Emotion Recognition and Sentiment Analysis for Relationship Improvement.

A.I. Bataduwaarachchi
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
ayeshisuranga@gmail.com

N.K.A.S. Dilshan
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
shakiladnanayakkara@gmail.com

H.G.C. Dilini
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
chanikadilini6@gmail.com

K.P.D. Thimesha
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
damidu2014@gmail.com

P. Shalindi
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
shalindi.p@sliit.lk

S.P. Vidanaarachchi
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
samitha.v@sliit.lk

D.I De Silva
Faculty of Computing
Sri Lanka Institute of Information Technology
New Kandy Rd, Malabe, Sri Lanka
dilshan.i@sliit.lk

***Abstract*— Humans can produce hundreds of expressions that differ in meaning during communication. Emotion is the most important factor that humans being well mentally. Many people are suffering from various mental problems. This study investigates the development of a mobile application aimed at improving users' emotional well-being through innovative technology. The application comprises AI-enhanced audio-based emotion recognition, image processing for emotion detection, sentiment analysis implementation, and AI-powered virtual reality (VR) character design. Using advanced AI algorithms, the application analyzes users' voices and facial expressions to identify their emotions in real-time. Subsequently, sentiment analysis is performed to gain deeper insights into users' emotional states. Based on the analysis results, personalized recommendations are provided, including mini-games for relaxation and contact information for counselors in cases of severe depression. Moreover, the application features an AI-powered VR character designed to foster relationships and provide support to users. This character serves as a virtual companion, offering guidance and assistance, particularly seeking to strengthen their bond. We aim to leverage technology to promote emotional wellness.**

Keywords–emotions, sentiment analysis, VR environment, text analyzing, image processing, audio recognition ResNet50, Random Forest Classifier Unity, Convai.

I. INTRODUCTION

In everyday actions, humans display a range of emotions, and these emotions have a great impact on human lives and relationships. Depression affects more than 280 million people worldwide [1]. Women are 50% more likely to suffer from depression than men. More than 700,000 people die by suicide every year. In the rapidly changing digital world, mobile applications offer unprecedented opportunities to integrate modern technologies for emotional well-being, changes in relationships, and many other complex human needs. In this context, research focuses on creating a mobile application using modern technological solutions to reduce people's stress and improve human emotional well-being, as well as building and strengthening meaningful relationships. Emotion recognition and sentiment analysis have appeared as powerful tools to understand and respond to users' emotional states in real time. In recent years, convolutional neural network (CNN) has often been used for sentiment analysis [2]. Stress can be assessed in a variety of ways, including mental, physiological, emotional, and physical activities [3].

The one of objective of research is to improve audio-based emotion recognition using AI and image processing for emotion recognition, to analyze emotions and thereby provide advice to reduce individual stress, and to provide a comprehensive platform to improve and manage interpersonal relationships through a virtual reality (VR) character. The application uses CNN to analyze users' facial expressions, allowing them to detect their emotions in real time. Based on sentiment analysis, it works to gain a deeper understanding of individual users' emotional states and deliver recommendations and interventions specifically tailored to them. A field of study called sentiment analysis facilitates the examination of people's beliefs, attitudes, feelings, and emotions [2]. It suggests mini-games for users to relax and reduce their stress levels and may include access to professional counseling services for users experiencing severe emotional distress. VR characters powered by AI technology have been created to interact in the virtual environment through a VR application as an assistive software. It avoids the complexity of interpersonal relationships and provides guidance and support to strong relationships. This character not only provides practical advice but also facilitates the improvement of emotional connections and communication between partners, especially for those starting new relationships. However, not much recent computer vision research has focused on sentiment analysis of visual data [2]. The main objective of research is to study how the relationship between technology and mental health can positively impact the lives of mobile app users. Using deep learning and VR technology, hope to help people better understand and control their emotions while developing stronger, more satisfying relationships.

