From Sound to Map: Predicting Geographic Origin in Traditional Music Works*

Daniel Kostrzewa[0000−0003−2781−3709] and Paweł Grabczyński

Department of Applied Informatics,
Silesian University of Technology, Gliwice, Poland
daniel.kostrzewa@polsl.pl

Abstract. Music is a ubiquitous phenomenon. In today's world, no one can imagine life without its presence, and no one questions its significance in human life. This is not a new phenomenon but has been prevalent for hundreds of years. Therefore, an automated approach to understanding music plays a nontrivial role in science. One of the many tasks in Music Information Retrieval is the categorization of musical compositions. In this paper, the authors address the rarely explored topic of classifying traditional musical compositions from different cultures into regions (continents), subregions, and countries. A newly created dataset is presented, along with preliminary classification results using well-known classifiers. The presented work marks the beginning of a long and fascinating scientific journey.

Keywords: Traditional Music · Music Information Retrieval · Machine Learning · Classification.

1 Introduction

Living on Earth, we are surrounded by the experience of "multiple worlds". We perceive nuances in songs, dances, instruments, and languages, all indicative of the complexity and richness that envelops us. Diversity manifests itself, among other aspects, in the traditional music of virtually all ethnic groups inhabiting the Earth.

In the public domain, terms such as traditional music, ethnic music, and folk music are often used interchangeably. To understand the topic addressed in this work, it is essential to introduce some distinctions. Traditional and ethnic music refers to the music inherent to a specific ethnic group, passed down from generation to generation, and performed on traditional instruments. Folk music, while inspired by ethnic music, is just its stylization [13]. Contemporary instruments unrelated to a particular region are often employed in folk compositions.

It is challenging to generalize the sound of traditional music. The term refers to various musical forms and styles, dependent on culture and region. In some

* This work is partially funded by statutory research funds of Department of Applied Informatics, Silesian University of Technology, Poland.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

cases, even geographically distant areas may share similar musical cultures; for example, certain traditions in the United States and Canada trace their roots back to Great Britain and Ireland [2].

Nowadays, people performing traditional music are often touring artists who share their musical heritage with others [9]. However, the noticeable increase in interest among listeners is counterbalanced by the number of people abandoning their musical traditions in favor of imitating other styles. Urbanization serves as the primary cause of this phenomenon. Even if individuals migrating from rural to urban areas initially maintain their cultural identity, ultimately, Western lifestyle may displace native traditions [1].

Each ethnic group has developed its distinctive musical style, reflecting diversity in the use of specific instruments, musical scales, or the application of unconventional rhythms. The production and propagation of sound waves are physical phenomena that can be described using numbers and mathematical formulas. In machine learning, numerical features contribute to creating a model capable of capturing subtle differences between the music of different ethnic groups. However, the classifiers can serve not only to recognize the origin of composition but also to analyze songs, offering a chance for a deeper understanding of diversity and the relationships between the various worlds on Earth.

In this context, the main goal and primary contribution of this work is predicting the geographic origin of traditional musical compositions based on sound analysis with the use of classic machine learning methods. The collateral goals are to describe the newly created dataset and perform preliminary results obtained by standard classification methods.

The remainder of the paper is as follows. Section 2 presents the related work and state-of-the-art of the task undertaken in this paper. The dataset prepared for the experiments is thoroughly described in Section 3. Section 4 outlines the methods used for the classification, while the outcomes of the experiments are shown in Section 5. Section 6 concludes the work and describes future work.

2 Related work

The prediction of the geographic origin of musical compositions remains a niche topic, with only two papers presenting a comprehensive approach to the issue [7, 21]. The direct influence on the development of this work was an article by Fang Zhou et al. [21]. The researchers addressed the problem of predicting the geographic origin of musical compositions and utilized machine learning methods for this purpose. They mentioned the use of booklets accompanying CDs for data labeling. It can be assumed that the researchers created a dataset (published on UCI Machine Learning Repository, [3]) based on songs from their own collection. The dataset consists of 1059 examples from 33 countries, with each instance having geographical coordinates corresponding to the capital city of the country from which the composition originates. Unfortunately, determining the recording region is somewhat generalized in cases where clear information is lacking.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

Upon analyzing the data, it was discovered that the classes are not balanced; for instance, traditional music from Belize is represented by 11 examples, while India has 69. Additionally, the class sizes are relatively small. Feature extraction was performed using the MARSYAS program, creating vectors of 68 features for each composition while maintaining the program's default settings. Statistics were based on the entire composition (rather than its parts), and the model's behavior was not verified on feature subsets.

The identification of these issues and the desire to tackle this challenging classification task were the reasons for delving into the topic of predicting the geographic origin of musical compositions.

