MUSIC SENTIMENT DETECTION PLATFORM

Saurabh Chalise[1], Sudip Rana[2]

[1] Department of Electronics and Computer, Thapathali Campus, IOE, TU
[2] Assistant Professor, Department of Electronics and Computer, Thapathali Campus, IOE, TU

Abstract - *The consumption of content has significantly increased in the internet age, with music being one of the most profitable and widely studied forms of media. Various factors, such as sex, music preference, and individual personality, have been considered in the development of a Music Sentiment Detection Platform (MSDP). Deep Neural Networks (DNN) have been employed for classification, enabling near-accurate prediction of music emotions. Russell's Two-Dimensional Emotion Model has been used to identify musical characteristics and categorize songs. After classifiers have been trained using the dataset, a collection of unknown songs has been evaluated to assess the model's accuracy. Musical features, including tempo, pitch, rhythm, and harmony, have been analyzed to classify songs into emotional categories. This platform has been shown to be highly beneficial for music streaming services, content creators, and therapeutic applications by enabling personalized music recommendations based on listeners' emotional states. Furthermore, dynamic mood detection has been supported, assisting users in discovering music that aligns with or alters their current emotional experience. Through its automatic recognition of emotional nuances, the platform has offered a novel approach to enhancing real-time user interaction with musical content.*

***Key Words*: AI, DNN, MSDP, ML, Model, Arousal and Valance, Music sentiment, Music**

1. INTRODUCTION

Music has always played a significant role in human history, influencing culture and emotions across civilizations. Today, with easy access to vast music libraries, modern listeners interact with music daily. As the volume of musical content continues to grow, accurately categorizing music based on emotional responses has become essential. Music is a powerful medium capable of evoking strong emotions, and understanding its impact is critical in various domains, including content personalization and emotional well-being. In this digital age, there is a growing interest in understanding how music influences emotions and how this knowledge can be applied to develop consumer products tailored to individual emotional needs.

Artificial intelligence (AI) and advanced computing have been increasingly employed to explore this phenomenon. Many studies have established a correlation between music and emotional responses. By leveraging machine learning, this project aims to deepen the understanding of this relationship. The project seeks to simulate the brain's response to music by composing music autonomously based on given inputs, showcasing the application of AI in creating music that reflects specific emotional states. This work attempts to reveal the intricate connection between music and emotions, offering new insights into how technology can replicate and understand human emotional experiences through music.

1.1 Motivation

Scrutinizing the correlation between music and invoked emotion has been our major motivation behind the project. Invoked emotion varies from individual to individual. But regardless of anything, tags generated by MSDP can be used to further investigation listening habits of individuals and can contribute to better the overall listening experience. For a long time, psychologists have been attempting to comprehend how music affects and preserves emotion. But due to factors such as age, sex, cultural background, musical taste, etcetera proper studies have been hard to perform. With advanced technology, we hope to compile larger data faster and give proper results. On other hand, Music emotion is biased from person to person. So, predicting the emotion of the music is a little ambiguous. This project aims to predict the emotion of music using a machine learning model. If the whole process were to be done manually it would take a long time and the probability for human error is also high. To accurately compute the data, we need the help of AI to reduce the errors and increase accuracy.

2. LITERATURE REVIEW

Research in music emotion analysis has progressed significantly, yet there remains room for improvement. Music evokes not only emotional responses but also physiological effects, prompting the use of artificial intelligence (AI) to evaluate these emotional movements. Machine learning (ML) approaches typically encompass three main tasks: data preparation, training, and evaluation. Various ML models have been utilized to identify the emotions associated with music, including Deep Neural Networks (DNNs), Linear Regression, Random Forests, and Support Vector Machines (SVMs). Research has compared these models regarding their flexibility and interpretability, highlighting the performance and generalizability of each [1]. Emotional judgments and physiological responses are collected from participants who listen to different audio pieces, assessing the emotions evoked without distinguishing between "perceived" and "felt" emotions. This evaluation is often structured around the Valence-Arousal (VA) Emotion Model, which categorizes emotions along two axes: Valence (ranging from pleasant to unpleasant) and Arousal (from calm to excited) [2].

