recognition dataset (explained in Section 3.1). Gulati et al. [20] proposed a novel feature for mode recognition in the context of Hindustani and Carnatic ragas, called the time-delayed melody surface (TDMS), which we will use in this study. Authors reported that TDMS-based models outperform PCD-based models of [10] in the raga recognition task.

Yeşiler et al. [14] used a Multilayer Perceptron (MLP) on pitch distribution of first and last sections together with overall distributions using a feature vector of 159 attributes (53-TET * 3 octaves). The highest accuracy reported is 75.6% on the OTMM recognition dataset with additional insights on relevant segments for better discriminability. Demirel et al. [11] advocate the advantage of using chroma features for the mode recognition task as it discards the need for automatic melody extraction of polyphonic audio, hence getting away with the imperfections thereof. Authors created makam templates from annotated data and used template matching using support vector machine (SVM) classifiers. The best performing model achieved an accuracy of 77% on the OTMM recognition dataset, which, to our knowledge, is the state-of-the-art in makam recognition applied to the OTMM corpus. Other works are not strictly mode recognition but use the framework of representation-cum-distance-measure for discriminating between allied raga-pairs [34, 35]. Section 4 discusses aspects we borrow from this methodology to quantify the classification errors from a clustering viewpoint.

3. METHODOLOGY

In [10], mode recognition is formally defined as classifying the mode of an audio fragment from a discrete set of modes. In the context of OTMM, the problem reduces to classifying the makam. Given that the mode recognition framework is already established and that we are benchmarking on literature applied to OTMM, we save some real estate assuming that the data (pre)processing and partial feature extractions will be exactly reproduced.

3.1 The OTMM Corpus and the Dataset

Considering the lack of open data sources for makam music, the CompMusic project gathered audio recordings, music scores and relevant metadata, and published in the public domain the *Dunya Ottoman-Turkish Makam Music Corpus* [5, 36], which is currently the most representative corpus for OTMM available for computational research purposes. From the corpus, [10] curated a test dataset of audio recordings with annotated makam and tonic, called the *Ottoman-Turkish makam music recognition dataset*. The dataset covers 20 commonly performed makams [5] composed of 1000 audio recordings. A single makam is performed in each recording (i.e. there are 50 recordings per makam). To the best of our knowledge, this dataset is the largest and the most comprehensive dataset for the evaluation of automatic makam recognition. Finally, the dataset has been used by other researchers [10, 11, 14] to demonstrate their methods, including the current state-of-the-art makam recognition approach [11].

We use the latest version of the dataset. [6] We use the pre-computed melody time-series (termed as "predominant melody" by [36]) provided in CompMusic Dunya. [7] The pitch is detected at a hop-size to sampling-rate ratio of 0.023 that translates to 23 ms intervals for 44.1 kHz sampled audio [37]. To compare across performances, it is crucial to normalise the melody with respect to the tonic frequency. For this study, we use manually curated tonic frequencies linked from the *Ottoman-Turkish tonic dataset*. [8] To avoid the effect of nonlinearity in the logarithmic Hz scale, we normalise the pitch time-series to a log-linear cents scale.

3.2 Feature Extraction and Modeling

The next step is to synthesise derived features from the raw predominant melody. These mid- or high-level features can be interpreted and mapped to musicological inferences. We follow an approach akin to Krumhansl's [38] to compute the histogram of pitch samples to construct the pitch-class distribution. The pitch values are octave-folded (0 — 1200 cents) and quantised into p bins of equal width. The bin centre is the arithmetic mean of the adjacent bin edges. The salience of each bin is proportional to the accumulated duration of the pitches within that bin. A probability distribution function is constructed where the area under the histogram sums to unity. Even though we use the equivalent of a PD method attributed to the high bin resolution, we converge to a PCD [18]. The PCD configuration is given in Section 4. The first row of Figure 1 shows PCDs computed from each Mahur, Rast, and Acemşiran in the dataset, with the average pitch at each bin drawn as a dashed line.

The next step is to construct a two-dimensional surface based on the concept of delay coordinates (also termed phase space embedding) [20]. The time-delayed melody surface (TDMS) is a compact representation that captures both the tonal and the temporal characteristics of melody, is robust to octave errors, also partially nullifies the relevance of melody transcription. We experiment with different parameters (See Section 4). The second row of Figure 1 shows TDMS averaged from the TDMS of all recordings in Mahur, Rast, and Acemşiran makams in the dataset. The horizontal and vertical trajectories indicate pitch transitions between the pitch classes. The isolated square shape-formation indicates a separation between the higher and lower tetrachords in the course of the melodic progression. In both PCD and TDMS features, makams Rast and Mahur are similar to a high degree. The PCD of makam Acemaşiran bears relatively small differences from the prior two; however, the TDMS representation manages to significantly differentiate itself via dif-

[5] Namely: Acemaşiran, Acemkürdi, Bestenigar, Beyati, Hicaz, Hicazkar, Hüseyni, Hüzzam, Karcığar, Kürdilihicazkar, Mahur, Muhayyer, Neva, Nihavent, Rast, Saba, Segah, Sultanıyegah, Suzinak, and Uşşak.