Access to musical examples is also provided by some online archives. In this manner, Kedyte et al. [7] acquired compositions to create their dataset. Unfortunately, the source of the recordings is no longer available. The researchers focused solely on the area of the United Kingdom and, after preprocessing, obtained 10055 examples. The method of determining the geographical coordinates of points was not discussed in the paper.

The next two papers [6,18] were built upon [21] and [7]. However, topics that combine music and machine learning are gaining popularity. Research includes determining the musical style of a composition [16] or distinguishing "Western" music from that influenced by other cultures [5, 10].

The application of different classifiers, hyperparameter optimization, and result analysis are standard stages in the machine learning process. In the experiment by Zhou et al. [21], the best result was achieved using the Random Forest classifier, while Kedyte et al. [7] built a neural network. Schedl et al. [18] propose a different hybrid approach. Two separate methods were used for prediction, and their combined results yielded higher effectiveness. The first method was based on typical feature analysis extracted from compositions using the KNN classifier. The second method involved data mining for data obtained from the Internet. The highest-rated websites related to queries about the song's name, biography, and origin were retrieved using the Bing Search API. The content of these pages was combined into one document for each composition. Subsequently, based on the list of country names, the frequency of each country's occurrence in the document was determined, and the one with the highest score was selected.

Although the prediction of data geolocation appears in various contexts, such as determining the origin of a photo, the topic of traditional music has not been developed for almost a decade. The lack of adequate datasets and the difficulty in acquiring materials may contribute to the limited interest in this area. Furthermore, no work has been found that frames the issue as a classification problem. Researchers approach the task in the context of regression, attempting to determine the geographical coordinates of a musical composition. The model's effectiveness is measured by the average distance error between points on the sphere (orthodrome). The following results were achieved: 3113 km [21], 1825 km [18], and 114 km for the United Kingdom [7]. Assessing the results of the mentioned models is challenging. In the case of [21] and [18], the average

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

distance between all points in the dataset could serve as a reference point, but such information was not provided in the respective studies.

3 Dataset

The work on this issue can be divided into two stages: data acquisition and feature extraction and the selection of classifiers and their optimization. While several datasets are available online, they mostly focus on the traditional music of individual countries. The dataset prepared as part of the work [21] is likely the only publicly available collection presenting data from multiple regions worldwide. Its drawbacks are small size (1059 compositions), significant class imbalances, and a "mechanical" approach to feature extraction. For this reason, a decision has been made to create a new dataset addressing the described problems.

The process of data acquisition began with gathering information about the traditional music of each country. Analyzing the List of Intangible Cultural Heritage maintained by UNESCO served as a good starting point. The elements on the list represent culturally significant phenomena for each country/region, with music being a crucial part. Another knowledge repository on traditional music is the Naxos Music Library platform [14], enabling the search for albums and recordings from around the world. The platform also provides insights into some booklets accompanying the records. However, the most substantial information was obtained directly from the websites of record labels specializing in traditional music. Among the notable labels are Smithsonian Folkways Recordings [19], Ocora [15], VDE-Gallo Records [20], and Maison Des Cultures Du Monde [12]. Most of the recordings are available on streaming services. Selected tracks were used to create the dataset following the permissible use defined in the Copyright and Related Rights Act.

The collected data (Fig. 1) underwent the following verification: checking the validity of including a recording in the dataset (e.g., distinguishing traditional from folk music) and ensuring adherence to the specified duration limits for each track, ranging from 30 seconds to 15 minutes. Longer soundtracks were omitted to avoid excessive memory load. The resulting database comprises 12,860 recordings from 44 countries worldwide. The highest number of examples were gathered from the Democratic Republic of the Congo and Siberia (both with 334 examples), while the fewest were from Scotland (227 examples). Additionally, each track received a label indicating its region (continent)and subregion of origin. The division into regions and subregions was done according to the United Nations geoscheme. There are more significant differences in class sizes, especially at the regional level.

Despite the dataset being created, it still requires thorough verification from ethnomusicologists. Therefore, it is not yet publicly available. Undoubtedly, work on it will continue, and ethnomusicologists have expressed interest in contributing to the construction of this dataset. This marks the first part of developing a

