Received: 29 December 2023 | Revised: 7 March 2024 | Accepted: 27 March 2024

DOI: 10.1002/eng2.12894

RESEARCH ARTICLE

Engineering Reports Open Access WILEY

Elevating metaverse virtual reality experiences through network-integrated neuro-fuzzy emotion recognition and adaptive content generation algorithms

Oshamah Ibrahim Khalaf[1] | **Dhamodharan Srinivasan[2]** | **Sameer Algburi[3]** | **Jeevanantham Vellaichamy[4]** | **Dhanasekaran Selvaraj[5]** | **Mhd Saeed Sharif[6]** | **Wael Elmedany[7]**

[1]Department of Solar, Al-Nahrain Research Center for renewable Energy, Al-Nahrain University, Baghdad, Iraq

[2]Department of Computer and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore, India

[3]Al-Kitab University, College of Engineering Techniques, Kirkuk, Iraq

[4]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,

[5]Department of Electronics and communication Engineering, Sri Eshwar College of Engineering, Coimbatore,

[6]Intelligent Technologies Research Group, Computer Science and DT, ACE, UEL University, London, UK

[7]College of Information Technology, University of Bahrain, Sakhir, Bahrain

Correspondence
Mhd Saeed Sharif, Intelligent Technologies Research Group, Computer Science and DT, ACE, UEL University, Way, London, UK.
Email: s.sharif@uel.ac.uk

Abstract

Interactions between individuals and digital material have completely changed with the advent of the Metaverse. Due to this, there is an immediate need to construct cutting-edge technology that can recognize the emotions of users and continuously provide material that is relevant to their psychological states, improving their overall experience. An inventive method that combines natural language processing adaptive content generation algorithms and neuro-fuzzy-based support vector machines natural language processing (SVM-NLP) is proposed by researchers to meet this demand. With this merging, the Metaverse will be able to offer highly tailored and engaging experiences. Initially, a neuro-fuzzy algorithm was developed to identify people's emotional moods from their physiological reactions and other biometric information. Fuzzy Logic and Support Vector Machine work together to manage the inherent ambiguity and unpredictability, which results in a more exact and accurate categorization of emotions. A key component of the ACGA is NLP technology, which uses real-time emotional data to dynamically modify and personalize characters, stories, and interactive features in the Metaverse. The novelty of the proposed approach lies in the innovative integration of neuro-fuzzy-based SVM-NLP algorithms to accurately recognize and adapt to users' emotional states, enhancing the Metaverse experience across various applications. The proposed method is implemented using Python software. This adaptive approach significantly enhances users' immersion, emotional involvement, and overall satisfaction within the augmented reality environment by tailoring information to their responses. The findings show that the SVM-NLP emotion identification algorithm based on neuro-fuzzy, has a high degree of accuracy in recognizing emotional states, which holds promise for creating a Metaverse that is more emotionally compelling and immersive. Stronger human–computer interactions and a wider range of applications, including virtual therapy, educational resources, entertainment, and social media networking, might be made possible by integrating SVM-NLP. These sophisticated systems are around 92% accurate in interpreting the emotions.

KEYWORDS

emotion recognition, feelings of users, fuzzy logic, high accuracy, humans and computers, Metaverse, NLP, social media, SVM

1 | INTRODUCTION

Identifying feelings from language signals is called speech emotion recognition, and it is crucial for improving interaction between humans and computers. In order to fully understand the dialogue and react effectively, it is helpful to be aware of one's emotions at the moment that communication occurs.[1] There still exists a lack of universal solution to this challenge, and this aspect of human–computer interaction is still not fully resolved, with the exception of a small number of programs. The method of recognizing feelings is fluid and focuses on the individual's emotional state, thus all individual's behaviors are accompanied by an assortment of feelings.[2] In overall, there are several ways that people express their emotions. The right understanding of these feelings is crucial for effective interaction. These feelings, nevertheless perform a significant influence in determining behavior among people, and emotional detection in daily life.[3] Every day, numerous emotions are felt by people. For instance, when there is an aspect to be glad about, anything to be fearful of the emotions should come out at that time. Feelings have frequently been the subject of research. The demand to enhance interactions between humans and machines has grown as people depend more and more on computers to carry out their daily tasks. Feeling can be challenging for a machine to perceive and manufacture because it lacks sensible information.[3] As a result, a lot of study has been done on recognizing emotions. There are three primary modalities for emotion detection: facial movements, voice, and text. Emotion detection in text involves automatically tagging text with a sense of emotion from a predetermined list of expression labels. Text serves as the primary channel for communications between people and computers through messages via email, text, message boards, conversations, consumer evaluations, Internet blog posts, and additional social networking sites like YouTube, Twitter, and Facebook. Moreover, applications of recognizing feelings in text can be discovered in many different domains where the technique is necessary to understand and analyze emotions, including business, psychology, teaching, and many more.[4]