[6] **dlfm2016-fix1** — `https://zenodo.org/record/4883680`

[7] `https://dunya.compmusic.upf.edu`

[8] `https://zenodo.org/record/260038`

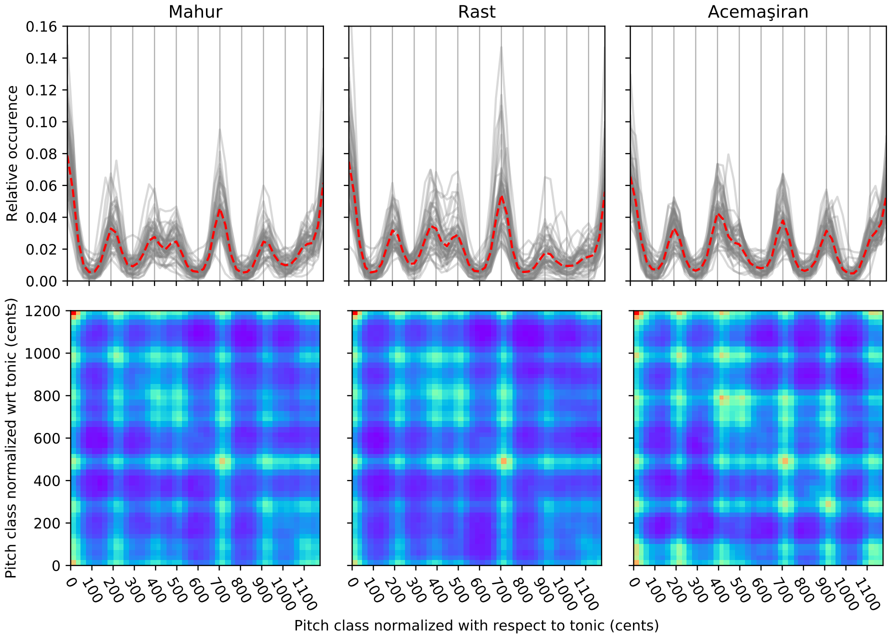

Figure 1. PCD (top row) and TDMS (bottom row) representations of Mahur (left column), Rast (middle column), and Acemaşiran (right column) makams.

ferences in melodic progression, captured through the delay coordinates. For example, in the TDMS representation, makams Mahur and Rast show clear transitions between 6^{th} (900 cents), 7^{th} scale degrees (1100 cents), and the tonic pitch class, whereas makam Acemşiran exhibits a progressions between 3^{rd} (400 cents) and 6^{th} scale degrees (900 cents).

4. EXPERIMENTS

Akin to the title of the paper, one of the main goals of the study is to treat makam recognition as a task and not a goal in itself. Our experiments are divided into two parts. The former addresses the 'goal'-oriented aspect, i.e., the best combination of features and classifiers in order to achieve the *optimal* accuracy. We also stress on intuitions why certain configuration of train-test partitioning would make more sense than others or why certain classifier is meant to 'learn' and not be totally data-proximity dependent. We report and discuss a subset of the results relevant to the optimal settings; the full experimental results are made available. [9]

For PCDs, we use the empirical "optimal" parameters for OTMM reported by [10], namely a bin resolution of 25 cents and Gaussian smoothing applied using a kernel width of 25 cents. We set the bin resolution of TDMS to 25 cents to compare with PCD, and grid-search time delay indices ($\in \{0.25, 0.5.1, 1.5, 2.5, 5\}$ seconds), compression exponents ($\in \{0.1, 0.25, 0.5, 0.75, 1\}$) and Gaussian smoothing kernel widths ($\in \{0, 12.5, 25, 50\}$ standard deviation in cents) in the classification experiments below to find the optimal configuration for OTMM.

4.1 (Supervised) Classification

In line with the objective of supervised learning, i.e., to model the intra-class similarity and inter-class differences, the inherent 'goal' is to maximise classification accuracy. However, we carefully choose the feature set and classifier in order to suit the music theory. That is to say, we aim to incorporate knowledge constraints into mainstream data-driven computational models. We restrict ourselves to k-nearest neighbors [10, 12, 13, 19, 20], support vector machine [11, 12], multilayer perceptron [14] and logistic regression [12] that were extensively used in past mode recognition work.