II. LITERATURE REVIEW

The 19th century marked the beginning of scientific knowledge and research on emotions and their expression, largely thanks to the writings [12]. To collect audio samples of the Short-Term Energy, Pitch, and Mel-frequency cepstral coefficients (MFCC) coefficients in the feelings of anger, pleasure, and frustration, suggested a system. As a result, identified emotions. In-depth characteristics of the speaker, such as sound, energy, and pitch, were also recognized. Train and test sets are manually created from the entire ravdess dataset [10]. In emotion recognition using audio, contains multiple layers which extracts audio features from the spectrogram. Facial expressions, both spontaneous and nonspontaneous. Most face emotion databases have been compiled by acting out the emotion. The timing and temporal dynamics of these nonspontaneous facial expressions will vary from those of spontaneous facial expressions. The database's resolution, the majority of the current facial expression has an excellent resolution. However, many applications of emotion identification need low-resolution photos or videos, such as video conferencing. This issue is solvable by subsampling the original dataset [11]. Research has shown that positive attitudes toward a person with a mental disorder can be enhanced in a simulated virtual reality environment [4]. The author has created several flowcharts to create favorable environments and opportunities for mental health treatment by involving psychologists. A script has been created that includes how to spend a normal day of a depressed person and the opposite profile of a non-depressive person. This script contained the registry of the person's actions, as well as their thoughts and feelings as the day progressed. By creating their own flowchart version, the author has introduced a game under two levels, depressive and non-depressive. A simple level prototype environment has been created and in these two levels, the user has been allowed to gain an understanding of the mental problem and identify the difference. Everything in the environment was modeled with Unity. The proposed system utilizes a high-level virtual reality environment and covers a large area rather than being limited to a single building. The user has been provided with a system that can deal with diverse characters and engage in activities. This paper focuses on providing a suitable virtual environment to develop the social skills of people with autism [5]. The research is conducted in several sessions by dividing the people into two groups under two age groups (adults, and children). All the volunteers must then show proof of their autism spectrum disorder (ASD) diagnosis from the Autism Diagnostic Observation Schedule (ADOS). Location, objectives, and scenarios have been created separately in the VR environment for the two related categories. The VR intervention also involves a coach in each session to provide feedback to the participants. Based on the results this research can conclude that VR-based intervention definitely has an impact on the social skills of ASD volunteers [5].

However, through the proposed system, the attention has been focused on the development of human relations beyond the development of social skills. It can be shown that a virtual therapist with special knowledge of human relations has been created in the proposed system to provide that awareness. Overall, through the proposed system, the focus is not only on developing social skills and social cognition but also on building human relationships, not limited to one mental problem. This provides a suitable environment for the proper use of social skills.

Researchers in the United Kingdom assessed the effectiveness of a wearable gadget (Bio Beam) in conjunction with a mobile app (Bio Base) in lowering anxiety and enhancing the well-being of college students [7]. Over a 4-week period, the intervention-which included biofeedback and individualized therapy content-showed significant reductions in anxiety levels and increases in reported well-being. The results persisted at the 2-week follow-up. Furthermore, the study showed that using the digital intervention for four weeks, there was a significant drop in depression levels. These results highlight how digital platforms may be used to successfully address mental health issues among college students. In a similar vein, mobile app-delivered cognitive behavioral therapy (CBT) has gained popularity as an approachable and successful way to reduce symptoms of anxiety and depression in a variety of populations, including college students. Users can customize their therapy experience to fit their schedules and tastes thanks to the flexibility provided by mobile apps [8].

Researchers investigated the dose-duration effect of mobile app usage on symptoms of stress, anxiety, and depression in a retrospective examination of app users [9]. Using the mobile app, the majority of participants who were suffering from these symptoms showed remarkable improvements, with a notable decrease in Depression Anxiety Stress Scale-21 (DASS-21) subdomain scores. All these results point to the potential of digital mental health therapies in addressing the rising rates of anxiety and depression among college students and other general populations. Mobile apps are useful tools in promoting mental well-being and provide prompt support to people facing psychological distress because of their accessibility, customization, and efficacy. Personalized therapies catered to individual needs are made possible by the integration of proven assessment instruments, such as the DASS-21 questionnaire, into digital platforms.

