Analysis of Frequency-Dependent Features in Music Genre Differentiation

By Joshua Threlkeld

Abstract

Music, at its most banal definition, is organized sound. Yet, the boundaries between genres undoubtedly resist simple classification. Genre differentiation and genre prediction are challenging to describe qualitatively even for professional musicians. What happens when that intuition is formalized: when the acoustic properties of music are reduced to numbers and a machine is prompted to do what a listener does instinctively?

Thus, music genre classification serves as a benchmark task in machine learning. (Peeters et al., 2011) In this work, I aim to analyze the effectiveness of frequency-domain audio features, such as short-time Fourier transform (STFT), spectral centroid, spectral bandwidth, and Mel-frequency cepstral coefficients (MFCCs) to differentiate music genres. Using a publicly available Kaggle dataset (Hajoori, Kaggle, 2018), I investigate how these attributes encode genre structure and assess their limitations in separating genres. Critically, I aim to emphasize how these features interact with one another; how individual audio data points connect with one another to synthesize perceivable, unequivocal genre structure.

Introduction

Machine learning in audio classification relies on compressions of raw audio files into localized, frequency-contingent spectral representations. These have become standard audio descriptors due to their respective accuracy. (Peeters et al., 2011) In spite of this, this ubiquity fails to underscore why these descriptors work, as well as their limitations in certain contexts.

Rather than optimizing a model for predictive accuracy, I aim to explore how these feature representations encode genre classification. From a more "macro-level" viewpoint, these characteristics are useful in understanding broader auditory classification. More pertinently, these same feature extraction principles appear broadly across biomedical signal processing, where the stakes of misclassification extend to mental health diagnostics in voice signal data to seizure detection in EEG data.

Dataset

I used Hajoori's music dataset containing audio samples labeled by genre and data on STFT, spectral centroid, spectral bandwidth, and MFCCs. The dataset includes audio samples with varying degrees of frequency-domain audio feature overlap. As such, misidentification patterns could be evaluated in addition to feature separability.

The dataset contains 1,000 audio samples distributed evenly across ten genres with each clip standardized to 30 seconds. (Hajoori, Kaggle, 2018)

Relevant Frequency-Dependent Features

The following features were extracted to categorically differentiate the audio clips.

Short-Time Fourier Transform (STFT)

The STFT converts a segment of audio into a time-frequency representation by computing frequency content within brief, overlapping windows, revealing how spectral structure evolves over time.

Equation 1) $F\{x[n]w[n - m]\}$

Equation 1 takes the Fourier transform F{...} of a short segment w[n-m] of a full audio waveform x[n]. This yields the frequencies present in a specific window of time. When repeatedly applied across overlapping windows, a full time-frequency representation is produced.

The STFT produces a complex-valued matrix of dimensions $time\ steps \times frequency\ bins$, where each entry represents the spectral energy at a given time and frequency. Magnitude spectrograms were derived from this matrix to capture temporal development. While the STFT produces a 2D time-frequency matrix, a single summary statistic — mean magnitude across the spectrogram — was extracted per sample for use as a model feature. (Durak & Arikan, 2003)

Spectral Centroid

Spectral centroid was extracted as a proxy for perceptual brightness. (Caetano et al., 2019) It reflects the weighted mean frequency of the spectrum over time.

Spectral Bandwidth

Spectral bandwidth describes the frequency spread surrounding the spectral centroid. In music terminology, spectral bandwidth encapsulates the variation in timbre across different sounds, or, in this study, genres. (Caetano et al., 2019)

Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs are the most standard metric for classifying music genres. More generally, MFCCs have very high predictive accuracy for assigning audio to different categories. (Tzanetakis & Cook, 2002) Essentially, MFCCs utilize the same time-frequency concept as STFTs, which is that audio is broken into short periods to find the Fourier Transform. MFCCs differ, though, in that this Fourier Transform is passed through the Mel scale[1].

Modeling Approach

The aim of this approach is to evaluate feature effectiveness without compromising interpretability. As such, classifiers like logistic regression, random forest, and support vector machines were applied to determine individual, as well as combined, contributions to genre separability. Given the dataset's balanced class distribution, no resampling was required.

Results

[1] The human ear is more sensitive to low frequencies than high frequencies. To accommodate for this, the Mel scale places more emphasis on the detail of low frequencies, while simultaneously compressing high frequencies, thus creating a particularly *human* representation of audio.

STFT

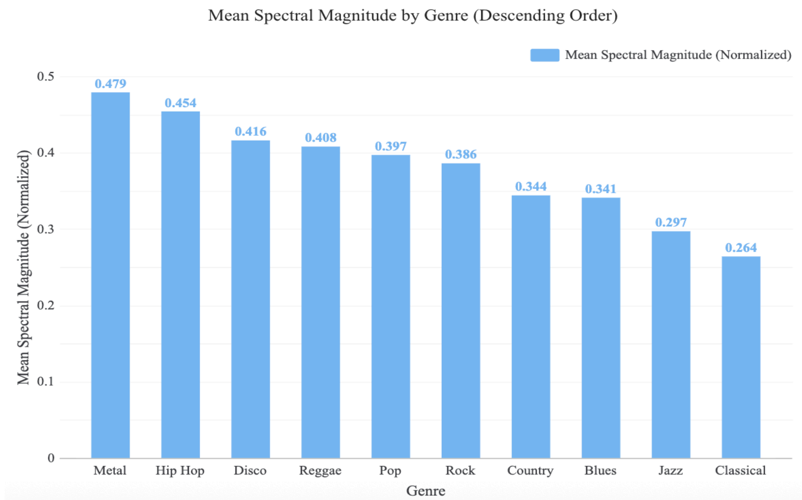


Fig. 1 Average STFT magnitude per genre. For each track, the STFT magnitude was averaged across time and frequency to obtain a single scalar value, then averaged across all tracks within each genre. Genres are ordered by decreasing mean magnitude, providing a visual representation for signal energy.

