

OPEN SET RECOGNITION FOR MUSIC GENRE CLASSIFICATION

Zhanlin Liu
Warner Music Group
Kevin.Liu@wmg.com

Julien De Mori
Warner Music Group
julien.demori@wmg.com

Kobi Abayomi
Seton Hall University
kobi.abayomi@shu.edu

ABSTRACT

Music genres, prototypically, are poorly defined. Recent work suggests a path - via open set recognition - to identify unknown genres solely on features extracted from audio. We explore segmentations of known and unknown genre classes using the open source GTZAN and FMA datasets; for each, we begin with ‘best-case’ closed set genre classification, then we apply open set recognition methods. This allows us, on audio samples from unknown genres, to establish a baseline capacity for novel genre detection. This design, as well, can illustrate interaction between genre labeling and open set recognition accuracy.

1. INTRODUCTION

Aural boundaries between music genres, *e.g. labeled sets* – Classical, Jazz, Hip-hop, Rock, alternative, Experimental, etc – are imprecise and poorly defined. The antecedents for the ill defined labels differ: Classical music is characterized by a European origin [1]; so-called Pop is a function of a time-indexed neighborhood of popularity - local audience affinities. Experimental music, in contrast, can be defined *against* this context. This paper explores these boundaries and their time-dependent definitions via a comprehension of the labeled sets of genre as functions of measurable aural features implicitly capturing instrumentation, harmony, melody, chord progressions, vocal qualities, etc.

The GTZAN [2] and Free Music Archive (FMA) [3] datasets have been used to benchmark the genre identification task via aural features. Genre classification performance on this data also illustrates the potential utility deep learning models provide on this classification problem [4–7], commonly via supervised learning on known, labeled sounds.

A traditional supervised genre model, thus, can only classify music as of a known type - i.e. as a genre already established and observable in the labeled training data. This precludes classification as a novel type, *e.g.* predicting a particular sound or sound collection comprising an unknown class. A classifier’s ability to identify sounds in exclusion from *every* class suggests its potential to detect a new, novel genre. Colloquially, sounds from this

novel class should be unrecognizable from any extant labeled class.

Open set recognition (OSR) has been demonstrated to be useful for detecting unknown classification categories in various domains of application, including image classification [8–10], acoustic event classification [11, 12], and text classification [13, 14]. Machine learning algorithms such as Support Vector Machines [12, 15, 16], Nearest Neighbour classifiers [17] and Sparse Representation [18] can be modified to address this OSR problem. Performance of these traditional machine learning models is unsatisfactory in big data settings: more specifically when the number of observations in each class is large. Recent advances in deep learning based solutions for OSR using SoftMax with threshold and OpenMax [19], though, suggest a more scalable approach to novel genre identification.

This paper offers a method for the novel genre identification problem that both

- classifies audio into existing music genres (those available in the labeled training data), and
- classifies audio into new music genres (those *not* present in the labeled training data).

We also conduct an experimental analysis and exploration of the various factors of this methodology that impact open set genre classification accuracy.

The remainder of the paper is organized as follows. Section 2 introduces the OSR automatic system for music genre classification and presents how the closed set classifier and open set classifier are constructed. Section 3 illustrates the experimental setup as well as the results on both the GTZAN dataset and the FMA dataset. Section 4 concludes the paper with a discussion on the results and future research directions.

2. METHODOLOGY

Based on the definition in [20], in a classification task there are *known known classes* (KCCs), *known unknown classes* (KUCs), *unknown known classes* (UKCs), and *unknown unknown classes* (UUCs). In this paper, we formulate the problem of open set recognition for music genre classification with only KCCs and UUCs, where the KCCs represent the closed set genre labels that are observed in training the deep classification model and UUCs represent the open set genre labels that do not appear in the training process. Colloquially, the closed set model knows nothing of the

arXiv:2209.07548v2 [eess.AS] 14 May 2024



© Z. Liu, J. De Mori, and K. Abayomi. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Attribution: Z. Liu, J. De Mori, and K. Abayomi, “Open Set Recognition for Music Genre Classification”, in *Proc. of the 23rd Int. Society for Music Information Retrieval Conf.*, Bengaluru, India, 2022.

