THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 1 (2024)
https://doi.org/10.62019/abbdm.v4i1.117.

ASIAN BULLETIN OF BIG DATA MANAGEMENT
ISSN (Print): 2959-0795
http://abbdm.com/ ISSN (online): 2959-0809

Song's & Non-Song's Pattern Classification Using Neural Network

Areeza Bhanbhro*, Abeer Mukhtiar, Muniba Memon, Akhtar Hussain Jalbani, Sunbul Bhangwar, Mohsin Ali Khokhar

Chronicle

Article history

Received: February 24, 2024
Received in the revised format: March 2, 2024
Accepted: March 3, 2024
Available online: March 4, 2024

Areeza Bhanbhro, Abeer Mukhtiar, Muniba Memon, Akhtar Hussain Jalbani, Sunbul Bhangwar, and Mohsin Ali Khokhar are currently affiliated with the Department of Information Technology Quaid-e-Awam University of Engineering Science and Technology, Nawabshah, Pakistan.

Email: bhanbhroareeza@gmail.com
Email: qureshiqur12@gmail.com
Email: muniba@quest.edu.pk
Email: jalbaniakhtar@gmail.com
Email: bhangwarsunbul@gmail.com
Email:hokharmohsinali851@gmail.com

Corresponding Author*

Abstract

The problem of pattern classification between musical or non-musical song audio snippets and non-songs is the main subject of this paper. The machine learning approach accomplishes the goal of automatically interpreting recordings to see if they are songs or not. There are two types of songs: musical and non-musical. This is due to the possibility that it could be difficult to discern between non-musical songs and non-song patterns. Thus, it is necessary to have such systems that differentiate between non-song patterns and songs, whether or not they are melodic. The extraction and selection of characteristics have made advantage of the preprocessing phases. The features that were obtained were pitch, intensity, length (duration), tempo, and sample rate. The Back-propagation Multi-layer Perceptron Neural Network model is used for both dataset testing and model training. This technique will distinguish between a musical or non-musical song and a non-musical piece (such conversation and speech). Training data for the dataset included recordings of speeches and dialogues in Hindi and Urdu for non-song audio files, as well as audio of various male and female singers from Pakistan and India for song audio files in Hindi and Urdu. When the classifier was evaluated using a range of audio samples, 90% of its classifications for songs and non-songs were accurate.

Keywords: Song (musical or non-musical), Non-song, Back-propagation Multi-layer Perceptron Neural Network.

INTRODUCTION

One of the most popular uses of machine learning is the classification of audio patterns. The process of automatically categorizing audio signals into one or more specified groups is known as audio pattern classification. Numerous uses for this activity exist, including sound event detection and voice recognition. Song and non-song pattern classification is the main goal of this study. One subtask of audio classification is the categorization of song (musical or non-musical) and non-song patterns, which includes the differentiation between song records and non-song recordings (like speech or dialogue). The goal is to automatically determine if an audio clip falls into the song or non-song category. Machine learning algorithms and feature extraction serve as the foundation for both song and non-song classification techniques. With this method, a set of properties is taken out of an audio stream. A machine learning system, like a neural network, uses these features as input and learns to distinguish between songs and non-songs in the recordings. AI-based audio classification systems learn to distinguish between various audio signal types by analysing big audio datasets and applying machine learning algorithms. Phase 1 of this approach consists of data collection and preprocessing; Phase 2 involves training and testing the data using a machine learning algorithm; and Phase 3 involves interpreting the findings to determine if the audio is a song or non-song clip.

LITERATURE REVIEW

A great deal of work has been done in the past on music information retrieval systems, and machine learning helps to accomplish this. Based on previous study, one of the researchers collected variables such as tempo, RMS, song duration, frequency, dynamic range, and tonality using RNN for singer and genre detection (Yasrab, 2023) . The study uses MFCC as a feature extraction technique and K-NN as a classification method to classify music and determine whether it is pop or RnB (Rhythm and Blues) genre (Ramadhana). It is advised that the discrete wavelet transform's mean and variance be used in the study of automatically classifying speech and music using neural networks. As a classifier, they used Multi-Layer Perceptron (MLP) Neural Networks (Khan, 2004). For their research, they made use of a music dataset that included ten distinct genres. The system is trained and categorised using deep learning. Convolution neural networks are utilised in this instance for both training and classification. For sound samples, the Mel Frequency Cepstral Coefficient (MFCC) is utilised as a feature vector (Vishnupriya).