Genre	**Mean STFT**	**STDEV for STFT**
Metal	**0.479**	0.0518
Hip Hop	0.454	0.0471
Disco	0.416	0.0471
Reggae	0.408	**0.0592**
Pop	0.397	0.0569
Rock	0.386	0.0447
Country	0.344	0.041
Blues	0.341	0.0584
Jazz	0.297	0.0544
Classical	0.264	0.0441

Tab. 1 Mean and standard deviation of per-track STFT magnitude by genre.

Figure 1 captures the spectral density between genres. A higher mean STFT corresponds to greater noise frequency (i.e., less silence in a time slice). The endpoints — metal/hip-hop and jazz/classical — separate themselves. This data alone could be indicative of why metal/hip-hop feel more high-energy, whereas jazz/classical feel more low-energy. Still, within these low-energy genres, it should be noted that there were still outliers, represented by jazz and blues having comparatively high standard deviations. Therefore, on its own, STFT is most aptly viewed as an indicator of a music genre as opposed to a qualifying factor.

To this point, disco and reggae share similar STFTs, while blues and jazz have noticeably different STFTs. This is due to the fact that when classifying certain genres, such as blues and reggae, there is a greater emphasis placed on *qualification* than *quantification*. For example, blues songs are based upon the blues scale (a 12-bar chord progression), typically having particularly "soulful" emotional expression. These immensely open-ended qualifications challenge our ability to classify using one or two data-derived features. As a result, the standard deviation of blues music is higher than the other genres, calling into question STFT's effectiveness for blues: a trend that persists into other classifiers.

Spectral Centroid and Bandwidth

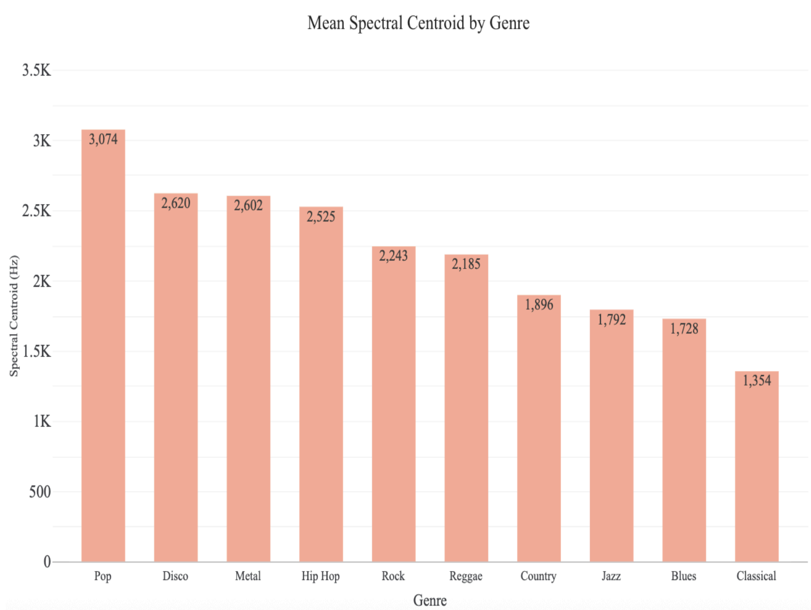


Fig. 2 Mean spectral centroid by genre. For each track, spectral centroid was averaged across time and then aggregated across tracks within each genre. Genres are ordered by decreasing centroid.

Genre	Mean Spectral Centroid (Hz)	STDEV of Spectral Centroid (Hz)
Pop	**3073.664**	582.062
Disco	2619.974	478.742
Metal	2602.175	368.584
Hip Hop	2524.614	479.309
Rock	2242.657	483.838
Reggae	2185.111	626.599
Country	1896.096	575.795
Jazz	1792.404	**680.607**
Blues	1727.655	515.546
Classical	1353.991	348.305

Tab. 2 Mean and standard deviation of per-track spectral centroid magnitude by genre.

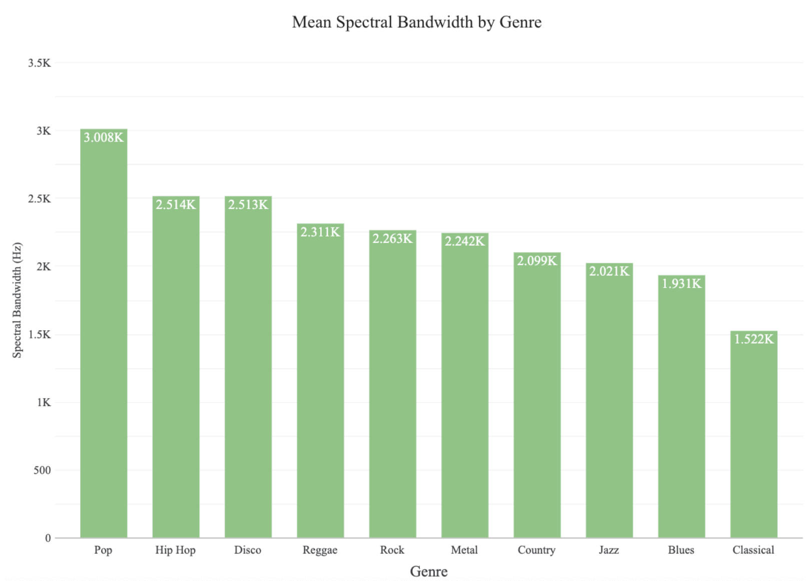


Fig. 3 Mean spectral bandwidth by genre. For each track, spectral bandwidth was averaged across time and then aggregated across tracks within each genre. Genres are ordered by decreasing bandwidth.

Genre	Mean Spectral Bandwidth (Hz)	STDEV Spectral Bandwidth (Hz)
Pop	**3008.243**	316.889
Hip Hop	2513.524	347.182
Disco	2513.371	357.455
Reggae	2311.494	390.729
Rock	2262.863	347.871
Metal	2242.496	201.35
Country	2099.471	445.243
Jazz	2020.99	**567.448**
Blues	1931.442	330.246
Classical	1521.704	254.299

Tab. 3 Mean and standard deviation of per-track spectral bandwidth magnitude by genre.

