SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index

ISSN: 2462-1943

Arabic Music Genre Identification

Moataz Ahmed[1,*], Sherif Fadel[1], Manal Helal[2], Abdel Moneim Wahdan[3]

[1] Computer Engineering Department, Arab Academy for Science, Technology, and Maritime Transport, Cairo, Egypt
[2] School of Physics, Engineering & Computer Science, Hertfordshire University, HATFIELD, United Kingdom
[3] Computer and System Department, Ain Shams University, Cairo, Egypt

ARTICLE INFO

Article history:
Received 12 October 2023
Received in revised form 20 December 2023
Accepted 22 February 2024
Available online 22 May 2024

Keywords:
MIR; Genre; Maqam Classification; Machine Learning

ABSTRACT

Music Information Retrieval (MIR) is one data science application crucial for different tasks such as recommendation systems, genre identification, fingerprinting, and novelty assessment. Different Machine Learning techniques are utilised to analyse digital music records, such as clustering, classification, similarity scoring, and identifying various properties for the different tasks. Music is represented digitally using diverse transformations and is clustered and classified successfully for Western Music. However, Eastern Music poses a challenge, and some techniques have achieved success in clustering and classifying Turkish and Persian Music. This research presents an evaluation of machine learning algorithms' performance on pre-labelled Arabic Music with their Arabic genre (Maqam). The study introduced new data representations of the Arabic music dataset and identified the most suitable machine-learning methods and future enhancements.

1. Introduction

In modern times, Music Information Retrieval (MIR) systems have gained significant popularity and are widely used in commercial tools such as query by humming, music similarity and recommendation systems, genre classification, music emotion recognition, music source separation, acoustic descriptions of Music, and Music transcription. These systems rely on various signal processing, machine learning, data mining, music theory, and cognitive psychology techniques to extract relevant information. However, the ultimate goal is to interpret the output mathematical representation meaningfully. This involves modelling the output to understand the information obtained from the signal processing techniques.

Music is information similar to any available data requiring retrieval and analysis to discover knowledge. This study is interested in MIR systems for Arabic Music for various application purposes. The first application of the outcomes of this research is for recommender tools that evaluate users' choices and group them by interest to recommend similar items for them [1]. The second application of the outcomes is for musical copyright, intellectual property protection, and plagiarism scoring [2].

* *Corresponding author.*
E-mail address: Moataz.amahmood@gmail.com

https://doi.org/10.37934/araset.46.1.187200

Identifying talents from the uploads available in the public web domain is another area of automatic MIR systems for all music types [3]. A third application of the outcomes of this research will be enabling the Web 5 objectives by evaluating users' moods and emotions [4]. A fourth application is in psychological music therapy, as Music can create tension or relaxation[4].

Worldwide music patterns refer to the various musical structures, rhythms, and styles found in Music pieces. These include chord progressions, time signatures, and specific instrumentations or arrangements. Some patterns are generic and do not identify the culture, while others are bound to a specific cultural or geographic region and can vary greatly depending on the musical traditions and genres of a particular country or region. The Arab world, centred on the map from North Africa to Central Asia, shares Arabic music characteristics and overlaps with Turkish Music. Arabic Music genres are called Maqams. Maqam is a system of melodic modes used in traditional Arabic Music in religious, secular, poetic, popular and folk music. Each Maqam is built on a unique musical scale between 5 and 9 notes, typically 7, with a specific pattern of intervals and a characteristic melodic phrase that distinguishes it from other maqams. These characteristics include a defined tonic, or "home" note, particular scale and specific rules for how the notes are used and ornamented in melodies, a set of melodic patterns, a hierarchy of pitches, melismatic vocals, tunable instruments to play microtonal intervals such as the quarter tones, typical Arabic percussion instruments, movement patterns, and emotional associations. The tunable instruments commonly used in Arabic Maqams include the oud (a stringed instrument similar to a lute), qanun (a type of zither), ney (a traditional Arabic flute-like instrument), and various percussion instruments emphasising improvisation in composition and performance, such as the use of ornaments and embellishments to provide variations on a melody. Maqams are typically named after the dominant pitch, which is the starting and ending note of the melodic mode. Arabic Music also uses a system of rhythmic modes called "iqa'at," which are based on specific patterns of beats.

In Turkish Music, Maqam is also a system of melodic modes, but it is somewhat different from Arabic maqams. Turkish maqams are different from Arabic by being often based on a collection of notes rather than a specific tonic, and they are characterised by a series of melodic and rhythmic patterns called usul. Turkish maqams are similar to Arabic in being characterised by a complex system of microtonal intervals, which allow for a greater range of expression and nuance in the Music than in Western Music. Going south in Africa, African Music often emphasises rhythm and percussion and may incorporate complex polyrhythms, call-and-response vocal patterns, and the use of percussion instruments such as drums and xylophones.