In the data preparation phase, audio samples are gathered and converted to a standardized format, followed by feature extraction through various algorithms. This involves using techniques like the spectral contrast algorithm and software tools such as PsySound and Marsyas. Emotional assessments from listeners serve as critical inputs, allowing researchers to train regression models that predict emotions based on audio features. One notable study by Yi-Hsuan Yang and colleagues in 2014 approached Music Sentiment Detection (MSD) as a regression problem, creating a database of popular music across genres, normalized to a uniform format [3]. They utilized the VA model, extracting features and collecting listener assessments to establish AV values, which helped visualize each music sample in a data space associated with its emotional impact. In contrast, another study by Tong Liu and collaborators employed Convolutional Neural Networks (CNNs) for emotion recognition, illustrating a different methodological approach while still relying on the VA model [4]. These studies underscore the significance of feature extraction and the VA model in accurately determining emotions in music. As computational demands have increased, researchers are increasingly turning to cloud services like AWS to handle the substantial processing power required for model training. This shift from high-performance local machines to scalable cloud solutions has made it more accessible for researchers and students, especially those with limited budgets. AWS enables rapid data processing and efficient model training, significantly reducing the time and costs associated with MSDP projects. Despite the advancements made in music sentiment detection, there is still potential for enhancing model performance and generalization. Ongoing research aims to refine algorithms and improve feature extraction techniques, striving to create more precise and adaptable models that can better capture the intricate emotional nuances found in music. The continued evolution in this field promises to deepen our understanding of how machines interpret the emotional effects of music, making music sentiment detection a vital area within AI and ML research.

In our work analysis, we recognized the necessity of using an existing dataset rather than collecting new data. Then, we opted to utilize a dataset from Deezer, featuring over 18,000 songs. We also implemented a user-friendly HTML interface for uploading audio. After evaluating different systems, we chose to work with Deep Neural Networks for their user-friendly features and efficient performance compared to other models.

3. METHODOLOGY

Music is composed of various elements such as rhythm, melody, chords, and timbre, which influence human emotional perception. Studies indicate that music eliciting similar emotions often shares comparable features. This relationship allows for the modeling of the brain's processing mechanisms using advanced machine learning techniques. Deep Neural Networks (DNNs) are particularly effective for classification tasks and can predict musical emotions with high accuracy, depending on the parameters and sample size [5]. In our Music Sentiment Detection Platform (MSDP), we utilize supervised learning with DNNs to forecast arousal and valence values, which we then map onto Russell's 2D Valence-Arousal plane.

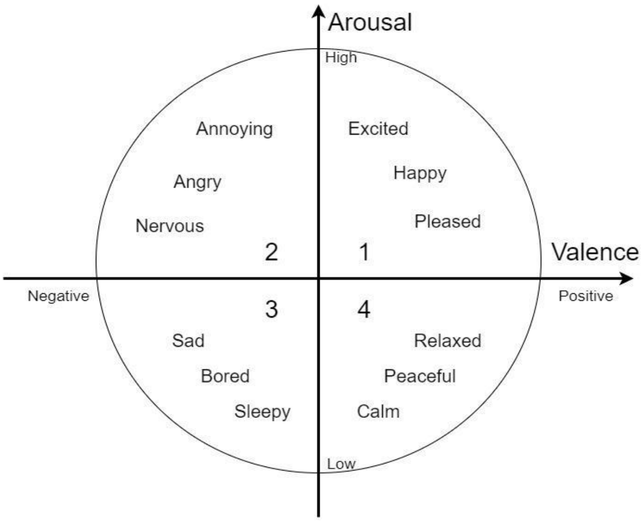


Fig-1: 2D Valence-Arousal Emotion Space

Our approach involves creating two separate DNN models dedicated to predicting arousal and valence, employing a consistent strategy across both models.

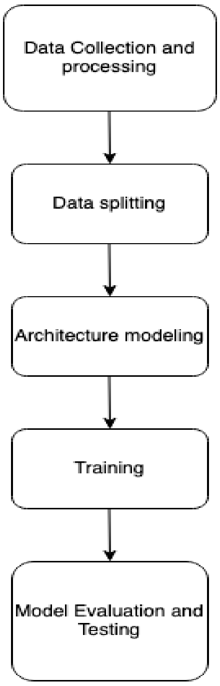
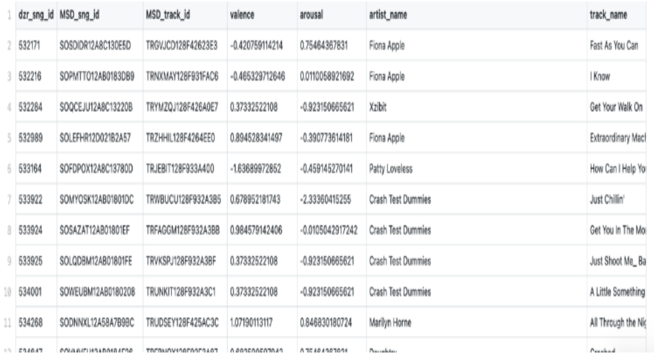


Fig-2: System Block Diagram

Data preparation is a critical first step in developing machine learning models, involving the selection or creation of an appropriate dataset. For our project focused on predicting arousal and valence in music, we chose the Deezer dataset [6], which offers relevant emotional data across 18,000 entries. We accessed audio files using the Deezer API [7], ultimately refining our collection to 13,793 usable files, divided into 8,635 training, 3,010 validation, and 2,148 test samples.