Mel-frequency cepstral coefficients are employed in this study to classify musical genres. Using an auto-associative neural network (AANN) and MFCC, songs are categorised according to criteria. Experimental results show that multilayer support vector machines perform well in music classification (Thiruvengatanadhan, 2018) .Using perceptual audio features—which are used to train classifiers to distinguish between the two gender classes—they hope to construct such a system in the proposed study. These characteristics include information based on tempo and pitch, transient energy, etc. As previous study had mostly focused on the pitch method, MFCC approach, etc., they decided to include a variety of these perceptual qualities in their work, which may be the first of its kind (Sengupta, 2017) . Chroma, Spectral Centroid, Spectral Rolloff, and ZCR are utilised in conjunction with MFCC, K-NN, and SVM Classifier as feature vectors in the research on music genre classification.

The system was trained using three classifiers: linear, polynomial kernel SVMs, and k-NN. They found that the accuracy rate (78%) of the polynomial kernel SVM is greater (Patil, 2017) . In this study of the deep learning technique for classifying vocalist voices in Vietnamese popular music, they outline a neural network architecture and use the voice's extracted Mel-Frequency Cepstral Coefficient (MFCC) as input data (Pham Van). This research on the development of machine learning for healthy and asthmatic voluntary cough noises aims to use machine learning to categorise cough sounds for children with asthma and children without asthma. The constant-Q cepstral coefficients and the mel frequency were recovered, along with other acoustic features. A training set and the Gaussian mixture model-universal background model (GMM-UBM) were used to develop a classification model that could distinguish between children's coughs that were "healthy" and those that were asthmatic (Hee, 2019). This study uses a multilayer perceptron (MLP) neural network to classify music genres while taking MFCCs into account. Assessing the effect of increasing the number of audio features in the MFCCs at the MLP's input on performance. The impact of several criteria, such as spectral bandwidth and zero-crossing rate, on the accuracy of genre classification and the results that show how similar and closely related various musical genres are to one another (Flederus) .This paper presents an automatic technique for classifying song signals into two categories: classical and non-classical/popular song. Based on MFCC, they have computed characteristics that correspond to the decomposed signals. Discrete Mel frequency co-efficients that are computed over short times intervals are found to co-occur, and properties are obtained to illustrate the signal pattern. RANSAC is the classifier that has been used (Ghosal, 2011).

This effort attempts to classify audio tracks according to their musical patterns, with the goal of retrieving the music clips based on listener choice. Seven main types—devotional, vigorous, folk, joyful, pleasant, melancholy, and drowsy—are taken into consideration for this work. In the training phase, forty music clips per category are mixed with vibrato-related features including jitter and shimmer, and in the testing phase, fifteen clips per category are combined with mel-frequency cepstral coefficients (MFCCs). To produce a 19-dimensional feature vector, statistical pitch values such as min, max, mean, and standard deviation are calculated and added to the jitter, shimmer, and MFCCs.Feedforward backpropagation neural networks (BPNNs) are used as classifiers because they are good at mapping nonlinear interactions (Sharma, 2016) .

METHODOLOGY

A song's musical or non-musical & non song's pattern classification entails gathering information, preprocessing (i.e., selecting and extracting features from a diverse dataset), selecting the best machine learning model, training and assessing the model, and maybe deploying the end result for use in real-world applications, such as speech classification etc.

Figure 1.
Proposed Methodology

Data Acquisition

Collecting a collection of audio clips that includes non songs & songs samples. The dataset is built using conversation from movies and podcasts, particularly for audio content that isn't music. It also includes songs from Pakistan and India.

Preprocessing

Preprocessing includes format setup in the.wav format, data cleaning, and reducing the length of each dataset audio file to no more than 15 seconds. Preprocess the audio data to identify the most informative features that can be used to differentiate between musical and non-musical song & not songs audio clips . Using the Python programming language, extract features such as tempo, duration, intensity, pitch, and sample rate that can be used for classification.