Going up to Russian Music, overlapping Europe and Asia, the Music tends to be more classical, with a strong emphasis on harmony and melody. Russian Music can employ more complex rhythms than other Western Music using instruments like piano, violin, and balalaika, but are still simpler than Arabic Music and tends to rely more on regular meters and strong downbeats. As we go more western in Europe, then to the USA, Canada and Australia, the Music tends to be simpler, with known genres. Classical Music uses harmonic progressions such as sonata form in symphonies, concertos, and operas using instruments such as the violin, piano, and orchestra. Other simpler Western Music genres include jazz, blues, rock, country, pop and hip-hop, using popular instruments such as the guitar, bass, and drums. Western Music structure is often characterised by melody, harmony, rhythm, and form. The Western diatonic scale using half-step intervals has traditionally used a system of major and minor keys that define the harmony in the melody. The rhythm is based on a system of regular meters, such as 4/4, 3/4, or 6/8. This is called the system of tonality using the chords, major/minor scales, and their progressions. Going down to South American Music, the Music features are still simpler than the Arabic Music, offering a wide variety of rhythms and instruments depending on the region but often includes percussion instruments such as the maracas, the cajón, and the guiro, as well as string instruments like the guitar and the charango.

Going more east, Persian Music uses a system called Dastgah that focuses on melody and tends to use a simpler rhythmic structure than Arabic Maqams. Going to further east, Indian Music uses ragas, a different set of scales, notes, and intervals usually known for its intricate rhythmic patterns emphasising the use of percussion instruments than Arabic Maqams do. Indians also use tunable stringed instruments, such as the sitar and sarod, analogous to the oud and qanun the Arabs use.

Reaching out to the far east, Chinese Music is much simpler than Arabic Music and uses various pentatonic scales that use a system of tuning called intonation, which divides the octave into pure fifths and fourths. Heading towards Japan, Japanese Music is a more complex Music genre than neighbouring countries the Japanese inherit from, but it is still simpler than Arabic Music. Some Japanese Music examples include shōmyō (Buddhist chanting), gagaku (court music), and shakuhachi (bamboo flute). They also use pentatonic scales with subtle dynamics and poetic expression, incorporating natural sounds and environmental elements.

This worldwide musical pattern comparison emphasises the complexity of Cultural Arabic Music. A finer-grained comparison of this variation can deep dive into the different cities and how the geography of the place affects the musical patterns and the cultural messages exchanged with the Music. Similar to the well-developed text-to-image generative models like Dall-E and Stable Diffusion, we can develop generative lyrics-to-music generative models if we define the mathematical characteristic of a given genre or Maqam [5].

1.1 Literature Review

1.1.1 Western music genre classification

In Western Music, the smallest interval between pitches that are adjacent to one another is a half-step (above/below). While in non-western Music, you can find micro-intervals, intervals smaller than a semitone [6]. For Western music classification and clustering, the following research outcomes were identified. A previous study identified the directions used in MIR systems and the challenges expected in the future [7]. This study listed the music features used to describe a music piece and how it affects its processing. Other studies the properties of music genres and proposes a set of features to represent texture, rhythm structure, form and strength to use in the proposed genre identification algorithm and statistical pattern recognition classifie [8].

Various approaches have been employed to classify Western music genres for music identification and classification, such as the Naïve-Bayes approach [8], Decision Trees [9], Support Vector Machines (SVMs)[10], Nearest-Neighbour (NN) classifiers [11], Gaussian Mixture Models [12], Linear Discriminant Analysis (LDA) [13], Hidden Markov Models (HMM) [14] [15], Multi-layer Perceptron Neural Nets [16], and self-organising maps neural networks [17]. Also, combinations of the different algorithms were attempted to classify musical instruments, such as Gaussian Mixture Models and support vector machines [18].

1.1.2 Turkish music genre classification

The Turkish and Persian music genres are closer to Arabic music genres, but the exact genres (Maqams) are different. Turkish and Western Music scales function similarly within the scale and have more or less the same sizes in whole-tones and half-tones. Other musical measures include the Pythagorean comma and Pitches. The Pythagorean comma is a musical term used to describe an interval that measures one-eighth tone. This interval can either increase or decrease the whole-tones and half-tones [19]. Pitches are measured by the number of commas needed to produce quarter-tones, and even three-quarter-tones, as illustrated in Figure 1.

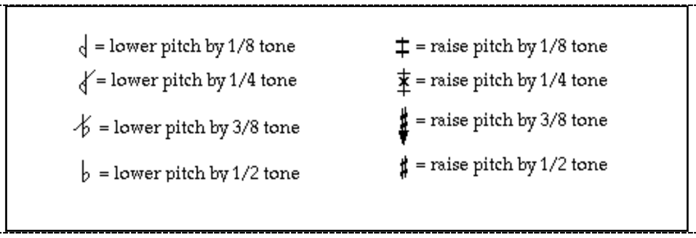


Fig. 1. tones lengths

The Turkish Maqams produce musical pieces for a given scale, and Maqam uses tetrachords and pentachords joined together. The Maqam is defined based on the pitches and the general direction of the melodic flow, creating their own rules of composition not based on scales only [9]. Figure 2 illustrates some Turkish Maqams. Microtonal intervals are considered a distinctive characteristic of maqam music, in which more than 12 unequally spaced intervals form the octave, unlike Western Music, which uses equal temperament. Thus, different representation methods are needed for eastern Music genres, such as the accidental system that can handle microtonal pitch intervals identified [10].