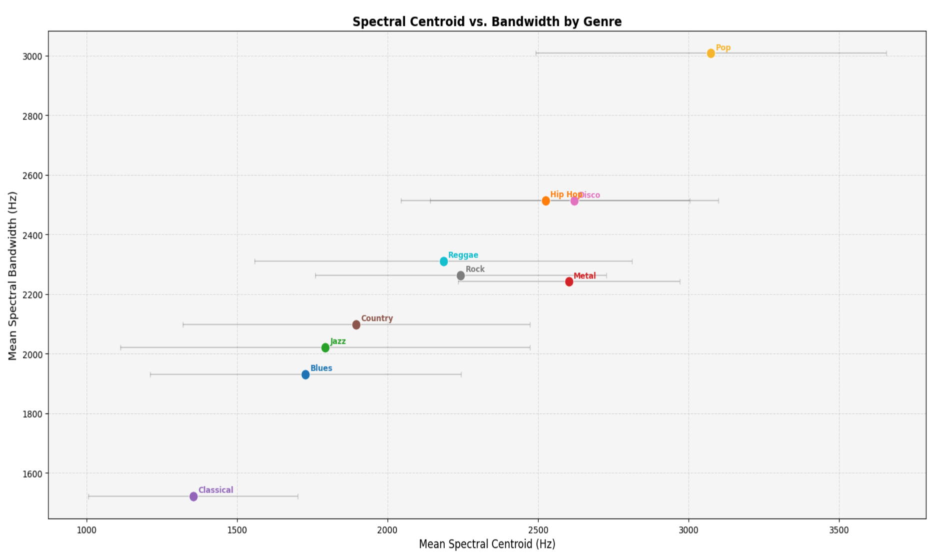


Fig. 4 Mean spectral centroid versus mean spectral bandwidth by genre. Each point represents a genre and error bars indicate standard deviation. Genres with higher spectral centroid tend to exhibit greater bandwidth, reflecting a broader distribution of higher-frequency content.

In Figure 4, pop and classical are clear outliers. Pop's high bandwidth and spectral centroid indicate that it is very bright ("sharp" or "intense"), yet spectrally diverse. Intuitively, these metrics lead one to assume that pop will have a high F1 score, as it is clearly separable in regards to spectral centroid and bandwidth. Conversely, classical's low bandwidth and spectral centroid indicate that it is very dark ("full-sounding") and spectrally homogenous. This indicates that classical will have a similarly high F1 score to pop. Clustered in the center are blues, jazz, country, metal, rock, reggae, hip hop, and disco. These genres' error bars have significant overlap, which makes it nearly impossible to draw any meaningful conclusion(s). This overlap is not entirely surprising. While spectral centroid and bandwidth capture timbral characteristics — which many of these genres share similarities in — they differ significantly in rhythm and structure. A blues guitar and a country guitar playing the same note at the same volume may sound different in tone, but that timbral distinction alone does not make the genres separable. Rather, genre identity is (highly, if not at least somewhat) constituted by rhythm and structure, and features that cannot encode time cannot fully encode genre.

MFCC

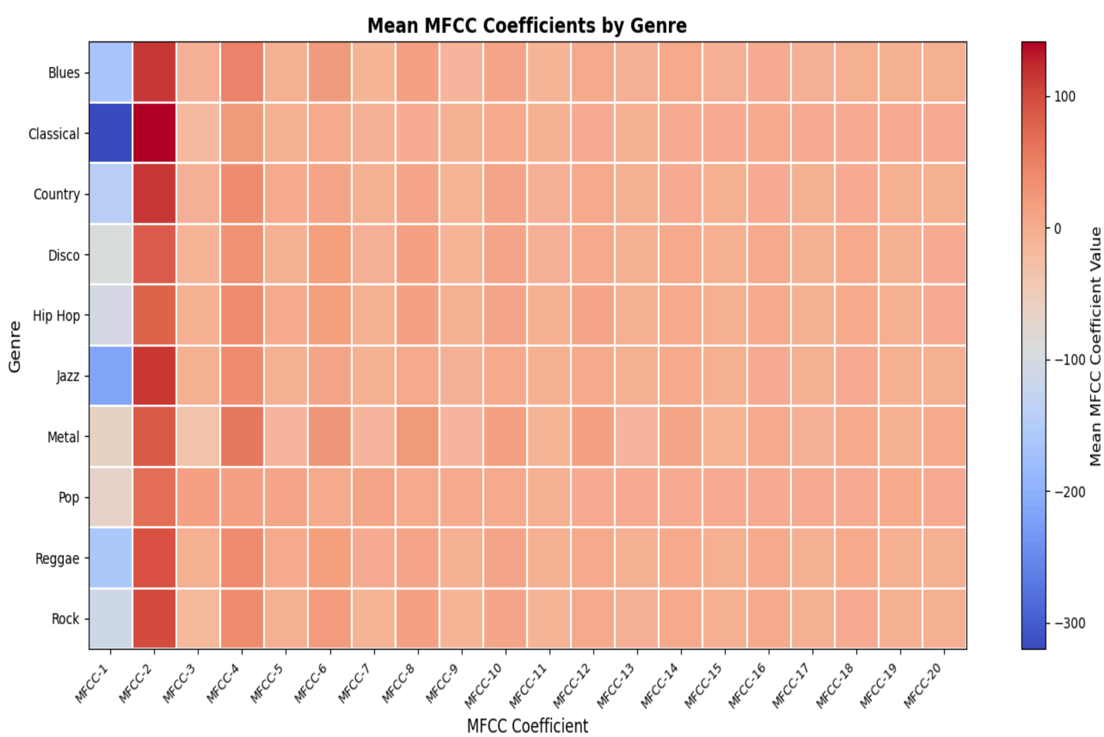


Fig. 5 Mean MFCC coefficients per genre, computed by averaging across time within tracks and then across tracks within each genre. Genres with more distinct coefficient profiles (e.g., classical, metal) tend to achieve higher F1 scores in downstream classification, suggesting that separable MFCC patterns contribute to model performance.