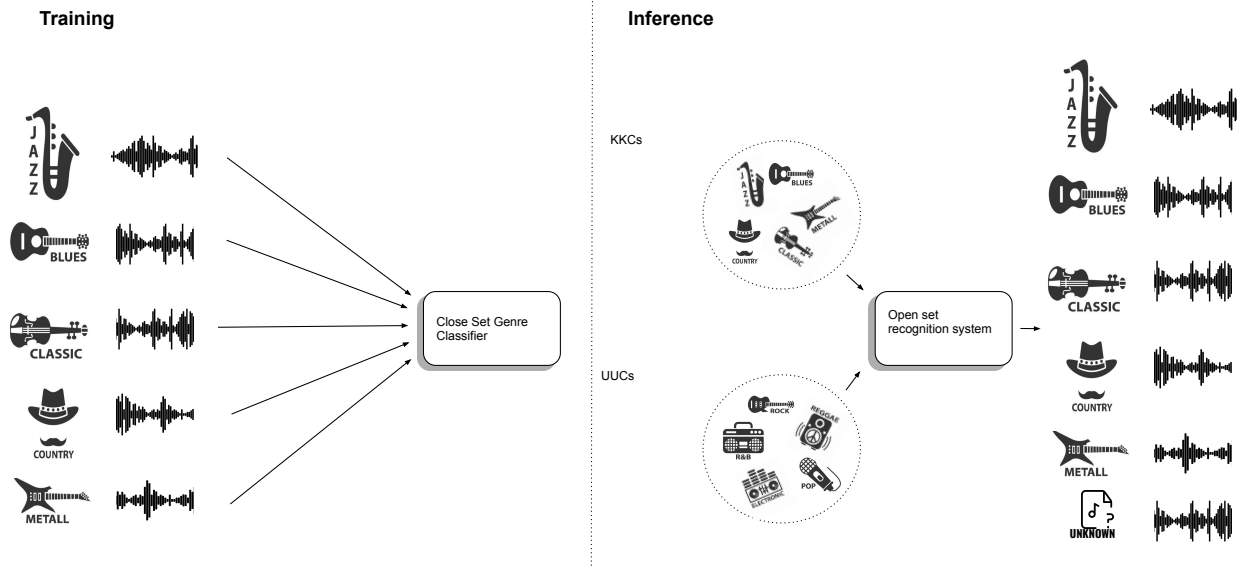


Figure 1: Training and inference processes for the OSR pipeline for genre classification. In the training stage (left), a closed set classifier is trained on the KCCs. In the inference stage (right), an OSR system classifies samples into the KCCs and an additional class for cases in which the sample is deemed unlike any of the KCCs

open set data (UUCs) that it runs inference on after training. Figure 1 shows the framework of the open set recognition pipeline including training and inference steps. As illustrated, a closed set genre classifier is trained only on samples from the KCCs. An OSR algorithm is trained and utilized in the inference process to classify music audio as one of the KCCs or an unknown class. In this section, we will illustrate the closed set genre classifier and the open set recognition model used in this paper.

2.1 Closed Set Genre Classifier

Kong et al. proposed a set of architectures in 2020 [5] that utilize transfer learning and fine-tuning of model weights trained on the AudioSet dataset [21]. CNN14, which is one of the proposed model architectures, has achieved state-of-the-art performance in multiple audio classification tasks including genre classification on the GTZAN dataset. Therefore, we propose using a CNN14 architecture and transfer learning as the closed set genre classifier. The model architecture using transfer learning based on the CNN14 architecture is illustrated in Table 1.

2.2 Open Set Classifier

SoftMax with threshold and OpenMax are two methods commonly used to address the open set recognition problem. In this section, we describe the SoftMax with threshold method and OpenMax method in detail.

The softmax function is often used as the activation function in the output layer for DNNs in multi-class classification tasks. After applying the softmax function, each

Log-mel spectrogram
$(3 \times 3 @ 64, BN, ReLU) \times 2$
Pooling 2×2
$(3 \times 3 @ 128, BN, ReLU) \times 2$
Pooling 2×2
$(3 \times 3 @ 256, BN, ReLU) \times 2$
Pooling 2×2
$(3 \times 3 @ 512, BN, ReLU) \times 2$
Pooling 2×2
$(3 \times 3 @ 1024, BN, ReLU) \times 2$
Pooling 2×2
$(3 \times 3 @ 2048, BN, ReLU) \times 2$
Global pooling
FC 2048 ReLU
FC N ReLU

Table 1: Transfer learning in the CNN14 architecture. BN represents the batch normalization, FC represents a fully connected layer, and N represents the number of KCCs. The model input takes a batch of 30-second audio waveforms.

node of the output layer can be regarded as the probability of belonging to the corresponding class. It follows that a test sample is from an unknown class if the softmax function produces low probabilities for all known classes. By thresholding the output probabilities on the output layer, a test sample can be identified as an unknown class.

