International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Mood Based Music Recommendation Using Voice Analysis

GVL Narayana [1], Vysyaraju Bhavya [2], Sanjeevi Sravya[3], Ragolu Nagalakshmi[4], Talasu Nikitha[5],

[1]Assistant Professor, Information Technology, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh, India
[2,3,4,5]UG Students, Information Technology, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh, India

Abstract
Music has a significant impact on people's ability to control their emotions, reduce stress, and increase productivity in today's fast-paced digital world. However, it can frequently be frustrating and time-consuming to choose music that fits a person's mood. In order to solve this, we are creating a mobile application that uses voice analysis to suggest music based on the user's present emotional state. Simply speaking into the microphone on their device allows users to send voice input to the AWS backend for safe and effective processing. The system extracts emotional features by analyzing speech patterns and voice tones using the Natural Language Toolkit (NLTK). After that, machine learning models categorize the mood into groups like energy, calmness, sadness, and happiness. The app immediately suggests music that fits the mood once it has been identified.

Keywords: Voice Analysis, Mood Detection, Machine Learning, Natural Language Tool Kit (NLTK), Amazon Web Services (AWS), Music Recommendation, Emotion Classification,

1. Introduction

Music has become very important to every age group in today's fast-paced and emotion-driven atmosphere. Normally, under the pressure of fast-paced appeals and emotionally dynamic lifestyles, most people spend time listening to music. Music helps humans cope and helps in relaxation, concentration, and improving overall mental health. Music is well capable of having a tie with emotional interference. Therapy through music can be applicable both at the level of self and socially. However, what is rarely achieved is the selection of good music for a particular time because moods change throughout the day and the individual's music library is growing larger and larger.

Most users always complain that they cannot find songs usually matching their moods and would spend time scrolling their playlists or searching for them on the Internet. Traditional music streaming apps depend much on manual selection to provide content and genre-based recommendation for songs. Hence, we are formulating a prototype for a mood-based app for music recommendation by voice analysis, AI-enabled mobile application for smoothening and customizing the journey to discover music. In this app, a user talks into their device microphone, and their voice input gets analyzed in real-time.

The audio data is secured for transmission to AWS cloud backend processed by the Natural Language Toolkit (NLTK). This tool extracts cues of emotion exhibited by the user's speech, including tone, pitch, speed, and sentiment features which go through machine learning algorithms to classify the mood into fixed emotion states like happiness, sadness, calmness, and energy.

Once the mood is detected, the system dynamically searches and immediately provides song recommendation relevant to the detected mood from a cloud-based music library. It is a fast, responsive, and scalable application and hooked up to deliver a pleasant experience. In addition, over a period, it learns from interactions with users and feedbacks to make improvements on its accuracy to give personalized suggestions with each usage. This way, according to the appropriately emotive music, users would not have to invest time searching.

The Expected Outcome is That the app is to recognize your emotions through your voice and then suggest music that really fits how you're feeling. It makes choosing music easy and offers a smooth, personalized listening experience. As you keep using it, you'll find yourself feeling more connected to the music and more happy with the recommendations and the app learns from your feedback, getting better and smarter over time. Overall, it shows how AI and emotion detection can come together to create a music experience that's more intuitive and enjoyable.

2. Literature Survey

Hu and Ogihara created the NextOne Player, a music recommendation system that focuses on user interaction data instead of analyzing the actual songs. It uses a concept called the 'Forgetting Curve' along with time series analysis to keep song suggestions feeling fresh and relevant. As a result, it helped reduce the number of songs skipped and made the listening experience more enjoyable by customizing suggestions to each person's habits.[1]

Jun et al., developed a music recommendation system that analyzes beat structures and uses k-medoids clustering to group songs based on their mood patterns. They also use the Smith-Waterman algorithm to compare mood sequences, and they factor in what users prefer. This way, their system can suggest music more accurately and find relevant songs more quickly compared to older methods.[2]9

De Prisco et al., created Moodify, a system that uses reinforcement learning to suggest music that helps users shift from how they’re feeling now to a different emotional state. They based it on a go-explore strategy and tested it on a million Spotify playlists. Users generally liked it, and it provided emotion-based music suggestions that felt accurate and personalized.[3]