III. METHODOLOGY

As shown in Fig. 1 the proposed solution aims to provide an intelligent approach for stakeholders, and researchers to identify depression that can affect humans and improve human emotional well-being as well as build and strengthen meaningful relationships.

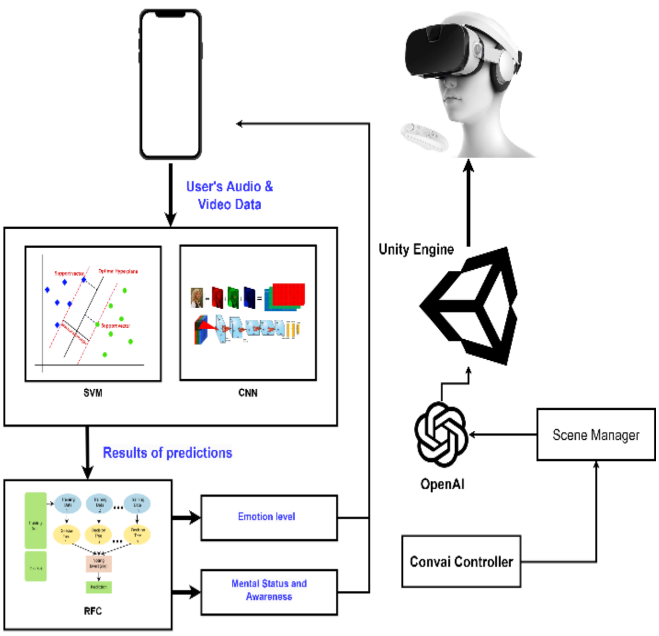


Fig. 1. Overall system diagram of the proposed solution.

The registered users of the system can upload the video call feed. The video is sent to the Amazon Web Services (AWS) back-end server. These videos are processed in the server by the designed CNN models for emotion identification. CNN architecture is defined using the Keras

API from TensorFlow for AI-enhanced audio-based sentiment recognition. The RestNet50 CNN architecture has been defined to recognize an emotion using image processing. Emotional severity is determined using the Random Forest Classifier model. Then it shows the remedies that can be given to the user. Ultimately designing a VR character to improve relationships allows users to express their thoughts, and practice social skills.

A. Identifying user emotions.

1. Data collection and processing

Deep learning (DL) models were trained using user video call feed data. Emotions were identified by voice and image processing. Video data had to be broken down into image frameworks to train DL models. A total of 1818 images were obtained. For image classification, 70% was used for training and 30% for testing. Identified 6 emotion classes. In audio classification, audio files exist in the form of wave files. Can not give this wave file to a direct machine-learning model to classify the emotion. Therefore, convert this wave file to a 2D spectrogram. Physical visible images can be obtained between frequency and amplitude. 1440 spectrograms were obtained and 80% of them were used for training and 20% for testing. To evaluate the Depression, we used DASS – 21 questionnaires (Depression, Anxiety and Stress Scale – 21 questionnaires) to scale the depression level of the people aged between 16 and 30 based on various factors. This questionnaire was tested for over 500 participants, and the participants were: school students, university students, students from villages, and people aged between 16 and 30 who can be found on social media. This scale comprises 21-item questionnaires, and it is divided into 3 subscales which are Depression Subscale (We are using only the depression scale here), Anxiety Subscale, and Stress Subscale.

Humans experience depression, anxiety and stress in everyday life. The questions are based on the day-to-day depression, anxiety and stress one feels. Accordingly, if someone experiences the symptoms, you can answer based on how often that feeling is experienced (not at all, applies to me to some extent, or sometimes, applies to me to a significant degree or a good degree. Some of the time, often or often apply to me).

