Multi-Model Approach to Recommend Personalized Music Playlist

Hansi Pabasara Sumanasekara
Faculty of Computing
SLIIT
Malabe, Sri Lanka
hansisumanasekara@gmail.com

Chamli Malshan Gunasekara
Faculty of Computing
SLIIT
Malabe, Sri Lanka
malshan.guanasekara@gmail.com

Sahan Dhananjaya
Faculty of Computing
SLIIT
Malabe, Sri Lanka
email address or ORCID

Kalpana Fernando
Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
kalpanafernando0421@gmail.com

Thusithanjana Thilakarathna
Faculty of Computing
SLIIT
Malabe, Sri Lanka
thusithanjana.t@sliit.lk

Dharshana Kasthurirathna
Faculty of Computing
SLIIT
Malabe, Sri Lanka
dharshana.k@sliit.lk

***Abstract*—In today's modern world, music holds a pivotal role in the lives of individuals, offering diverse benefits and experiences. Despite the existence of numerous systems for recommending personalized music playlists, there remains room for enhancement to elevate the user experience. This paper proposes a novel approach that leverages multi-modal analysis, integrating selfie images, surrounding images, and voice clips to extract user age, gender, current weather, and current emotion. By amalgamating these inputs, the system aims to curate a personalized music experience tailored to enhance user emotions and overall satisfaction. This research contributes to the evolving landscape of personalized music recommendation, offering insights into the integration of diverse modalities to enrich user experiences and foster emotional well-being.**

***Index Terms*—Music Recommendation Systems, CNN, RNN, Classification Models, Audio Signal Analysis**

I. Introduction

In today's digital age, personalized recommendation systems have become integral to various aspects of our lives, including entertainment, health, education and more. Among these, music recommendation systems hold a significant place, catering to the diverse tastes and preferences of users. To help users overcome the dilemma of choosing songs and prevent them from sticking to the same music repeatedly, platforms such as Spotify and Apple Music utilize advanced music recommendation algorithms [2]. These algorithms have been explored and discussed by various researchers [10] [14] [13].However, enhancing the accuracy and personalization of such systems remains an ongoing challenge. In this paper, we propose a groundbreaking approach — the Multi-Model Approach to Recommend Personalized Music Playlists —integrated within our innovative mobile application, MeloWave. Our research introduces several novel components to augment the efficacy of music recommendation systems. Firstly, we delve into the realm of user-specific analysis, utilizing image processing techniques to predict age and gender profiles. Leveraging Convolution Neural Network (CNN) models, inspired by seminal works [3], we extract valuable insights from user images to tailor music recommendations to individual preferences more accurately. Furthermore, we extend our investigation to consider contextual factors influencing users' music choices. Integrating weather prediction data into our recommendation framework, inspired by the work of Verma et al. We recognize the impact of environmental surroundings on music preferences, thus ensuring recommendations align with users' current moods and situations. Emotion detection is another pivotal aspect we address, aiming to enhance recommendation precision. By employing Long Short-Term Memory (LSTM) models and the RNN algorithm, we delve into the intricate realm of vocal responses to discern users' emotional states accurately. This allows MeloWave to curate playlists that resonate with users on a deeper emotional level. Finally, we explore the dynamic nature of users' emotional states concerning machine-generated playlists. We investigate how users' emotional responses evolve in response to playlist recommendations, facilitating the continuous refinement of our recommendation algorithms. Through the integration of these innovative components, MeloWave represents a significant advancement in the realm of music recommendation systems, offering users a seamless and deeply personalized listening experience. In the subsequent sections, we delve into the methodologies, experimental setups, results, and implications of our multi-model approach, showcasing its efficacy and potential for revolutionizing music recommendation systems.

Identify applicable funding agency here. If none, delete this.

II. Literature Review

Music recommendation has been a major aspect of life these days. There are many platforms that we can use to listen to music. Some of them are our traditional and oldest: YouTube, Spotify, iTunes, and many others. There are so many mechanisms to give filtered recommendations, like content-based filtering, collaborative filtering, and hybrid filtering [11].