Past studies used different cross-validation (CV) techniques such as leave-one-out CV [13], 10-fold CV [10, 19, 20] and nested k-fold CV [11, 14]. In our initial experiments, we compared nested 10-fold CV, 10-fold CV (without any unseen test set), and 10-times repeated shuffle split CV with 10% of the recordings reserved as test set for each repetition. We used stratified splits in all our experiments to keep the makam classes balanced and repeated each experiment 10 times. We report the mean & standard deviation of classification accuracy reported on the test set. We also compute a confusion matrix for each test set and aggregate it across all test sets in each repeated experiment per model. We report the results of the 10-times repeated shuffle split CV in the rest of the Section. Similar to [20], we observed that TDMS is robust in different time delay indices (between 0.5 and 2.5 seconds), kernel width (less than 25 standard deviations in cents) and compression exponent (above 0.25). For the rest of the experiments, we report results for TDMS with an "optimal" configuration of 1-second time-delay index, 12.5 cents of smoothing kernel width and a compression exponent of 0.5. Table 1

[9] https://sertansenturk.com/work-research/ismir-2022-makam/

Model	PCD	TDMS
Support Vector Machine	$71.0 \mp 3.2\%$	$\mathbf{77.2 \mp 3.5\%}$
Multilayer Perceptron	$70.9 \mp 3.8\%$	$74.6 \mp 5.0\%$
k-nearest Neighbors	$68.2 \mp 3.4\%$	$70.2 \mp 3.1\%$
Logistic Regression	$66.8 \mp 3.9\%$	$75.5 \mp 4.1\%$

Table 1. Average $\mp$ standard deviation of classification accuracy for all feature-classifier combinations.

shows the mean and standard deviation in classification accuracy for PCD and TDMS using different models. In sum, TDMS with SVM works the best at par with the current state-of-the-art [11]. TDMS consistently performs better than PCD; statistical significance results are kept out of the scope of this work.

The associated confusion matrix for the optimal performing system is shown in Figure 2. Makams {Acemaşiran, Hicaz, Hüzzam} and {Bestenigar, Rast, Uşşak} are examples of highly discriminable and highly confused pairs respectively. The inferences from the confusion matrix are, however, limited to qualitative evaluation of the confused cases and a count of them. In the next Section, we propose a new approach to compute the pairwise distances in a clustering scenario which facilitates a quantitative evaluation of the proportion and magnitude of the confusions. This is, in a way, a manifestation of the recognition task wherein we model the inherent complexity in the data rather than the limitations of the method.

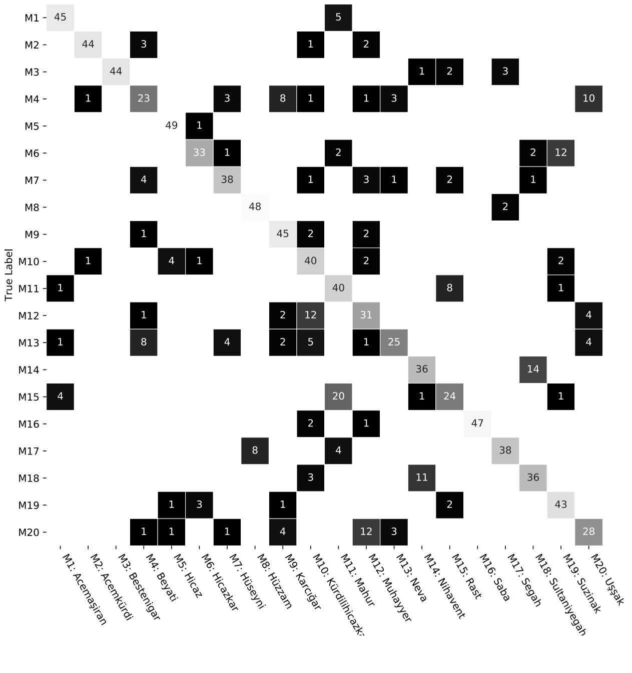

Figure 2. Aggregated confusions matrix for the optimal performing system: TDMS with SVM.

4.2 (Unsupervised) Clustering

To address the second half of the experiments, our focus moves to the 'task'-oriented results. This, we believe, is the highlight of the current contribution in terms of gaining/reconfirming musicological knowledge/concepts from the computational model that can feed back into the pedagogy and practices to further enrich the repertoire.

A typical mode recognition study would stop at this point after the goal accuracy is achieved [10, 11, 20]. Even though we report comparable results to that of the current state-of-the-art accuracy, we are keen on evaluating how much the representation-cum-distance-measure attribute to musicological insights. Over and beyond the error analysis, we stress validating the gap and reinforcing other possible avenues, so the complexity of the musical characteristics is not attributed to the shortcomings of computational models. One such way is to disregard the makam labels and study the melodic similarity space of relevant predictor features. Through these methods, we aim to verify whether the machine learning models indeed 'learn' what they are intended for. We present three complementary and supplementary retrieval scenarios to corroborate the 'task' details and bridge the gap that the best classification could achieve.