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

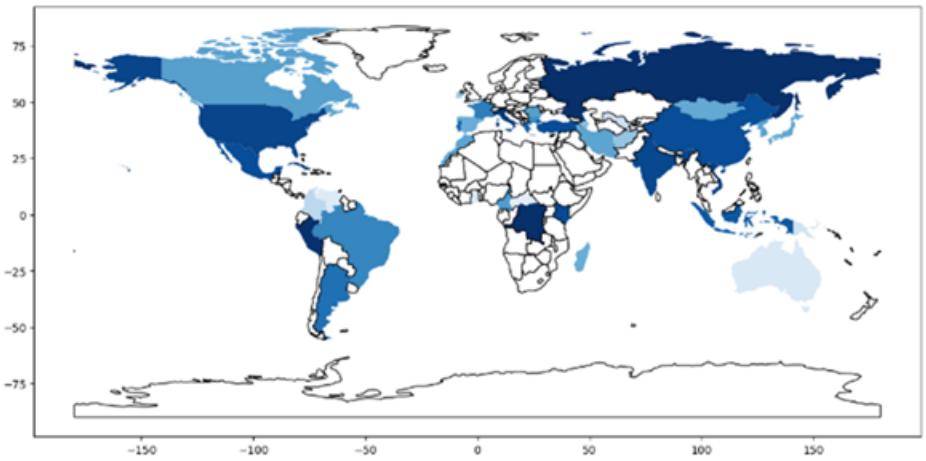

Fig. 1. A map illustrating the content of the dataset created for the purposes of this study (countries); the darker the color is, the more examples are present.

comprehensive dataset. The second equally crucial element is the extraction of numerical features.

Using audio files directly in analytical tasks would be impractical due to their size. Therefore, input data undergo transformations. Values calculated in this way describe characteristic elements of music, such as timbre, melody, and rhythm. Furthermore, feature extraction helps eliminate irrelevant information, improving the efficiency of machine learning algorithms.

When discussing feature extraction, it is essential to consider the duration of examples. Analyzing the entire composition is a rare practice, with only [21] mentioning it. Some papers propose cutting a portion of the recording, e.g., a 10-second [7] or a 30-second snippet [5,16]. Typically, the beginning of the track is analyzed. However, an algorithm can be employed to choose a suitable segment, such as one with the highest spectrum energy (the culmination point of the composition). In this study, the feature extraction process from an audio file began with cutting a 30-second segment of the composition, possessing the highest energy. The assumption was made that the strongest signal is synonymous with the most characteristic excerpt.

A total of 323 features were extracted from each audio recording, which can be categorized into the following groups: timbre (219 features), melody (72 features), and rhythm (25 features). The next two attributes (album and title) serve to identify the track during dataset creation. The last 5 categories are labels: latitude, longitude, region (continent), subregion, and country of origin. The geographical coordinates point to the capital city of the country. Exceptions are the coordinates of Korea and Siberia. Due to data acquisition from both North and South Korea, coordinates of the Korean Demilitarized Zone were used. In the case of Siberia, coordinates indicate the central part of the area.

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

Attributes were calculated using pre-existing functions from the Librosa library. To better understand the data distribution, the following statistics were applied for most features: mean, median, standard deviation, maximum value, minimum value, skewness, and kurtosis. Extracting a large number of features aims to minimize the impact of values that might inaccurately characterize a specific composition.

The dataset employs various methods for signal description in both the time and frequency domains. Features belonging to the first group can be directly derived from the audio file, providing a straightforward way to analyze the signal. To obtain attributes from the frequency domain, it is necessary to transform the audio signal, for example, into a Mel spectrogram.

One of the parameters computed in the time domain is the Zero-Crossing Rate (ZCR). This indicator defines the frequency of changes in the signal value from positive to negative and vice versa. Additionally, ZCR can indicate the noise in the recording—usually, it achieves much higher values if the signal is noisy. The standard deviation [4] is a particularly useful statistic describing ZCR. Different musical instruments (and human voices) produce sound in their own distinctive ways, allowing the prediction of the likely distribution of this value.

Next time domain features are connected with the rhythm. The *librosa.onset_detect* function detects the beginning of the sound, called onsets, by analyzing the envelope of the signal. A sudden increase in amplitude can indicate the occurrence of a sound. The number of detected onsets was used to calculate the frequency of events per second. Another valuable parameter is the onset_intervals. This array contains the time differences between consecutive onsets. This information helps determine if there are rhythmic patterns in the composition or if the rhythm undergoes frequent changes. Kurtosis was used to assess rhythm variability. This statistic provides insights into the frequency of extreme values and the ratio of the value distribution to the normal distribution. In addition to detected onsets, there is the onset_strength function. This is an example of a method for signal description in the frequency domain. The function calculates the spectrum difference between adjacent frames. Information about the strength of successive sounds can be helpful in classifying a recording. Instruments characterized by a sharp sound achieve higher onset_strength values. An example is the djembe, one of the most popular African drums. The sound produced by the djembe depends on the way it is hit and the part of the membrane that is hit. The onset_strength value is usually higher for percussion instruments. In contrast, Japanese court music, gagaku, exhibits a slow tempo and the occurrence of long tones, often performed in unison.