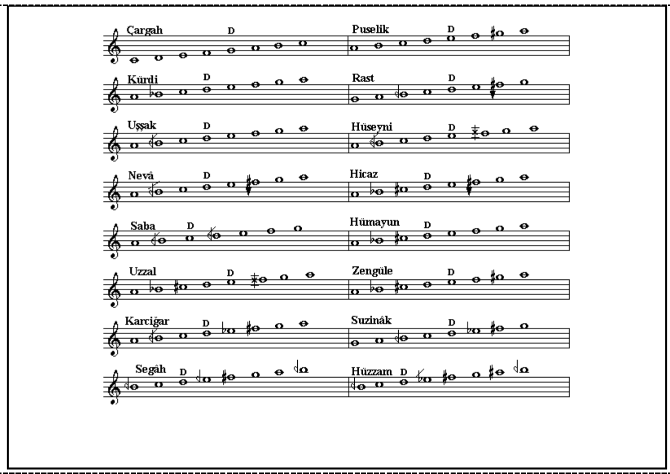


Fig. 2. Turkish Maqams

Another representation system is the Arel–Ezgi (AE) notation (Arel (1991) theory) [10], in which the octave is formed of 24 main notes. Figure 3 illustrates Arel–Ezgi (AE) accidentals and the interval sizes it employs on a full octave in comparison to the 12-TET equally tempered notation. Since Pythagorean tuning defines Arel's notation as being very close to the 12-TET tone and equally spaced [11], thus, it can represent the Arabic Music Maqams because of the employment of quarter-tone in various maqams.

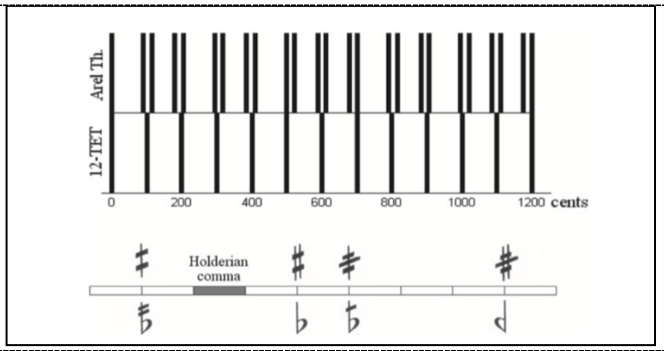
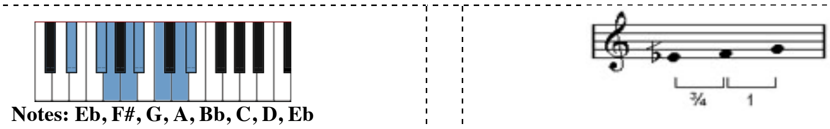


Fig. 3. 24 pitches of the Arel theory with respect to the 12-TET tones (shown in one octave), on the bottom, the accidentals (showing interval sizes within a whole tone) used to represent these pitches on the staff

1.1.3 Arabic music

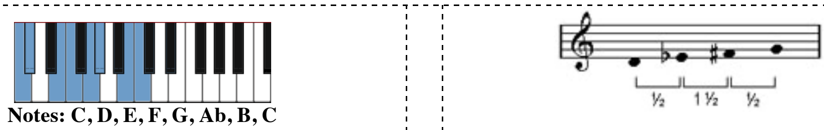
The composition of Traditional Arabic Music is based on inherited rules for ordering and distancing notes. In Western classical Music, Major and Minor modes are the closest equivalent measures. The Arabic Music octave comprises more than 24 bins, which are a combination of whole-tones, half-tones, and quarter-tones. Arabic and Turkish maqams are defined based on the employed pitches and melodic direction. Some maqams share similarities in both cultures but have different names, such as Arabic Bayyati, which is similar to the Turkish Huseyni and Ussak (Bayati). Similarly, Nahawand is similar to the Turkish Puselik and the Western minor scale, while Ajam Ashiran is similar to the Turkish Cargah and the Western major scale. The Arabic Hijaz and Turkish Hicaz are comparable to the Kurd and Kurdi [9].

There are approximately 300 different maqams, which are fundamentally a combination of around 30 basic maqams. This study focused only on nine maqams: Hicaz, Huseyni, Rast, Nihavend, Kurdili Hicazkar, Saba, Segah, Huzzam, and Ussak. Although these maqams share many characteristics, they also have several differences. The primary factor distinguishing one Maqam from another is the distance between tones [12]. As illustrated in Figure 4 and Figure 5 , scales define the different intervals. ¾ is the notation for the quarter tones intervals. Other examples include the Eb (Maqam Mustar), in which the intervals are ¾ - 1- ¾ - ¾ - 1 - 1 - ¾, as illustrated in Figure 4.



(a) (b)

Fig. 4. Eb (Maqam Mustar) (a) Mustar Maqam notes (b) Mustar(Sikah) trichord, starting on E



(a) (b)

Fig. 5. C (Hijaz Kar Maqam) (a) Hijaz Maqam notes. (b) Hijaz tetrachord, starting on D

Various methods have been employed aiming at increasing the efficiency of maqam identification. The chroma features (Harmonic pitch class profiles (HPCP) [13]) were extracted as the main features representing music pieces in previous studies [18]. A template matching method was applied against the ground truth samples, and trained classifiers were also used to compare the results. Other work focused on the tonal features of the music piece and extracting the corresponding pitch class distribution [14]. They applied data mining techniques to conduct a comparative analysis of Music audio recordings.

The Arel theory was employed to classify Turkish Music into maqams using a feature set of pitch class distributions and pitch frequency histograms [15]. While other studies applied a hierarchical framework using new maqam features such as n-gram encoding [16]. An initial decision is estimated, and as the melody progresses, a final decision is developed by template matching from knowledge-based statistical rules such as a Gaussian Mixture Model and the Maximum Likelihood Rule approach with a modified Mahalanobis distance measurement.