Alternatively, OpenMax, which was proposed in [19] by Bendale et al, is regarded as pioneering work in deep open-set classifiers. It uses the extreme value theorem (EVT)

[22] and defines a per-class compact abating probability (CAP) model to reject unknown inputs with a threshold.

To obtain OpenMax scores, activation vectors for each training instance \mathbf{x} are computed using the trained closed set classifier. After calculating the activation vectors for each training instance, $\mathbf{v}_1(\mathbf{x}), \dots, \mathbf{v}_N(\mathbf{x})$, where N is the number of KKC, can be obtained based on the activation vector for each training instance and its corresponding label. $\mu_j, j = 1, \dots, N$ referred to as the mean activation vector (MAV), can be computed for each class separately by taking the mean over only correctly classified training examples. Afterwards, a Weibull distribution $\rho_j = (\tau_j, \lambda_j, \kappa_j)$, where $\tau_j, \lambda_j, \kappa_j$ are the parameters of the Weibull distribution, can be fitted to each class on the largest distances between the MAV class and positive training instances using the libMR [8] FitHight function. We use the euclidean distance as the measure of distance in this paper. After obtaining the activation vectors for each class, calculating the MAV, and fitting the Weibull distributions, the OpenMax probability can be estimated using the procedures in Algorithm 1 [19]. Three hyper-parameters are used in the OpenMax algorithm. Tail size (η) is used during the parameter estimation phase for fitting the Weibull distribution by holding out a small set of data samples. α is the parameter used in revising the probabilities for the "top" classes. ϵ is the threshold used to support the rejection of uncertain inputs.

Algorithm 1 OpenMax probability estimation with rejection of unknown or uncertain inputs.

Require: Activation vector $\mathbf{v}(\mathbf{x}) = \mathbf{v}_1(\mathbf{x}), \dots, \mathbf{v}_N(\mathbf{x})$.

Require: means μ_j and libMR models $\rho_j = (\tau_i, \lambda_i, \kappa_i)$

Require: α , the number of "top" classes to revise

1: Let $S(i) = \text{argsort}(\mathbf{v}_j(\mathbf{x}))$; Let $w_j = 1$

2: **for** $i = 1, \dots, \alpha$ **do**

3: $w_{S(i)}(\mathbf{x}) = 1 - \frac{\alpha-i}{\alpha} e^{-\left(\frac{\|\mathbf{x}-\tau_{S(i)}\|}{\lambda_{S(i)}}\right)^{\kappa_{S(i)}}}$

4: **end for**

5: Revise activation vector $\hat{\mathbf{v}}(\mathbf{x}) = \mathbf{v}(\mathbf{x})\mathbf{w}(\mathbf{x})$

6: Define $\hat{\mathbf{v}}_0(\mathbf{x}) = \sum_i \mathbf{v}_i(\mathbf{x})(1 - w_i(\mathbf{x}))$

7:

$$\hat{P}(y = j|\mathbf{x}) = \frac{e^{\hat{v}_j(\mathbf{x})}}{\sum_{i=0}^N e^{\hat{v}_i(\mathbf{x})}}$$

8: Let $\mathbf{y}^* = \text{argmax}_j P(y == j|\mathbf{x})$

9: Reject input if $\mathbf{y}^* == 0$ or $P(y = \mathbf{y}^*|\mathbf{x}) < \epsilon$

3. EXPERIMENTAL ANALYSIS

This section illustrates the experimental setup, including the datasets, the selection of classes in the closed and open space for each experiment and a discussion of the results.

The experimental design and setup are also aimed at providing insights into the impact that *data quantity*, *closed and open set selection*, *model selection*, *data label quality* have on the performance of the OSR music genre classifier.

3.1 Datasets

We evaluate the performance of the OSR algorithm for music genre classification on two benchmark datasets: the GTZAN dataset and the FMA dataset. Each dataset is split into a training set, an evaluation set, and a test set for each experiment. The training set and the evaluation set only contain samples from the KKC while the test set also contains samples from genres in the UUCs. The KKC and UUCs are defined in Table 3. More detail on the training and test split for each experiment is presented in Table 2.

The GTZAN dataset consists of 1000 single-label audio files uniformly distributed among 10 genre labels. Each audio sample is a 30-second mp3. In this paper, we split the GTZAN dataset using the data split proposed in [23]. The genre labels in the closed and open sets for Experiments 1-4 are illustrated in Table 3. Notice from Table 2 that the training set and evaluation set has a smaller number of samples than the split proposed in [23] due to the fact that the training and evaluation sets only contain genres in the closed set labels.