Wang came up with a deep learning approach that combines deep belief networks and probabilistic graphical models to make music recommendations better, whether you're just starting out or have lots of listening history. Their model handles feature extraction and recommendations all in one go, and the results show it works really well, outperforming older methods and doing especially better when there's little or no initial data.[4]

Bayle et al., created Kara1k, a karaoke dataset used for recognizing cover songs and analyzing singing voices. It includes 1,000 cover versions and their original songs, along with details like metadata and audio features. Their research found they could identify cover songs with about more accuracy. They also pointed out that figuring out the singer's gender becomes tricky when background music is playing.[5]

Han et al., created a music recommendation system that considers your emotional state. Using a model called the Emotion State Transition Model (ESTM), it can both predict how you might feel and suggest music to help influence your mood. The system combines reasoning based on a structured knowledge base, techniques like Support Vector Machine (SVM) classifier to deliver personalized music suggestions.[6]

Agrawal et al., created a mood-based song recommendation system that uses a CNN along with OpenCV's Haar Cascade classifier to recognize emotions from facial expressions. The system sorts emotions into seven categories and then suggests songs through the Spotify Web API. It was able to correctly identify emotions about 76% of the time when tested on the FER2013 dataset.[7]

Zhou et al., created a new way to change the emotion in someone's voice without messing up who the person is or what they're saying. They tested their method, using a database of emotional speech, and looked at two model - Cycle GAN-EVC and VAWGAN-EVC. The results showed that their approach made the speech sound more expressive and clearer than before.[8]

Burch et al., created SEND/RETURN, a virtual world where multiple users can interact together. This space combines music suggestions with visual effects that change based on mood, all built with Unreal Engine 5. The system uses clustering techniques and a type of AI called CNN trained on Mel-spectrograms to figure out the mood of songs and generate matching visual effects in real-time, making the experience more immersive and engaging.[9]

Kim et al., developed a music recommendation system that combines emotional insights with collaborative filtering to make playlists more satisfying for users. Their method sorts speech emotions into six groups—neutral, happy, sad, angry, surprised, and bored—using a Support Vector Machine (SVM) with a radial basis function (RBF) kernel.[10]

3. Existing System

Face Recognition system systems detect emotions through facial expressions. They use image processing and machine learning techniques to identify facial movements, such as the shape of the mouth, eyes, and eyebrows. some disadvantages are FER systems require a clear view of the face, which might not always be feasible, especially in crowded or poorly lit environments. Additionally, they are less effective for users who consciously hide their emotions.

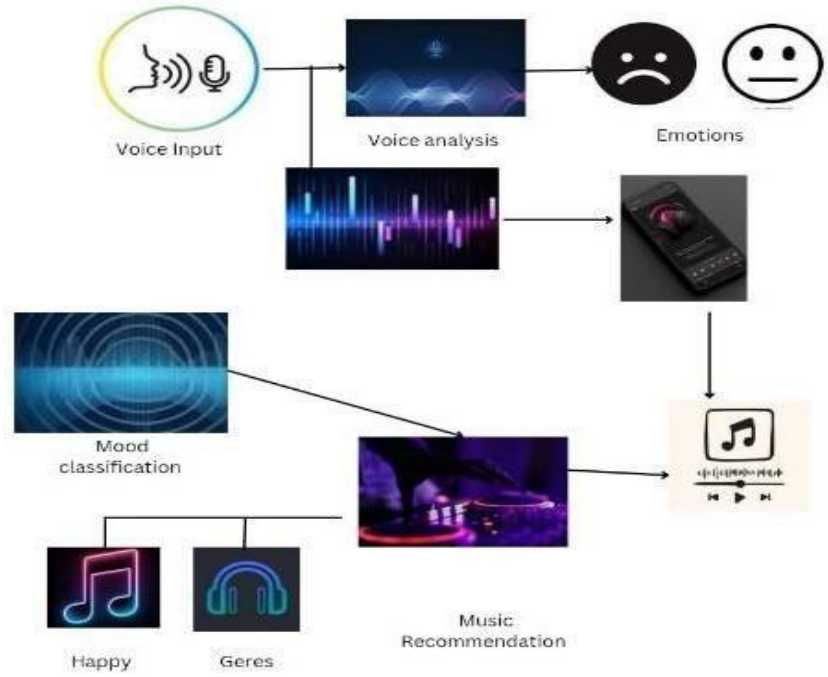
Wearable devices, such as smartwatches, monitor physiological data like heart rate, body temperature, or skin conductivity to infer emotional states. These sensors help detect stress or relaxation, providing insights into a user's emotional condition. some disadvantages are Wearables may have limited accuracy

due to individual differences in how emotions manifest physiologically, and they may not detect emotions accurately without continuous use.