Hierarchical clustering: In the presence of theoretical grouping of makams, yet not having a prescription on the counts, it is practically impossible to set a k for a k-means clustering algorithm. However, hierarchical clustering seems to offer a dynamic solution in such scenarios. Here, each element is treated as distinct clusters at the lowest threshold, whereas there is a single giant cluster at the highest threshold. We present, in part of Figure 3, the dendrogram representation to capture the melodic similarity/grouping space across the 20 makams obtained from the hierarchical clustering of the TDMS features averaged from the 50 recordings per makam using Canberra distance. We use the Canberra distance to contrast with the hierarchical clustering reported in [14, see Figure 2] that was calculated from averaged pitch distributions, and keep the empirical experimentation of different distance metrics out of the scope. The groupings are shown in different colours, and the relative height where the tree elements merge is indicative of the normalised threshold. At the highest threshold, makam Saba isolates itself from the rest 19; this is indicative of the distinct nature of the pitch distribution captured through TDMS. Makam-pair (Rast, Mahur) show a very low distance, indicating high similarity in the feature space. A high distance between the (Hüzzam, Hicaz) pair, as shown in other figures, is also evident.

Cluster purity matrix: One supplementary way to capture all pairwise distances is through an unconventional method of computing a distance matrix with the salience function of cluster purity. This is broadly a homogeneity measure that evidences the quality of clustering. Any value close to 1 indicates perfect clustering, while 0.5 signifies random clustering. Out of the $\binom{20}{2} = 190$ distinct makam-pairs, 115 pairs show a cluster purity score ≥ 0.85, 73 other pairs bear cluster purity values in the range of (0.5,

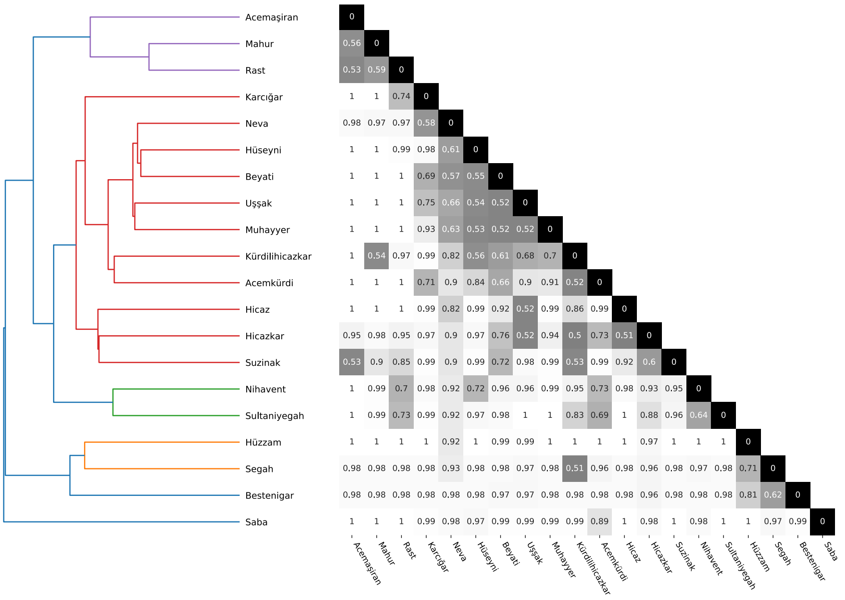

Figure 3. Left: dendrogram representation to capture the melodic similarity space across the 20 makams obtained from the hierarchical clustering of the TDMS. Right: the cluster purity matrix capturing the homogeneity of pairwise distances.

0.85), and none below a score of 0.5. We plot the pairwise purity scores partly in Figure 3, this representation has an intuitive inverse proportionality with the confusion matrix. The makam indices obtained from the dendrogram are aligned with that of the current matrix. It is intuitive to follow that the row corresponding to makam Saba (which is the most distinctive) has got the highest cluster purity (mode=1) with many other pairs, whereas the aforementioned confusable pairs yield a random clustering. This visualisation provides a complementary view of what is aggregated in the dendrogram.

Query retrieval score: The third scenario we present complementary to the clustering is the receiver-operated characteristics (ROC). We use the same two makam-pairs in the context of a query search for all possible matching and non-matching pairings out of each makams-pair. The ROCs in Figure 4 show the true positive rate versus the false positive rate achieved in the detection of non-matching makam pairs for the PCD (we consciously chose PCD over TDMS to introduce diversity) representations for four unique distance measures, inspired from [34]. The subplots correspond to makam-pairs (Hüzzam, Hicaz) and (Rast, Mahur), which are examples of highly discriminable and highly confused pairs, respectively. The ROC curves, the area under the curve (AUC) and equal error rate (EER) clearly indicate a better retrieval for the former, while the latter almost grazes the diagonal. We have inferences on why certain distance metric works better, but discussion on their relative performance is not directly related to the main narrative of this work and hence omitted [34, 39].