For all experiments, the choice of KKC and UUC was made based on prior experiments and the evidence that improving the accuracy of the closed set classifier yields an improved open set classifier [24]. We chose genres for KKC that are better defined by the properties of their audios, whose embeddings have relatively lower mean distance to the genre's centroid, whilst also having experiments with balanced amounts of data in the open and closed sets.

The FMA dataset contains three different splits including FMA small, FMA medium, and FMA large. The descriptions can be found in [3]. We evaluate the model performance on the FMA dataset in three different experiments (index at 2, 3, and 4), where Table 2 describes the sample size for the training, the evaluation as well as the test set and Table 3 illustrates the choices of KKC and UUCs for each. We illustrate how to split data into the training set, the evaluation set, and the test set as follows:

1. Experiment 2 uses 8 genres in the FMA small as KKC to train and evaluate the closed set classifier and inference is run on UUCs from the remaining 8 genres in the valid split of the FMA medium dataset.
2. Experiment 3 redefines the KKC and UUC while still using the FMA small and medium datasets. The evaluation and test sets use samples from both the FMA small and medium datasets. The samples of KKC are split evenly for evaluation and testing.
3. Experiment 4 is the same as Experiment 3, with training and inference on audio samples from all FMA datasets including the FMA large split. Note that we only use audio sample with genre labels in the FMA large split.

3.2 Objective Metrics

The accuracy defined in (1) is used to evaluate the model performance. Here TP_k and FN_k are the number of true

Data	Training set	Evaluation set	Test set
GTZAN	226	103	290
FMA small & medium	6400	800	2505
	15040	940	1565
FMA large	26045	1516	3188

Table 2: This table contains the data sources as well as the number of samples for the training, evaluation, and test sets for each experimental setup. The size of the training and evaluation sets are based only on the audio samples in the KKC’s defined in Table 3 respectively.

positives and false negatives corresponding to class k . In the closed set model training process, $k \in [1, N]$, where N is the number of KKC’s. In the open set recognition process, $k \in [0, N]$ with UUC’s indexed at 0.

$$\frac{\sum_k TP_k}{\sum_k (TP_k + FN_k)}. \quad (1)$$

3.3 Implementation Details

As discussed in the previous section, we use transfer learning based on the CNN14 [5] architecture as the model backbone for the closed set classifier. In addition, the model weights are fine-tuned based on a task on the AudioSet dataset. For each experiment, we set the maximum epoch to 50. We compare the model performance by freezing or not freezing the CNN14 model weights. Moreover, we also compare the model performance with and without mixup augmentations for closed set classifier training. We find that the best model performance achieved for the GTZAN dataset occurs when the pre-trained model weights are frozen, and without mixup augmentation. However, better closed set model accuracy is achieved by not freezing the model weights and including mixup augmentation in the FMA experiments. All experiments use the Adam optimizer with a learning rate of 0.0001.

For the OSR task, we use grid search for $\alpha \in [N/2, N]$ and η ranges between 20 and 40 with increments of 5. The best model performances are presented in the next section.

3.4 Results

Table 3 shows the best performance for each experimental setup with the optimally identified hyper-parameters including α , η , and ϵ . In the following section, we elaborate on the results summarized in the table.

3.5 Discussion

We discuss the results and observe how they vary with the *rejection threshold* ϵ , *open set model*, *training data size*, *open and closed set class selection*, *data quality*, and observations on how well defined specific genres are based on the labels and audio features in these datasets.

Figure 2 plots the accuracies of OpenMax and SoftMax for different rejection thresholds ϵ . In plot (b), we notice that the highest accuracy occurs at the lowest ϵ amongst

experiments. Note that for all experiments the function of accuracy appears concave with respect to the rejection threshold. The optimal accuracy reached in Experiment 2 is the lowest among all experiments, consistent with the finding in [24] that an closed set classifier performance is correlated to the accuracy on the open set recognition task.

Comparing the performance of SoftMax and OpenMax with threshold we note that in plot (a) Figure 2, SoftMax with threshold has the highest optimal accuracy, though OpenMax performs best in lower ϵ regimes. In (b), SoftMax with threshold outperforms OpenMax but they converge to a similar maximum accuracy. In both (c) and (d) OpenMax consistently obtains higher accuracy than SoftMax indicating that OpenMax likely performs best when the closed set classifier has been trained on a larger dataset. Inspecting the confusion matrices in Figure 3 we also see that OpenMax is slightly less prone to misclassifying Country as unknown.

