Journal of Computer Science
https://dergipark.org.tr/en/pub/bbd
https://doi.org/10.53070/bbd.945894

Journal of Computer Science
ISSN: 2548-1304
Volume:6 , Issue:3, pp:133-144,2021
@Anatolian Science

Music Emotion Recognition with Machine Learning Based on Audio Features

Mehmet Bilal ER*[1], Emin Murat ESİN[2]

[1]Harran University, Dept. Of. Computer engineering, Şanlıurfa, Turkey

[2]Maltepe University, Computer engineering, Faculty of Engineering and Natural Sciences, İstanbul, Turkey

(bilal.er@harran.edu.tr, muratesin@maltepe.edu.tr)

Received: Jun. 3, 2021 Accepted: Jul. 7, 2021 Published: Dec. 1, 2021

Abstract— Understanding the emotional impact of music on its audience is a common field of study in many disciplines such as science, psychology, musicology and art. In this study, a method based on audio features is proposed to predict the emotion of different samples from Turkish Music. The proposed method consists of 3 steps: preprocessing, feature extraction and classification on selected music pieces. As a first step, the noise in the signals is removed in the pre-process and all the signals in the dataset are brought to the equal sampling frequency. In the second step, a 1x34 size feature vector is extracted from each signal, reflecting the emotional content of the music. The features are normalized before the classifiers are trained. In the last step, the data are classified using Support Vector Machines (SVM), K-Nearest Neighbor (K-NN) and Artificial Neural Network (ANN). Accuracy, precision, sensitivity and F-score are used as classification metrics. The model is tested on a new 4-class dataset consisting of Turkish music data. 79.30% accuracy, 79.31% precision, 79.13 % sensitivity and 79.03% F-score are obtained from the proposed model.

Keywords : *Music emotion recognition, audio feature extraction, SVM, ANN, K-NN*

1. Introduction

Music is a branch of art in which emotions are expressed with sounds (Hevner, 1936). Due to the variety and richness of music content, it has been the subject of research in many studies in different fields. Most of the work done in the field of music signal processing; It covers applications such as classifying musical genre, determining the musical instruments used, and revealing musical similarities such as notes and chords. Recently, there has also been increased interest in studies aimed at automatically analyzing and recognizing the emotional content of a piece of music (Lin, Liu, Hsiung, & Jhang, 2016). Music emotion classification systems; It may be utilized actively for a variety of reasons, including organizing personal music collections and constructing repertoire recommendation systems, music therapy, and emotional illness treatment. In this study, it is aimed to classify the emotional content of music with the approach presented by using the data in Turkish music. For this reason, a new dataset consisting of selected samples from Turkish music is prepared and the music emotion recognition process is performed by applying feature-based machine learning methods. It is thought that, thanks to the dataset and method presented in the literature, it will contribute to the creation of more collaborative work in the field of engineering and music with datasets consisting of Turkish music samples.

The main contributions of this work are as follows:

4-class dataset consisting of Turkish music data is created.

The effective power of machine learning in recognizing musical emotions has been demonstrated. Acoustic signal patterns expressing the emotional content of music have been removed as a feature.

Within the framework of this paper, section 2 includes previous studies on the classification of musical emotions. In section 3, the proposed methodology is introduced. In section 4, the dataset is introduced and also experimental results are given. Section 5 is devoted to the presentation and discussion of the results of the study.