[1] Recommender systems were created to bridge that gap between information gathering and analysis by filtering all available data to offer only what is most important to the user. Some research has found that content-based filtering similarity results reach up to eighty percent similarity for the song and fifty percent similarity for the artist, which means this type of filtering works well for our recommendation system. Therefore, there are already tested and proven machine learning algorithms in use in recommender systems [3]. This proposed system will take this approval another step ahead and use user emotions, surroundings, and many other inputs to enhance the accuracy of the recommender system along with the content, collaborative, and hybrid filtering algorithms. The cold start problem comes into play with these recommendation algorithms, and it was a notable problem found in the above algorithms. The ”cold start” problem means that when a user first logs in to the system, there is no user history or input for these algorithms to run. This problem has also been addressed by various research projects so far and has many kinds of solutions. For instance [6], this research mitigates the cold-start problem by using matrix factorization and spatial information for users with few restaurant visits in the past. The research [7] is using micro-services, to classify the cold start problem to smaller services like Geo services, user-profile services etc. to mitigate above problem. This approach is also another successful aspect to mitigate the cold start problem. The performance of collaborative filtering by utilizing the rating information of common users who belong to multiple domains and [12] is discussing about both the sparsity (inadequate rating information) and the cold start (no rating information at all) problems. In this proposed system, we are hoping to address this problem by predicting a user profile with the use of image processing, object detection, feature extraction, etc. Then the main challenge is to accurately extract and identify the sensitive details of the user by using a selfie. So far, there has been research conducted emphasizing this matter [3] [4]. One of the studies we found has [4] three neural network-based models to detect age, gender, and emotion, respectively, and depending on this combination, a personalized playlist has been suggested. In this case, only those combinations of inputs are sent to the recommender system, and in our research, we are predicting a user profile at the very beginning, which will be combined with many other inputs like surroundings and voice-based emotion detection to enhance accuracy and personalization. Moving to emotions, the human voice is a strong tool for communicating emotions, with nuances in pitch, tone, rhythm, and other acoustic characteristics. Voice-based emotion recognition research has advanced greatly, with studies utilizing approaches such as Mel-Frequency Cepstral Coefficients (MFCCs), prosodic characteristics, and deep learning models [9] [5]. The results of these studies highlight the possibility of properly recognizing emotional states in real time from speech signals. Despite advances in voice-based emotion recognition as well as customized recommendation systems, there is still a research gap in the seamless integration of these technologies for real-time engagement. The integration challenge has several dimensions, including creating an intuitive user interface for voice input recognition [5], algorithms that generate emotionally coherent playlists [8], and responding contextually to users' emotional cues [9]. The problem of creating emotionally cohesive playlists that represent customer's emotional experiences is yet unsolved. Existing research focuses on mood-boosting recommendations [8] however, there is a gap in research on keeping emotional consistency while respecting user preferences. With the combination of age, gender, current weather, and the emotional status of the user, the proposed system can provide a tailored music experience to the user, enhancing user emotions and satisfaction.

III. Methodology

”Melowave” is a mobile application developed for personalized music recommendations utilizing machine learning and artificial intelligence. Data gathering was crucial for us to identify user groups, age groups, weather impact, and emotional impact of the user to ensure the success of this system. We evaluated to collect data and identify factors, which are mentioned below *(fig. 1) (fig. 7)*.

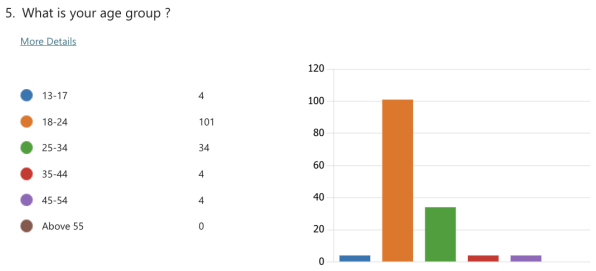

Fig. 1. Identify the interest towards music on different age groups

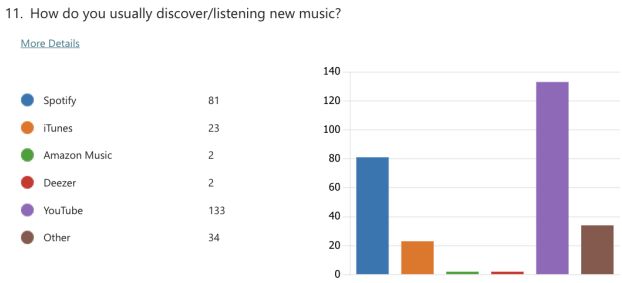

Fig. 2. The platforms users are utilizing to listen to music

By the survey we have proved *(fig. 3)* that music is impacting people's feelings, energy levels, moods, and health. Also, we have identified that a large proportion of people are using 'YouTube' to listen to music according to our survey. *(fig. 2)*.