Regarding the impact of dataset size, Table 3 shows that both ACC_C and ACC_{Ot} accuracies are higher for the FMA large experiments than those on the FMA medium. In addition, genre classes with too few samples are unlikely to provide rich enough data to the closed set genre classifier to learn meaningful embeddings that would also be effective in the open set task. Specifically, the model struggles to effectively classify the genres with limited samples in training. For example, we have extremely low precision on Blues and Country and can see from Table 4 that these classes have the smallest training dataset sizes, which is likely impacting the model precision.

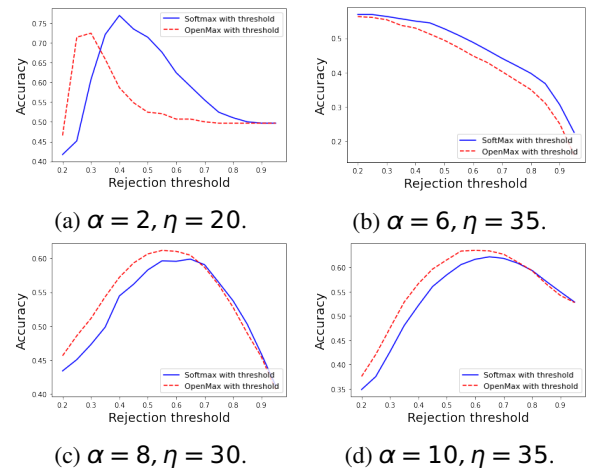


Figure 2: Accuracy vs. the rejection threshold for each experiment. (a) shows the result on Experiment 1. (b) and (c) present results using the FMA small & medium dataset on Experiment 2 and Experiment 3, respectively (d) shows the result on Experiment 4.

Experiments 2 and 3 illustrate the impact that the choice of closed and open set classes has on the OSR model accuracy. Colloquially, less a clearly defined genres like Pop and Experimental were exchanged with Jazz and Classical in the open set. The hypothesis that the classes moved to the open set in Experiment 3 are more challenging to define based on audio features is consistent with the higher

Index	Data	KKCs	UUCs	ACC_C	ACC_S	ACC_{St}	ACC_{Ot}
1	GTZAN	Blues, Classical, Country, Jazz, Metal	Disco, Hip-hop, Pop, Reggae, Rock	0.829	0.417	0.769	0.724
2	FMA small & medium	Hip-hop, Pop, Folk, Experimental, Rock, International, Electronic, Instrumental	Jazz, Soul-RnB, Blues, Spoken, Country, Classical, Old-Time Historic, Easy Listening	0.609	0.556	0.556	0.556
3	FMA small & medium	Hip-hop, Rock, Electronic, Instrumental, Jazz, Blues, Spoken, Country, Classical, Old-Time Historic	Pop, Folk, Experimental, International, Soul-RnB, Easy Listening	0.717	0.431	0.597	0.608
4	FMA large	Hip-hop, Rock, Electronic, Instrumental, Jazz, Blues, Spoken, Country, Classical, Old-Time Historic	Pop, Folk, Experimental, International, Soul-RnB, Easy Listening	0.718	0.343	0.622	0.635

Table 3: The table shows the model performance based on different datasets and different KKC and UUC selections. ACC_C is the closed set test accuracy - accuracy on KKCs. ACC_S is the overall accuracy using the softmax activation on the entire test set, including unknown classes. ACC_{St} and ACC_{Ot} are the accuracies for the test set using SoftMax with thresholds and OpenMax with thresholds respectively.

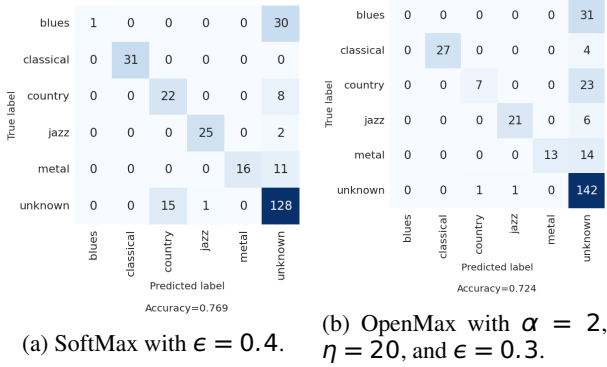


Figure 3: Confusion matrix for the GTZAN test split using the SoftMax with threshold and the OpenMax with threshold.

