Trends in Computer Science and Smart Technology*, vol. 3, no. 4, pp. 294–304, May 2022, doi: 10.36548/jtcsst.2021.4.005.

- [11] L. Pallotta, M. Neri, M. Buongiorno, A. Neri, and G. Giunta, "A Machine Learning-Based Approach for Audio Signals Classification using Chebychev Moments and Mel-Coefficients," in *2022 7th International Conference on Frontiers of Signal Processing (ICFSP)*, Paris, France: IEEE, Sep. 2022, pp. 120–124. doi: 10.1109/ICFSP55781.2022.9924832.
- [12] K. Zaman, M. Sah, C. Direkoglu, and M. Unoki, "A survey of audio classification using deep learning," *IEEE Access*, vol. 11, pp. 106620–106649, Jan. 2023, doi: 10.1109/access.2023.3318015.
- [13] S. Akinpelu and S. Viriri, "Speech Emotion Classification: A Survey of the State-of-the-Art," in *Pan-African Artificial Intelligence and Smart Systems*, T. M. Ngatched Nkouatchah, I. Woungang, J.-R. Tapamo, and S. Viriri, Eds., Cham: Springer Nature Switzerland, 2023, pp. 379–394. doi: 10.1007/978-3-031-25271-6_24.
- [14] "R Studio IDE", *Posit Software, PBC*. Available: <https://posit.co/downloads/>