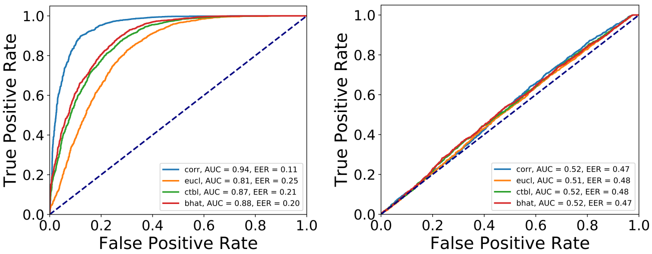

Figure 4. ROCs obtained using correlation (corr), euclidean (eucl), city-block (ctbl), and Bhattacharyya (bhat) distance from PCDs. Left: (Hüzzam, Hicaz) and right: (Rast, Mahur) makam-pairs.

5. CONCLUSION

We employed methodologies that combine domain knowledge and data-driven optimisations with a view to understanding the makam recognition 'task' in depth. We report comparable accuracy (77.2%) with the state-of-the-art [11] using the newly adapted TDMS feature with SVM. This achievement evidences the credential in our approach that provides us with a solid ground to argue the critique on goal- versus task-oriented approaches by comparing and contrasting. We have reported only the best-performing configuration in this paper [10] which will eventually be expanded to include temporal features and sequence models. In sum, we advocate that good supervised learning performance is a necessary but insufficient condition for a computational representation-cum-distance-measure to be considered informative for all purposes. As future work, we will incorporate convolutional networks and transformers to reproduce makam recognition on OTMM corpus to understand the potential of deep learning and the trade-off between goal and task involved. The application of such an approach aids in understanding music and other forms of sound culture and developing methodologies for cross-cultural mapping and comparing these materials.

[10] Array of alternate configurations: `https://sertansenturk.com/work-research/ismir-2022-makam/`

6. ACKNOWLEDGEMENTS

This research is part of project "Computationally engaged approaches to rhythm and musical heritage: Generation, analysis, and performance practice," funded through a grant from the Research Enhancement Fund at the New York University Abu Dhabi.

7. REFERENCES

[1] H. S. Powers and R. Widdess, *India, subcontinent of*, 2nd ed., ser. New Grove Dictionary of Music. Macmillan, London, 2001, ch. III: Theory and practice of classical music, contributions to S. Sadie (ed.).

[2] H. S. Powers, F. Wiering, J. Porter, J. Cowdery, R. Widdess, R. Davis, and A. Marett, "Mode. grove music online," 2008.

[3] K. K. Ganguli, "How do we 'See' & 'Say' a raga: A Perspective Canvas," *Samakalika Sangeetham*, vol. 4, no. 2, pp. 112–119, Oct. 2013.

[4] T. Çevikoğli, "Klasik Türk müziğinin bugünkü sorunları," in *Proceedings of International Congress of Asian and North African Studies (ICANAS 38')*, Ankara, 2007.

[5] B. Uyar, H. S. Atli, S. Şentürk, B. Bozkurt, and X. Serra, "A corpus for computational research of Turkish makam music," in *Proceedings of the 1st International Workshop on Digital Libraries for Musicology (DLfM 2014)*, London, UK, 2014, pp. 1–7.

[6] B. Bozkurt, R. Ayangil, and A. Holzapfel, "Computational analysis of Turkish makam music: Review of state-of-the-art and challenges," *Journal of New Music Research*, vol. 43, no. 1, pp. 3–23, 2014.

[7] K. K. Ganguli and P. Rao, "On the perception of raga motifs by trained musicians," *The Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2418–2434, 2019.

[8] ——, "A study of variability in raga motifs in performance contexts," *Journal of New Music Research*, vol. 50, no. 1, pp. 102–116, 2021.

[9] S. T. Madhusudhan and G. Chowdhary, "DEEPSRGM-sequence classification and ranking in Indian classical music with deep learning," in *Proceedings of the 20th International Society for Music Information Retrieval Conference (ISMIR 2019)*, 2019, pp. 533–540.

[10] A. Karakurt, S. Şentürk, and X. Serra, "MORTY: A toolbox for mode recognition and tonic identification," in *Proceedings of the 3rd International workshop on Digital Libraries for Musicology (DLfM 2016)*, New York, NY, 2016, pp. 9–16.

[11] E. Demirel, B. Bozkurt, and X. Serra, "Automatic makam recognition using chroma features," in *Proceedings of the 8th International Workshop on Folk Music Analysis (FMA 2018)*. Thessaloniki, Greece: Aristotle University of Thessaloniki, 2018, pp. 19–24.

[12] G. K. Koduri, S. Gulati, P. Rao, and X. Serra, "Raga recognition based on pitch distribution methods," *Journal of New Music Research*, vol. 41, no. 4, pp. 337–350, 2012.

[13] A. C. Gedik and B. Bozkurt, "Pitch-frequency histogram-based music information retrieval for Turkish music," *Signal Processing*, vol. 90, no. 4, pp. 1049–1063, 2010.