Model-Picking

Choosing a machine learning model that is appropriate for the categorization task. For audio categorization, back-propagation multilayer perceptron neural networks are the model of choice. The Structural Equation Model for confirmatory analysis in Pakistan's environment, as one variable is dependent on another.

Table 1.
Songs Data Description

S. No	Singers	Song Names	Songs Category
1	Kaifi Khalil	"Kahani Suno 2.0"	With music
2	Ali Zafar	"Laila o Laila"	With music
3	Arijit Singh	"Phr aur kya chahiye"	With music
4	Lata Mangeshkar	"Jaane kyun log"	with music
5	Ahmed Jahanzeb	"Ishq Murshid ost"	with music
6	Arijit Singh	"Phir lay aya Dil"	Song with music
7	Shreya Ghoshal	"Barish "	with music
8	Lata Mangeshkar	"Tere Bina Zindagi se"	with music
9	Ali Sethi	"Chandni Raaten"	without music
10	Arijit Singh	"Saware"	Song without music
11	Weshaal Meshra	"Pehle bhi Ma"	without music
12	Weshaal Meshra	"ek mulakat"	without music
13	Momina	"Aya na tu"	without music
14	Yashal	"Sajna"	without music
15	Atif	"Zindagi"	without music

Training & Testing of Model

Utilizing the labels, train the chosen model on the audio features dataset to obtain the total error network graph. In order to assess the accuracy and robustness of the classifier, fresh audio clip features that have not been trained on the model are also added to test and evaluate its performance.

DESCRIPTION OF DATA

There are 400 audios in all, including speeches and dialogues as well as songs that are either musical or not. There are 200 audios total—100 with musical songs and another 100 with non-musical songs—and 200 with non-songs. Each audio clip has a maximum duration of 15 seconds and is in the high-fidelity .wav format. The dataset is built using conversation from movies and podcasts, particularly for audio content that isn't music. It also includes songs from Pakistan and India. The dataset is available in Hindi and Urdu, and it includes both male and female voice samples. For audio that isn't a song, the information is gathered from conversation from silent films and podcasts.Using machine learning techniques, various features were retrieved from the.wav file format. Next, using a Python programme, normalised all those features by creating a text file with their labels. "Neuroph studio" has been used for neural networks, which aids in model training and appropriate output generation. The performers whose music is used to compile the dataset.

PATTERN CODE

Table 2 below displays the Patterns for songs that are either musical or non-musical and non-song. A song's code, whether musical or non-musical, is "1," and a non-song is "0."

Table 2.
Song & Non-song Pattern codes

S. No	Audio	Pattern
1	Song (musical or non-musical)	1
2	Non-song	0

NEURAL NETWORK ARCHITECTURE

The neural network architecture is depicted in Figure 2 and is comprised of three layers: layer 1 is the input layer, wherein five neurons are acquired for the input from audio data; layer 2 is the hidden layer, wherein four hidden neurons are used; and the last layer is the output layer, which consists of one neuron, since the project's output will be a song, whether musical or not. coding pattern is either "1" or, as Table 2 illustrates, "0" for the non-song audio. non-musical).

Trian Test Separation (Split)

Thirty percent of the dataset is designated for testing and seventy percent is designated for training in the machine learning division. It is ensured that a sizable portion of the data is used to train the model and assist it in finding the division. Underlying patterns in the dataset by employing this split strategy. However, the testing data, which makes upto

30% of the dataset, is set aside especially for assessing the performance of the model.