ACC_C in experiment 3. Both ACC_{St} and ACC_{Ot} are also higher, providing further evidence that a ‘best-case’ closed set classifier will provide the optimal accuracy when including the open set in the task.

The t-SNE plots of the activation vectors of each song in Figure 4 also illustrate how much more densely concentrated around their centroid some genres like Classical are. In clear juxtaposition to this, Pop labeled samples are far less concentrated and overlap significantly with the spaces predominantly populated by other genres.

Looking more closely at specific genres, we have already noticed that most audio samples labeled as Blues are misclassified as unknown in the open set recognition task, as shown in Figure 3. This is consistent with research found in [25] that there are manually mislabeled audio samples between Blues and Country. In addition, there are mislabelled data points for the Metal genre label. Plot (a) in Figure 4 illustrates a number of the outliers for different classes, such as the Blues activation vectors near the Metal cluster. The combination of limited data points labelled Blues, outliers, and highly distributed samples could be resulting in probabilities of all KKCs, thus resulting in all Blues samples being labeled as unknown.

Overall, the plots in Figure 4 indicate how distributed and less tightly defined the UUCs are than KKCs. In part this is to be expected because the activation vectors visualized are optimal representations learned from audio to separate the samples of each KKC. Pop and Hip-hop are sig-

nificantly overlapping, and are also distributed over Classical, Jazz and Metal clusters. However, even some of the UUCs are more tightly distributed than others. For example, the Classical samples in plot (e) of Figure 4 are noticeably nearer to each other than those for Blues. A genre that can be succinctly defined in terms of the features in its audio is likely to have higher precision, even if belonging to the set of UUCs.

Genre	Training sample	Precision	Recall
Hiphop	2939	0.407	0.667
Rock	11386	0.581	0.914
Electronic	7649	0.254	0.860
Instrumental	1578	0.031	0.091
Jazz	462	0.387	0.176
Blues	88	0.000	0.000
Spoken	276	0.597	0.254
Country	158	0.000	0.000
Classical	1069	0.784	0.453
Old-time Historic	440	0.556	0.789
Unknown	None	0.822	0.627

Table 4: This table contains the training data size, precision and recall for each genre in the FMA large experiment. It also includes the precision and recall for the UUCs that were not present in training.

4. CONCLUSION & FUTURE WORK

This paper establishes a baseline capacity and viable approach to novel genre detection by applying open set recognition methods to embeddings learned from ‘best-case’ closed set genre classifiers. We propose a system for open set recognition on music genres, which is both capable of classifying audio into known genres and also detecting previously unseen genres. We conduct a full experimental analysis of its performance on experiments designed from open source benchmark datasets and find that the open set accuracy can be affected by factors including the rejection threshold, the number of training samples, the choice of open set recognition model, the selection of KKCs and UUCs, and the data label quality, influenced by the qualitative definition of genres.