[14] F. Yeşiler, B. Bozkurt, and X. Serra, "Makam recognition using extended pitch distribution features and multi-layer perceptrons," in *Proceedings of the 15th Sound and Music Computing Conference (SMC 2018)*. Limassol, Cyprus: Cyprus University of Technology, 2018, pp. 249–253.

[15] P. Dighe, P. Agrawal, H. Karnick, S. Thota, and B. Raj, "Scale independent raga identification using chromagram patterns and swara based features," in *Proceedings of IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2013, pp. 1–4.

[16] P. Dighe, H. Karnick, and B. Raj, "Swara histogram based structural analysis and identification of Indian classical ragas." in *Proceedings of 14th International Society for Music Information Retrieval Conference (ISMIR 2013)*, 2013, pp. 35–40.

[17] S. Gulati, J. Serra, V. Ishwar, S. Şentürk, and X. Serra, "Phrase-based rāga recognition using vector space modeling," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*. IEEE, 2016, pp. 66–70.

[18] P. Chordia and A. Rae, "Raag recognition using pitch-class and pitch-class dyad distributions." in *Proceedings of 8th International Society for Music Information Retrieval Conference (ISMIR 2007)*, 2007, pp. 431–436.

[19] P. Chordia and S. Şentürk, "Joint recognition of raag and tonic in north Indian music," *Computer Music Journal*, vol. 37, no. 3, pp. 82–98, 2013.

[20] S. Gulati, J. Serrà Julià, K. K. Ganguli, S. Sentürk, and X. Serra, "Time-delayed melody surfaces for rāga recognition," in *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR 2016)*, New York, NY, 2016, pp. 751–757.

[21] S. Abdoli, "Iranian traditional music dastgah classification." in *Proceedings of 12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 2011, pp. 275–280.

[22] P. Heydarian and D. Bainbridge, "Dastgàh recognition in Iranian music: Different features and optimized parameters," in *Proceedings of 6th International Conference on Digital Libraries for Musicology (DLfM 2019)*, New York, NY, USA, 2019, p. 53–57.

[23] L. Cimarone, B. Bozkurt, and X. Serra, "Automatic dastgah recognition using Markov models," in *Proceedings of 14th International Symposium on Computer Music Multidisciplinary Research (CMMR 2019)*, Marseille, France, 10 2019, pp. 51–58.

[24] D. Huron and J. Veltman, "A cognitive approach to medieval mode: Evidence for an historical antecedent to the major/minor system," *Empirical Musicology Review*, vol. 1, no. 1, pp. 33–55, 2006.

[25] B. Cornelissen, W. Zuidema, and J. A. Burgoyne, "Mode classification and natural units in plainchant," in *Proceedings of the 21th International Society for Music Information Retrieval Conference (ISMIR 2020)*, 2020, pp. 869–875.

[26] K. K. Ganguli, S. Gulati, X. Serra, and P. Rao, "Data-driven exploration of melodic structures in Hindustani music," in *Proc. of the International Society for Music Information Retrieval (ISMIR)*, Aug. 2016, pp. 605–611, New York, USA.

[27] G. K. Koduri, V. Ishwar, J. Serrà, and X. Serra, "Intonation analysis of rāgas in Carnatic music," *Journal of New Music Research*, vol. 43, no. 1, pp. 72–93, 2014.

[28] K. K. Ganguli, A. Rastogi, V. Pandit, P. Kantan, and P. Rao, "Efficient melodic query based audio search for Hindustani vocal compositions," in *Proc. of Int. Soc. for Music Information Retrieval (ISMIR)*, Oct. 2015.

[29] K. K. Ganguli, A. Lele, S. Pinjani, P. Rao, A. Srinivasamurthy, and S. Gulati, "Melodic shape stylization for robust and efficient motif detection in hindustani vocal music," in *National Conference on Communications (NCC)*, 2017.

[30] K. K. Ganguli and P. Rao, "Discrimination of melodic patterns in Indian classical music," in *Proc. of National Conference on Communications (NCC)*, Feb. 2015.

[31] S. Şentürk, G. K. Koduri, and X. Serra, "A score-informed computational description of svaras using a statistical model," in *Proceedings of 13th Sound and Music Computing Conference (SMC 2016)*, Hamburg, Germany, 2016, pp. 427–433.

[32] J. C. Ross, A. Mishra, K. K. Ganguli, P. Bhattacharyya, and P. Rao, "Identifying raga similarity through embeddings learned from compositions' notation." in *Proc. of the International Society for Music Information Retrieval (ISMIR)*, 2017.

[33] S. Gulati, J. Serra, K. K. Ganguli, and X. Serra, "Landmark detection in Hindustani music melodies," in *Proc. of Int. Computer Music, Sound and Music Computing*, 2014, pp. 1062–1068, Athens, Greece.