2. Related Works

The proposed method is revaluated by using 800 music pieces in the dataset and success close to 86.3% is achiever (Lu, Liu, & Zhang , 2006). A comprehensive evaluation of the different stages of content-based music emotion recognition system training is presented with a regression approach by Huq ret al.160 features are extracted from the audio signals and tested with different regressions in three categories. Linear Regression, Regression Trees and Radial Based Functions are used as methods. In addition, testing is done using both feature selection and without feature selection (Huq, Bello, & Rowe, 2010). Schmidt rand Kim used the Deep Belief Networks based on regression to extract features directly from the spectrogram. The system has been shown to be easily applicable both to the problem of recognizing certain musical emotions and to any regression-based sound feature learning problem (Schmidt & Kim, 2011). The effect of musical characteristics on emotion classification has been extensively studied by Song et al. On the Last.FM website, a dataset of 2904 songs labeled with the words "happy", "sad", "angry" and "comfortable" is collected and various sound characteristics are extracted using standard algorithms. For classification, the dataset is trained using SVM with a polynomial and radial-based function kernel, and these are tested by applying 10-fold cross validation. It is observed that the spectral features show a better performance according to the results obtained (Song, Dixon, & Pearce, 2012). In 2015, a new method is proposed by Panda ret al. To combine standard and melodic features extracted from sound signals rand to recognize musical emotions. A new sound dataset is prepared by the authors to classify musical emotions. For each data in the dataset, 253 standard and 98 melodic features are extracted, and emotion recognition is performed using various classification algorithms. In addition, feature selection is used. According to the experimental results, it is observed that the melodic features perform better than the standard features. The best result is obtained as 64% F criterion with ReliefF feature selection and SVM (Panda, Rocha, & Paiva, 2015). Ren and colleagues used SVM to classify the emotions of music in several different datasets. They proposed a two-dimensional model to extract acoustic frequency and modulation frequency features (Ren, Wu , & Jang , 2015). A new approach to musical emotion detection based on the audio signal and the lyrics of a part has been presented by Delbouys et al. Traditional feature-based approaches are used and a new model based on deep learning is proposed. The performance of both approaches is compared in a database containing 18,000 audio files with valence rand arousal values. The presented model performed better in arousal prediction (Delbouys, Hennequin, Piccoli, Royo-Letelier, & Moussallam, 2018). A new dataset consisting of 124 Turkish music samples, each 20 seconds long, is introduced and experiments are performed on this dataset. In the proposed method (long short term memory) LSTM and (convolutional neural network) CNN are used together. 99.19% accuracy is obtained from the method. (Hizlisoy, Yildirim, & Tufekci, 2021)

3. Materials and Methods

In this study, a method based on audio feature extraction and machine learning is proposed for the classification of emotional content in music. In the first step of the proposed method, the samples in the dataset are pre-processed. In the second stage, audio features in different categories are extracted from the pre-processed music data. In the last stage, the classification process is made using the extracted audio features and machine learning algorithms. A schematic representation of the proposed method is given in Figure 1.

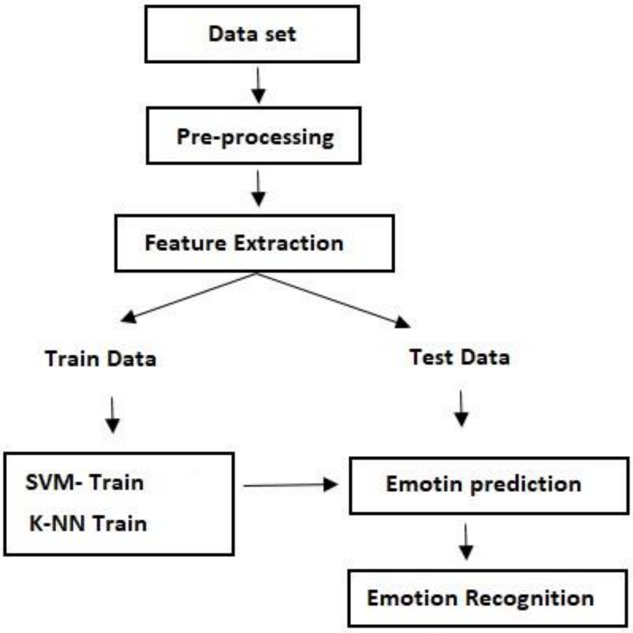

Figure 1. Proposed Method

3.1. Data Preprocessing

Preprocessing is a fundamental step before feature extraction and classification. The second order Butterworth filter is used in this study to eliminate the noise in music sound signals. In addition, the music recordings in the dataset are converted to mp3 format with a sampling frequency of 41100 Hz.

3.2. Feature Extraction

After the music in the dataset is emotionally tagged, the MIRtoolbox toolbox, which is widely preferred in music processing, is used to extract acoustic features. With the help of MIRtoolbox, tempo, loudness, tone, rhythm, etc. It can extract sound features from different groups such as. Feature vectors with a total of 34 feature values are obtained using MIRtoolbox.

Table 1. Features Used in the Experiment

Feature Name	Feature Type	Size
Root Mean Square Energy	Mean	1
Low energy	Mean	1
Tempo	Mean	1
Spectrum centroide	Mean	1
Spectral entropye	Entrpoy	1
Skewnesse	Mean	1
Zero Crossing ratee	Mean	1
MFCC	Mean	13
Attacktime	Mean, Standart	2
Chromagram	Mean	12

Root Mean Square - RMS: It is used to calculate the power of sound systems. The energy of the sound signal x can be calculated by taking square root of sum of the squares of amplitude values. (Lartillot, 2018). Figure 2 shows the RMS energy curve of a signal that is framed with a certain number of examples.