These systems analyze text input from users (e.g., social media posts, chats, or reviews) to detect emotions such as happiness, sadness, or anger. They use natural language processing (NLP) algorithms to analyze word choice, sentence structure, and context. some disadvantages are Text-based systems may not account for sarcasm or context, leading to potential misinterpretation of emotions. They also require users to express their emotions through text, which may not always be accurate.

4. Methodology

The Proposed methodology for "Mood Based Music Recommendation Using Voice Analysis "project works by having the user speak into the device's microphone, and then sending that recording to an AWS server for analysis. Using tools like the Natural Language Toolkit (NLTK), the system picks up on emotional cues in the voice—like tone, pitch, and what's being said. Then, it uses machine learning to figure out the emotion—such as happiness, sadness, or high energy. Once it knows the mood, it pulls up songs that fit that feeling from an online music library. The system also keeps learning from user feedback, helping it get better and more accurate at suggesting music over time.



The process begins when the user speaks into their device's microphone. This voice input is captured as an audio signal and then converted into a digital format. The digitized audio is securely transferred to the AWS backend, which serves as the processing and storage centre for the application. This step ensures that the raw voice data is ready for in-depth analysis.

Once the audio reaches the backend, it undergoes audio signal processing. Here, the system uses advanced techniques like spectrograms**,** pitch**,** and tone analysis to extract emotional features from the

voice. These features help identify vocal patterns that indicate the user's emotional state, such as intensity, rhythm, and speech pace.

With the extracted features, the system uses Natural Language Toolkit (NLTK) and emotion recognition algorithms to interpret the user's mood. The emotional states are identified and categorized into common moods like happiness, sadness, calmness, and energy**.** This stage is crucial as it forms the foundation for selecting the most suitable music.

After detecting the general emotion, the system uses machine learning models to classify the mood more precisely. This classification helps in determining the best type of music that fits the current state of the user. The emotional label is then used to filter and fetch songs from a curated database of music.

Based on the classified mood, the system instantly recommends music that aligns with the user's emotional state. These songs are selected from a cloud-based music library and presented through the app interface. The recommendation can be further refined into genres or themes (e.g., upbeat for happy, soft instrumental for calm) for better personalization. Over time, the app learns from user feedback and preferences, making future recommendations even more accurate.

5. Designing The System

Architecture And Components

The diagram illustrates the complete workflow of the Mood-Based Music Recommendation System, beginning with the setup of the song application and integration of the microphone. It then captures the user's voice input, which is sent to the AWS backend for secure processing. Using Natural Language Toolkit (NLTK), the system performs voice analysis to extract emotional features. These features are classified into moods such as happy, sad, or calm, and the music recommendation engine suggests songs that match the detected mood. This ensures a personalized and emotionally intelligent listening experience for the user.

Song Application Setup: The first step in creating this app involves building the basic setup. It includes designing an easy-to-use interface and adding the necessary features like audio recording, mood detection, and music streaming. The goal is to make the app interactive, simple to navigate, and good at capturing voice data smoothly.

Microphone Integration: In this step, the app asks for permission to access your microphone. Once granted, it can record your voice in high quality. This setup lets you speak directly into your device, making it easy to use your voice for mood detection right from the app.

Voice Input: With voice input, we just speak naturally into the app, and it records what you say in real time. It picks up different parts of your speech like your tone, pitch, energy which can tell a lot about how you're feeling. After recording, this voice data is ready to be sent to the server for processing.