Speech emotion identification is a significant issue that is attracting more attention from academics owing to its wide range of programs, including auditory monitoring, e-learning, medical investigations, recognizing lies, recreation, video games, and customer service centers. Sophisticated machine learning approaches still find it extremely difficult to solve this challenge. The difficulty in selecting the perfect characteristics constitutes one of the factors contributing to such a meek performance. A machine learning model's performance may also be significantly impacted by the presence of noise from behind recordings of sound, such as actual human voices. Although user satisfaction in platforms requiring human–machine interactions, such as those in the fields of Artificial Intelligence (AI) or Mobile Health (mHealth), might be considerably improved by the development of competent emotional voice recognition models. In fact, the capability to identify feelings from audio recordings has the capacity to mimic these feelings, may have a significant influence on the area of AI. After utilizing such frameworks, many automated systems in the mHealth sector might greatly increase their efficiency. The technological specifications for emotional recognition of speech techniques are also modest.[4] Human-Computer Interaction (HCI) has numerous practical uses, and speech emotion recognition (SER) is a rapid and common way for people to communicate and exchange information with computers.[5] Choosing an accurate way for extracting prominent and distinctive characteristics from speech signals in order to describe the emotional state of someone who speaks from their acoustical content remains an important task for feature extraction. Moreover, speech comes most naturally to humans when it comes to express themselves. Researchers rely on it so heavily to understand how crucial it is. When other forms of communication, like texts or emails gets utilized, emojis have been popular in text messaging, which is not unexpected that these messages may be misread and want to convey feeling in text just like doing in conversations.[6] Someone might be able to recognize their feelings by using SER. There is no denying that emotions play a part in many facets of life. Feeling has been defined in a variety of ways up to this point. Sentiments are states of mind with either a neutral or positive emotional polarity.[6] Numerous actions in society, preserves positive connections for plotting influence in politics, are supported by the capacity to make sense about feelings. Emotions play a crucial function upon human perception and cognition in addition to playing a significant influence in every individual's everyday life and social relationships.

The clear importance of human feelings in rational decision-making is demonstrated by neurological research and investigations into practical brain functioning. According to another investigation, emotional computing is related to the identification of emotions. Furthermore, the investigation regarding how machines perceive and identify feelings is known as affective computing. A person who is emotionally unbalanced could be less capable of carrying out routine chores. Therefore, it is obvious that feelings have a significant impact on a person's life.[7] A number of recent studies are not content with merely identifying an individual utterance, phrase, or article employing standard feeling classification algorithms in order to more effectively adapt to everyday scenarios involving applications, such as smart homes, mental health treatment, educational tools, and so on.[8] Today's research is focused on the emotive content evaluation of voice signals. Speech is the most common means for people to convey data, hence it is important to consider how people communicate with computers. On the other hand, HCI heavily relies on social media, written language, and additional semantic information modalities, all of which are quite beneficial. Feelings are the most important component of human communication, and they can be used to make conclusions about interpersonal gestures, paralanguages, and other things.[9] As a result, the speech signal is an effective method for the quickest interaction between HCI, which effectively detected human behavior. One of the most rapidly developing research areas nowadays is recognizing feelings in speech signals. In this area, scientists have created techniques to inherently identify feelings in speech signals. Once suggested, the concept of SER will be extensively utilized in the domains of health and education. Researchers are developing a wide range of methods over the last decade in order to render the SER effective and trustworthy for real-time applications, as the SER serves a significant role in the HCI. Recognizing the expressive indications and psychological realities from speech has been difficult in the last 10 years. The earlier techniques consistently replicate the same mood for a certain party since a sudden shift in emotion is improbable. As a result, these techniques frequently fail when something different happens. Using ML techniques like deep learning enables the accurate detection of user emotions in the Metaverse, enhancing immersion. Reinforcement learning optimizes adaptive content generation, providing personalized experiences. Together, they create more engaging and emotionally resonant Virtual Reality (VR) interactions, enriching the Metaverse. The enhancement of VR experiences is a crucial undertaking in the developing Metaverse, since it creates an emotionally charged and immersive digital world. To drive this advancement, a novel strategy that makes use of the integration of SVM-NLP technology is utilized. When neuro-fuzzy-based emotion identification and adaptive content production algorithms are combined, a new way of thinking is created that improves VR experiences in the Metaverse. The integration of SVM-NLP and neuro-fuzzy processes creates a complex interaction that opens up new possibilities for the intersection of emotions, physiological signals, and content creation.[10] This enhances the Metaverse environment by providing deeper emotional connectedness and more engagement.