Others applied the concept of counting the occurrences of each fundamental repeated frequency from the music signal and extracted the scale from the pitch class histogram [17]. While other studies applied the approach of maqam detection by analyzing the pitch histogram of Maqam of only Turkish Music [20]. Alternatively, other studies computed the Maqam based on the Pitch curve [19]. This is not considered a highly accurate detection method. Although rhythm contributes to the maqam estimation process, it is considered an add-on and out of the scope of maqam detection. This paper did not compute the Tonic intervals. They are estimated from pitch curve analysis.

Moreover, other work presented the concept of real-time maqam estimation and concentrated on music pieces generated from the Ney instrument (end-blown flute) only [21]. This concept assumes that the musician will create the music piece in only one Maqam, which may need at least one minute to detect. In addition to using pre-labelled data (knowing the Maqam in advance), the author considered that the first chord, “Gens”, in the Maqam describes the whole Maqam, which is not necessary to be valid in all music pieces.

Other work used a Turkish dataset considering the selection of polyphonic instrumental songs only in order to avoid the singer’s gender effect or any other variations that may cause any noise [22]. MFCCs and Delta-MFCCs were the features extracted. The study experimented with multiple classifiers, including probabilistic neural networks (PNN) and deep belief networks (DBN). Several approaches applied similar concepts. However, very few focused on Arabic Music as a specific kind of Music and with small datasets.

Generally, detecting the Maqam requires studying several aspects. These aspects include the specification of the scale as an ordered list of intervals that can be specified in a logarithmic scale and the overall melodic progression. Although the composer enjoys some level of freedom in transitioning from one flavour to another, general rules exist for each Maqam on which transitions to other flavours can be done. Inappropriate transitions may deteriorate the maqam harmony. In the coloured notes in Figure 6, green encodes the leading note, red encodes the tonic and blue encodes the dominant. The first scale is composed of a pentachord followed by a tetrachord, and the second is composed of a tetrachord followed by a pentachord. While most music theoreticians would specify the first note of the second n-chord (marked as blue) as the dominant, the function of a dominant as a recently introduced Western term indicating an emphasis note is open to discussion. Other studies detect the ending note to discriminate the maqams in the music piece [23]. Some music pieces may contain a note that does not exist in the whole Maqam, which must be detected and considered an error.

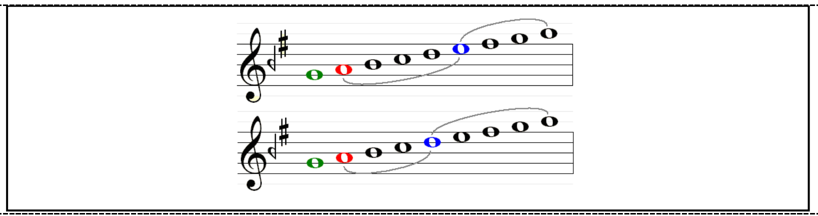


Fig. 6. Leading (top) and dominant (bottom) notes of Maqam

2. Methodology

2.1 Dataset

Arabic Music Maqams differ from the Western music genres in various ways. Various research groups have qualified and quantified the different characteristics as follows. First, we will identify Eastern music types and how different they are from Arabic Music. We collected a dataset of 97 music pieces of Arabic (primarily Egyptian) songs belonging to 7 maqams (Arabic Genres). The labelling of the maqams is done by an expert music teacher, which is traditionally done by ear among educators and talented artists. The dataset contains 7 songs in the bayati maqam c major, 4 songs in the bayati maqam g major, 3 songs in the ajam c major, 3 songs in the ajam f major, 16 songs in the Nahawand maqam c major, 6 songs in the rast maqam c major, 7 songs in the rast maqam B#, and 7 maqams with its major basic scale are used as fingerprints to calibrate the songs to.

2.2 Feature Extraction

Music data are described at the time level and the frequency level. The time level described the rhythm. The frequency level is described as the pitch feature, which is the level from high to low tone, and the orthogonal timbre feature, which is the shape of the music sound. The following paragraph explains the feature space used to represent the audio musical pieces, the transformation performed on them, and its effect on modelling the musical content.

Short Time-Frequency Transform (STFT) is used for the time-frequency analysis of audio signals. It decomposes the shorter windows of longer time samples into the set of active frequency bins active in every time frame. The complex values represent the phase offset of the basic sinusoids in each frequency, and the absolute value represents the energy present at each frequency in each time sample [24].

Chroma vectors are computed using the fixed window STFT and variable window Constant Q Transform (CQT) with binning strategies applied on the amplitude/phase against log frequency. Originally, they represented a 12-bin representing the pitch classes, encoding harmony and suppressing perturbations in octave heights, loudness and timbre. Chroma vectors are good representations for equally-tempered scale music. The CQT Chroma extracts the time-varying frequency spectrum. We created a 36-bin Chroma vector to accommodate the unequal-tempered Arabic Music [24].