Traditional music is often based on musical scales, which are sequences of sounds arranged according to fixed patterns, defining the distances between consecutive tones. One of the earliest scales confined within an octave was the pentatonic scale, employed in ancient Chinese, Greek, and Peruvian music, known for its smooth sound, free of dissonances. "Western" music has been based on the seven-tone scale (the eighth tone is a repetition of the first, an octave higher) for several centuries. This relationship is reflected in Fig. 2. The sound material

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

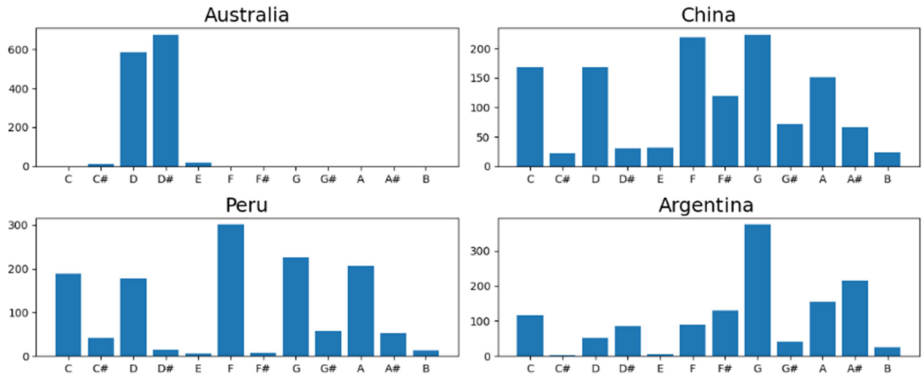

Fig. 2. Frequency of occurrence of individual pitch classes. Sources: Aboriginal Songs of the Northern Territory – Devilman (Australia), Chinese Masterpieces of the Pipa and Qin – Wild Geese Descending (China), Traditional Music of Peru, Vol. 5 – Cusqueño (Peru), Tango de Buenos Aires – Balada para Mi Muerte (Argentina).

of traditional Chinese and Peruvian music is based on the pentatonic scale, as evidenced by the dominance of tones belonging to this scale. Argentine tango is constructed on a seven-note minor scale, where the G tone predominates in this example due to the composition being in the key of G minor. The last example features a solo performed on the didgeridoo. The sound produced by the instrument resembles a buzzing, oscillating between adjacent tones (in this case, D and D#). The obtained frequencies were transformed into tones and normalized to one octave. The sorted array of results was also used to determine the chroma contrast coefficient. This ratio represents the sum of the six most frequently occurring pitch classes divided by the sum of the remaining pitch classes [10]. In principle, "Western" compositions should have a lower chroma contrast value compared to compositions based, for example, on the pentatonic scale, as the seven-tone scale includes more naturally occurring tones. The extreme case is represented by didgeridoo recordings, where typically only 4 pitch classes are detected. Calculating chroma contrast would involve division by 0, leading to the automatic setting of the value to 1000.

In addition to the described features, essential information regarding timbre includes Mel-Frequency Cepstral Coefficients (MFCC, Fig. 3) [11] and harmonics. It is worth discussing these parameters more precisely, constituting as much as 44% of all attributes in the created dataset. MFCC coefficients are obtained by analyzing the Mel spectrogram. They undergo further transformations, including Discrete Cosine Transform (DCT), resulting in a simplified representation of sound. Several processing stages follow the creation of the Mel spectrogram, including pre-emphasis and windowing (differentiation and segmentation), fast Fourier transform (processing the segment into a frequency spectrum and modification according to the Mel scale), Mel-scale filtering (using triangular filters), normalization (subtracting the mean and normalizing the variance of each co-

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

efficient). Using the Mel scale, which concentrates higher frequencies, MFCC is calculated using DCT.

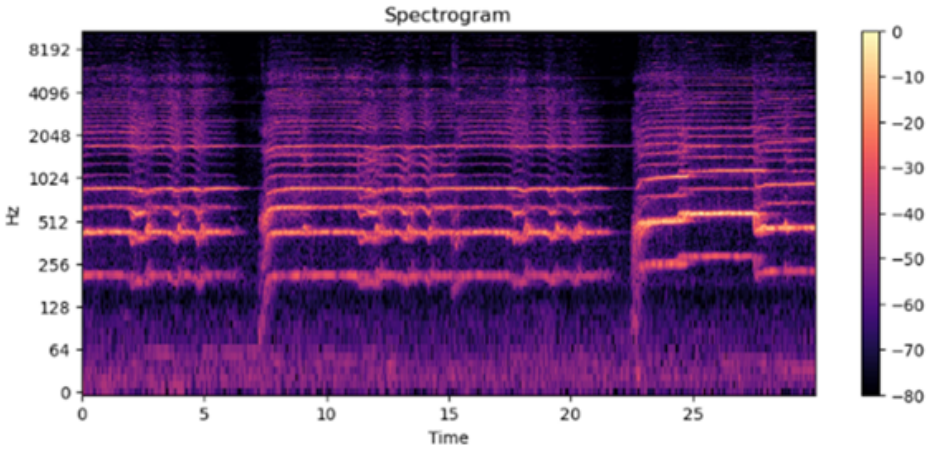

Fig. 3. Spectrogram of the song on the Mel scale. Source: The Gagaku – Kashin.