In Fig. 5, MFCC-1 and MFCC-2 stand out the most visually. MFCC-1 measures the loudness of the sample, regardless of the distribution (dynamic contrast). Jazz and classical stand out as the most negative, which makes sense given their respective dynamic ranges; both go through periods of intense volume juxtaposed with periods of quietness. It should be noted that MFCC-1 and STFT independently analyze similar qualities, and that they are converging on a similar conclusion is encouraging. MFCC-2 encodes the distribution of low-frequency and high-frequency balance across the audio sample. (Müller, 2015) That is to say, a positive MFCC-2 indicates high-frequency dominance, whereas a negative MFCC-2 indicates low-frequency dominance. In practice, instruments like hi-hat cymbals or electric guitars have very high frequencies and are staples of rock. Conversely, woodwind instruments have very low frequencies and are staples of jazz and classical.

Models

Utilizing a multinomial logistic regression model achieved an accuracy of 67.5% and a macro-averaged F1 score of 0.67 across the ten genres. This model's accuracy is substantially greater than the projected 10% baseline, indicating that frequency-domain features encode meaningful information for classifying genres. Nevertheless, the logistic regression model's accuracy varied significantly by genre. Classical (F1 score of 0.93) and metal (F1 score of 0.86) were classified very accurately, suggesting that these spectra, on their own, encode distinctiveness. Conversely, reggae (F1 score of 0.50), disco (F1 score of 0.56), and rock (F1 score of 0.56) displayed noticeable ambivalence. Reggae was most frequently misclassified as hip-hop and disco, rock as disco, and country as rock: patterns that reflect shared rhythmic foundations rather than spectral confusion. These results suggest that while frequency-domain features capture broad energetic and timbral distinctions, they are insufficient for cleanly separating genres with similar orchestration or rhythmic structures.

Opting for a random forest classifier achieved an accuracy of 64.5% with a macro-averaged F1 score of 0.64 across the ten genres. Similarly to the logistic regression model, performance varied notably by genre. Classical (F1 score of 0.90), metal (F1 score of 0.79), and blues (F1 score of 0.78) proved to be the most reliably classified genres. Blues' comparatively strong performance — producing only six misqualifications — despite its qualitatively open-ended structure stands apart from the other mid-tier genres and invites closer examination in the discussion. In contrast, rock (F1 score of 0.38), disco (F1 score of 0.55), and hip-hop (F1 score of 0.56) exhibited lower performance and higher rates of misclassification. The low F1 scores for rock and hip-hop could be attributed to their decade-spanning relevance. That is to say, both genres have been popular for long enough for legitimate sub-genres (such as indie, alternative, soft rock) to develop, making the defining criteria more fluid. Rock was most frequently misclassified as disco, pop, and country. In coordination with its F1 of 0.38, this reflects how fluidly rock's spectral profile bleeds into neighboring genres. Overall, despite the model's ability to capture nonlinear feature interactions, the aggregate performance remained moderate, suggesting that the selected frequency-domain features provide only partial separability across all genre categories.

The support vector machine achieved an accuracy of 67.5% and a macro-averaged F1 score of 0.68, matching the logistic regression model in accuracy but showing slightly improved balance across genres. While classical and metal remained highly separable, rock was most frequently misclassified as disco, and reggae as hip-hop and disco. The fact that this pattern is consistent across all three models points to shared rhythmic foundations rather than spectral overlap. Most tellingly, this confusion matrix showed spectral inseparability that no increase in model complexity resolved.

Across all three models, genre-level F1 scores reveal a consistent pattern of high performance for spectrally distinct genres and low performance for rhythmically defined ones.

Genre	Logistic Regression F1 score	Random Forest F1 score	SVM F1 score
Blues	0.69	0.78	0.74
Classical	0.93	0.90	0.87
Country	0.58	0.59	0.70
Disco	0.56	0.55	0.52
Hip hop	0.65	0.56	0.70
Jazz	0.67	0.63	0.67
Metal	0.86	0.79	0.90
Pop	0.74	0.64	0.78
Reggae	0.50	0.60	0.47
Rock	0.56	0.38	0.41

Tab. 4 Per-genre F1 scores across logistic regression, random forest, and support vector machine models.

The following confusion matrices identify the specific misclassification patterns underlying these F1 scores.