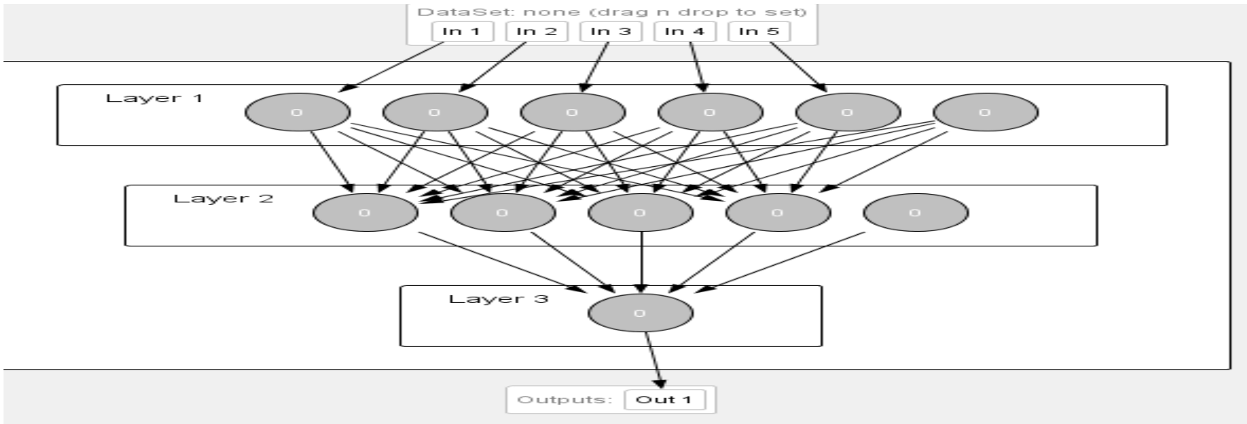

Figure 2.
Neural Network Architecture

Total Network Error Graph

Figure 3 below illustrates the total training error rate of the network graph following training of the dataset by loading it into a model (neural network). This rate is 0.009.

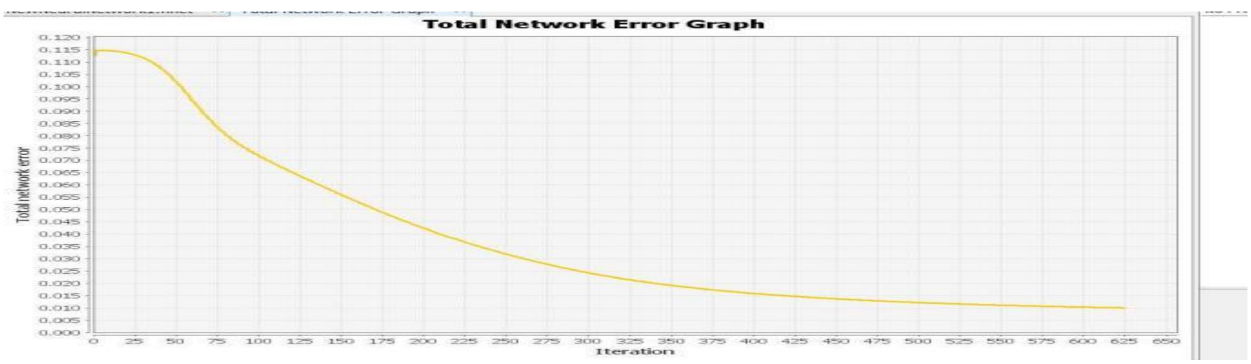

Figure 3.
Training Error Graph

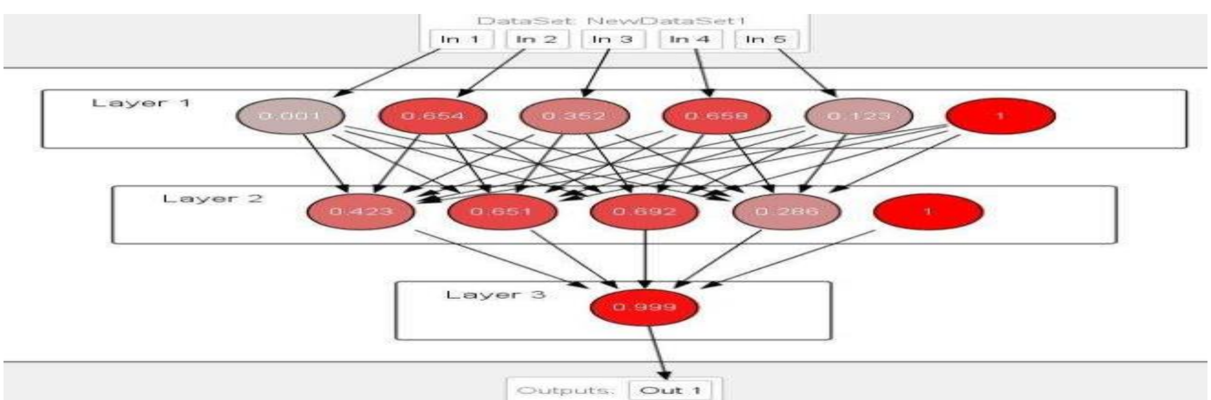

Figure 4.
Testing Results of Musical song

OUTCOMES & DISCUSSION

Musical Song Experimental Outcomes

The experimental findings of the tested model, which included features from an audio sample of a musical song, are displayed in Figure 4. Lata Mangeshkar's performance of "Dil Deewana Bin Sajna ke" is the song being examined. Given that a song's pattern code is displayed as "0.999," which becomes "1" when rounded off, the model correctly predicted the audio results for the song (see Table 2). The model predicts the following outcomes, which are displayed below.