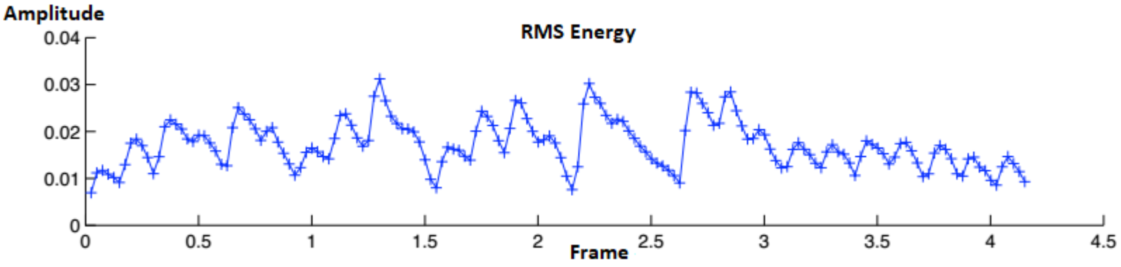

Figure 2. RMS Energy Curve (Lartillot, 2018)

Low energy: The energy curve can be used to evaluate the transient energy distribution to see if the signal remains constant over time or if some frames are different from others. One way to estimate the low energy ratio is to calculate the percentage of frames showing less energy than average (Tzanetakis & Cook, Musical genre classification of audio signals, 2002). Figure 3 presents a visualization of this feature. The selected part of the energy curve expresses the lower-than-average energy value. In Figure 3, there are some rare frames that contain particularly high energy, and for this reason, we can see that most of the frames are below the average RMS.

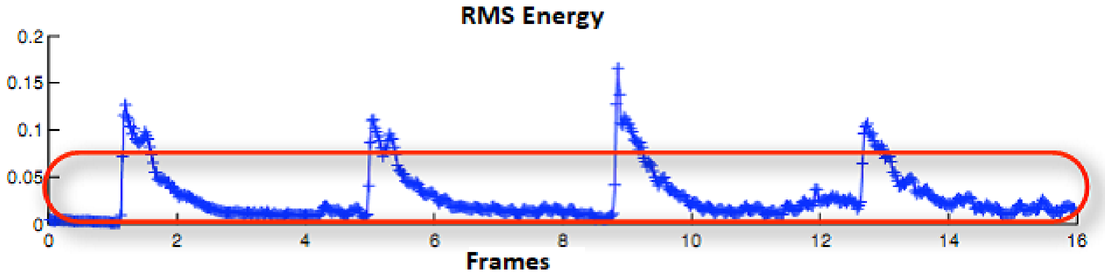

Figure 3. Display of Low Energy Amount in the Signal (Lartillot, 2018)

Tempo: The repetition of the selected unit in one minute determines the speed of the music. It is bpm, which is a unit used to measure the tempo in music. The tempo can be estimated by detecting periodicity from the event detection curve. The classical paradigm for tempo estimation is based on the determination of periodicity (Davies & Plumbley , 2007). The method that gives extremely accurate results in tempo estimation is autocorrelation. Autocorrelation determines the similarities of a given signal at different times. It performs the correlation of a signal window selected as reference with other windows. Figure 4 shows the peaks of a periodic signal with autocorrelation.

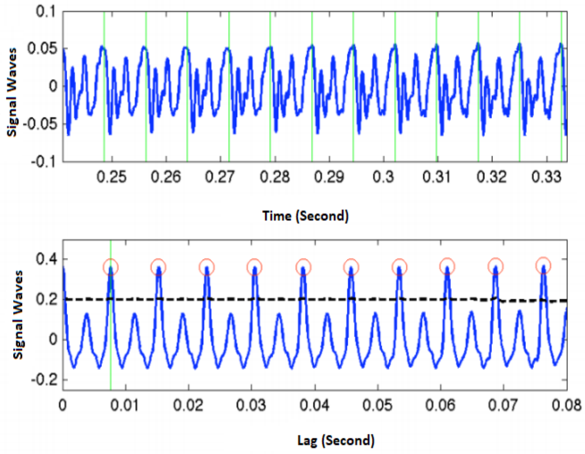

Figure 4. Peak Points of the Periodic Signal

Spectrum centroid: The statistical mean of the spectral distribution can be expressed as the average obtained over centroids calculated from separate frequency sub bands. Centroid is a measure of spectral shape and higher centroid values relate to sounds with higher frequencies. It has been demonstrated by spectral centroid user experiments that there is an important perceptual feature in the characterization of instrument timbre (Tzanetakis & Cook, Musical genre classification of audio signals, 2002).