As [25] suggested, there are issues with labeled genre

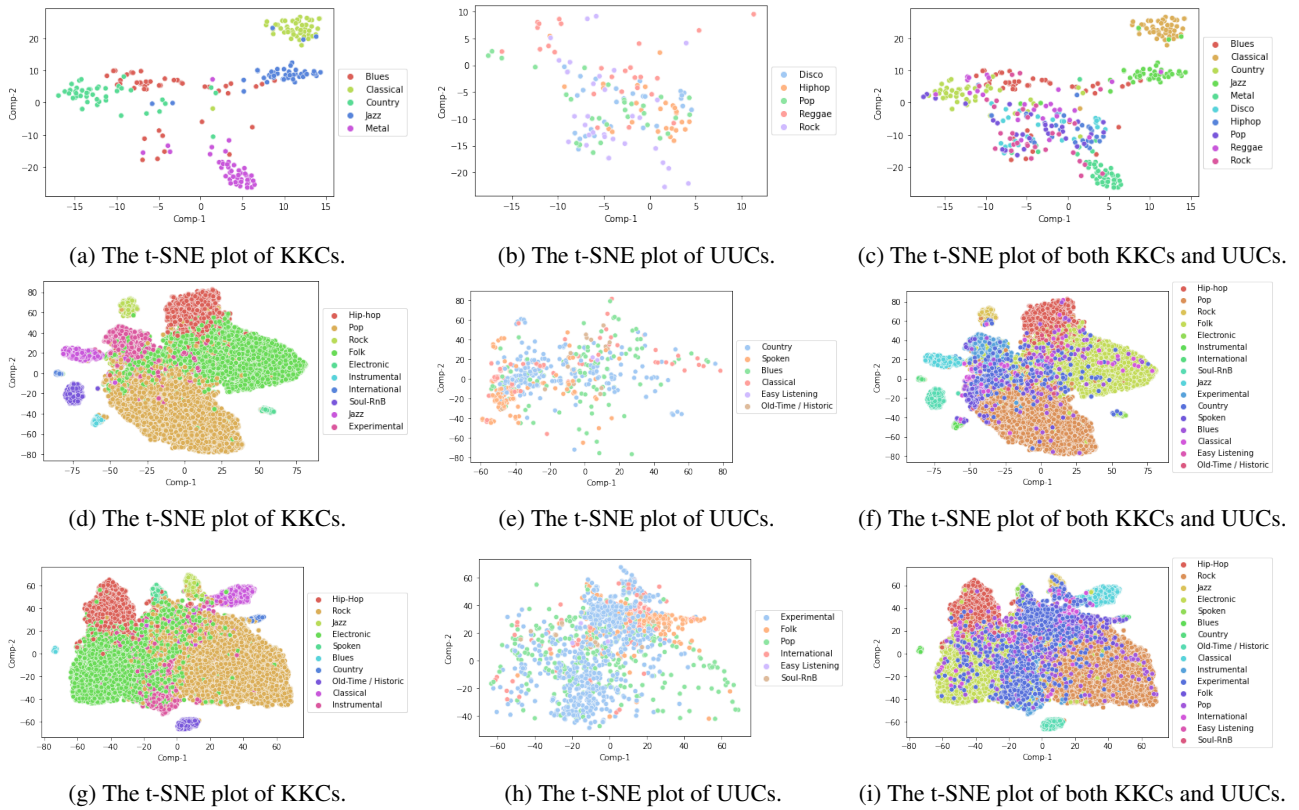


Figure 4: The t-SNE plots. (a)-(c) are the t-SNE plots for Experiment 1. (d) - (f) are the t-SNE plots for Experiment 2. (g) - (i) are the t-SNE plots for Experiment 4. The t-SNE plots for Experiment 3 is not presented as it is similar to the plot for Experiment 4 due to the fact that they have the same KKC and UUCs.

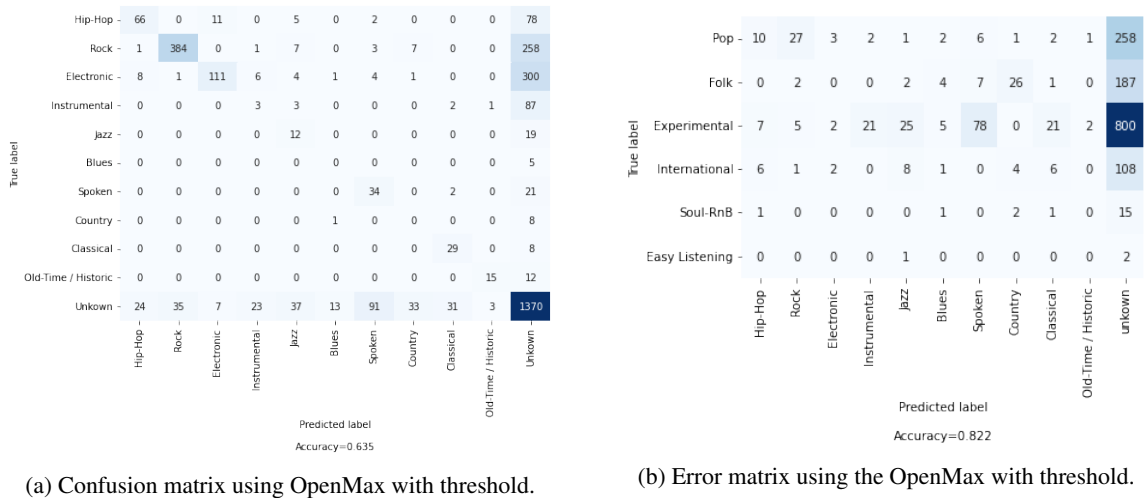


Figure 5: Confusion matrix and error matrix for the experiment using the FMA large dataset. (a) is the confusion matrix showing how each genre category in the test split is predicted. (b) is the error matrix illustrates on how the unknown genre categories predicted incorrectly into close set genre categories or correctly as the unknown genre category.

tags, even those provided by experts in the GTZAN dataset. There are similar caveats and complications in the artist provided labels in the FMA dataset [26]. A possible research direction is to better define music genres based on audio features and design machine learning systems to automatically correct mislabeled genres, leveraging unsupervised approaches to learning groups of similar music.