Non-Musical Song Experimental Outcomes

The experimental results of the model under test, which examined an audio sample of a non-musical song, are displayed in Figure 5. The song "Dill Sambhal ja Zara" by singer Arijit Singh is the only one heard in this audio clip. Given that the pattern code for non-musical songs is "1," the model correctly predicted the outcomes for the non-musical song audio. The model predicts the following outcomes, which are displayed below.

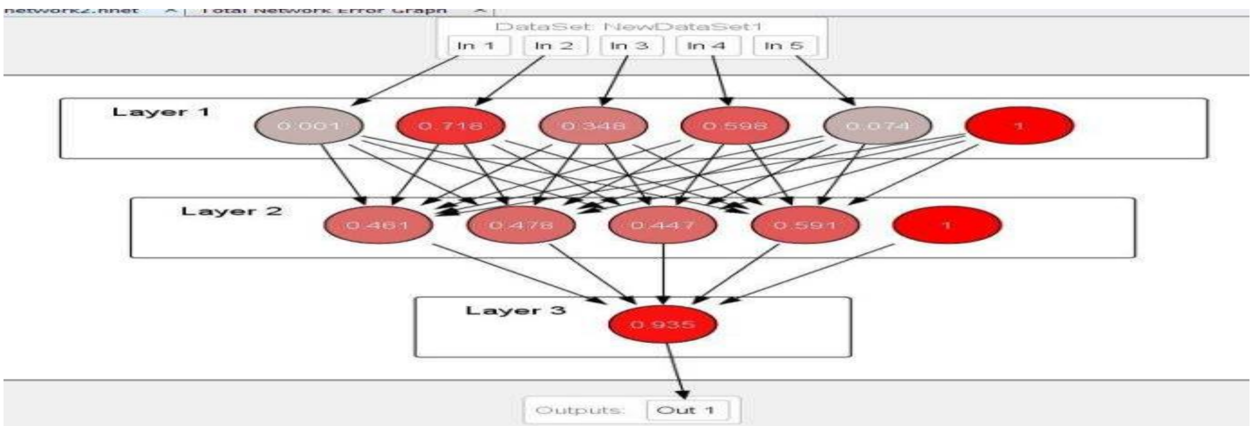

Figure 5.
Testing Results of Non-Musical Song

Non- Song Experimental Outcomes

The experimental results of the model, which was tested using a non-song audio sample, are displayed in Figure 6. This audio clip features "Rabia Mughni," the founder of a Pakistani podcast, speaking normally. This non-song audio excerpt is from the podcast Gup Shap featuring FUCHSIA. Given that the non-song audio's pattern code is "0," the model correctly anticipated the results for the non-song audio. Refer to Table 2. The model predicts the following outcomes, which are displayed below.

Accuracy of Testing

The performance of this multi-layer perceptron neural network model is good. It is crucial to the outcomes of training and testing because, as Figure 7 below illustrates, when we tested several audio files that weren't fed to the model for training, the model matched the patterns, predicted the right answers, and gave us an accuracy of 90%.