[34] K. K. Ganguli and P. Rao, "On the distributional representation of ragas: experiments with allied raga pairs," *Transactions of the International Society for Music Information Retrieval*, vol. 1, no. 1, 2018.

[35] ——, "Towards computational modeling of the ungrammatical in a raga performance." in *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR 2017)*, Suzhou, China, 2017, pp. 39–45.

[36] S. Şentürk, "Computational analysis of audio recordings and music scores for the description and discovery of Ottoman-Turkish makam music," Ph.D. dissertation, Universitat Pompeu Fabra, Barcelona, Spain, December 2016.

[37] H. S. Atlı, B. Uyar, S. Şentürk, B. Bozkurt, and X. Serra, "Audio feature extraction for exploring Turkish makam music," in *Proceedings of 3rd International Conference on Audio Technologies for Music and Media (ATMM 2014)*, Ankara, Turkey, 2014.

[38] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch.* Oxford University Press, New York, 1990, ch. 4: A key-finding algorithm based on tonal hierarchies, pp. 77 – 110.

[39] K. K. Ganguli, "A corpus-based approach to computational modeling of melody in raga music," Ph.D. dissertation, Indian Institute of Technology Bombay, Mumbai, India, 2019.

A DATASET FOR GREEK TRADITIONAL AND FOLK MUSIC: Lyra

Charilaos Papaioannou[1], Ioannis Valiantzas[2], Theodoros Giannakopoulos[3], Maximos Kaliakatsos-Papakostas[4], Alexandros Potamianos[1]
[1] School of ECE, National Technical University of Athens, Greece
[2] Department of Music Studies, National and Kapodistrian University Of Athens, Greece
[3] National Center for Scientific Research - Demokritos, Greece
[4] Athena RC, Greece
cpapaioan@mail.ntua.gr

ABSTRACT

Studying under-represented music traditions under the MIR scope is crucial, not only for developing novel analysis tools, but also for unveiling musical functions that might prove useful in studying world musics. This paper presents a dataset for Greek Traditional and Folk music that includes 1570 pieces, summing in around 80 hours of data. The dataset incorporates YouTube timestamped links for retrieving audio and video, along with rich metadata information with regards to instrumentation, geography and genre, among others. The content has been collected from a Greek documentary series that is available online, where academics present music traditions of Greece with live music and dance performance during the show, along with discussions about social, cultural and musicological aspects of the presented music. Therefore, this procedure has resulted in a significant wealth of descriptions regarding a variety of aspects, such as musical genre, places of origin and musical instruments. In addition, the audio recordings were performed under strict production-level specifications, in terms of recording equipment, leading to very clean and homogeneous audio content. In this work, apart from presenting the dataset in detail, we propose a baseline deep-learning classification approach to recognize the involved musicological attributes. The dataset, the baseline classification methods and the models are provided in public repositories. Future directions for further refining the dataset are also discussed.

1. INTRODUCTION

Traditional music of Greece is under-represented in available datasets, despite the fact that this music offers unique perspectives that combine characteristics of Western and Eastern music. The development of a traditional Greek music dataset is interesting to study in its own right, while it is also expected to provide a more complete picture of the music in the Mediterranean and Europe. Computational methods have been extensively studied in the past decades for carrying out ethnomusicological studies, constituting the field of Computational Ethnomusicology. A review of such methods is presented in [1], where it is evident that there are many benefits in compiling datasets of traditional music that can readily be used in computational models.

Several such datasets have been developed. For instance, an enormous database of Dutch melodies and songs that allows studying multiple musicological aspects is presented in [2]. Similarly, a dataset of Indian art music was presented in [3], where various MIR-related tasks were adjusted and applied therein. The employment of standard MIR tools has been evaluated in Arab-Andalusian music [4] and Flamenco music with the COFLA dataset [5]. Iranian Dastgah classification has been studied with MIR tools in [6]. A dataset of Georgian vocal music from historic tape recordings of three-voice songs was presented in [7], where the aim was to preserve cultural heritage and also to apply and examine existing MIR methods (e.g., for pitch and onset detection) in data of this form. Such datasets also provide opportunities for studying idiosyncratic musical instruments, e.g., in the case of Chinese music [8], where many instruments are employed that are not related to the ones used in Western music.

In several cases, off-the-shelf MIR tools are not well-suited for tasks that include traditional musical types. For instance, studying melodic similarity of Turkish non equally-tempered makam phrases in symbolic format required the development of a novel representation based on MIDI [9]. Similarly, rhythmic attributes of makam music with new models that integrate note transition rules for this music had to be developed for lyrics-to-audio alignment [10]. In addition, a novel method, based on audio signal processing, was created for identifying asymmetric rhythms in Greek traditional music [11].