For the music recommendation system, we have employed various tools and technologies, which are outlined below:

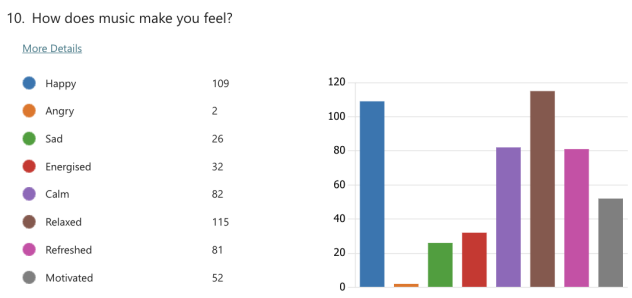

Fig. 3. Fluctuation of emotions by listening to music

A. Tools and technologies

1) ***Python and google colaboratory****:* We have utilized Python in conjunction with Google Colaboratory to create machine-learning models for our proposed system. Specifically, we chose Python for its ability to leverage essential libraries like NumPy, pandas, TensorFlow, PyTorch, and others to build and train machine learning models for our system.

We chose Google Colaboratory as our development platform for its comprehensive support, including free access to GPUs and CPUs that accelerate model training.

Utilizing Google Colaboratory alongside Python, we efficiently conducted data preprocessing, analysis, tuning, and optimization. These tools were crucial in evaluating and validating our models, allowing for easy visualization of metrics such as loss and accuracy. After model training and evaluation, we effortlessly deployed and integrated these models into our mobile application using the same platform and programming language.

2) ***Kaggle****:* We have chosen Kaggle, a well-known platform for data science projects, for its extensive educational content related to data science and machine learning. . We have selected a few datasets from Kaggle to develop our models, which are mentioned below.

Weather-prediction

- https://www.kaggle.com/datasets/vijaygiitk/multiclass-weather-dataset/

Age prediction

- https://susanqq.github.io/UTKFace/

Gender prediction

- https://susanqq.github.io/UTKFace/

Voice emotion recognition

- https://www.kaggle.com/datasets/ejlok1/toronto-emotional-speech-set-tess
- https://www.kaggle.com/datasets/ejlok1/cremad/data
- https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio
- https://www.kaggle.com/datasets/ejlok1/surrey-audiovisual-expressed-emotion-savee

3) ***TensorFlow****:*
4) ***Flutter****:*
5) ***Visual Studio Code****:*
6) ***FireBase****:*

B. System overview

In the proposed solution for a multi-model music recommendation system, we have identified four main sub-components. The System overall Diagram is illustrated below *(fig. 4)*.

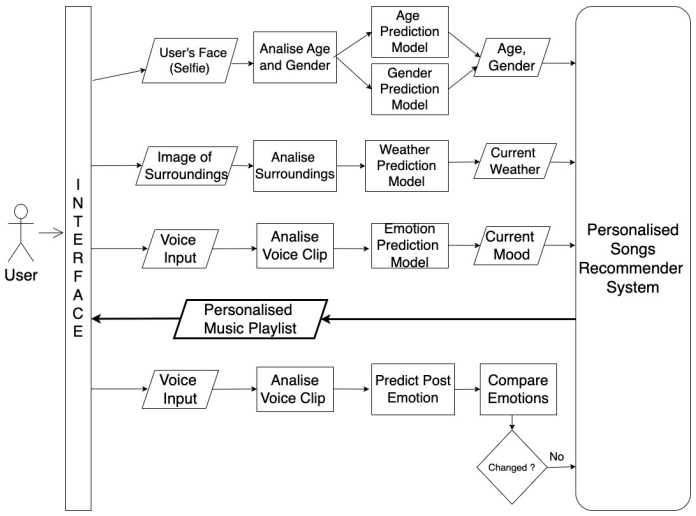

Fig. 4. System Overview Diagram

As the above diagram *(fig. 4)* illustrates, we have four main sub-components to achieve the proposed system. The first component is to retrieve an image of the user for analysis and extract the user details like age and gender. It will be passed to the recommender system as an input.

The second component is gathering an image of the surroundings to identify the current weather and the extracted weather will be sent to the recommender system as another input.

The third component is to gather a voice input from the user to extract the current emotion of the user which will be sent to the recommender system as the last input.

After getting all the above inputs, the music recommendation models generate a tailored playlist for the enhancement of the entertainment of the particular user accordingly. Moreover, there is an evaluation of the post-emotions of the user to enhance the music recommendation in the future for the particular user.