Mel-frequency Cepstral Coefficients (MFCC) represent an equally-spaced Mel Scale representation. It is considered the closest approximation to the human auditory system. It is computed by the log of the power of the STFT converted to the Mel scale and then taking the Discrete Cosine Transform (DCT) amplitudes as the MFCCs [25]. Harmonic Pitch Class Profile (HPCP) is an extension of the PCP. It is an approach for Chroma feature estimation representing the pitch content of polyphonic music signals mapped to a single octave [13]. The explanation of each feature used in this experiment is illustrated in Table 1.

Table 1
Feature Extraction Methods Applied

Feature	Explanation
Chroma STFT	Chromagram from power spectrum
Chroma CQT	Constant-Q chromagram
MFCC	Mel-frequency cepstral coefficients
HPCP	Harmonic Pitch Class Profiles
Chroma Cens	Chroma Energy Normalized[26]

The classifiers applied in this study are listed with their parameters in Table 2. They are SVC: support vector classifier; LR: logistic regression; LDA: linear discriminant analysis; DT: decision tree; KNN: K-nearest neighbour; NB: Naive Bayes; SGD: stochastic gradient descent; and MLP: multi-layer perceptron. The dataset used in this study was pre-processed by removing the noise by extracting the harmonic elements from the old records and rescaling it using MinMax Scaler [27].

Table 2
Parameters of the classifiers

Classifier	Parameter
SVC	C=10
LR	C=10
LDA	n_components=7
DT	min_samples_split=2, random_state=none
KNN	n_neighbors=5
NB	var_smoothing=1e-09
SGD	alpha=0.0001, l1_ratio=0.15
MLP	alpha=0.0001,learning_rate='constant', learning_rate_init=0.001

3. Results and Discussion

3.1 Clustering Methods and Results

We applied several clustering techniques as follows. K-means, GaussianMixture, Spectral Clustering, MiniBatchKMeans, MeanShift and DBSCAN on the features selected, we also used a combination of some features. We used the metrics in Table 3 to measure our clustering algorithms used with various feature selections described in Table 1.

We observe that the best overall clustering performance is obtained with the K-means clustering algorithm using the Chroma CQT feature in Table 4. The results obtained in Table 5 reported the second-best clustering-performing algorithm while a combination of Chroma CQT, Chroma STFT, and HPCP was used. The K-means algorithm showed performance up to 54% in terms of homogeneity of the songs belonging to a single class, while the v_measure score showed 53%, which describes the percentage of music pieces in the same cluster.

Table 3
Clustering metrics

Metric	Explanation
homogeneity	A clustering algorithm satisfies homogeneity if all the clusters it forms contain only data points that belong to a single class [28].
completeness	Completeness is achieved by a clustering algorithm when all the data points that belong to a specific class are part of the same cluster [28].
v_measure	The harmonic mean between homogeneity and completeness is a measure used to evaluate the overall performance of a clustering algorithm in terms of both homogeneity and completeness [28]
adjusted_rand	By examining all pairs of samples and keeping track of the pairs that are assigned to the same or different clusters in both the predicted and true clusterings, the Rand Index calculates a measure of similarity between the two clusterings [29].
adjusted_mutual_info (AMI)	The Adjusted Mutual Information (AMI) score is a modified version of the Mutual Information (MI) score that takes chance into account. It compensates for the fact that the MI score tends to be higher when there are more members in two clusterings, even if they don't share more information [30].

Table 4
Results of the clustering in terms of homogeneity (H), completeness (C), v_measure (V), adjusted_rand(R), adjusted_mutual information (M) for Chroma CQT

	(H)	(C)	(V)	(R)	(M)
KMeans+	0.50	0.50	0.50	0.24	0.27
kmeansnorm	0.54	0.53	0.53	0.26	0.31
kmeansrandom	0.54	0.51	0.52	0.24	0.29
AffinityPropagation	0.00	1.00	0.00	0.00	0.00
MeanShift	1.00	0.51	0.68	0.00	0.00
MiniBatchKMeans	0.53	0.54	0.54	0.26	0.33
SpectralClustering	0.45	0.41	0.43	0.11	0.14
DBSCAN	0.00	1.00	0.00	0.00	0.00
Birch	0.43	0.42	0.42	0.13	0.15
GaussianMixture	0.47	0.61	0.53	0.23	0.33

Table 5
Results of the clustering in terms of homogeneity (H), completeness (C), v_measure (V), adjusted_rand(R), adjusted_mutual information (M) for a Combination of Chroma CQT, Chroma STFT, and HPCP using horizontal concatenation

	(H)	(C)	(V)	(R)	(M)
KMeans+	0.43	0.48	0.45	0.20	0.21
kmeansnorm	0.44	0.46	0.45	0.17	0.22
kmeansrandom	0.45	0.45	0.45	0.16	0.21
AffinityPropagation	0.00	1.00	0.00	0.00	0.00
MeanShift	1.00	0.51	0.68	0.00	0.00
MiniBatchKMeans	0.39	0.41	0.40	0.10	0.14
SpectralClustering	0.52	0.47	0.50	0.19	0.23
DBSCAN	0.00	1.00	0.00	0.00	0.00
Birch	0.33	0.39	0.35	0.11	0.10
GaussianMixture	0.46	0.54	0.50	0.25	0.29

We found also that highly distinguishable maqams are Bayati_C, and Nahawand_G. The maqams that are very much closer together are agam_c, agam_f. This is illustrated in the PCA plot of the clusters generated by the best-performing clustering method in Figure 7.