The key contribution of the established outline is given as follows:

- The enormous dataset makes possible was to generate and analyze emotion identification algorithms for a wide variety of facial movements and emotional states using neuro fuzzy-based SVM-NLP Adaptive Content Generation Algorithm.
- Preprocessing is a method used to reduce distortion and unnecessary components of frequency and improve the overall quality of physiological data.
- A thorough knowledge of emotions is attained by utilizing Mel frequency cepstral coefficient (MFCC) features to represent emotional traits and K-means clustering to organize emotional data. This is especially useful in the context of VR and physiological signal analysis.
- For ambiguity and intricate relationships in emotional data, the neuro-fuzzy algorithm is used. Support vector machine (SVM) offers reliable and effective classification, leading to accurate and comprehensible emotion detection.
- The creation of emotionally relevant and thematically relevant content is made possible by NLP, which improves the entire Metaverse VR experience.

The remaining section of this research are broken down into the following sections. Section 2 clearly depicts about the literature works. The proposed Neuro fuzzy-based SVM-NLP Adaptive Content Generation Algorithm is discussed in the Section 3. The results and discussions take place in Section 4. Finally, the conclusion is discussed in Section 4.5.

2 RELATED WORKS

2.1 Multimodal deep learning models for emotion recognition

The study Comparing Recognition Performance and Robustness of Multimodal Deep Learning Models for Multimodal Emotion Recognition examines the application of multimodal deep learning algorithms for emotion recognition that make use of various data techniques (such as audio, video, and text) to identify emotions said by Liu et al.[10] To efficiently handle and combine data from diverse sources for increased emotion identification accuracy, the investigation will probably propose and compare several algorithms for deep learning. Large, varied, balanced, and multimodal emotion datasets might be difficult to find, and deep learning models' real-time application may be hampered by their computational difficulty. However, the finding is crucial because it enhances multimodal emotion detection technology, improving the reliability and accuracy of emotion recognition methods. The research has potential applications in emotional computing, interaction between humans and computers, and social robots because these fields require the development of systems that are sensitive to human emotions in order to be more effective and compassionate. Mustaqeem and Kwon[11] discussed in the paper that Speech is the main form of human interaction, has a lot of potential for HCI when used with microphone detectors. Quantifiable identification of emotions from voice signals is a developing area of HCI research that uses these sensors. It can be used in many different contexts, such as human-robot interaction, augmented reality behavior evaluation, medical care, and emergency call centers, where it is vital to ascertain the speaker's emotional state from speech. The preciseness of SER as compared to current techniques is improved in this research, and the computing difficulty of the suggested SER model is also reduced. The method extracts prominent and discriminative features from improved voice signal spectrograms with the aid of a deep stride convolutional neural network (DSCNN) architecture and an ordinary nets approach. The exact border recognition of phrases is nevertheless hampered by massive concatenated feature fusion, which is a consequence of data sparseness.

In the contemporary social media landscape, the conventional approach of merely sharing images is insufficient to capture attention.[12] Many platforms now encourage users to accompany their images with captions or songs. Given the vast array of songs available, selecting the most fitting lyrics for a given image becomes a challenge. To address this, it presents a novel multimodal model designed to recommend lyric captions that harmonize with input images. The model incorporates deep learning and object detection techniques to generate captions and word tags from images. Subsequently, a convolutional neural network detects the emotion of the image, categorizing it for an appropriate emotional match with song lyrics. Finally, a natural language processing (NLP) model refines the match among captions, word tags, and lyrics, presenting users with suitable lyric alternatives to enhance their social media posts. Although our model is experimented with and evaluated on Thai lyrics, its adaptability extends to lyrics in other languages, showcasing its versatility and broad applicability.