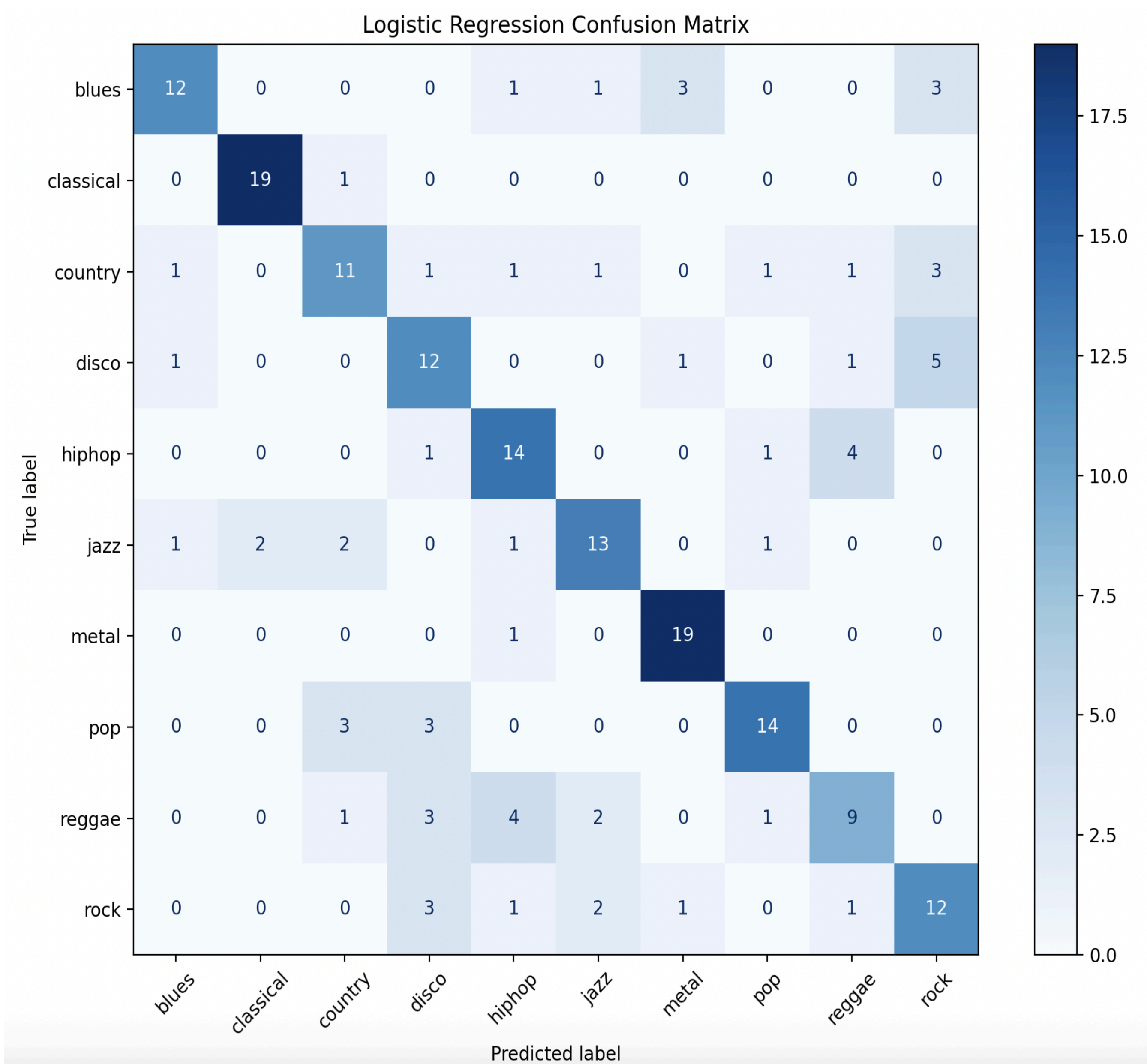


Fig. 6 Logistic regression confusion matrix across ten genres (n=200 test samples). Diagonal values represent correct classifications; off-diagonal values represent misclassifications. Classical and metal show the strongest diagonal concentration, while rock and reggae show the most dispersed off-diagonal patterns.

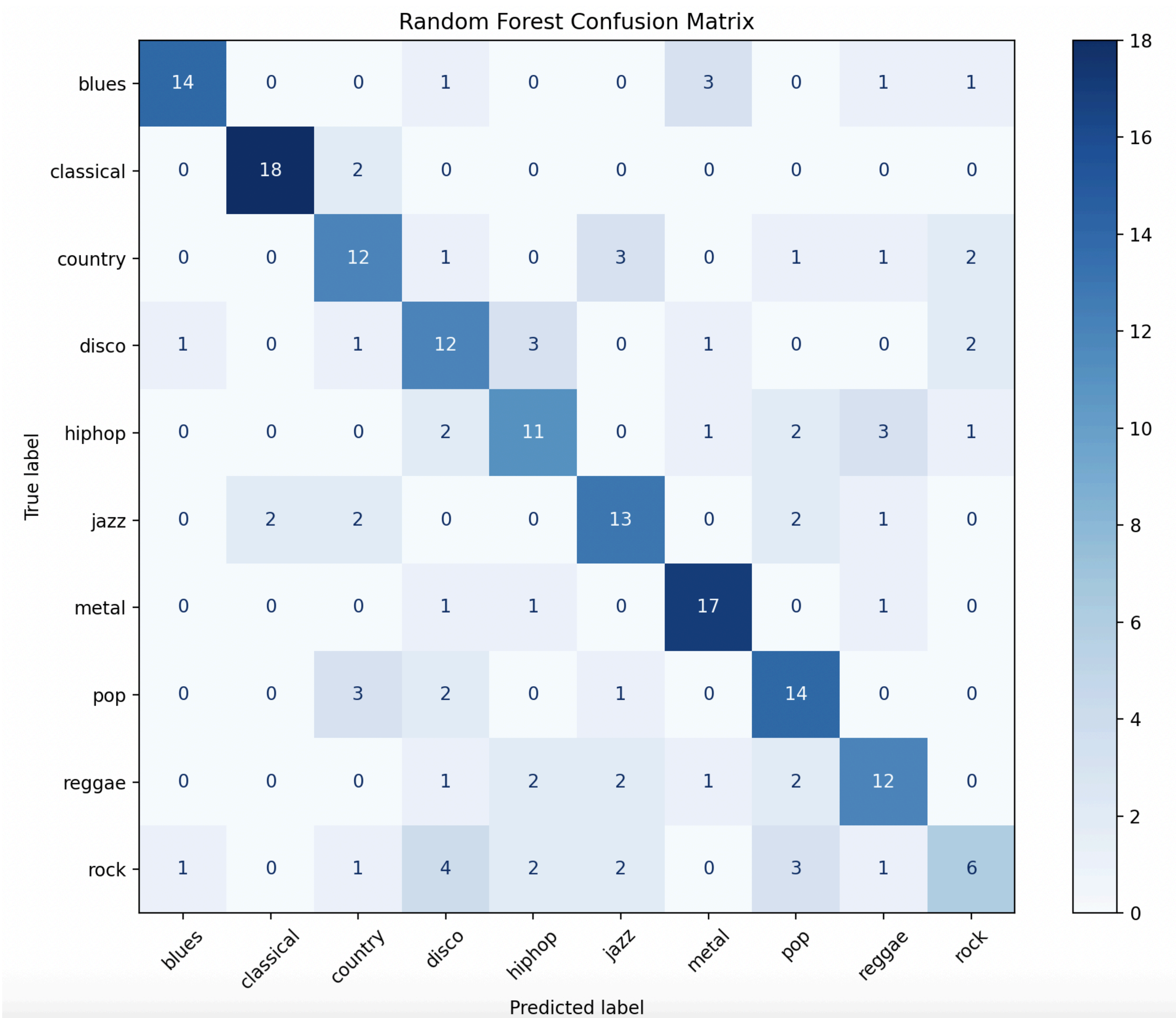


Fig. 7 Random forest confusion matrix across ten genres (n=200 test samples). Blues achieves its strongest performance across all three models here (14/20), consistent with the combinatorial timbral signature argument developed in the discussion. Rock remains the most dispersed genre with misclassifications spread across disco, hip-hop, jazz, and pop.