Spectral Entropy: The entropy of the spectral distribution is obtained by averaging over the spectral entropies computed in separate frequency subbands (Toh, Togneri, & Nordholm, 2005).

Skewness: It is the coefficient of skewness of the average spectral distribution obtained from measurements of skewness calculated in separate frequency subbands (Lidy & Rauber, 2006).

Zero-crossing Rate: Specifies the number of times the waveform is displaced in a window. The X-axis shows how many times the signal has passed, it can be used as an indication of noise as well as frequency (Lartillot, 2018). Sound is also used to examine whether there are splits. Figure 5 shows the points where a signal crosses the x axis.

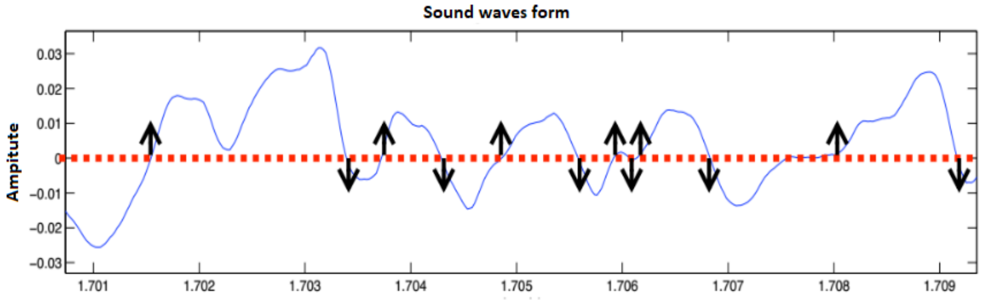

Figure 5. Representation of Points where a Signal Crosses the x-Axis (Lartillot, 2018)

Mel Frequency Cepstral Coefficients (MFCC): Mel is the measure of the frequency of a sound tone perceived by the human ear. This measured value does not correspond linearly with the physical frequency of the sound tone because a human ear perceives sound frequencies in a nonlinear way (Chauhan & Desai, 2014). As a result of the studies, it has been observed that the measurements are linear up to 1 kHz and a logarithmic increase in higher values. The Mel scale reflects how people hear the tone of their voice. MFCC is widely used in voice recognition. MFCC is a good way to distinguish speakers by imitating the frequency selectivity of the human ear (On, Pandiyan, Yaacob, & Saudi, 2006). The steps to be applied in order to derive the MFCC are given in Figure 6.

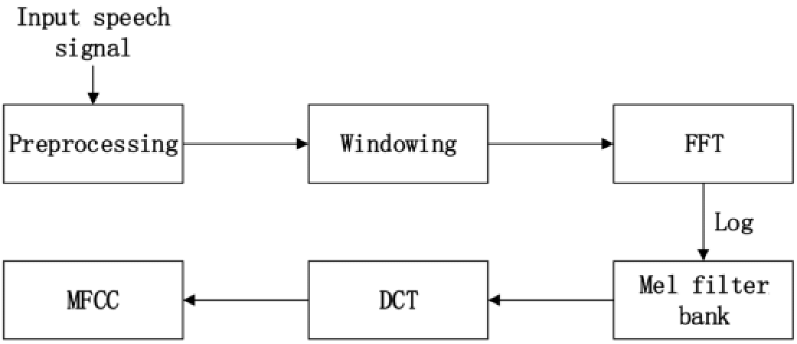

Figure 6. MFCC Extraction Steps

Conversion between Mel scale (M) and frequency scale (Hz) can be done using equations 1 and 2 below. MFCC is calculated according to equation 3.

$$m = 2595 log_{10}(1 + \frac{f}{700}) \qquad (1)$$

$$f = 700(10^{\frac{m}{2595}} - 1) \qquad (2)$$

$$MFCC_i = \sum_{k=1}^{20} X_k . Cos\left[i.\left(\frac{k-1}{2}\right).\frac{\pi}{20}\right] \qquad i = 1,2,\dots M \qquad (3)$$

Attack Time: It is an estimate of the time it takes for a signal to rise to its peak. The way to define and calculate this property is to estimate the temporal duration of the phase interval in which the amplitude of the signal rises (Lartillot, 2018).

Chroma: It shows the energy distribution around each note. Each note has a specific frequency range. It calculates the energy density in the frequency ranges of these notes (Lartillot, 2018).