MFCC parameters play a significant role in identifying musical genres [8]. By extracting characteristic audio features, they enable the differentiation between music styles such as jazz, rock, or classical music. They are also employed in streaming services for automatic music sorting and serve as input data for neural networks used in music genre classification.

4 Methods

The modest size of the feature-extracted CSV file (approximately 55 MB) contributed to the acceptable (in minutes) computation time required for evaluating results across various models. In the pursuit of identifying the optimal classifier, a diverse set of algorithms was systematically tested. While some approaches were discarded based on insights from literature and personal observations, the primary objective was to assess numerous classifiers and compare their performance comprehensively. The following algorithms were employed in the experiment.

Dummy Classifier - serves as a reference point for other classifiers. Predictions are made based on a selected strategy, bypassing the analysis of input data. The parameter *strategy* is set to *uniform*, allowing each class to be chosen with equal probability.

Gaussian Naive Bayes Classifier - a method based on Bayes' theorem. For each class, the probability of the occurrence of individual features is calculated. The class with the highest score is chosen as the model's prediction. Gaussian Naive Bayes assumes that data are independent of each other, meaning the occurrence of one feature has no impact on the presence of another. Additionally,

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

the algorithm assumes that values for each class follow a normal distribution. While this method may be suitable for specific tasks such as text classification or spam filtering, its inability to detect correlations between features limits its effectiveness in complex classification tasks.

Decision Tree Classifier – a classifier employing a decision tree-based approach. The model's structure resembles a tree, with nodes representing individual features, branches denoting decisions made by the algorithm, and leaves presenting the selected classes. Based on successive attributes, the algorithm divides the data into progressively smaller subsets. The order of features chosen for set partitioning is determined by measures such as the Gini index or entropy. The advantage of using a decision tree lies in the ease of interpreting results. The model allows for tree visualization and analysis of the most significant features identified by the algorithm. Moreover, it performs well with both numerical and categorical features. However, decision trees are inherently unstable, a small change in the data can lead to an entirely different tree structure. For complex classification tasks, it might be necessary to utilize advanced algorithms like Random Forest or Gradient Boosting, which are also based on decision trees.

Stochastic Gradient Descent Classifier – the classifier's name does not directly refer to the model but to the training method. The stochastic gradient descent (SGD) method is applied using a linear classifier as an example. SGD continuously updates the model's parameters in the direction where the values of the loss function decrease. In contrast to the Gradient Descent method, where updating a parameter in a given iteration requires processing the entire training set, Stochastic Gradient Descent allows the use of randomly selected samples. The method is efficient for large datasets as it does not necessitate processing all data at once.

Logistic Regression Classifier – despite the term regression in the model's name, it is more commonly used for classification than regression. The algorithm operates by employing the logistic (sigmoidal) function, transforming any value into a range between (0, 1). Each feature carries a weight determining its influence on the prediction. Various optimization methods are used to minimize the difference between expected and predicted classes. Although logistic regression is mainly associated with binary classification, choosing the parameter *multi_class='multinomial'* enables its application to multiclass classification. The ease of interpreting the model's results is a significant advantage of logistic regression (weights of individual features). The method is effective in simple tasks, especially when a linear relationship exists between variables. However, for more complex data, the application of logistic regression may be ineffective.

Gradient Boosting Classifier – a technique involving the sequential training of simple models (such as decision trees) and their optimization through the evaluation of the previous predictor. The residual error, i.e., the difference between the expected and predicted values, is calculated for each decision tree. In successive iterations, the algorithm creates new decision trees that consider the errors made in previous iterations. The direction of changes is determined by the gradient of the loss function. The Gradient Boosting Classifier often achieves

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

very good results in many cases, frequently outperforming classifiers like Random Forest [17]. It handles complex data and outliers well. However, the downsides of this technique include high computational complexity and the risk of overfitting when inappropriate parameters are used.

K-Nearest Neighbors Classifier – the algorithm is based on the principle of similarity among data points located closely in the feature space. The model determines the distance between a given instance and the remaining points in the dataset. Subsequently, a specific number of nearest points (neighbors) are selected, and through voting, the most frequently occurring class is identified. KNN requires feature scaling to prevent large values from dominating computations. The model performs well in tasks with multiple classes and nonlinear decision boundaries. On the other hand, the algorithm is highly sensitive to imbalanced datasets, where the dominance of one or a few classes can impact prediction accuracy.