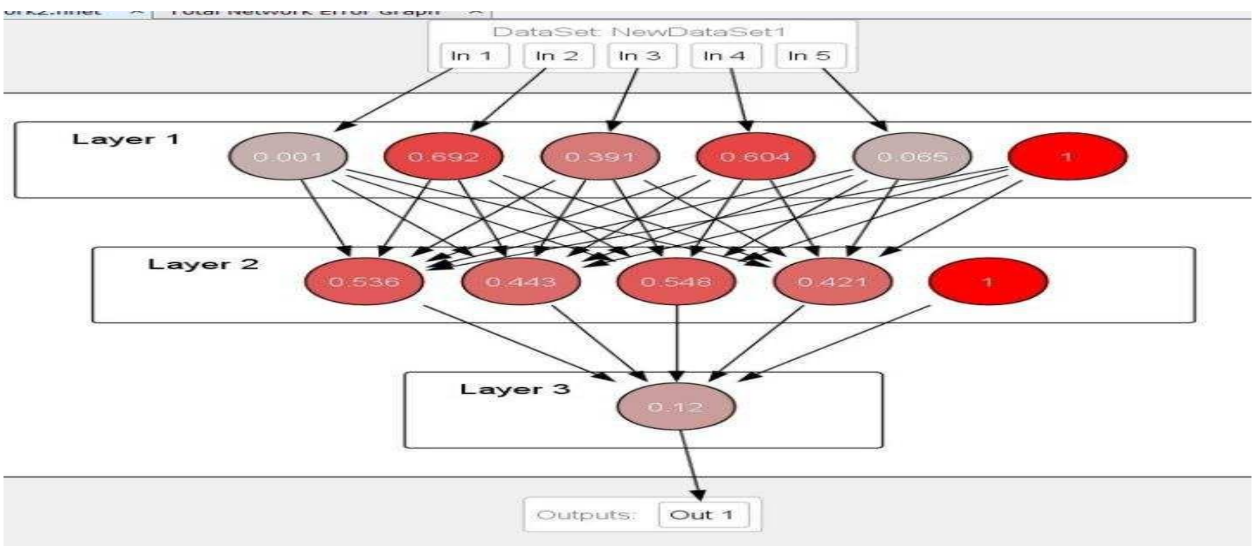

Figure 6.
Testing Results of Non-song

Overall Accuracy: 90%

Class	n (truth)	n (classified)	Accuracy	Precision	Recall	F1 Score
1	60	59	90%	0.93	0.92	0.92
2	30	31	90%	0.84	0.87	0.85