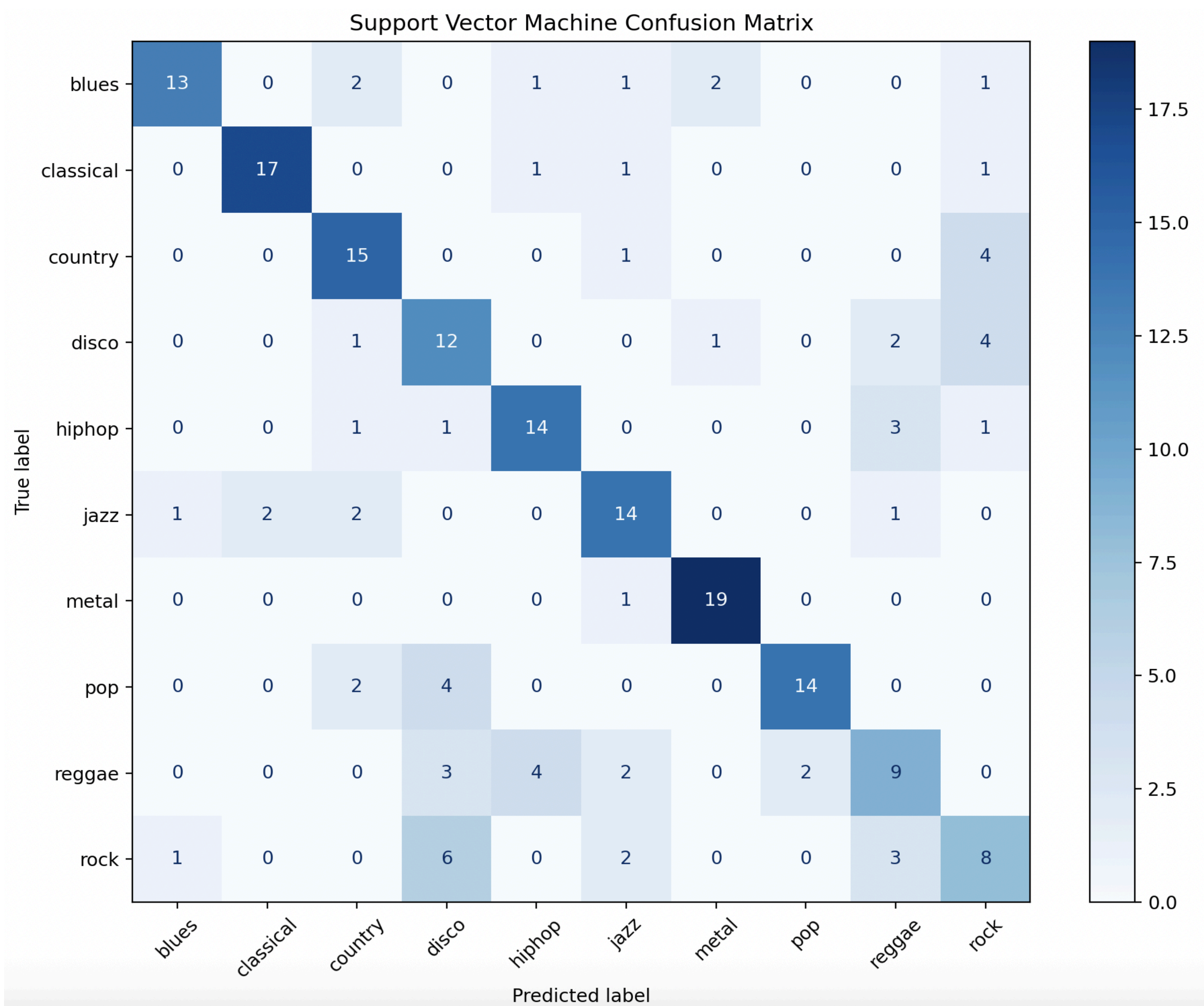
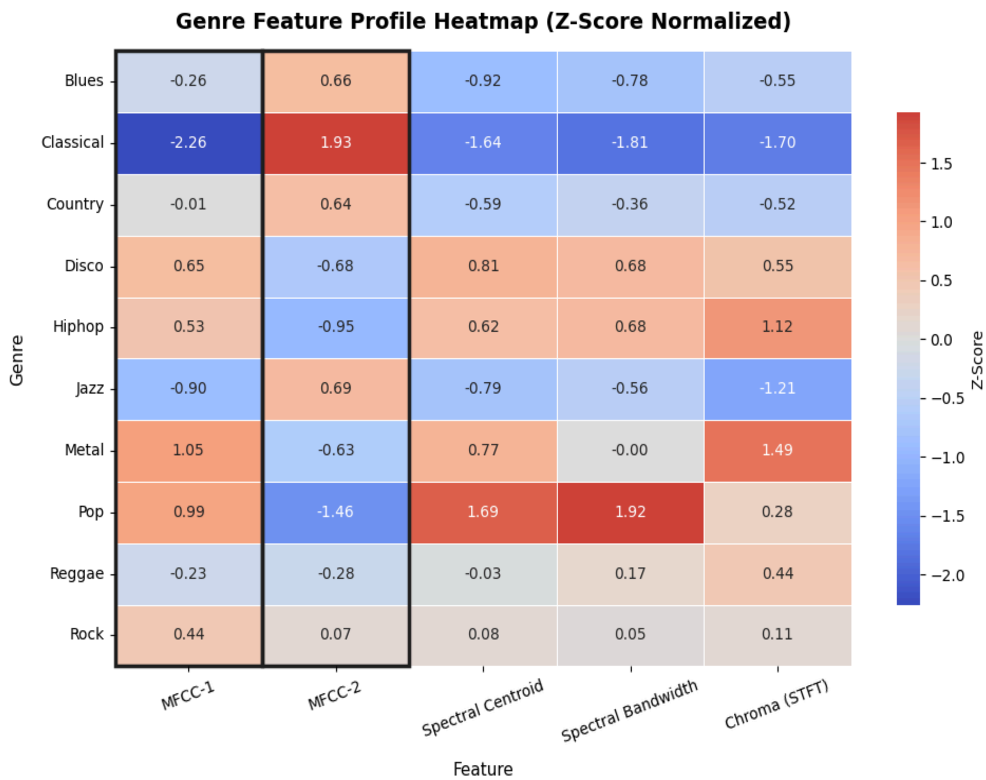