Random Forest Classifier – a method classified under ensemble learning, which involves combining multiple simple models to obtain a more accurate final model. RF is based on decision trees, and the algorithm creates a "forest" consisting of a specified number of these trees. Each tree is independent and trains on different subsets of data (using bootstrap sampling), allowing for greater diversity in models and preventing overfitting. Then, similar to KNN, the most frequently occurring class selected by individual decision trees is chosen. The algorithm handles outliers effectively. In multiclass classification, RF achieves significantly better results than a single decision tree. The use of a large number of trees helps minimize the negative impact of outliers.

Support Vector Machine Classifier – the classifier's task is to find a suitable hyperplane that separates individual classes. The optimal hyperplane is maximally distant from all classes. Support vectors, points in the feature space closest to the hyperplane, play a crucial role in determining its position. The removal of a specific support vector changes the position of the hyperplane. If the data is not linearly separable, it undergoes transformation into a higher dimension where linear separation is achieved. SVM performs well with data featuring a large number of features. The ability to apply different kernel functions (for transforming data into higher dimensions) allows the model to be used for various purposes. However, interpreting the classifier's operation is generally more challenging than in the case of, for example, logistic regression.

Linear Support Vector Machine Classifier – this classifier is also based on the support vector method and employs a linear kernel function. Unlike SVM, the kernel function cannot be changed. The method is effective for data that can be separated using a line or hyperplane, especially when dealing with a higher number of features. Linear SVC is a faster classifier but struggles more with complex patterns and feature relationships than SVM. It should be assumed that the application of the linear version of the SVM algorithm may be doomed to failure due to the characteristics of the data.

Neural Network – neural networks process data in a manner similar to the human brain, primarily manifested in their structure. The human brain is com-

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

posed of neurons connected by synapses, and communication between neurons occurs through electrical signals controlling synapses. Artificial neural networks feature interconnected neurons as well, and each connection has a weight adjusted during the machine learning process.

For constructing the network, the Keras library was utilized. A simple architecture was implemented, comprising an input layer (128 neurons, ReLU activation function), two hidden layers (128 and 64 neurons, ReLU activation function), and an output layer (the number of neurons dynamically determined based on the number of classes, softmax activation function). Compilation parameters were set as: optimizer - Adam, loss function - sparse categorical crossentropy. To prevent overfitting and enhance the model's performance, an early stopping callback was employed. This callback monitored changes in the validation accuracy metric with a patience parameter set to 10. At the end of each epoch, the model was compared with the best and, if necessary, overwritten.

5 Experiments

At the initial stage of the experiment, a preliminary assessment of classifier results was conducted. The task involved determining the origin of a musical piece, categorized into three levels: region (continent), subregion, and country. Calculations were performed for three categories: region, subregion, and country. All features were utilized in the analysis, having been previously scaled. Standardization was necessary as the value ranges of certain attributes significantly varied, which could adversely affect the performance of classifiers like SVM. The dataset was divided into training and testing sets in a 70:30 ratio.

For evaluating the model's quality, cross-validation was employed. The dataset was divided into 10 subsets. After evaluating each subset, the average and standard deviation of the accuracy metric were computed. Table 1 shows the quantitative results (average accuracy and its standard deviation) for the selected classifiers with the default parameters.

The Dummy Classifier achieved a result close to the probability $P = 1/n$, where n represents the number of classes. The Dummy Classifier was applied as a reference point for more complex algorithms. Gaussian Naive Bayes, Decision Tree, and Stochastic Gradient Descent yielded average results, placing them towards the lower end of the classification effectiveness ranking. Logistic Regression and Linear SVC performed the task with very similar effectiveness. Both classifiers are linear models that seek a hyperplane to separate classes. Gradient Boosting and Random Forest Classifier required the most time to find a solution. Random Forest achieved better results than Gradient Boosting, especially in country classification. However, the Gradient Boosting result is below expectations, and overfitting could be a reason, as this classifier is less resistant to it than Random Forest.

K-Nearest Neighbors and Support Vector Machine Classifier achieved good results, especially SVM (the best result in country classification). The relatively high effectiveness of KNN suggests that data points close to each other in the

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

Table 1. Comparison of the average accuracy and its standard deviation of the classifiers.

Classifier	Region (continent)	Subregion	Country
Dummy Classifier	16.12% (1.10%)	5.38% (0.57%)	2.51% (0.56%)
Gaussian Naive Bayes	36.40% (1.19%)	20.78% (1.23%)	21.97% (1.40%)
Decision Tree	40.24% (1.68%)	24.23% (0.89%)	18.14% (0.54%)
Stochastic Gradient Descent	49.86% (1.77%)	32.64% (1.61%)	28.76% (1.32%)
Logistic Regression	57.23% (1.67%)	40.31% (1.78%)	36.84% (1.10%)
Linear SVC	57.37% (1.27%)	39.76% (1.64%)	36.29% (1.23%)
Gradient Boosting	59.55% (1.44%)	38.21% (1.12%)	33.35% (0.81%)
Random Forest	60.10% (0.81%)	47.38% (1.64%)	49.47% (1.81%)
K-Nearest Neighbors	64.06% (1.47%)	50.29% (1.51%)	45.82% (1.11%)
Support Vector Machine	66.60% (1.11%)	52.93% (1.42%)	**51.39%** (2.05%)
Neural Network	**70.41%** (0.83%)	**54.77%** (1.36%)	49.47% (0.99%)

feature space may belong to the same class. The best-performing model in the experiment was the neural network, which achieved the highest score for region and subregion classification.