From the user's perspective, the 'Melowave' mobile application initiates with a window prompting the user to sign into a Google account. Upon completion, the application prompts the user to take a selfie using the camera *(fig. 4)*.

Subsequently, it progresses to the next step in the stepper, which again utilizes the camera to capture the surroundings. In the third and final step, the stepper prompts the microphone, enabling the user to record a voice clip *(fig. 6)*.

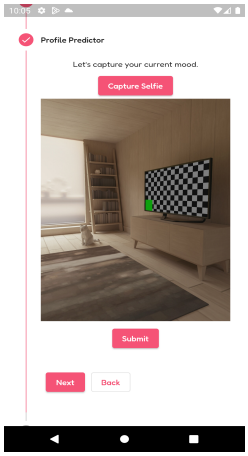

Fig. 5. Profile Predictor UI

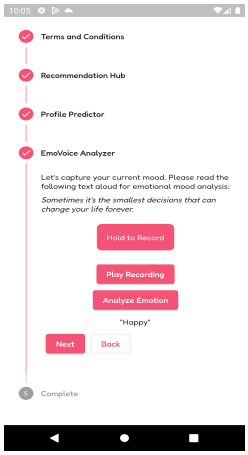

Fig. 6. Voice Recorder UI

After clicking the ”next” button, the trained and evaluated machine learning models, including the age prediction model, gender prediction model, weather prediction model, and emotion prediction model, will run and produce outputs. These outputs comprise the user's age, gender, current weather conditions, and current emotion. Additionally, there will be an option to edit the predicted outputs. Subsequently, the music recommender model will execute and recommend a personalized music playlist tailored to the user's age, gender, current weather, and current emotions. The primary functionality will include these features, with potential scenarios where the user may not be satisfied with the playlist or their mood may not change or enhance due to the generated playlist. In such instances, the post-emotion classification model will intervene. The user can record three voice clips and submit them to the system. Upon submission, the post-emotion classification model will analyze the emotion of the user and compare it with previous voice clips and emotions. If the emotion has not improved, the music recommender algorithm will execute again, generating a new personalized song playlist for the user. Furthermore, our proposed mobile application includes various external features to enhance the user experience.

1) ***Gender identification through image processing****:* As the first component extracting user gender, we have used the CNN algorithm, image processing, deep learning methods, and the pre-trained model – “EfficientNetB3” from the Kaggle. The following images *(fig. 7)* convey the accuracy levels of the Gender Classification model after evaluation.

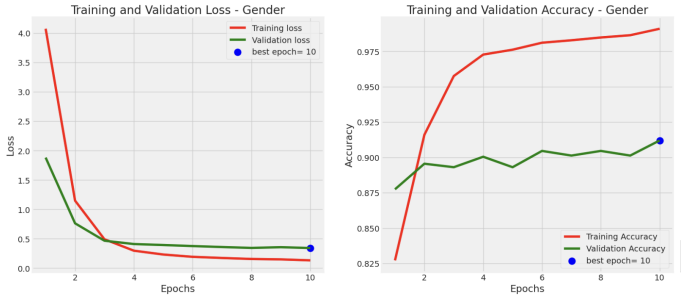

Fig. 7. Loss and the Accuracy of Gender Prediction Model

2) ***Age identification through image processing****:* Same as the Gender classification model, here also we have used the CNN algorithm, image processing, deep learning methods, and the pre-trained model – “EfficientNetB3” from Kaggle. The Confusion matrix of the Age prediction model is as follows *(fig. 8)* .

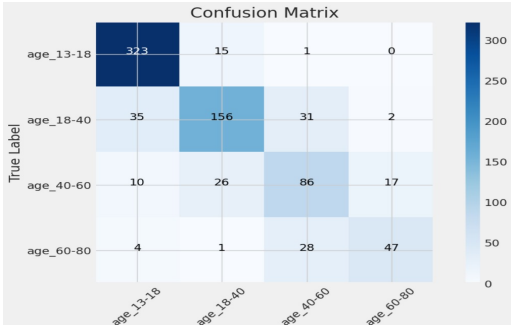

Fig. 8. Confusion matrix of Age prediction model

3) ***Weather classification using an image****:* Moving to weather prediction using an image, the CNN algorithm, image processing, and deep learning methods have been used. The graphical representation of the model evaluation is displayed below *(fig. 10)*.