Above data collection had some blank spots because some of the participants of the form did not fully complete all of the questions in the form. So, cleaning those blank spots is the first step of preprocessing. We need to convert the text data set to a numerical data set. Random forest classification was used to do this. A total of 464 datasets were obtained and used for sentiment analysis, 80% for training and 20% for testing. Since the generated images are of different sizes, they were resized into the same dimensions. Pre-processing techniques were used to improve accuracy and reduce the complexity of the data set.

2. Training the detection models

A transfer learning-based CNN was used to train some models. Two models were trained to recognize AI-enhanced audio-based emotions, three models were trained to recognize an emotion using image processing, and two models were trained to assess depression. Created Deep Learning Model and Support Vector Classification machine (SVC) learning models were trained in emotion recognition using AI-enhanced audio-based emotions. Finally, based on the test accuracies, the most appropriate model was selected for each condition as shown in Table 2. When an audio is given, it is classified, and getting an emotion is called audio classification. A machine learning model cannot classify an audio directly.

TABLE I. SUMMARY OF DATA SAMPLES

Purpose	**Number of Data**		
	Total	***Training***	***Validation***
Emotion recognition using audio data	1440	1152	288
Emotion recognition using video data	1818	1273	545
Sentiment analysis Implementation	464	371	93

TABLE II. S**ELECTING THE** BEST ARCHITECTURE

Purpose	**Tested Architectures**	**Best Architecture**
Emotion recognition using audio data	SVC (Support vector classification) ML Model, Created deep learning model	Created deep learning model
Emotion recognition using video data	VGGNet19, VGGNet19, ResNet50	ResNet50
Sentiment analysis Implementation	Xgboost, Random Forest Classifier	Random Forest Classifier

Audio files exist in the form of wave files. Therefore, need a format that a model can understand. Spectrogram format was used for that. A wave file contains frequencies and amplitudes. These parameters can be converted to a visually appealing 2D representation. These spectrograms exist in the form of images. Then this spectrogram is passed to a CNN model. Somewhere between time and Hz, represents the trinity of sound. You can take care of the trinity and Hz of the sound by means of 2D images. RAVEE dataset exploration was used for this and 8 emotions were identified. Those are Sad, Happy, Calm, Neutral, Angry, Disgust, Fear and Suprise. Then cleans the data set and prepares it as required by the model. First, this data set was tested by SVC machine learning. But the accuracy was very low. Something as complex as emotion recognition using AI-enhanced audio-based emotions cannot be grasped by a normal machine learning model. Therefore, a deep learning model had to be used.The data set is placed in an audio training file path, where 1152 images are used for training and 288 images are used for testing. Transfer learning cannot be used. A spectrogram cannot be classified by transfer learning. Therefore, a unique architecture is designed. That is, have to get a from-scratch neural network deep learning model. Therefore, create all up convolutional layers and batch normalization in this model. In transfer learning, it is possible to train the model with a very low number of epochs, but this cannot be done. Since it is not optimized, you have to train through a lot of epochs. Finally, emotion classes can be obtained with good accuracy. In emotion recognition using image processing, VGGNet 16, VGGNet 19, ResNet50 models were trained. Finally, based on the test accuracies, the most appropriate model was selected. There are many frames in a video and they are analyzed one by one to detect the emotion. The emotion with the highest number of emotions detected will be the emotion in the overall video. Basically, transfer learning is used in image processing. We use Resonet50 from TensorFlow Keras. 70% of the total images are used for training and 30% for testing. 6 emotions are identified. Those are Anger, Fear, Neutral, Sad, Smile and Suprise. we need to define the model and add some transfer learning layers to tune it for specific use. Then the model can be compiled. Then, the emotion recognition performance is tested for several epochs. Finally, the behavior of the model can be obtained from the conclusion matrix.

It is necessary to carry out a sentiment analysis to confirm the emotion obtained through video and audio. For this, a model has to be trained using the survey data set obtained by the DASS-21 questionnaire. The data needs to be cleaned and for that the void spaces must be removed. Data gets blank values when inspecting the set. They are removed and the data set is created. Then the data should be converted to a numerical format that the model can understand. That is, the text should be converted into a number format, i.e. N-coded. Then that data is split into training and testing and fitted to the random forest classifier model.