At the second stage of the experiment, classifiers with the highest accuracy metric were selected for further optimization: K-Nearest Neighbors, Random Forest, Support Vector Machine, and Neural Network. Suboptimal sets of hyperparameters were identified using grid search. The search included values for the following parameters: KNN (*n_neighbors* - the number of neighbors, *weights* - the weight function, *algorithm* - the algorithm used to compute the nearest neighbors), Random Forest (*n_estimators* - the number of trees in the forest, *max_features* - the number of features considered when finding the best split, *bootstrap* - specifies whether the tree is built from the whole dataset or a sample of data), SVM (C - regularization parameter, *kernel* - the type of kernel used in the algorithm, *gamma* - kernel coefficient), Neural Network (*optimizer* - the algorithm adjusting model parameters during training, *activation* - activation function, *batch_size* - the number of samples processed before updating the model, *neurons* - the number of neurons). Experiments also involved changes in network architecture, such as adding hidden layers or dropouts.

All classifiers improved their results by approximately 3-4% (Table 2). The most significant changes in effectiveness increased by about 10%. Each model presents a different approach to the classification problem. KNN makes decisions based on local similarity in the feature space, Random Forest utilizes decision trees, SVM finds the optimal hyperplane separating classes, and neural networks search for complex patterns using neurons. Despite their differences, all classifiers achieved satisfactory results.

Table 2 shows average accuracy before and after hyperparameter optimization for the selected classifiers as well as a set of found hyperparameter values. The best values are in bold.

The presented experiment allowed for the selection of the best classifiers. In the category of region (continent) and subregion, the SVM algorithm emerged

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

Table 2. Comparison of the average accuracy after hyperparameters optimization.

Classifier	Parameters	Accuracy before opt.	Accuracy after opt.
Region (continent)			
Random Forest	n_estimators = 500, max_features = 40, bootstrap = False	60.10%	66.43%
K-Nearest Neighbors	n_neighbors=4, weights= 'distance', algorithm = 'auto'	64.06%	68.12%
Support Vector Machine	C = 0.1, gamma = 1, kernel = 'poly'	66.60%	**72.42%**
Neural Network	optimizer = 'adam', activation = 'relu', batch_size = 32, neurons = 512, 256, epochs = 50	70.41%	71.02%
Subregion			
Random Forest	n_estimators = 500, max_features = 60, bootstrap = False	47.38%	57.67%
K-Nearest Neighbors	n_neighbors=3, weights= 'distance', algorithm = 'auto'	50.29%	58.04%
Support Vector Machine	C = 0.1, gamma = 1, kernel = 'poly'	52.93%	**60.26%**
Neural Network	optimizer = 'adam', activation = 'relu', batch_size = 32, neurons = 512, 256, epochs = 50	54.77%	60.06%
Country			
Random Forest	n_estimators = 500, max_features = 10, bootstrap = False	49.47%	**60.03%**
K-Nearest Neighbors	n_neighbors=3, weights= 'distance', algorithm = 'auto'	45.82%	54.28%
Support Vector Machine	C = 0.1, gamma = 1, kernel = 'poly'	51.39%	57.62%
Neural Network	optimizer = 'adam', activation = 'relu', batch_size = 32, neurons = 512, 256, epochs = 50	49.47%	55.78%

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

as the winner, achieving average accuracies of 72.42% and 60.26%, respectively. For the task of country prediction, Random Forest performed the best, achieving an average accuracy of 60.03%.

6 Conclusions and future work

The prediction of the geographic origin of musical compositions is an area that is still waiting to be thoroughly explored. This paper presents a comprehensive approach to the problem, from acquiring musical compositions and creating a dataset to applying various classifiers.

The obtained results can be considered satisfactory, especially the effectiveness of predicting the country of origin of a composition at a level of 60% for 44 classes. A holistic approach is needed to characterize the sound, involving musical features such as timbre, melody, rhythm, and ethnomusicological knowledge.