Figure 7.
Model Accuracy Testing

Cross Validation

In the interesting topic of pattern classification, where we attempt to distinguish between songs and non-songs, there exists an efficient method called cross-validation that aids in enhancing the accuracy of models. Fundamentally, cross-validation is like putting a model through a rigorous rehearsal run before its big show. Rather of training the dataset only once and hoping for the best results, it is divided into smaller groups, called "folds." Since we've selected a k value of five in this case, our dataset will be divided into five sections, acting as five mini-concerts for our model to learn from.

```
Total Network Error Graph | crosValNet-0.nnet | Test Results
Input: 0.0011; 0.7643; 0.3429; 0.5343; 0.1129;  Output: 0.3437; Desired output: 0;  Error: 0.3437;
Input: 0.001; 0.7697; 0.3299; 0.5429; 0.0628;  Output: 0.3564; Desired output: 0;  Error: 0.3564;
Input: 0.001; 0.7914; 0.3303; 0.5086; 0.0764;  Output: 0.204; Desired output: 0;  Error: 0.204;
Input: 0.0011; 0.7622; 0.3451; 0.54; 0.0915;  Output: 0.3126; Desired output: 0;  Error: 0.3126;
Input: 0.0012; 0.6821; 0.3945; 0.6097; 0.0857;  Output: 0.4274; Desired output: 0;  Error: 0.4274;
Input: 0.0011; 0.7508; 0.3499; 0.5555; 0.0727;  Output: 0.3308; Desired output: 0;  Error: 0.3308;
Input: 0.0011; 0.7565; 0.3595; 0.5335; 0.1175;  Output: 0.2539; Desired output: 0;  Error: 0.2539;
Input: 0.0011; 0.7492; 0.3561; 0.5405; 0.1409;  Output: 0.3688; Desired output: 0;  Error: 0.3688;
Input: 0.0012; 0.683; 0.3903; 0.6129; 0.0743;  Output: 0.4471; Desired output: 0;  Error: 0.4471;
Input: 0.001; 0.7955; 0.332; 0.5029; 0.0643;  Output: 0.1575; Desired output: 0;  Error: 0.1575;
Input: 0.0011; 0.7558; 0.3455; 0.5465; 0.1042;  Output: 0.3802; Desired output: 0;  Error: 0.3802;
Input: 0.0011; 0.7682; 0.3446; 0.529; 0.1058;  Output: 0.2863; Desired output: 0;  Error: 0.2863;
Input: 0.001; 0.7896; 0.3296; 0.4987; 0.1391;  Output: 0.2826; Desired output: 0;  Error: 0.2826;
Input: 0.0012; 0.7095; 0.3784; 0.5848; 0.1072;  Output: 0.4218; Desired output: 0;  Error: 0.4218;
Input: 0.0011; 0.7728; 0.3467; 0.5182; 0.1189;  Output: 0.2473; Desired output: 0;  Error: 0.2473;
Input: 0.0011; 0.7809; 0.3439; 0.5131; 0.0936;  Output: 0.1919; Desired output: 0;  Error: 0.1919;
Input: 0.0011; 0.7738; 0.3439; 0.524; 0.0919;  Output: 0.236; Desired output: 0;  Error: 0.236;
Input: 0.0013; 0.6483; 0.4149; 0.6285; 0.1121;  Output: 0.4961; Desired output: 0;  Error: 0.4961;
Input: 0.001; 0.8191; 0.3096; 0.4719; 0.1028;  Output: 0.1893; Desired output: 0;  Error: 0.1893;
Input: 0.0011; 0.7696; 0.3452; 0.5105; 0.1671;  Output: 0.3225; Desired output: 0;  Error: 0.3225;
Input: 0.0014; 0.643; 0.3994; 0.6453; 0.1034;  Output: 0.6859; Desired output: 0;  Error: 0.6859;
Input: 0.0011; 0.8155; 0.3127; 0.4791; 0.0874;  Output: 0.1796; Desired output: 0;  Error: 0.1796;
Input: 0.001; 0.8295; 0.295; 0.4657; 0.0897;  Output: 0.2066; Desired output: 0;  Error: 0.2066;
Input: 0.0011; 0.7877; 0.3277; 0.5104; 0.1079;  Output: 0.2863; Desired output: 0;  Error: 0.2863;
Input: 0.0014; 0.6389; 0.4083; 0.6376; 0.136;  Output: 0.6659; Desired output: 0;  Error: 0.6659;
Input: 0.0013; 0.7336; 0.3676; 0.5623; 0.1024;  Output: 0.334; Desired output: 0;  Error: 0.334;
Input: 0.001; 0.8363; 0.2974; 0.4446; 0.1207;  Output: 0.1655; Desired output: 0;  Error: 0.1655;
Input: 0.0011; 0.8217; 0.3151; 0.4628; 0.1063;  Output: 0.1424; Desired output: 0;  Error: 0.1424;
Input: 0.0013; 0.7441; 0.3687; 0.5507; 0.0844;  Output: 0.2274; Desired output: 0;  Error: 0.2274;
Input: 0.0013; 0.7027; 0.3826; 0.5925; 0.0934;  Output: 0.4095; Desired output: 0;  Error: 0.4095;
Input: 0.0014; 0.6344; 0.435; 0.624; 0.1376;  Output: 0.4027; Desired output: 0;  Error: 0.4027;
Input: 0.001; 0.8001; 0.3227; 0.4968; 0.0941;  Output: 0.2184; Desired output: 0;  Error: 0.2184;
Input: 0.0013; 0.643; 0.4008; 0.6443; 0.1038;  Output: 0.6726; Desired output: 0;  Error: 0.6726;
Input: 0.0011; 0.7663; 0.3537; 0.517; 0.1426;  Output: 0.2541; Desired output: 0;  Error: 0.2541;
Input: 0.0011; 0.7663; 0.3537; 0.517; 0.1426;  Output: 0.2541; Desired output: 0;  Error: 0.2541;
Input: 0.001; 0.7517; 0.3222; 0.5731; 0.0515;  Output: 0.5794; Desired output: 1;  Error: -0.4206;
Input: 0.0012; 0.7036; 0.3711; 0.5944; 0.1176;  Output: 0.5664; Desired output: 1;  Error: -0.4336;
Input: 0.001; 0.7413; 0.3235; 0.5174; 0.2796;  Output: 0.8003; Desired output: 1;  Error: -0.1997;
Input: 0.0012; 0.603; 0.3859; 0.6945; 0.0718;  Output: 0.8987; Desired output: 1;  Error: -0.1013;
Input: 0.0012; 0.6357; 0.3814; 0.6641; 0.0969;  Output: 0.8472; Desired output: 1;  Error: -0.1528;
Total Mean Square Error: 0.1011970256511038
Output | Notifications
```