B. Identification of Sentiment Analysis

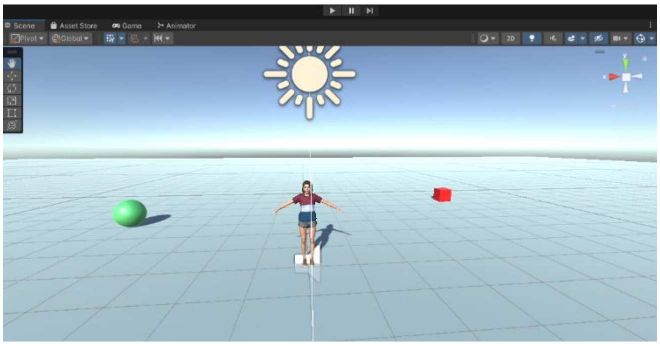
Identifying the emotion, the DASS-21 questionnaire determines the level of depression of the respective user. If the level of depression is in a normal state, some mini-games consisting of several levels are suggested to reduce the level. If the level of depression is not in a normal state, information about the nearest doctor will be provided and the National Mental Health Helpline will also provide support.

C. Virtual reality environment implementation.

The development was carried out using Unity, a powerful game development platform, chosen for its versatility in creating complex VR environments, simulating real-world-like experiences, and its compatibility with various VR hardware and software. The primary objective of this part was to develop a virtual reality (VR) application and environment that can provide an immersive environment for users to express their thoughts, practice social skills, and interact with a virtual therapist.

Natural language understanding and interaction, the system integrates OpenAI and it's connected to the Unity Engine. A unity script file is configured using the OpenAI API key and the connection is established. It provides the general knowledge required by the characters through a larger knowledge system, enhancing the capability of nuanced conversation, and making interactions more engaging and lifelike. To achieve realistic voice interactions within the VR environment AWS service was plugged into the unity engine. Amazon Polly is a cloud service by Amazon Web Services, that converts text into spoken audio. Using AWS Polly with Unity engine achieves a lifelike voice controller in the environment. The creation of responsive, animated characters that are capable of expressing emotions used the "Readyplayer.me". Lip sync, eye tracking, and facial expression functions were added to it. As shown in Fig. 2 infuse these characters with personality and depth, APIs from Convai and InworldAI were integrated. Information has been injected into the respective script files coming from APIs.

Fig. 2. Convai sample scene



The VR controller can activate the microphone when needed and directly chat with any NPC. Characters can show a range of emotional responses similar to real-world interactions by giving them a new shape. Each NPC character has a trigger area included. As soon as the user approaches the character, the trigger area will activate its life stage. A Popup prompt provides the user with the necessary instructions. The prompt provides tips about the task and how the user should proceed in the simulation. The character is given a specific background and set of instructions. Characters are meticulously crafted using a special knowledge base based on their behavior, appearance, and knowledge. That way VR character is trained separately to act and react concerning their designed roles in the software. All environmental systems have been designed with attention to psychological colors aiming to evoke specific emotional responses and create a soothing atmosphere conducive to mental health improvement. Multiple Unity assets were used for the creation of the virtual environment.

A conversation with an NPC end, the conversation is analyzed through a deep learning model. BERT, a pre-trained model, has been used for this. Unlike previous models, BERT reads the entire sequence of words at once. This allows the model to understand the context of a word based on all of its surroundings. This pre-training involves learning by predicting words in a sentence and predicting the relationships between sentences. Pre-training, BERT can be fine-tuned with just one additional output layer to create state-of-the-art models. The complete conversation is analyzed through the model and a prediction is made. Based on that, it is hoped to allow the user to self-diagnose himself. It is possible to judge the progress achieved in the development of abilities by the results of training in the training environment several times.