Soft-shell crab production demands continuous visual inspections by aquacultural farmers to identify molting completion, preventing the new shell from hardening.[13] This study proposes an automated solution using a vision-based deep learning approach, focusing on the challenge of detecting crabs within obstructive box baskets. Initial attempts using transfer learning and fine-tuning a pre-trained VGG-16 model achieved 89% accuracy, but sensitivity to image bias led to false feature learning. Subsequently, the study employed Faster R-CNN object detection techniques, achieving crab detection with an average precision ranging from 83% to 91%. By translating crab counting results into image classification, accuracy remained consistently high at 98%–99%. This innovative approach presents a robust solution for automating visual inspections in soft-shell crab production, promising increased efficiency without compromising accuracy.

The Bosniak renal cyst classification system, widely utilized for assessing renal cyst complexity, presents a dilemma as approximately half of patients undergoing surgery for Bosniak category III exhibit benign rather than malignant cysts upon pathology analysis, resulting in unnecessary surgical risks.[14] To address this issue, this study employs deep learning techniques to explore alternative analytics methods for precise binary classification of CT images, distinguishing between benign and malignant tumors. The proposed approach involves two key steps: segmenting kidney organs or lesions from CT images and subsequently classifying the segmented kidneys. The study introduces a novel method utilizing 2.5D ResUNet and 2.5D DenseUNet for efficient extraction of intra-slice and inter-slice features during kidney segmentation. Experimental models, trained on the Kidney Tumor Segmentation (KiTS19) challenge dataset and validated in different training environments, achieve high mean kidney Dice scores exceeding 95% on the KiTS19 validation set. Despite a drop in performance on abdomen CT images from four Thai patients, where the best mean kidney Dice score is 87.60%, the study offers promising insights into enhancing precision in renal cyst classification through advanced imaging techniques.

2.2 | Advancements in speech emotion detection

Identifying emotions from recordings of voices and providing information about the well-being of individuals are the goals of decades of research into speech emotion detection, a fascinating subject in human–computer interaction said by Bhavan.[15] This study focuses on classifying emotions using three corpora: the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS), the Indian Institute of Technology Kharagpur Simulated Emotion Hindi Speech Corpus, and the Berlin EmoDB. Researchers compressed and evaluated the data using spectral characteristics to get the key feature set. Researchers provide a feasible approach for this job that uses a bagged ensemble with support vector machines and a Gaussian kernel for improved performance. The outcomes demonstrate the success on the previously mentioned datasets. Cimtay and Ekmekcioglu[16] exposed that the electroencephalogram (EEG), which has an advantage compared to visual or spoken cues for emotion recognition, is appealing in investigations on emotion recognition because of its immunity to human deceitful behaviors. Yet, a major obstacle to subject-independent accuracy in EEG-based emotion identification is the diversity in EEG patterns between people and over time. This study uses pretrained convolutional neural network (CNN) architecture to improve subject-independent recognition efficiency. By removing erroneous observations inside the range of predictions of emotions, a median filter is also used to increase classification accuracy. With regard to the potential applications, facial emotion recognition (FER) has attracted a lot of interest in the scientific community said by Akhand et al.[17] Mapping different expressions on the face to their corresponding state of mind is the main objective of FER. Feature extraction and emotion recognition are typically the two main phases in FER. Due to its capacity for automatically obtaining information from images, Deep Neural Networks, particularly CNN, are currently used extensively in FER. Standard depth CNNs, on the other hand, lack the capacity to successfully extract emotion data from images with a high resolution. Additionally, the majority of approaches now in use only consider frontal views, ignoring perspective views, which are crucial for a functional FER system that can operate from various angles.

VR holds substantial promise across various domains, particularly in business and psychology, by offering realistic, safe, and controllable simulations for research, training, and enriched consumer experiences.[18] In the Metaverse, where virtual representations of people, known as avatars, play a pivotal role, the meticulous design of their non-verbal behaviors becomes crucial for a truly immersive experience. This paper focuses on understanding how users perceive avatar non-verbal behaviors, including body posture, facial expression, and head movement, in VR contexts. Through an experiment involving 125 participants, the study validates the perception of emotional valence and arousal levels in VR, yielding a library of nonverbal behaviors corresponding to different emotional states. Additionally, the research explores the impact of low-end versus high-end VR headsets and photo-realistic versus cartoon avatars. The findings offer valuable insights for designing realistic, challenging, and interactive virtual audiences, enhancing the overall VR experience.