Figure 8.
Data Validation

SUMMARY

In summary, classifying song and non-song patterns is a crucial task in audio classification, having a variety of applications in music streaming services, voice recognition, and other areas. The classification of song (musical or non-musical) and non-song audio patterns is valued in different fields. The entire paper discusses how a system that can distinguish between audio snippets that are songs and those that aren't is designed to work. The project's goals are to gather and acquire audio samples, both musical and non-musical, & non songs in order to classify patterns. OR to use web crawling to collect audio data from streaming services, including songs and non-songs. In order to prepare the data for selection and feature extraction. The multilayer perceptron neural network model needs to be trained. to assess the outcomes that reliably distinguish between audio recordings with a musical quality (songs) and those without (non-songs).

Although researchers have already attempted a variety of approaches to teach computers these distinctions, more work needs to be done to ensure that computers can distinguish between these sounds accurately. In this work, the results of feature extraction and machine learning algorithms are encouraging. The features that were recovered from the audio include tempo, duration, sample rate, pitch, and intensity. The dataset was trained using a Back-Propagation Multi-layer Perceptron Neural Network Model, which helped the system learn to match audio data properties like songs and non-songs. Large data sets have also made it possible to create and assess a variety of algorithms for the classification of song and non-song patterns. But thanks to the advancement of complex machine learning algorithms, AI-based song classification systems—whether they are musical or not—and non-song pattern classification systems have demonstrated astounding accuracy in recognising and categorising audio. The experiment produced good results in classifying audio files as songs or non-songs, with an accuracy of 90% on the test set.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.

Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.

Availability of data and material: In the approach, the data sources for the variables are stated.

Authors' contributions: Each author participated equally to the creation of this work.

Conflicts of Interests: The authors declare no conflict of interest.

Consent to Participate: Yes

Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Flederus, D. R. (n.d.). Enhancing music genre classification with neural networks by using extracted musical features (Bachelor's thesis, University of Twente). Retrieved from UNIVERSITY OF TWENTE STUDENT THESES: https://essay.utwente.nl/80549/

Ghosal, A. C. (2011). Song classification: Classical and non-classical discrimination using mfcc co-occurrence based features. . In Signal Processing, Image Processing and Pattern

Recognition: International Conference, SIP 2011, Held as Part of the Future Generation Information, (pp. 179-185). jeju island korea.

Hee, H. I. (2019). Development of machine learning for asthmatic and healthy voluntary cough sounds: a proof of concept study. . Applied Sciences,, 2833.

Khan, M. K.-K. (2004). Automatic classification of speech and music using neural networks. In Proceedings of the 2nd ACM international workshop on Multimedia databases, (pp. 94-99).

Patil, N. M. (2017). Music genre classification using MFCC, K-NN and SVM classifier. International Journal of Computer Engineering In Research Trends,, 43-47.

Pham Van, T. T. (n.d.). Deep Learning Approach for Singer Voice Classification of Vietnamese Popular Music. . arXiv e-prints,, 2102.

Ramadhana, Z. H. (n.d.). Classification of Pop And RnB (Rhythm And Blues) Songs With MFCC Feature Extraction And K-NN Classifier. Jurnal Elektronik Ilmu Komputer Udayana p-ISSN, 2301, 5373.

Sengupta, S. Y. (2017). Classification of male and female speech using perceptual features. In 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), (pp. 1-7).

Sharma, R. S. (2016). Audio songs classification based on music patterns. In Proceedings of the Second International Conference on Computer and Communication Technologies: IC3T 2015, (pp. 157-166). Springer India.

Thiruvengatanadhan, R. (2018). Music genre classification using mfcc and aann. International Research Journal of Engineering and Technology (IRJET).

Vishnupriya, S. &. (n.d.). Automatic music genre classification using convolution neural network. In 2018 international conference on computer communication and informatics (ICCCI) ., (pp. 1-4).

Yasrab, Z. I. (2023). Singer and Genre Recognition using Deep Learning. Pakistan Journal of Emerging Science and Technologies (PJEST),, 1-12.