IV. RESULTS AND DISCUSSION

Emotion identification, sentiment analysis and VR character design were trained using several architectures (Table 1), and the best architecture was selected among them by considering the loss value. The loss value indicates how well or poorly a model performs in each iteration of optimization. The testing accuracies for emotion recognition using audio and video, and sentiment analysis are given in Table 3.

TABLE III. TESTING ACCURACIES FOR EACH ARCHITECTURE

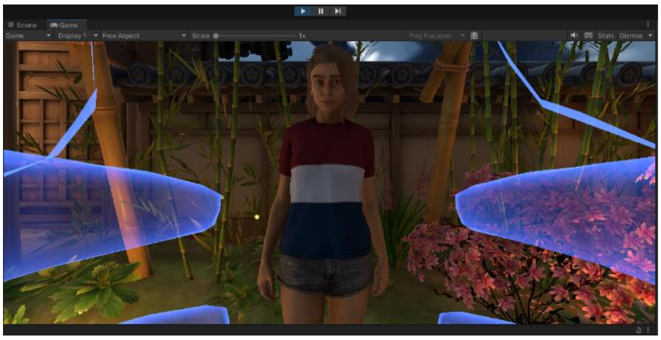
DL Architectures	Accuracies
SVC	17.43%
Created DL Model	88.07%
ResNet50	99.45%
VGGNet16	69.21%
VGGNet19	58.01%
Xgboost	69.89%
Random Forest Classifier	68.81%

Image sizes 224×224 dimensions were used to evaluate the performance of the models. Images with lower dimensions resulted in poor extraction of essential features, whereas images with higher dimensions require better computational power (higher GPU) to train models. Therefore, images with dimensions of 224 × 224 were used to train the models with the best performance and accuracy. In addition, a good fit was achieved by adjusting batch sizes (1, 32), epochs (1‑405), learning rates (0.001), steps per epoch, validation steps, and Adam optimizer. The adjustment process is critical because even minor changes have a significant impact on the training process. During the training process, CNNs tend to overfit; therefore, a dropout regularization technique was used with various dropout rates to prevent the overfitting of the neural network. Fine-tuning network parameters, highest testing accuracies were considered to select the best architectural models. The AI enhanced audio-based emotion recognition classification model using the support vector classifier architecture achieved 17.43% training accuracy and 15.64% testing accuracy. Furthermore, the created deep learning model achieved training and validation accuracies of 98.61% and 88.07%, respectively. Emotion recognition using image processing, VGGNet19 architecture-based model achieved 71.29% training and 58.01% validation accuracy. VGGNet16 architecture-based model achieved 73.61% training and 69.21% validation accuracy and ResNet50 architecture-based model was selected which resulted in 100% training and 99.45% validation accuracy. Also, Xgboost architecture-based model with 64.58% training and 69.89 % validation accuracy. Fig 3 illustrates the confusion matrices related to audio-based emotion identification (a), image-based emotion identification (b) which help to evaluate the performance of the trained models. The precision of this technology allows timely identification emotion, managing depression level. Collaboration and dataset growth can improve model accuracy. The model for Depression analysis is trained using two architectures (table 2) and the best architecture was selected among them by considering the accuracy of depression level predictions of the user. The accuracy of models indicates how well or poorly a model performs analysis of answers given by the users. The testing accuracies for depression analysis are given in the table.

A. *Virtual reality environment*

As shown in Fig. 4 using the Unity Engine, the environment has been designed to resemble the real world. Ability to talk to virtual characters, and perform activities using a VR headset.

Fig. 3. Perform environment



Characters in the environment have been brought to life by OpenAI and Convai. In the beginning, it was planned to create a chatbot using the RASA framework. The RASA chatbot was intended to connect to the Unity engine and empower the characters, but due to the limitations of the chatbot, Convai was used. This is because Convai offers a combination of the highest quality. Convai has provided a suitable environment to interact with the characters.