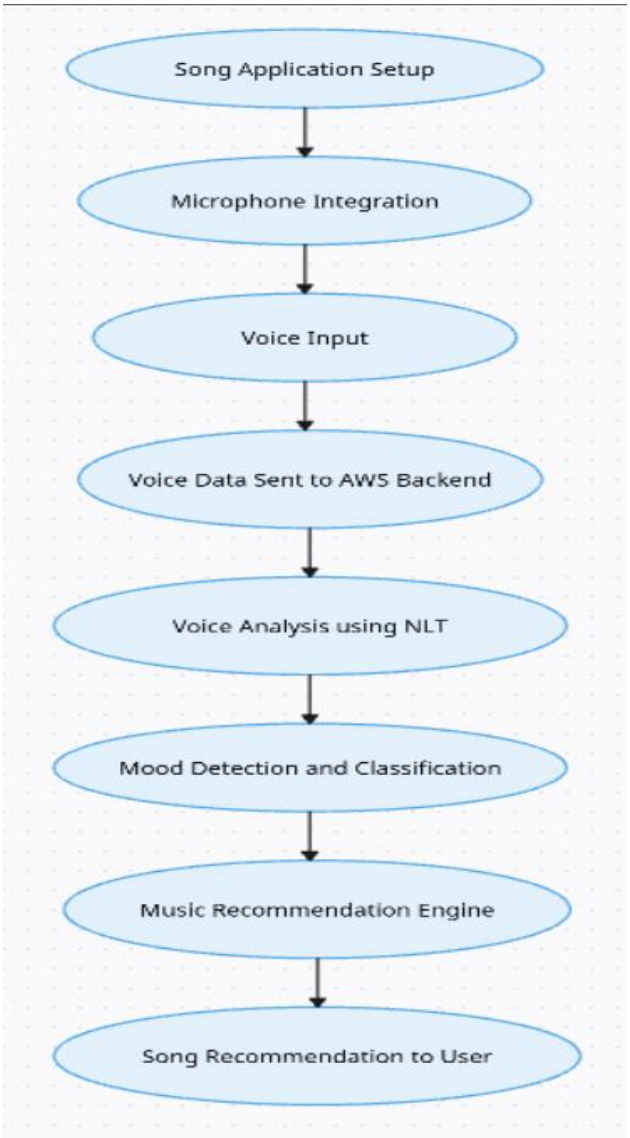
Voice Data Sent to AWS Backend: After we speak, our voice input is securely sent to a cloud server run by Amazon Web Services (AWS). This server is responsible for storing the data, providing the computing power needed, and quickly analyzing your audio for any necessary processing.

Voice Analysis using NLT: Using the Natural Language Toolkit (NLTK) alongside other audio processing tools, this system can analyze voice inputs. It looks at important emotional features like tone, speaking speed, the feelings conveyed in speech, and the words used (if there's a transcript). All these details help discover the emotional tone behind a person's voice.

Mood Detection and Classification: Mood detection and classification help figure out how someone is feeling, like whether they're happy, sad, calm, or full of energy. This part is really important because it guides us in choosing the right kind of music that matches their mood.

Music Recommendation Engine: The Music Recommendation Engine picks songs based on how we are feeling. It searches through online music platforms like YouTube Music or cloud-based song collections, looking for tracks that match certain moods or genres. The system aims to connect the emotional vibe you're experiencing with music that fits that tone, making your listening experience more personal and enjoyable.

Song Recommendation to User: Finally, the app shows you a personalized list of songs that match your mood. These recommendations pop up right on the screen, making it easy for you to discover music that fits how you're feeling. Over time, the app gets smarter by learning what you like and value based on your feedback, so it can suggest songs that are even more in tune with your taste.


Song Application Setup
Microphone Integration
Voice Input
Voice Data Sent to AWS Backend
Voice Analysis using NLT
Mood Detection and Classification
Music Recommendation Engine
Song Recommendation to User