4) ***Emotion extraction through voice clips****:* Emotion extraction through voice clip has been done using an emotion classification model which is consisting an RNN algorithm, Deep learning methods, and LSTMs.The accuracy of the trained model is given below *(fig. 10)*.

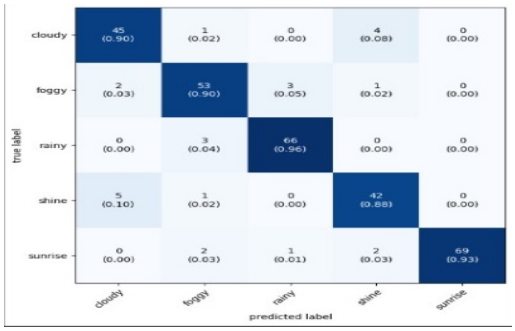

Fig. 9. Confusion matrix of weather prediction model

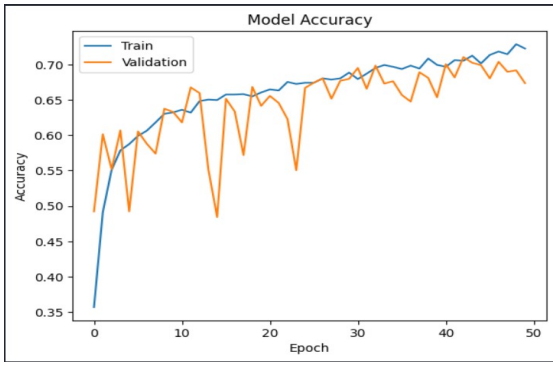

Fig. 10. Accuracy of Emotion Detection model

REFERENCES

[1] Jonas Theon Anthony, Gerard Ezra Christian, Vincent Evanlim, Henry Lucky, and Derwin Suhartono. The utilization of content based filtering for spotify music recommendation. In *2022 International Conference on Informatics Electrical and Electronics (ICIEE)*, pages 1–4, 2022.

[2] Ce1ma and P. Lamere. If you like the beatles you might like...: a tutorial on music recommendation. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 1157–1158. ACM, 2008.

[3] P. Darshna. Music recommendation based on content and collaborative approach and reducing cold start problem. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pages 1033–1037, Coimbatore, India, 2018.

[4] J. Jayakumar and P. Supriya. Cnn based music recommendation system based on age, gender and emotion. In *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, pages 1356–1359, Coimbatore, India, 2022.

[5] K. S. Krupa, G. Ambara, K. Rai, and S. Choudhury. Emotion aware smart music recommender system using two level cnn. In *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, pages 1322–1327, Tirunelveli, India, 2020.

[6] E. Mayrhuber and O. Krauss. User profile-based recommendation engine mitigating the cold-start problem. In *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–6, Maldives, Maldives, 2022.

[7] Elisabeth Mayrhuber and Oliver Krauss. User profile-based recommendation engine mitigating the cold-start problem. In *2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, pages 1–6, 2022.

[8] V. Mounika and Y. Charitha. Mood-enhancing music recommendation system based on audio signals and emotions. In *2023 International Conference on Inventive Computation Technologies (ICICT)*, pages 1766–1772, Lalitpur, Nepal, 2023.

[9] K. R. Nambiar and S. Palaniswamy. Speech emotion based music recommendation. In *2022 3rd International Conference for Emerging Technology (INCET)*, pages 1–6, Belgaum, India, 2022.

[10] H.-S. Park, J.-O. Yoo, and S.-B. Cho. A context-aware music recommendation system using fuzzy bayesian networks with utility theory. In *International conference on Fuzzy systems and knowledge discovery*, pages 970–979. Springer, 2006.

[11] J. Singh. Collaborative filtering based hybrid music recommendation system. In *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, pages 186–190, Thoothukudi, India, 2020.

[12] Pradeep Kumar Singh, Pijush Kanti Dutta Pramanik, Garima Ahuja, Anand Nayyar, Vaibhav Pandey, and Prasenjit Choudhury. Mitigating sparsity and cold start problem in collaborative filtering using cross-domain similarity. In *2020 8th International Conference on Orange Technology (ICOT)*, pages 1–6, 2020.

[13] B. B. Sree and S. Demissie. A context-aware music recommendation. *International Journal of Research*, 5(7):497–504, 2018.

[14] X. Wang, D. Rosenblum, and Y. Wang. Context-aware mobile music recommendation for daily activities. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 99–108. ACM, 2012.