We could also employ few-shot learning to identify music genres for which we have limited training data. Furthermore, genres emerge over time, and a practical system that detects the emergence of novel genres from audio could be trained on samples of genres that are segmented into KKC and UUCs based on when they were released, allowing the trained model to provide a realistic measure of our ability

to identify if music is of a new genre.

5. REFERENCES

- [1] Gramophone. A brief history of classical music. [Online]. Available: <https://www.gramophone.co.uk/features/article/a-brief-history-of-classical-music>
- [2] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on speech and audio processing*, vol. 10, no. 5, pp. 293–302, 2002.
- [3] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson, "Fma: A dataset for music analysis," *arXiv preprint arXiv:1612.01840*, 2016.
- [4] A. Ghildiyal, K. Singh, and S. Sharma, "Music genre classification using machine learning," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 2020, pp. 1368–1372.
- [5] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "Panns: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.
- [6] C. Liu, L. Feng, G. Liu, H. Wang, and S. Liu, "Bottom-up broadcast neural network for music genre classification," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7313–7331, 2021.
- [7] S. Chillara, A. Kavitha, S. A. Neginhal, S. Haldia, and K. Vidyullatha, "Music genre classification using machine learning algorithms: a comparison," *Int. Res. J. Eng. Technol.(IRJET)*, vol. 6, no. 05, pp. 851–858, 2019.
- [8] W. J. Scheirer, A. Rocha, R. J. Micheals, and T. E. Boulton, "Meta-recognition: The theory and practice of recognition score analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1689–1695, 2011.
- [9] L. Neal, M. Olson, X. Fern, W.-K. Wong, and F. Li, "Open set learning with counterfactual images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 613–628.
- [10] S. Kong and D. Ramanan, "Opengan: Open-set recognition via open data generation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 813–822.
- [11] J. Naranjo-Alcazar, S. Perez-Castanos, P. Zuccarello, A. M. Torres, J. J. Lopez, F. J. Ferri, and M. Cobos, "An open-set recognition and few-shot learning dataset for audio event classification in domestic environments," *arXiv preprint arXiv:2002.11561*, 2020.
- [12] D. Battaglino, L. Lepauloux, and N. Evans, "The open-set problem in acoustic scene classification," in *2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2016, pp. 1–5.
- [13] V. M. Venkataram, *Open set text classification using neural networks*. University of Colorado Colorado Springs, 2018.
- [14] L. Shu, H. Xu, and B. Liu, "Doc: Deep open classification of text documents," *arXiv preprint arXiv:1709.08716*, 2017.
- [15] W. J. Scheirer, L. P. Jain, and T. E. Boulton, "Probability models for open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 11, pp. 2317–2324, 2014.
- [16] L. P. Jain, W. J. Scheirer, and T. E. Boulton, "Multi-class open set recognition using probability of inclusion," in *European Conference on Computer Vision*. Springer, 2014, pp. 393–409.
- [17] P. R. Mendes Júnior, R. M. De Souza, R. d. O. Werneck, B. V. Stein, D. V. Pazinato, W. R. de Almeida, O. A. Penatti, R. d. S. Torres, and A. Rocha, "Nearest neighbors distance ratio open-set classifier," *Machine Learning*, vol. 106, no. 3, pp. 359–386, 2017.
- [18] H. Zhang and V. M. Patel, "Sparse representation-based open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1690–1696, 2016.
- [19] A. Bendale and T. E. Boulton, "Towards open set deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1563–1572.
- [20] C. Geng, S.-j. Huang, and S. Chen, "Recent advances in open set recognition: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3614–3631, 2020.
- [21] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2017, pp. 776–780.
- [22] R. L. Smith, "Extreme value theory," *Handbook of applicable mathematics*, vol. 7, pp. 437–471, 1990.
- [23] C. Kereliul and B. Sturm, "dnn-mgr," <https://github.com/coreyker/dnn-mgr>, 2016.
- [24] S. Vaze, K. Han, A. Vedaldi, and A. Zisserman, "Open-set recognition: A good closed-set classifier is all you need," in *International Conference on Learning Representations*, 2021.
- [25] B. L. Sturm, "The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use," *arXiv preprint arXiv:1306.1461*, 2013.
- [26] C. Zhang, Y. Zhang, and C. Chen, "Songnet: Real-time music classification," 2019.