6. Results and Discussion

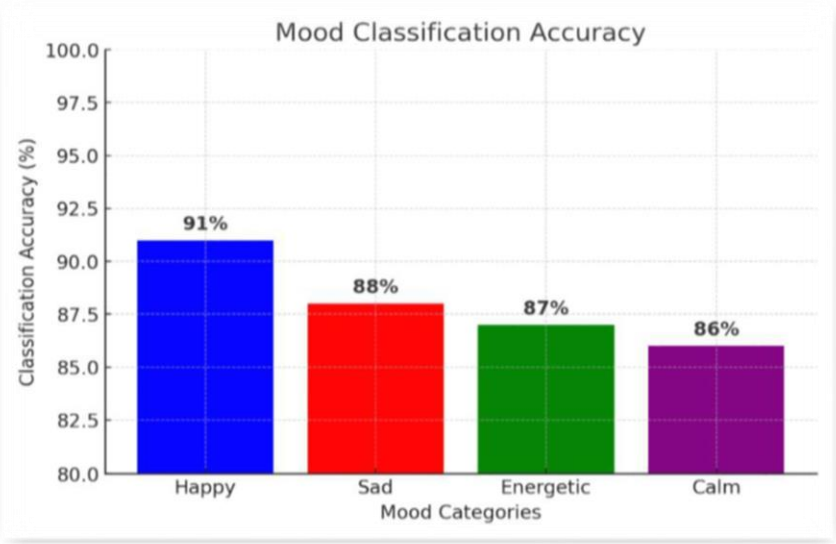
Mood Classification Accuracy:

The system was tested using multiple voice inputs corresponding to different emotions: Happy, Sad, Energetic, and Calm. The accuracy was determined by comparing the predicted moods with actual user emotions. Every time someone spoke, the system analyzed their voice to figure out how they were feeling. Then, it compared these guessed emotions with what the users actually intended to express. This helped us see how well the system was doing at recognizing moods.

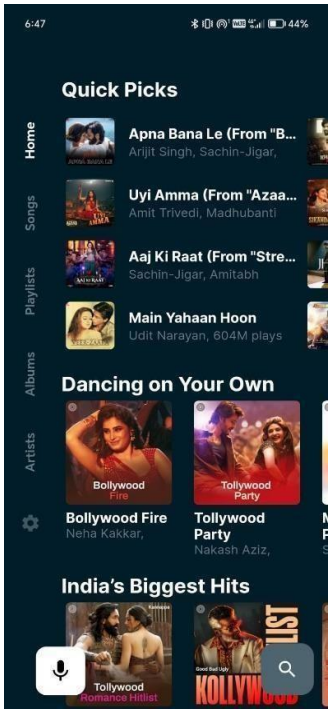
Table 1 : Emotion Detection Accuracy Summary

Mood	Accuracy (%)
Happy	91%
Sad	88%
Energetic	87%
Calm	86%

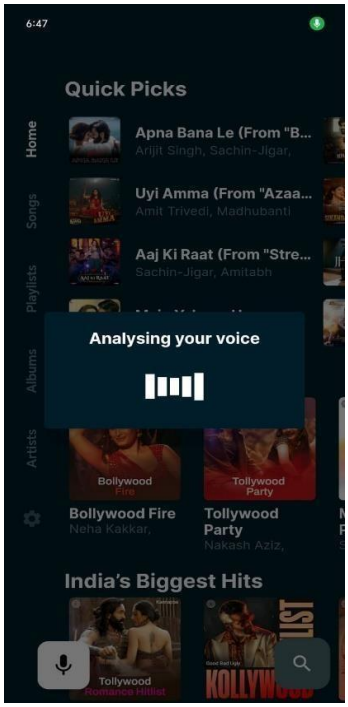
Figure 1 **:** Emotion detection Accuracy



When you open the app, you'll see a simple home screen showing personalized song picks and popular playlists. To get started, just speak into the microphone, and the app will listen and analyze your mood using AWS and NLTK. It then figures out if you're feeling happy, sad, energetic, or calm. Based on that, it suggests songs and playlists that match your mood. As you listen and give feedback, the app learns more about your preferences and keeps improving its suggestions, making your music experience more personal and emotionally tuned-in over time.