dzr_sng_id	MSD_sng_id	MSD_track_id	valence	arousal	artist_name	track_name
532171	SOSDIDR12A8C130E5D	TRGVJCD128F42623E3	-0.420759114214	0.75464367831	Fiona Apple	Fast As You Can
532216	SOPMTTO12AB0183DB9	TRNXMAY128F931FAC6	-0.465329712646	0.0110058921692	Fiona Apple	I Know
532284	SOQCEJU12A8C13220B	TRYMZQJ128F426A0E7	0.37332522108	-0.923150665621	Xzibit	Get Your Walk On
532989	SOLEFHR12D021B2A57	TRZHHIL128F4264EE0	0.894528341497	-0.390773614181	Fiona Apple	Extraordinary Macl
533164	SOFDPOX12A8C13780D	TRJEBIT128F933A400	-1.63689972852	-0.459145270141	Patty Loveless	How Can I Help Yo
533922	SOMYOSK12AB01801DC	TRWBUCU128F932A3B5	0.678952181743	-2.33360415255	Crash Test Dummies	Just Chillin'
533924	SOSAZAT12AB01801EF	TRFAGGM128F932A3B8	0.984579142406	-0.0105042917242	Crash Test Dummies	Get You In The Mo
533925	SOLQDBM12AB01801FE	TRVKSPJ128F932A3BF	0.37332522108	-0.923150665621	Crash Test Dummies	Just Shoot Me_ Ba
534001	SOWEUBM12AB0180208	TRUNKIT128F932A3C1	0.37332522108	-0.923150665621	Crash Test Dummies	A Little Something
534268	SODNNXL12A58A7B9BC	TRUDSEY128F425AC3C	1.07190113117	0.846830180724	Marilyn Horne	All Through the Ni

Fig-3: Deezer mood detection dataset

To extract essential audio features, we utilized the Python package Librosa, focusing on spectral centroid, spectral Rolloff, and Mel-frequency cepstral coefficients (MFCC) [7]. The model architecture was optimized using Bayesian optimization via the Keras Tuner library, leading to a valence model with 20 hidden layers (learning rate: 0.001) and an arousal model with 32 hidden layers (learning rate: 0.0001).

During training, we employed the Adam optimizer and mean squared error as the loss function, using a batch size of 256 over 20,000 epochs. The model's accuracy was evaluated with an unseen test set, providing insights into its performance in predicting emotional dimensions based on the audio features extracted from the dataset.

The dataset containing the tag and sample will be then fed to the model. Model using different architecture will be tested and the best performing architecture will be used. With the help of a microphone, we will be able to record sounds to further analyze and process the information.

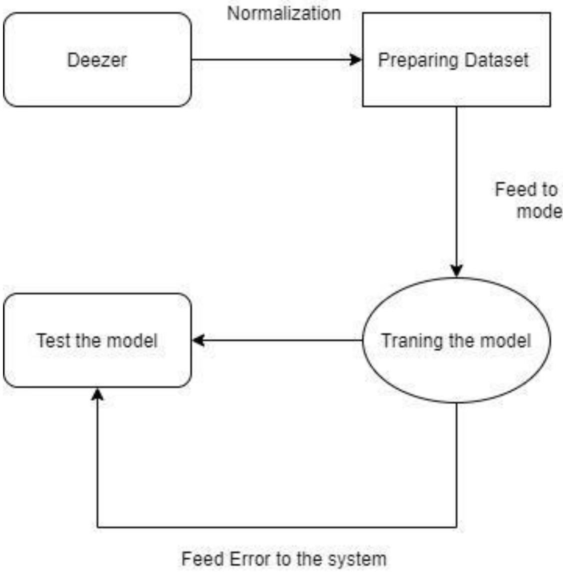


Fig-4: Working Principle

4. RESULTS AND ANALYSIS

Two Deep Neural Network models were trained on the Deezer dataset, aimed at predicting arousal and valence from audio features. The arousal model achieved a Mean Squared Error (MSE) of 0.61%, while the valence model performed slightly better with an MSE of 0.48%.

The formula for Mean Squared Error (MSE) can be expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - p_i)^2$$

To optimize the model's performance, hyperparameter tuning was conducted using both Random Search and Bayesian Optimization algorithms. Bayesian Optimization proved more effective, yielding superior parameters compared to those obtained through Random Search. In an experiment, careful attention must be given to the parameters. Without proper adjustment and optimization, the outcomes may not reflect the best possible results, potentially leading to issues with the overall findings. We conducted 10,000 iterations on music samples. When using a very small learning rate (e.g., 0.001), the learning process becomes excessively slow, and recognition tends to be inconsistent. On the other hand, a high learning rate (e.g., 0.1) results in instability and may even degrade performance. Our test was conducted on 30 voluntary participants and overall theme of this musical is predominantly sad. We collected the specific emotions of participants which is shown in below table.