3.3. Classifiers

Support Vector Machine: It is one of the best machine learning techniques that can be distinguished by linear and non-linear lines derived from Vapnik's statistical learning theory (Cristianini & Ricci, 1992). This technique can be used for both classification and regression analysis and is a computational learning method for classifying small samples (Widodo & Yang, 2007). In SVM, first of all, input data is taken to a higher dimensional area and the most suitable distinctive hyper plane between two classes is created in this area. It is claimed that the wider the gap between two classes in SVM, the more successful the classification will be. SVM that separates the two classes in the optimal hyper plane is given in Figure 7.

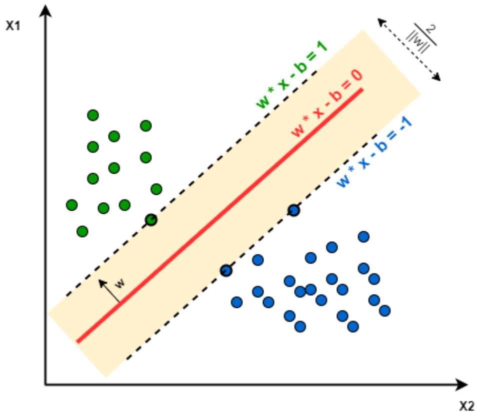

Figure 7. SVM

SVM solves a constrained quadratic optimization problem based on inherent risk minimization, creating the optimal hyperplane *f (x) = 0* between datasets (Schölkopf & Smola, 1990). The input vector $\{x_i$, i = 1, ..., n} belongs to one of the two classes $y_i\{-1, 1\}$, the hyperplane is defined as:

$$w_0.x + b_0 = 0 \qquad (4)$$

Here w is the weight vector, x is the input vector, b is a bias. For a given w and b, the data can be linearly separated in the following cases:

$$w.x_i + b \geq 1 \quad if\ y_i = 1 \qquad (5)$$
$$w.x_i + b \leq 1 \quad if\ y_i = -1 \qquad (6)$$

The kernel method is used to solve a nonlinear problem with a linear classifier. The input data is transformed into a high dimensional space with the Φ function. K core function:

$$k(x, x') = (\Phi(\mathrm{x}), \Phi(x')) \qquad (7)$$

-Polynomial:

$$k(x_i, x_j) = (x_i.x_j + 1)^d \qquad (8)$$

-Radial Basis Function:

$$(x, y) = e^{-\gamma||(x-x_i||^2} \qquad (9)$$

K-Nearest Neighbor (K-NN): K-NN algorithms aiming to find a data subset most similar to a query sample from a large-scale dataset; It is used as a basic component in a wide range of applications such as dimension reduction, model classification and image acquisition (Jégou, Matthijs, & Schmid, 2010). K-NN is an example-based learning algorithm based on the principle that samples in a dataset will often be found near other samples with similar characteristics. With this algorithm, there are k training points closest to this point data to classify a new point data. Classification is done by the majority vote of the neighbors; An element to be classified is distributed to the closest class among the closest neighbors measured by a distance function (Jégou, Matthijs, & Schmid, 2010). The formula of the K-NN algorithm is shown in equation 10.

$$x(x,y) = \sqrt{\sum_{j=1} w_j (x_j - z_j)^2} \tag{10}$$

In Figure 8, it is seen that the test sample is in the positive class when k = 1 or k = 3 is selected, and in the negative class when k = 2 is selected.

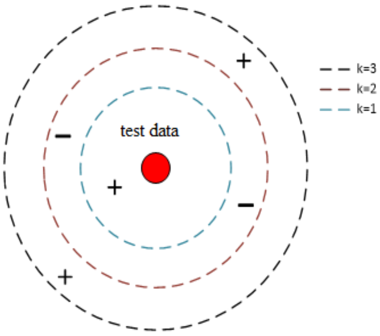

Figure 8. K-NN Diagram

Artificial neural network: ANN is inspired by the intelligent data processing capability of the human brain. ANN is motivated by mimicking the interaction of neurons in the brain with each other (Chien, ASCE, Ding, & Wei, 2002). ANN model, which is a mathematical model consisting of neurons and related neuron connections, is shown in Figure 9. The connections created between neurons used in network formation are associated with numerical values called weights. Each weight has a certain value that is transferred across the network and multiplied by data samples (Yang, Blond, Aggarwal, Wang, & Li, 2017). ANNs can be trained to recognize nonlinear relationships between input and output data without having knowledge about the problem. After training, the running time of ANN is extremely fast because it contains only a few simple, interconnected compute units. ANNs have features of model recognition, generalization and interpolation. Therefore, when an unknown input is applied to the trained network, they can produce a suitable output.