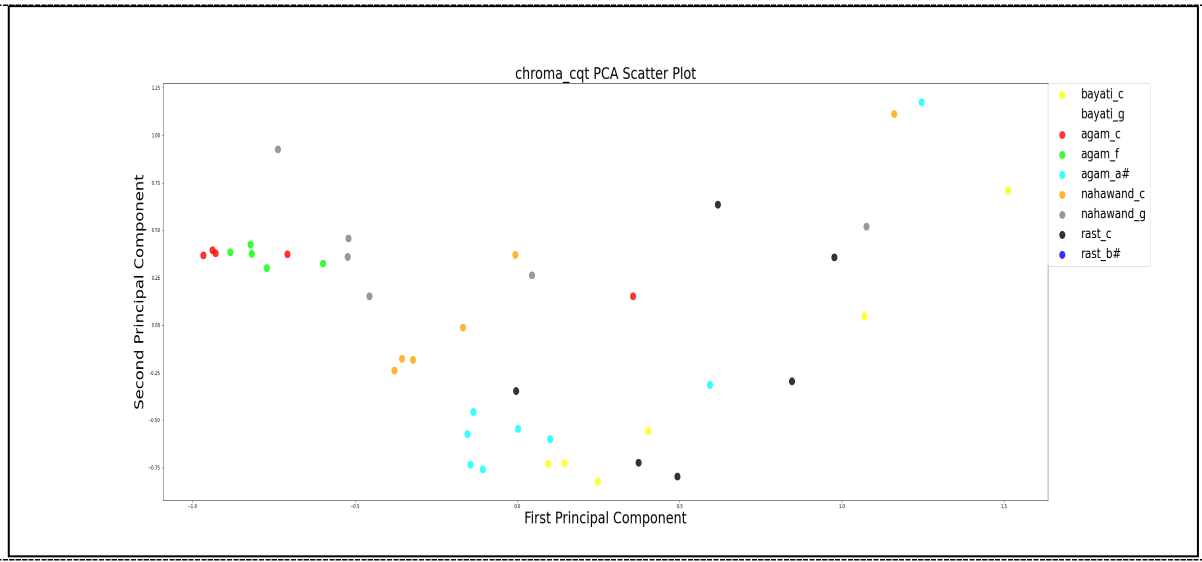


Fig. 7. PCA clustering for chroma_cqt feature

3.2 Classifications, Methods and Results

The same features were used to train multiple classifiers to find the best-performing method. We performed multiple experiments in multiple feature combinations and used the metrics described in Table 6 to measure the performance of the classifiers. Higher performance metrics are considered more efficient than lower values.

Table 6
Classification metrics

Metric	Explanation
f1_macro	Calculate the metric separately for each class and then compute the average, thereby treating all classes equally [31].
f1_micro	Combine the individual contributions of all classes to calculate the average metric [31].
f1_weighted	Compute the metrics for each label and determine their weighted average based on the support, which refers to the number of true instances for each label [31].
pres_macro	Compute the classifier's ability to correctly avoid labeling a sample as positive when it is actually negative for each label and determine their mean without considering label imbalance [32].
pres_micro	This refers to the calculation of specificity, which measures the ability of a classifier to correctly identify negative samples. It is computed by counting the total number of true negatives (samples correctly classified as negative), false positives (negative samples incorrectly classified as positive), and false negatives (positive samples incorrectly classified as negative) [32].
pres_weighted	Compute the metrics individually for each label and then average them by weighting each label's contribution based on its number of true instances. This modified 'macro' approach considers the imbalance in the labels, but the resulting F-score may not necessarily lie between precision and recall [32].
rec_macro	Involves calculating the classifier's capacity to identify all positive samples for each label and taking their unweighted average, without considering label imbalance [32].
rec_micro	This refers to the calculation of the classifier's overall performance in identifying all positive samples, regardless of the label. This is done by counting the total number of true positives, false negatives, and false positives [32].

rec_weighted	Compute the classifier's ability to correctly identify all positive samples and determine their average performance [32].

Although various classification techniques were tested, we finally report only the results obtained with MLP (Multi-layer Perceptron Classifier), as they yielded the best performance with a score of F-measure is 0.68 in Table 7.

Table 7
Results of the classification of Chroma CQT

	f1 macro	f1 micro	f1 weight	Pres macro	Pres micro	Pres weight	Rec macro	Rec micro	Rec weight
SVC	0.63	0.63	0.63	0.73	0.63	0.69	0.61	0.63	0.63
LR	0.58	0.59	0.59	0.69	0.59	0.69	0.58	0.59	0.59
LDA	0.46	0.46	0.47	0.56	0.46	0.63	0.49	0.46	0.46
DT	0.37	0.39	0.34	0.38	0.39	0.37	0.41	0.39	0.39
KNN	0.60	0.61	0.62	0.76	0.61	0.76	0.57	0.61	0.61
NB	0.26	0.28	0.27	0.49	0.28	0.51	0.27	0.28	0.28
SGD	0.39	0.46	0.41	0.45	0.46	0.44	0.44	0.46	0.46
MLP	0.68	0.67	0.68	0.80	0.67	0.78	0.68	0.67	0.67

Additionally, we did another test with a combination of various features using horizontal concatenation; we found the best performance was 0.61 as an F-measure using the SVC classifier, as illustrated in Table 8.