The “Lyra” dataset of traditional and folk Greek music aims to contribute to all the aforementioned fields, from data-driven (computational) ethnomusicology to shaping new directions of research for MIR tools. The majority of available datasets that concern Greek music are unsuitable

Attribution: C. Papaioannou, I. Valiantzas, T. Giannakopoulos, M. Kaliakatsos-Papakostas and A. Potamianos, “A Dataset for Greek Traditional and Folk Music: Lyra”, in *Proc. of the 23rd Int. Society for Music Information Retrieval Conf.,* Bengaluru, India, 2022.

for computational analysis due to their unstructured nature. An exception is the Greek Audio Dataset (GAD) [12], which includes 1000 pieces with audio content (YouTube links), lyrics and annotations for genre (coarse categories including traditional, pop, rock), mood (valence-arousal plane coordinates) and pre-computed audio features; this dataset was later expanded in the Greek Music Dataset (GMD) [13], consisting of 1400 audio pieces with the aforementioned data. However both the GMD and the GAD are not focused on traditional and folk music, while the quality of audio recordings is varying significantly between genres.

The Lyra dataset presented in this paper, in contrast to the GMD, includes mainly traditional and folk Greek music with fine-grained labeling, focusing on musicological aspects of interest. Additionally, the recording quality is homogeneous across all pieces. Musicological soundness and high quality content are ensured by the fact that data has been collected and annotated from a documentary series that was presented by academics in Greek television. Some baseline classification tasks that are of particular interest in this dataset are presented, namely classification of pieces in genres, instruments and places of origin. In all cases, the results show that computational analysis can provide useful insight about musicological relations and phenomena in traditional and folk music of Greece – and, possibly, of other places.

2. DATASET EXTRACTION AND DESCRIPTION

2.1 Challenges and Methods

Large amounts of clean data is fundamental for current AI models to achieve their full potential. In this paper, we walked all the way, from the "data in-the-wild" multimedia content of a TV show to a fully annotated dataset, through a combination of machine automation and human evaluation/annotation processes. The consistency of the dataset and the richness of information it provides are tested by developing and training models that perform three different classification tasks.

In the case of Greek traditional and folk music, there are few cases where metadata is combined with recordings in a structured manner. Additionally, there is a matter of quality of recordings as it is significantly affected by various factors, including the equipment used, the social occasion (e.g., during a festival or inside a studio) and the time period in which it took place, i.e., older recordings tend to be of lower quality.

An integration of dissimilar recordings, in terms of quality, can introduce significant deficiencies towards studying the musicological characteristics of world music with computational tools. In order to truncate the effect of the audio quality factor, we decided to incorporate the episodes from the Greek documentary series "To Alati tis Gis - Salt of the Earth" broadcasted by ERT (Hellenic Broadcasting Corporation), where primarily traditional and folk music is presented. The episodes were filmed during a 10 year period under strict production-level specifications, resulting to very clean and homogeneous audio content while significant wealth of information is provided by the presenter and the guests in the form of narrations between music performances.

The presented dataset consists of both the multimedia content and the annotations of interest. The multimedia content is provided as start and end timestamps that correspond to a single music piece, as parts of a longer episode, which is available online. Regarding the annotations, a taxonomy of labels is defined, based on the potential purposes of studies that might involve this dataset, considering also what metadata information can be retrieved either directly from the source or be integrated by volunteer annotators during the data collection process.

The study of Greek traditional and folk music involves knowledge about (i) the instrumentation, (ii) the genres, (iii) the places of origin and (iv) the way listeners perceive this music in terms of "danceability", among others. While musical instruments, genres and geography are semantically well-defined, the same can not be claimed for listeners' perception. Having at hand the multimedia content, i.e., audio and video, can be helpful to this end. Annotation about whether a music piece is being danced during its live performance can reveal cultural characteristics regarding the way this piece is perceived by the community, because body movements play an important role in music perception [14].

As a result, the taxonomy consists of (i) the musical instruments participating in the performance of each music piece (singing voice is considered an instrument), (ii) the musical genres and sub-genres that are identified by musicologists in Greek music, (iii) the places of origin and (iv) whether the music piece is being danced during its performance.

Volunteer annotators, students of the Department of Music Studies, undertook the task of separating each episode in music pieces and also labeling each one of them according to the specified taxonomy. A helper website was utilized where the respective category labels were added. An account was created for each annotator for the label assignment task. Every piece was labeled by two annotators and the final labels are the set of them where both annotators agree. At the end, the dataset that contains the aforementioned annotations along with the timestamps and the respective video id for each music piece was extracted from the database of the helper site.

2.2 Dataset description

Lyra dataset is organized into a single table where each row corresponds to a music piece while the columns include the various metadata information. Table 1 demonstrates the metadata categories.

Beginning with the simplest metadata categories, in terms of description, "id" is a unique identifier for each piece, generated by its title, replacing Greek with Latin characters and spaces with dashes. As expected, the number of unique values will be the same with the number of pieces, namely 1570. The same stands for "start-ts" and