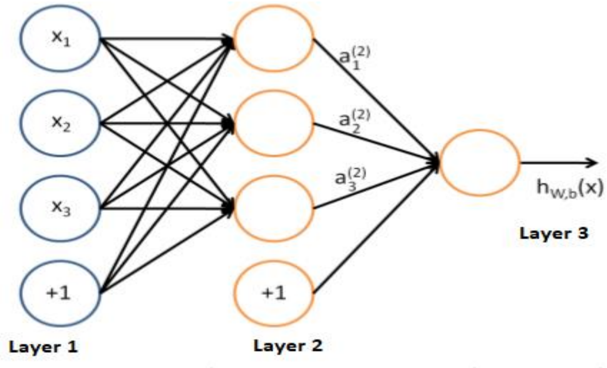

Figure 9. ANN

4. Experimental Applications

The scope of the experimental applications carried out for the study is stated below.

4.1. Dataset

In this research, a new dataset is prepared for music emotion. When we examine the studies in this field, we see that most researchers prepare their own datasets instead of using a common dataset, because it cannot be said that there is still a common dataset related to this field. There is still no consensus on which emotion pattern or how many categories of emotions should be used (Yang & Chen, 2012). In addition, the subjective nature of human perception makes it difficult to establish a common database. For this reason, a special dataset has been prepared by us for this study. Verbal and non-verbal musical works from different genres of Turkish music are taken to prepare the dataset. The dataset is designed as a discrete model and there are four classes in the dataset: tense, sad, happy and relaxing. In an experiment in which 13 people participated in order to determine the emotion tags of these works, the participants are asked to label the selected musical works with tense, sad, happy and relaxing emotion tags. 30-second episodes of the selected music are randomly cut to the participants and each participant labeled the music according to the emotions they felt. Then, emotion class of a music is determined according to these tags. The most labeled music parts are included in the class. For example, if a music is labeled 10 relaxing and 3 labeled happy, the music is included in the relaxation class. The experiment is conducted in 3 sessions and each participant listened to 500 music in total. In the database, 100 pieces of music have been determined for each class so that there is an equal number of samples in each class. The remaining music is not evaluated. In the original dataset, there are a total of 400 samples, 30-second records from each sample. The number of samples in each class in the dataset is given in Table 2.

Table 1. Classes and Number of Instances in the Dataset

Class	Number of Samples
Happy	100
Sad	100
Angry	100
Relax	100

4.2. Experimental Results

In this study, SVM, K-NN and ANN are used for the classification process. Before the classifiers are trained, the data are normalized by pre-processing. SVMs are trained using Polynomial and Radial Based Function (RBF) kernels. For K-NN, the number of K is chosen as 3. The data is divided in 2 different ways to determine the effect of the data size allocated for training and testing on the classifier. First, 70% of the data is used for training (30% for testing, secondly, 80% for training, 20% for testing. Performance criteria of the proposed method are evaluated according to Accuracy, precision, sensitivity and F-score classification metrics. Evaluation criteria in classification problems are made using a matrix with correct and incorrectly classified sample numbers for each class called confusion matrix. The concepts of FP, FN, TP and TN can be defined as follows:

- False positives (FP): samples of negative class, predicted positively.
- False negatives (FN): Negatively predicted samples that have a positive true class.
- True positives (TP): Correctly predicted examples of positive class.
- True negatives (TN): Samples of correctly predicted as belonging to the negative class.

$$Accuracy = \frac{|TN|+|TP|}{|FN|+|FP|+|TN|+|TP|} \tag{11}$$

Precision measurement evaluates the efficiency of the classifier for each class in binary problems. Precision, known as the true positive rate, is the ratio of predicted data from the positive class to the true positive data. Precision measurement is given in equation 12.

$$precision = \frac{|TP|}{|FN|+|TP|} \tag{12}$$

Sensitivity is a measure that predicts the probability that a positive guess is correct. Sensitivity measurement is given in equation 13.

$$sensitivity\ = \frac{|TP|}{|TP|+|FN|} \tag{13}$$

The F-score is a harmonious average of positive predictive ratio and sensitivity measures and is calculated as shown in equation 14.

$$F - score = \frac{2*|TP|}{2*|TP|+|FP|+|FN|} \tag{14}$$

In the first experiment, the classifiers are trained without normalizing the data, and in the second experiment, the classifiers are trained by normalizing the data. The classification results obtained from the data cannot be normalized are given in Table 3. When SVM is used with Polynomial kernel, the highest accuracy is 77.46%, when SVM is used with RBF core, the highest accuracy is 77.08%. In addition, the highest 75.76% accuracy is obtained with K-NN and the highest 77.37% accuracy with ANN. In terms of accuracy, the best classification performance is obtained with the SVM classifier. Confusion matrix for the best classification result obtained with non-normalized data is given in figure 10.