Fig. 8 Support vector machine confusion matrix across ten genres (n=200 test samples). Despite the SVM's greater representational capacity relative to logistic regression, accuracy and misclassification patterns are nearly identical.

Discussion

While each feature isolates one dimension of genre structure, viewing them in coordination reveals which genres truly share meaningful overlap.



Classical and metal occupy opposite ends of the feature space in the most diagnostically useful way. Classical is defined by a strongly negative MFCC-1 and strongly positive MFCC-2, indicating low overall energy with high-frequency dominance. Conversely, metal shows an elevated MFCC-1 with negative MFCC-2, reflecting high energy concentrated in low frequencies. These opposing profiles explain their consistently high F1 scores across all three models. The rhythmically-defined genres show no such distinctive cross-feature signature, clustering near zero across most features, which is precisely why frequency-domain representations fail to separate them.

Frequency-domain features encode timbral structure and variation. Be that as it may, they are blind to rhythmic structure and complexity, and this fundamental limitation predicts exactly which genres classify well and which don't. The convergence of all three models in the 65–67.5% accuracy range is not incidental. Rather, it reflects a ceiling imposed by the features themselves, not by model complexity. Most directly, the logistic regression and SVM achieve identical accuracy despite the SVM's substantially greater representational capacity. If the features encoded sufficient discriminatory information, the SVM and random forest models would have surfaced nonlinear patterns the logistic regression could not. That it did not point to one conclusion: the bottleneck is feature expressiveness, not model sophistication.

Overall, classical and metal are consistently high F1 scores, while reggae, disco, and rock are consistently low. Classical is especially distinctive, as it is characterized by wide dynamic range and low-frequency energy (consistent with its instrumentation). Metal is characterized by heavy distortion and high energy across compressed dynamic range. Blues is a notable exception in the random forest, achieving an F1 of 0.78 — its strongest performance across all three models. While blues is culturally defined by its 12-bar harmonic structure and expressive phrasing rather than a fixed instrumentation, these harmonic characteristics produce a consistently low-frequency spectral profile that the random forest surfaces through nonlinear feature interactions. The features are encoding what blues *sounds like* as an acoustic byproduct of its harmonic structure. Reggae, disco, and rock, however, have more fluid instrumentation and are encoded by rhythmic structure, making these consistently difficult to classify.

This pattern extends most directly into mental health diagnostics through voice analysis. Conditions such as depression and schizophrenia produce measurable changes in vocal acoustics, such as reduced spectral variability, flattened prosody, and shifts in formant structures (resonance frequency peaks of the vocal tract). Taguchi et al. (2018) provide a concrete illustration of these phenomena: among twelve MFCC features extracted from patient speech, only MFCC-2 significantly distinguished depressed patients from controls. Depressed speakers exhibited reduced energy in the 2000–3000 Hz range, with a large effect size (Cohen's $d = -1.72$), and MFCC-2 alone achieved 81.9% classification accuracy. This finding highlights both the diagnostic value of individual spectral coefficients and the specificity of where a meaningful signal resides in the frequency domain. Despite this clear correlation, frequency-domain features are only ancillary to mental health diagnostics through voice analysis. (Menne et al., 2024) The time variance of speech, including pause duration, rhythm irregularity, and the dynamic unfolding of pitch over time, carries significant diagnostic weight that static spectral summaries fail to encode. (Rapcan et al., 2010) Much like reggae or disco resisting classification due to their rhythmic and fluid structure, mental health conditions defined by altered speech dynamics are precisely the cases where frequency-domain features lose their discriminatory power.

Furthermore, in clinical contexts with spectrally stable diagnostic signals, frequency-domain features can achieve high discriminatory power. For example, conditions such as atrial flutter, characterized by its sawtooth waveform, or epilepsy, marked by spike-and-wave discharges in EEG, exhibit consistent frequency patterns that are well captured by Fourier-based representations (Zhang et al., 2018). These cases are analogous to “classical” and “metal” in music classification, where defining structures are encoded clearly in the spectral domain.
In practice, however, raw EEG signals are heavily contaminated by noise from muscle activity, electrical interference, and cardiac signals, making direct interpretation difficult. As a result, preprocessing steps such as denoising are essential to isolate meaningful neural activity before analysis. STFT-dependent techniques such as thresholding, spectral subtraction, and independent component analysis are commonly used to separate signal from noise, enabling more reliable downstream feature extraction. The limitations of frequency-domain approaches emerge when the diagnostic signal shifts from stable spectral structure to temporal complexity. Early neurodegenerative changes, for instance, do not present as distinct frequency signatures, but rather as disruptions in the temporal organization and complexity of brain activity. These patterns are not well captured by mean-based spectral features (Papaliagkas, 2025). This reinforces a broader principle: frequency-domain features perform best when the underlying signal is spectrally stable, and degrade in effectiveness when classification depends on temporal dynamics.