Table-1: Emotional feelings of participants.

No. of Participants	Emotions
1	Expectation, panic
2	Interest, disgust, pain, fear
3	Fear, sorrow
4	Pain
5	Fear, anger
6	Panic, sorrow
7	Boredom, sadness
8	Fear, sadness
9	Surprise, sadness, fear
10	disgust, fear, sadness, surprise

The outcomes assessed by the music emotion model are presented in below table.

Table-2: Test results of model

Name	Emotions
DNN Model	Pain, anger, surprise, Sadness, sadness, pain, disgust, surprise, surprise, anxiety, and so on

To enhance user interaction with the AI models, a decision was made to develop a simple Graphical User Interface (GUI). This GUI is intended to facilitate user interaction with the models on a website platform. The website already has a basic functionality that allows users to upload audio files, which are then processed by the trained models. The goal of this web interface is to make the model predictions accessible and user-friendly, offering an intuitive method for users to engage with the underlying machine learning system. With the combination of optimized models and a user-friendly web interface, the project aims to deliver accurate audio analysis in a straightforward manner.

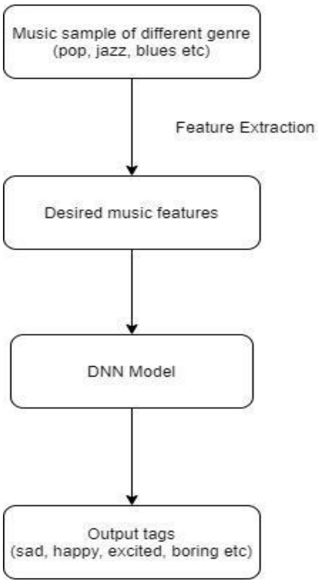


Fig-5: Outcomes

5. CONLUSIONS

In this study, we successfully demonstrated the potential of artificial intelligence and machine learning in recognizing emotions elicited by music. By building two distinct Deep Neural Network (DNN) models for predicting arousal and valence, we provided a framework that effectively maps emotional states in a two-dimensional arousal-valence (AV) plane. The hyperparameter tuning process, particularly through the use of Bayesian Optimization, resulted in improved model performance, surpassing other optimization techniques. Our research underscores the broader applications of emotion recognition in fields such as music theory, psychology, and AI, offering a deeper understanding of the intricate relationship between music and human emotion. The successful prediction of emotional states through machine learning opens up new avenues for interdisciplinary research, as this approach can be applied in various domains for emotion-based content analysis and user interaction systems.

Furthermore, the project presented several challenges, particularly in model training and optimization, but it was both an intellectually rewarding and technically stimulating endeavor. This research can serve as a foundation for future studies aiming to enhance emotion recognition systems, offering potential improvements in AI-based emotional interaction in music and other related fields.

REFERENCES

[1]. Vempala, Naresh N.; Russo, Frank A.;, "Modeling Music Emotion Judgments Using Machine Learning Methods," 5 January 2018. [Online]. Available:

https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02239/full. [Accessed 15 september 2020].

[2]. Citron, Francesca M.M.; Gray, Marcus A.; Critchley, Hugo D. ; Weekes, Brendan S. ; Ferstl, Evelyn C.;, "Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework," April 2014. [Online]. Available:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4098114/#:~:text=Emotional%20valence%20describes%20the%20extent,1997%3B%20Russell%2C%202003. [Accessed 15 september 2020].

[3]. Yang, Yi-Hsuan; Lin, Yu-Ching; Su , Ya-Fan; Chen, Homer H.;, "A Regression Approach to Music Emotion Recognition," April 2014. [Online]. Available: https://www.researchgate.net/publication/261471943_yangtaslp08_emo_regression. [Accessed 1 September 2020].

[4]. Lui, Tong; Han, Li; Ma, Liangkai; Guo, Dongwei;, "Audio-based deep music emotion recognition," 23 may 2018. [Online]. Available: https://aip.scitation.org/doi/pdf/10.1063/1.5039095. [Accessed 2 semtember 2020].

[5]. E. Allibhai, "Towards Data Science," [Online]. Available: https://towardsdatascience.com/building-a-deep-learning-model-using-keras-1548ca149d37.

[6].”GitHub," [Online]. Available: https://github.com/deezer/deezer mood detection dataset..

[7]. "Deezer," [Online]. Available: https://developers.deezer.com/login?redirect=/api..\