Table 3. Classification Results from Unnormalized Data

Model	Spliting Data at Different Rates for Training and Testing	Accuracy %	precision %	sensitivity %	F-score %
SVM	%70- %30	77.01	77.01	77.73	76.74
SVM	%80- %20	77.46	76.38	77.84	77.96
SVM (RBF)	%70- %30	75.30	77.87	76.26	76.87
SVM (RBF)	%80- %20	77.08	77.43	77.18	77.35
K-NN	%70 %30	75.24	76.00	76.34	76.25
K-NN	%80- %20	75.76	75.00	76.13	75.71
ANN	%70- %30	77.09	77.10	75.29	76.08
ANN	%80- %20	77.37	77.26	76.11	76.97

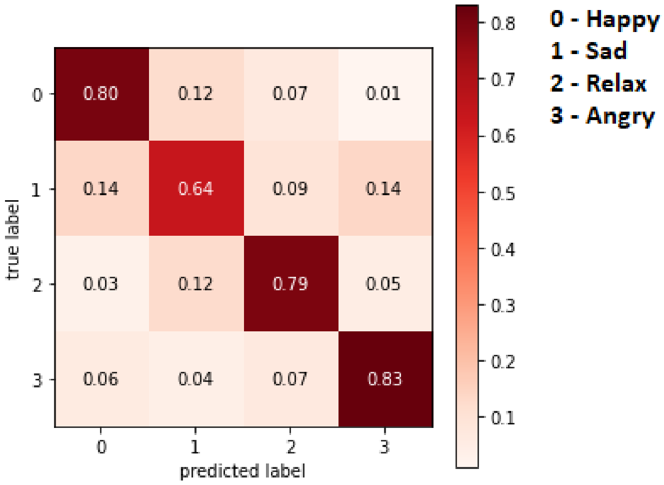

Figure 10. Confusion matrix of the best classification result obtained with non-normalized data

The classification results obtained after the data are normalized are given in Table 4. SVM has the highest 77.82% accuracy when used with the Polynomial kernel, and the highest 78.70% accuracy when used with the SVM, RBF kernel. In addition, the highest accuracy is 78.10% with K-NN and 79.30% with ANN. In terms of accuracy, the best classification performance is obtained with ANN. The confusion matrix for the best classification result obtained with normalized data is given in figure 11.

Table 4. Classification Results from Normalized Data

Model	Spliting Data at Different Rates for Training and Testing	Accuracy %	precision %	sensitivity %	F-score %
SVM	%70- %30	77.77	77.04	77.65	78.65
SVM	%80- %20	77.82	79.31	77.91	77.41
SVM (RBF)	%70- %30	76.92	77.90	77.06	77.69
SVM (RBF)	%80- %20	78.70	78.01	79.13	78.02
K-NN	%70 %30	77.61	78.29	78.38	76.50
K-NN	%80- %20	78.10	77.52	77.68	77.50
ANN	%70- %30	78.64	76.16	77.25	77.75
ANN	%80- %20	79.30	78.77	78.94	79.03

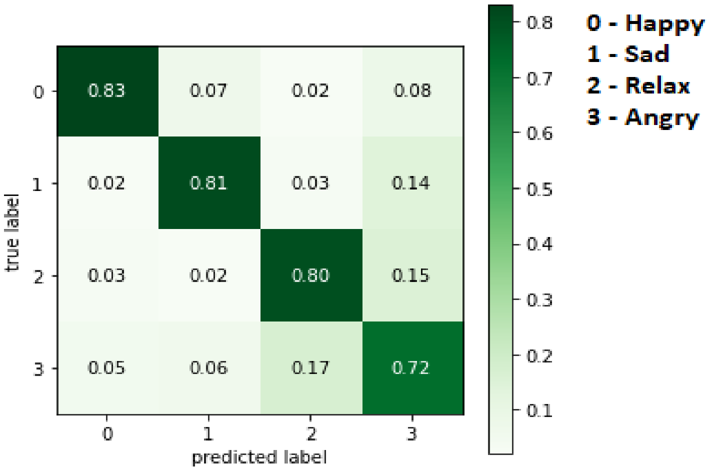

Figure 11. Confusion matrix of the best classification result obtained with normalized data

5. Conclusion

In this paper, a method based on machine learning and audio features is proposed to determine the emotions that music pieces will convey to the listener. The proposed method has been evaluated on a new dataset consisting of four classes. First, the music signals are pre-processed. Audio features such as energy, tempo, MFCC, attack time and chroma are extracted from the pre-processed signals in the next step. In the last stage, SVM, K-NN and ANN classifiers are trained by using audio features. Audio features are used both in normalized and non-normalized forms. The data are divided in two different ways for testing and training. It is divided into 70-30% in the first experiment and 80-20% in the second experiment, respectively for the test and training. The best classification success is obtained from 77.46% SVM without normalizing the acoustic features. After the features are normalized, it is seen that the best classification success is obtained from ANN as 79.30%.