Limitations

This dataset struggles in several areas. For one, STFT is condensed into one value (the mean). Typically, multiple points of STFT throughout an audio sample could be used to show the change in spectral energy across time. Using an average STFT is more aptly viewed as a proxy for overall spectral energy. Therefore, two audio samples with vastly different spectral shapes could have identical mean STFTs and, for the purposes of this paper, be classified as identical. This constraint is not incidental — the temporal-spectral resolution trade-off inherent to STFT has motivated active methodological development, including modified architectures designed specifically to minimize information loss at the feature extraction stage. (Ali et al., 2022)

Additionally, this dataset lacks a feature that differentiates between rhythms. While the dataset initially included tempo data, this feature was intentionally excluded from the present analysis to maintain focus on frequency-domain representations and avoid conflating distinct feature classes. The genres that proved most difficult to classify, mainly reggae, disco, and rock, are precisely those whose identity is most strongly rooted in rhythmic structure. Even so, tempo is still limited in describing rhythm, as it is more accurately a time/rate descriptor than polyrhythm. I propose that a feature measuring rhythmic entropy rather than speed/structure may provide the separability that spectral features alone cannot, and their inclusion represents a natural direction for future work.

Most importantly, the initial assignment of audio samples to genres relies on the human ear and inherently fuzzy definitions of music genres. In other words, the creator of this dataset assigned genres to these audio samples. For the purposes of this analysis, these assignments were accepted as true. However, it is feasible some of these audio samples overlapped between multiple genres and could have been better classified. This inherent human bias adds legitimate skepticism to these results, particularly genres with middle of the pack F1 scores.

Conclusion

Frequency-domain features provide a meaningful but incomplete basis for music genre classification. Across all three models, performance converged in the 65–67.5% accuracy range, a result that is not incidental. The fact that a linear model and a kernel-based model with substantially greater representational capacity achieved identical accuracy suggests that the bottleneck is not model intricacy, but feature expressiveness. These features simply do not encode what separates certain genres from one another.

The genre-level F1 scores make this concrete. Classical and metal, whose identities are anchored in stable timbral structure classified reliably and consistently. Reggae, disco, and rock, whose identities are rooted in rhythmic organization and fluid instrumentation, resisted classification across every model tested. The features knew what the music sounded like; they did not know how it moved.

Blues, whose strong random forest performance traces to the acoustic byproduct of its harmonic structure, is the clearest illustration of this principle: these features are reliable when the classificatory signal lives in sound, and unreliable when it lives in structure.

This limitation is not unique to music. As explored in the discussion, the same timbral-versus-temporal divide surfaces in biomedical signal processing: frequency-domain features reliably identify epilepsy subtypes where the diagnostic signal is spectrally stable, yet struggle with mental health conditions where it is embedded in time. The pattern identified in this music classification task thus reflects that frequency-domain features are reliable classifiers when structure is timbral, and insufficient ones when structure is rhythmic or dynamic.

Future work incorporating temporal-sensitive features such as rhythmic entropy, onset detection rates, or beat histograms would directly address the separability gap observed here. More broadly, this analysis suggests that the value of feature selection lies not only in maximizing predictive accuracy, but in understanding the structural properties a given feature can and cannot encode. In both music and neuroscience, that understanding is a prerequisite for building classifiers that are not just accurate, but interpretable and trustworthy.

Citations

Ali, O., Saif-ur-Rehman, M., Dyck, S., Glasmachers, T., Iossifidis, I., & Klaes, C. (2022). Enhancing the decoding accuracy of EEG signals by the introduction of anchored-STFT and adversarial data augmentation method. *Scientific Reports*, *12*(1), Article 4245. https://doi.org/10.1038/s41598-022-07992-w

Caetano, M., Saitis, C., & Siedenburg, K. (2019). Audio content descriptors of timbre. In K. Siedenburg, C. Saitis, S. McAdams, A. N. Popper, & R. R. Fay (Eds.), *Timbre: Acoustics, perception, and cognition* (pp. 297–333). Springer. https://doi.org/10.1007/978-3-030-14832-4_11

Durak, L., & Arıkan, O. (2003). Short-time Fourier transform: Two fundamental properties and an optimal implementation. *IEEE Transactions on Signal Processing, 51*(5), 1231–1242. https://doi.org/10.1109/TSP.2003.810293

Hajoori. Music features. Kaggle, 2018. Available at: https://www.kaggle.com/datasets/insiyeah/musicfeatures. Accessed February 23, 2026.

Menne, F., Dörr, F., Schräder, J., Tröger, J., Habel, U., König, A., & Wagels, L. (2024). The voice of depression: Speech features as biomarkers for major depressive disorder. *BMC Psychiatry, 24*, 794. https://doi.org/10.1186/s12888-024-06253-6

Müller, M. (2015). *Fundamentals of music processing: Audio, analysis, algorithms, applications.* Springer. https://doi.org/10.1007/978-3-319-21945-5

Papaliagkas, V. (2025). The role of quantitative EEG in the diagnosis of Alzheimer's disease. *Diagnostics, 15*(15), 1965. https://doi.org/10.3390/diagnostics15151965

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N., & McAdams, S. (2011). The Timbre Toolbox: Extracting audio descriptors from musical signals. *Journal of the Acoustical Society of America, 130*(5), 2902–2916. https://doi.org/10.1121/1.3642604

Rapcan, V., D'Arcy, S., Yeap, S., Afzal, N., Möller, H. J., & Reilly, R. B. (2010). Acoustic and temporal analysis of speech: A potential biomarker for schizophrenia. *Medical Engineering & Physics, 32*(9), 1074–1079. https://doi.org/10.1016/j.medengphy.2010.07.013

Taguchi, T., Tachikawa, H., Nemoto, K., Suzuki, M., Nagano, T., Tachibana, R., Nishimura, M., & Arai, T. (2018). Major depressive disorder discrimination using vocal acoustic features. *Journal of Affective Disorders*, *225*, 214–220. https://doi.org/10.1016/j.jad.2017.08.038

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing, 10*(5), 293–302. https://doi.org/10.1109/TSA.2002.800560

Zhang, M., Li, W., Du, Y., Wan, L., & Zhang, W. (2018). Epileptic seizure detection based on EEG signals and CNN. *Frontiers in Neuroinformatics, 12*, 95. https://doi.org/10.3389/fninf.2018.00095