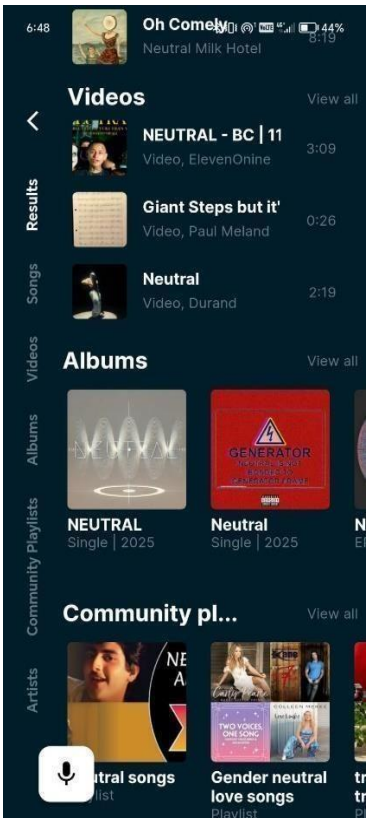


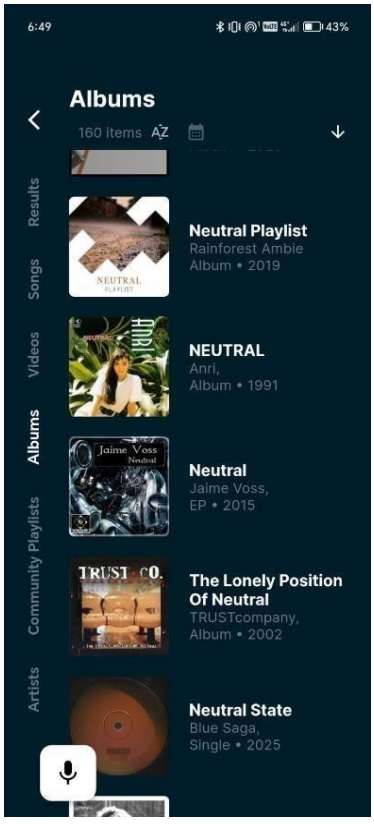
Home Page



Analysing Voice

Recommended Song





References

1. Y. Hu and M. Ogihara, “NEXTONE PLAYER: A MUSIC RECOMMENDATION SYSTEM BASED ON USER BEHAVIOR,” *Poster Sess.*, 2011.
2. J. Yang, “Personalized Song Recommendation System Based on Vocal Characteristics,” *Math. Probl. Eng.*, vol. 2022, pp. 1–10, Mar. 2022, doi: 10.1155/2022/3605728.
3. R. De Prisco, A. Guarino, D. Malandrino, and R. Zaccagnino, “Induced Emotion-Based Music Recommendation through Reinforcement Learning,” *Appl. Sci.*, vol. 12, no. 21, p. 11209, Nov. 2022, doi: 10.3390/app122111209.
4. X. Wang and Y. Wang, “Improving Content-based and Hybrid Music Recommendation using Deep Learning,” in *Proceedings of the 22nd ACM international conference on Multimedia*, Orlando Florida USA: ACM, Nov. 2014, pp. 627–636. doi: 10.1145/2647868.2654940.
5. Y. Bayle *et al.*, “Kara1k: A Karaoke Dataset for Cover Song Identification and Singing Voice Analysis,” in *2017 IEEE International Symposium on Multimedia (ISM)*, Taichung: IEEE, Dec. 2017, pp. 177–184. doi: 10.1109/ISM.2017.32.
6. B. Han, S. Rho, S. Jun, and E. Hwang, “Music emotion classification and context-based music recommendation,” *Multimed. Tools Appl.*, vol. 47, no. 3, pp. 433–460, May 2010, doi: 10.1007/s11042-009-0332-6.
7. S. Agrawal, S. Sharma, Y. Fulzele, R. Bhasarkar, and S. D. Pande, “Mood-based song recommendation system using CNN,” presented at the 4TH INTERNATIONAL SCIENTIFIC CONFERENCE OF ALKAFEEL UNIVERSITY (ISCKU 2022), Najaf, Iraq, 2023, p. 030003. doi: 10.1063/5.0187390.
8. K. Zhou, B. Sisman, and H. Li, “VAW-GAN for Disentanglement and Recomposition of Emotional Elements in Speech,” 2020, *arXiv*. doi: 10.48550/ARXIV.2011.02314.
9. C. Burch, R. Sprowl, and M. Ergezer, “A Multi-User Virtual World with Music Recommendations and Mood-Based Virtual Effects,” *Proc. AAAI Conf. Artif. Intell.*, vol. 37, no. 13, pp. 16063–16069, Jun. 2023, doi: 10.1609/aaai.v37i13.26908.
10. T.-Y. Kim, H. Ko, S.-H. Kim, and H.-D. Kim, “Modeling of Recommendation System Based on Emotional Information and Collaborative Filtering,” *Sensors*, vol. 21, no. 6, p. 1997, Mar. 2021, doi: 10.3390/s21061997.