References

Chauhan, P. M., & Desai, N. (2014). "Mel Frequency Cepstral Coefficients (MFCC) based speaker identification in noisy environment using wiener filter. 2014 International Conference on Green Computing Communication and Electrical Engineering (ICGCCEE), (s. 1-5). Coimbatore.

Chien, S., ASCE, M., Ding, Y., & Wei, C. (2002). Dynamic Bus Arrival Time Prediction with Artificial Neural Networks. Journal of Transportation Engineering, 128(5), 429-438.

Cristianini, N., & Ricci, E. (1992). Support Vector Machines. Encyclopedia of Algorithms. içinde

Davies, M. E., & Plumbley , M. (2007). Context-Dependent Beat Tracking of Musical Audio. IEEE Transactions on Audio, Speech, and Language Processing, 15(3), 1009-1020.

Delbouys, R., Hennequin, R., Piccoli, F., Royo-Letelier, J., & Moussallam, M. (2018). Music Mood Detection Based On Audio And Lyrics With Deep Neural Net. ISMIR 2018.

Hevner, K. (1936). Experimental Studies of the Elements of Expression in Music. The American Journal of Psychology, 48(2), 246-268.

Hizlisoy, S., Yildirim, S., & Tufekci, Z. (2021). Music emotion recognition using convolutional long short term memory deep neural networks. Engineering Science and Technology, an International Journal, 760-767.

Huq, A., Bello, J., & Rowe, R. (2010). Automated Music Emotion Recognition: A Systematic Evaluation. Journal of New Music Research, 39(3), 227-244.

Jégou, H., Matthijs, D., & Schmid, C. (2010). Product Quantization for Nearest Neighbor Search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(1), 117-1128.

Lartillot, O. (2018). MIRtoolbox 1.7.1 User's Manual. Jyväskylä, Finland.

Lidy, T., & Rauber, A. (2006). Computing statistical spectrum descriptors for audio music similarity and retrieval. MIREX 2006 - Music Information Retrieval Evaluation.

Lin, C., Liu, M., Hsiung, W., & Jhang, J. (2016). MUSIC EMOTION RECOGNITION BASED ON TWO-LEVEL SUPPORTVECTOR CLASSIFICATION. 2016 International Conference on Machine Learning and Cybernetics (ICMLC), (s. 375-386). Jeju.

Lu, L., Liu, D., & Zhang , H.-J. (2006). Automatic mood detection and tracking of music audio signals. in IEEE Transactions on Audio, Speech, and Language Processing, 14(1), 5-18.

On, C. K., Pandiyan, P., Yaacob, S., & Saudi, A. (2006). Mel-frequency cepstral coefficient analysis in speech recognition. 2006 International Conference on Computing & Informatic, (s. 1-5). Kuala Lumpur.

Panda, R., Rocha, B., & Paiva, R. (2015). Music Emotion Recognition with Standard and Melodic Audio Features. Applied Artificial Intelligence, 29(4), 313-334.

Ren, J.-M., Wu , M.-J., & Jang , J.-S. (2015). Automatic Music Mood Classification Based on Timbre and Modulation Features. IEEE Transactions on Affective Computing, 6(3), 236-246.

Schmidt, E. M., & Kim, Y. (2011). Learning emotion-based acoustic features with deep belief networks. 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, (s. 65-68). New Paltz.

Schölkopf, B., & Smola, A. (1990). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

Song, Y., Dixon, S., & Pearce, M. (2012). Evaluation of musical features for emotion classification. Proc. ISMIR, (s. 523-528).

Toh, A. M., Togneri, R., & Nordholm, S. (2005). pectral entropy as speech features for speech. In Proceedings of PEECS, (s. 22-25).

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing, 10(5), 293-302.

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21(6), 2560-2574.

Yang, Q., Blond, S., Aggarwal, R., Wang, Y., & Li, J. (2017). New ANN method for multi-terminal HVDC protection relayin. Electric Power Systems Research , 148, 191-201.

Yang, Y.-H., & Chen, H.-H. (2012). Machine recognition of music emotion: A review. ACM Trans. Intell. Syst. Technol, 3(4).