Collecting the necessary data proved to be a tedious task, requiring a thorough examination of each album and track. The effectiveness of classifiers depended on the quality of the acquired data and extracted features. On average, each country is represented by 292 compositions, ten times more than in the dataset of the previous work [21]. In some regions of the world, traditional music is highly diverse, so even the representation of several hundred recordings may be insufficient to capture this diversity. A possible solution is to divide countries into smaller areas, such as separating the northern and southern parts of India.

The created dataset was tested using ten different classifiers. The best results, after hyperparameter optimization, were achieved by K-Nearest Neighbors, Random Forest (best in the country category: 60.03%), Support Vector Machine (best in the region (continent): 72.42%, and subregion category: 60.26%), and Neural Network.

To improve results for region and subregion, it would be necessary to balance the data. For this purpose, new compositions from minority classes could be added to the dataset, or oversampling methods, such as the Synthetic Minority Over-sampling Technique (SMOTE), could be employed.

In further research, extracting information about the most significant features from the mentioned estimators would be valuable. This involves examining the paths created in the decision trees of the Random Forest classifier and identifying the most frequently repeated attributes that lead to correct predictions. Moreover, combining different methods, as demonstrated in [18], could contribute to performance improvement.

The attractiveness of the topic of this study arises from its interdisciplinary nature. Predicting the geographic origin of compositions combines machine learning, ethnomusicology, and geography. Such topics demonstrate that the only limit in tasks posed to machine learning is imagination.

References

1. Elbourne, R.: The study of change in traditional music. Folklore **86**(3-4), 181–189 (1975)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12

2. Frank, A.: That's the way I've always learned: The Transmission of Traditional Music in Higher Education. Ph.D. thesis, East Tennessee State University (2014)
3. Geographical Original of Music dataset website: (2014), https://www.archive.ics.uci.edu/dataset/315/geographical%2Boriginal%2Bof%2Bmusic, last visit 2024-02-10
4. Giannakopoulos, T., Pikrakis, A.: Introduction to audio analysis: a MATLAB® approach. Academic Press (2014)
5. Gómez, E., Haro, M., Herrera, P.: Music and geography: Content description of musical audio from different parts of the world. In: ISMIR. pp. 753–758 (2009)
6. Hamoen, J., Engbers, S.: Explaining the geographic origin of music (2021), www.jellehamoen.nl/wp-content/uploads/2021/07/Research-paper-REDI-Sharon-Engbers-Jelle-Hamoen.pdf
7. Kedyte, V., Panteli, M., Weyde, T., Dixon, S.: Geographical origin prediction of folk music recordings from the united kingdom. 18th International Society for Music Information Retrieval Conference (2017)
8. Kostrzewa, D., Kaminski, P., Brzeski, R.: Music genre classification: looking for the perfect network. In: International Conference on Computational Science. pp. 55–67. Springer (2021)
9. Littlefield, M.D.: Folk music in new england: A living tradition (2020)
10. Liu, Y., Xiang, Q., Wang, Y., Cai, L.: Cultural style based music classification of audio signals. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 57–60. IEEE (2009)
11. Logan, B., et al.: Mel frequency cepstral coefficients for music modeling. In: Ismir. vol. 270, p. 11. Plymouth, MA (2000)
12. Maison Des Cultures Du Monde website: http://www.boutiqueenligne.maisondesculturesdumonde.org/, last visit 2024-02-10
13. Metzig, C., Gould, M., Noronha, R., Abbey, R., Sandler, M., Colijn, C.: Classification of origin with feature selection and network construction for folk tunes. Pattern Recognition Letters **133**, 356–364 (2020)
14. Naxos Music Library World website: www.naxosmusiclibrary.com/world, last visit 2024-02-10
15. Ocora website: https://www.radiofrance.com/les-editions/collections/ocora, last visit 2024-02-10
16. Patil, N.M., Nemade, M.U.: Music genre classification using mfcc, k-nn and svm classifier. International Journal of Computer Engineering In Research Trends **4**(2), 43–47 (2017)
17. Piryonesi, S.M., El-Diraby, T.E.: Data analytics in asset management: Cost-effective prediction of the pavement condition index. Journal of infrastructure systems **26**(1), 04019036 (2020)
18. Schedl, M., Zhou, F.: Fusing web and audio predictors to localize the origin of music pieces for geospatial retrieval. In: Advances in Information Retrieval: 38th European Conference on IR Research, ECIR 2016, Padua, Italy, March 20–23, 2016. Proceedings 38. pp. 322–334. Springer (2016)
19. Smithsonian Folkway Recordings website: https://folkways.si.edu/, last visit 2024-02-10
20. VDE-Gallo Records website: https://vdegallo.com/, last visit 2024-02-10
21. Zhou, F., Claire, Q., King, R.D.: Predicting the geographical origin of music. In: 2014 IEEE International Conference on Data Mining. pp. 1115–1120. IEEE (2014)

ICCS Camera Ready Version 2024
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-63751-3_12