Table 8
Results of the classification for a combination of Chroma CQT, Chroma STFT, and HPCP using horizontal concatenation

	f1 macro	f1 micro	f1 weight	Pres macro	Pres micro	Pres weight	Rec macro	Rec micro	Rec weight
SVC	0.61	0.63	0.62	0.78	0.63	0.71	0.57	0.63	0.63
LR	0.53	0.54	0.54	0.67	0.54	0.64	0.52	0.54	0.54
LDA	0.53	0.54	0.56	0.62	0.54	0.65	0.53	0.54	0.54
DT	0.47	0.46	0.46	0.62	0.46	0.66	0.52	0.46	0.46
KNN	0.45	0.46	0.44	0.58	0.46	0.56	0.45	0.46	0.46
NB	0.20	0.22	0.15	0.40	0.22	0.27	0.23	0.22	0.22
SGD	0.52	0.52	0.55	0.80	0.52	0.72	0.46	0.52	0.52
MLP	0.50	0.52	0.51	0.70	0.52	0.62	0.48	0.52	0.52

Deep neural networks require a massive amount of data to produce efficient results. We trained a five layers Convoluted Neural Network (CNN), producing an average F1 score of 0.30, as shown in Table 9. Also, we trained four Recurrent Neural Networks (RNN) layers, resulting in an F1 score of 0.29, as shown in Table 10. We refined the performance by changing the architecture to four CNN layers, followed by two Recurrent Neural Networks (RNN) layers, to have an AUC score of 0.778.

Table 8
Results of the classification for Mel Spectogram and CQT for CNN

	f1 macro	f1 micro	f1 weighted	Pres macro	Pres micro	Pres weighted	Rec macro	Rec micro
Mel	0.02	0.07	0.01	0.01	0.07	0.00	0.13	0.07
CQT	0.25	0.30	0.22	0.28	0.30	0.24	0.35	0.30

Table 9
Results of the classification for Mel Spectogram and CQT for RNN

	f1 macro	f1 micro	f1 weighted	Pres macro	Pres micro	Pres weighted	Rec macro	Rec micro
Mel	0.14	0.17	0.13	0.15	0.17	0.16	0.22	0.17
CQT	0.29	0.28	0.24	0.37	0.28	0.27	0.31	0.28

4. Conclusions

This work compared various feature representations of digital Music to identify the most suitable transformation to describe Arabic Maqam for clustering and classification purposes. We conclude that the Chroma CQT is more efficient in representing the Arabic Music pieces for Maqam identification purposes. This is because Chroma CQT accounts for the unequal-tempered and as many frequency bins as required in Arabic Music. The most accurate clustering method is K-Means, while GaussianMixture and MiniBatchKMeans perform well. The most accurate classification method is MLClassifier, while SVC Classifier performed well. Using CNN and RNN did not perform as well as SVC because all neural network methods require much larger datasets than what was available in this study. Using mixed deep neural network RNN with CNN performed a higher AUC of 0.778.

Further research is required in the pre-processing of the dataset. To accurately evaluate the performance of neural network approaches, it is required to add more pre-labelled datasets. Visual representation of the maqams using the extracted features can add more insights into the definition.

Acknowledgement
This research was not funded by any grant

References

[1] Song, Yading, Simon Dixon, and Marcus Pearce. "A survey of music recommendation systems and future perspectives." In *9th international symposium on computer music modeling and retrieval*, vol. 4, pp. 395-410. 2012.

[2] De, Soham, Indradyumna Roy, Tarunima Prabhakar, Kriti Suneja, Sourish Chaudhuri, Rita Singh, and Bhiksha Raj. "Plagiarism Detection in Polyphonic Music using Monaural Signal Separation." In *INTERSPEECH*, pp. 1744-1747. 2012. https://doi.org/10.21437/Interspeech.2012-476

[3] Nichols, Eric, Charles DuHadway, Hrishikesh Aradhye, and Richard F. Lyon. "Automatically discovering talented musicians with acoustic analysis of youtube videos." In *2012 IEEE 12th International Conference on Data Mining*, pp. 559-565. IEEE, 2012. https://doi.org/10.1109/ICDM.2012.83

[4] Wiering, Frans. "Can humans benefit from music information retrieval?." In *Adaptive Multimedia Retrieval: User, Context, and Feedback: 4th International Workshop, AMR 2006, Geneva, Switzerland, July 27-28, 2006, Revised Selected Papers 4*, pp. 82-94. Springer Berlin Heidelberg, 2007. https://doi.org/10.1007/978-3-540-71545-0_7

[5] Yu, Yi, Abhishek Srivastava, and Simon Canales. "Conditional lstm-gan for melody generation from lyrics." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 17, no. 1 (2021): 1-20. https://doi.org/10.1145/3424116

[6] Farmer, Henry George. "Historical facts for the Arabian musical influence." *(No Title)* (1930).