Convai has provided a detailed backstory and relevant knowledge for NPCs. Connecting with the unity engine is done using an API key. It can include specific attributes for NPCs as per requirement connecting it through the Convai unity package. Using AWS Polly, we tried to provide high-quality voices for NPCs as if they were in the real world. But later, through the audio engine in the Convai, a real-life voice simulation was given to the NPC. At the end of the conversation with NPC, analysis is done using a deep learning model. A pre-training model called BERT has been used for this purpose. The entire conversation is analyzed through the model and a prediction is made. This prediction enables the user to self-study about himself.

V. CONCLUSION AND FUTURE WORK

In this paper, cutting-edge technologies such as convolution neural networks, image processing and 2D spectroscopy were used to identify emotions through voice and image processing, emotional severity through sentiment analysis. Emotion identification and severity of the emotion conditions were determined using RestNet-50, Keras API, and Random Forest Classifier models of overall accuracies between 75%-99%. In the future, existing sentiment recognition and analysis models will be expanded. There are various problems with emotion identification and sentiment analysis of audio and video data. Therefore, future researches should focus on developing computer vision techniques to better understand emotions expressed through facial expressions, body language, and audio expressions, and to obtain an accurate range of emotions. Likewise, it is necessary to explore the effectiveness of using VR technology, especially to improve and strengthen interpersonal relationships. Continuous refinement and optimization of deep learning algorithms for sentiment recognition and sentiment analysis is essential to improve accuracy and reliability when using app.

REFERENCES

[1] Arya, R. Kumari and P. Bansal "Predicting Depression and Mental Illness Using Machine Learning Algorithms," 2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI), Greater Noida, India, 2023.

[2] Y. S. Deshmukh, N. S. Patankar, R. Chintamani and N. Shelke, "Analysis of Emotion Detection of Images using Sentiment Analysis and Machine Learning Algorithm," 2023 5th International Conference on Smart Systems and Inventive.

[3] A. Bamanikar, R. V. Patil and L. V. Patil, "Stress & Emotion Recognition Using Sentiment Analysis With Brain Signal," 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2022.

[4] F. Morim, E. Oliveira, C. Braga and N. Rodrigues, "Virtual Reality Video Game for Depression Awareness," *2022 IEEE 10th International Conference on Serious Games and Applications for Health(SeGAH)*, Sydney, Australia, 2022.

[5] E. Halim, T. N. Muliawan, Y. Kurniawan, T. Pujadi, P. P. Halim and D. Kartawiguna, "Virtual Reality As A Social Learning Tool For Individuals With Autism," 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 2022.

[6] M. Chintapalli, N. Kanugula, L. Nagisetti, A. V. V, I. T. J. S and G. Bindu, "Investigation of Intelligent Methodologies in Prediction of Depression," 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2023.

[7] Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D Efficacy of the Digital Therapeutic Mobile App BioBase to Reduce Stress and Improve Mental Well-Being Among University Students: Randomized Controlled TrialJMIR Mhealth Uhealth 2020.

[8] Valinskas S, Nakrys M, Aleknavicius K, Jonusas J Sensa Mobile App for Managing Stress, Anxiety, and Depression Symptoms: Pilot Cohort Study JMIR Form Res 2023.

[9] Berg, D., & Perich, T. (2022). Use of mobile mindfulness apps in young adults with depression: Results from a cross-sectional survey. Professional Psychology: Research and Practice, 53(1), 42–49.

[10] P. Harish, J. Jeevagan, M. P. Selvan, A. V. A. Mary and S. Jancy, "Human Emotion Recognition by Audio Signals using MLP Classifier," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023.

[11] T. S. Gunawan, M. N. Muktaruddin, M. Kartiwi, Y. A. Ahmad, T. D. Rosahdi and Ulfiah, "Development of Video-Based Emotion Recognition System using Transfer Learning," 2022 IEEE 8th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Melaka, Malaysia, 2022.

[12] E. Singh and P. Nand, "Enhancing Facial Emotion Recognition through Deep Learning and Image Pre-processing," 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT), Dehradun, India, 2023.