[7] Casey, Michael A., Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney. "Content-based music information retrieval: Current directions and future challenges." *Proceedings of the IEEE* 96, no. 4 (2008): 668-696. https://doi.org/10.1109/JPROC.2008.916370

[8] Tzanetakis, George, and Perry Cook. "Musical genre classification of audio signals." *IEEE Transactions on speech and audio processing* 10, no. 5 (2002): 293-302. https://doi.org/10.1109/TSA.2002.800560
[9] Edward J Hines. "What Are Makams? Theory Joining Turkish, Arabic and Western Modes." http://www.hinesmusic.com/What_Are_Makams.html
[10] Gedik, Ali C., and Barış Bozkurt. "Evaluation of the makam scale theory of Arel for music information retrieval on traditional Turkish art music." *Journal of New Music Research* 38, no. 2 (2009): 103-116. https://doi.org/10.1080/09298210903171152
[11] Beken, Münir, and Karl Signell. "Confirming, delaying, and deceptive elements in Turkish improvisations." *Maqam traditions of Turkic Peoples. Berlin: Trafo* (2006).
[12] Farraj, Johnny, and Sami Abu Shumays. *Inside Arabic Music: Arabic maqam performance and theory in the 20th century*. Oxford University Press, 2019.
[13] Gómez, Emilia. "Tonal description of music audio signals." *Department of Information and Communication Technologies* (2006).
[14] Gómez, Emilia, and Perfecto Herrera. "Comparative analysis of music recordings from western and non-western traditions by automatic tonal feature extraction." (2008). https://doi.org/10.18061/1811/34105
[15] Gedik, Ali C., and Barış Bozkurt. "Pitch-frequency histogram-based music information retrieval for Turkish music." *Signal Processing* 90, no. 4 (2010): 1049-1063. https://doi.org/10.1016/j.sigpro.2009.06.017
[16] Ünal, Erdem, Bariş Bozkurt, and M. Kemal Karaosmanoğlu. "A hierarchical approach to makam classification of Turkish makam music, using symbolic data." *Journal of New Music Research* 43, no. 1 (2014): 132-146. https://doi.org/10.1080/09298215.2013.870211
[17] Six, Joren, and Olmo Cornelis. "Tarsos: a platform to explore pitch scales in non-western and western music." In *12th International society for music information retrieval conference (ISMIR-2011)*, pp. 169-174. University of Miami, Frost School of Music, 2011.
[18] Ioannidis, Leonidas, Emilia Gómez, and Perfecto Herrera. "Tonal-based retrieval of Arabic and Middle-East music by automatic makam description." In *2011 9th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 31-36. IEEE, 2011. https://doi.org/10.1109/CBMI.2011.5972516
[19] Lartillot, Olivier. "Computational analysis of Maqam music: from audio transcription to structural and stylistic analyses, everything is tightly intertwined." *The Journal of the Acoustical Society of America* 131, no. 4_Supplement (2012): 3329-3329. https://doi.org/10.1121/1.4708458
[20] Gómez, Emilia, and Perfecto Herrera. "Estimating The Tonality Of Polyphonic Audio Files: Cognitive Versus Machine Learning Modelling Strategies." In *ISMIR*. 2004.
[21] Al-Ghawanmeh, Fadi M., Mohammad T. Al-Ghawanmeh, and Mohammad W. Abed. "Real-Time Maqam Estimation Model in Max/MSP Configured for the Nāy." *International Journal of Communications, Network and System Sciences* 9, no. 2 (2016): 39-52. https://doi.org/10.4236/ijcns.2016.92004
[22] Kızrak, Merve Ayyüce, and Bülent Bolat. "A musical information retrieval system for Classical Turkish Music makams." *Simulation* 93, no. 9 (2017): 749-757. https://doi.org/10.1177/0037549717708615
[23] Bozkurt, Baris. "Features for analysis of Makam music." In *Serra X, Rao P, Murthy H, Bozkurt B, editors. Proceedings of the 2nd CompMusic Workshop; 2012 Jul 12-13; Istanbul, Turkey. Barcelona: Universitat Pompeu Fabra; 2012. p. 61-65*. Universitat Pompeu Fabra, 2012.
[24] McFee, Brian, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. "librosa: Audio and music signal analysis in python." In *SciPy*, pp. 18-24. 2015. https://doi.org/10.25080/Majora-7b98e3ed-003
[25] Sahidullah, Md, and Goutam Saha. "Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition." *Speech communication* 54, no. 4 (2012): 543-565. https://doi.org/10.1016/j.specom.2011.11.004
[26] Müller, Meinard, Frank Kurth, and Michael Clausen. "Audio Matching via Chroma-Based Statistical Features." In *ISMIR*, vol. 2005, p. 6. 2005.
[27] Salamon, Justin, and Emilia Gómez. "Melody extraction from polyphonic music signals using pitch contour characteristics." *IEEE transactions on audio, speech, and language processing* 20, no. 6 (2012): 1759-1770. https://doi.org/10.1109/TASL.2012.2188515
[28] Rosenberg, Andrew, and Julia Hirschberg. "V-measure: A conditional entropy-based external cluster evaluation measure." In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*, pp. 410-420. 2007.
[29] Hubert, Lawrence, and Phipps Arabie. "Comparing partitions." *Journal of classification* 2 (1985): 193-218. https://doi.org/10.1007/BF01908075

[30] Vinh, Nguyen Xuan, Julien Epps, and James Bailey. "Information theoretic measures for clusterings comparison: is a correction for chance necessary?." In *Proceedings of the 26th annual international conference on machine learning*, pp. 1073-1080. 2009. https://doi.org/10.1145/1553374.1553511
[31] Sasaki, Yutaka. "The truth of the F-measure." *Teach tutor mater* 1, no. 5 (2007): 1-5.
[32] Powers, David MW. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." *arXiv preprint arXiv:2010.16061* (2020).