This review explores the impact of VR on emotions, recognizing the potential of VR to create environments mirroring real experiences.[19] A thorough search, covering the Scopus database and other sources, resulted in the analysis of 16 studies meeting specific criteria, such as user interaction and the application of VR software development. Unity was prevalent in 81.25% of the articles, and HTC Vive featured in 37.5% of the research. The duration of VR experiences ranged from a minimum of 1 min to a maximum of 30 min. Notably, only 18.75% of the articles discussed theories of emotions, with a majority (56.25%) utilizing VR to influence users. Additionally, only one article aligned with ISO 25000 standards, highlighting the need for more comprehensive evaluation frameworks in the field of VR and emotions.

In this research, the focus lies on enhancing the expressiveness of avatars in shared virtual environments within the emerging Metaverse.[20] The study compares two real-time modalities for conveying expressions in virtual reality using realistic, full-body avatars through a user study. The first modality employs dedicated hardware, including eye and facial trackers, to map the user's facial expressions and eye movements onto the avatar model. The second modality utilizes an algorithm that approximates facial motion by generating lip and eye movements based on an audio clip. The user study involved participants observing avatars portraying various emotions, and the evaluation emphasized social presence and emotion conveyance. Results indicated that facial tracking using dedicated hardware was significantly superior in conveying sadness and disgust, while differences were less evident for happiness and fear. No significant distinctions were observed for anger and surprise. These findings contribute valuable insights into optimizing avatar expressiveness for enhanced user experience and social presence in shared virtual environments.

2.3 | Advances in facial emotion recognition with FERC

Mehendale[21] discussed in the paper that humans have historically had no trouble recognizing emotional expressions in facial expressions, but computer algorithms find this task to be quite difficult. Nevertheless, emotion recognition from images is now possible because to recent advances in computer vision and machine learning. The facial emotion recognition utilizing convolutional neural networks (FERC) method is a revolutionary technique that is introduced in this research. The foundation of the FERC model is a two-part CNN, with the first component concentrating on elimination of background from the image and the second part on extracting the facial feature vector. An expressional vector (EV) is used in the FERC concept to categories normal facial movements into five different categories. The use of face masks has increased significantly as a result of the Covid-19 outbreak, which is good for reducing infections yet raises questions about how it will affect social interaction. In this study said by Batbaatar et al.,[22] researchers investigate how face re-identification, recognition of emotions, and trustworthiness attribution alter when faces are observed without a mask, with a normal clinical face mask, and with an opaque face masks that recovers sight of the mouth area. In contrast to typical clinical face masks, which dull the feeling of trustworthiness, the outcomes show that transparent mask keep their capacity to recognize facial expressions and infer trustworthiness. Transparent face masks, curiously, do not appear to effect later face re-identification, showing an estrangement among systems that regulate emotional and identification processing, even while they have no impact on emotion trust and recognition attribution. Mastering face reading when a facial mask covers the lower half of the face is dependent on this research.

The earlier techniques consistently replicate the same mood for a certain party since a sudden shift in emotion is improbable.[23] As a result, these techniques frequently fail when something different happens. In order to address this, a novel Neuro Fuzzy-based SVM-NLP Adaptive Content Generation was implemented in this research.

3 | PROPOSED SVM-NLP FRAMEWORK

The dataset second-hand for the process of training and testing is the Affect Net dataset. This is employed to show the efficacy of the analysis and to find the emotions of peoples. Preprocessing is used to eliminate the unwanted noise distortions and improve specific qualities that are essential for the process. Correspondingly, the standard SVM-NLP technique is employed for the VR experience in the Metaverse to detect the emotions of humans. Furthermore, it is employed to attain a better accuracy value. Figure 1 shows the architecture of proposed SVM-NLP framework.

Figure 1 shows the architecture of proposed SVM-NLP framework. It involves steps like Steps like normalizing and FFT are used in MFCC to extract sound features. K-means helps identify patterns by segmenting data.

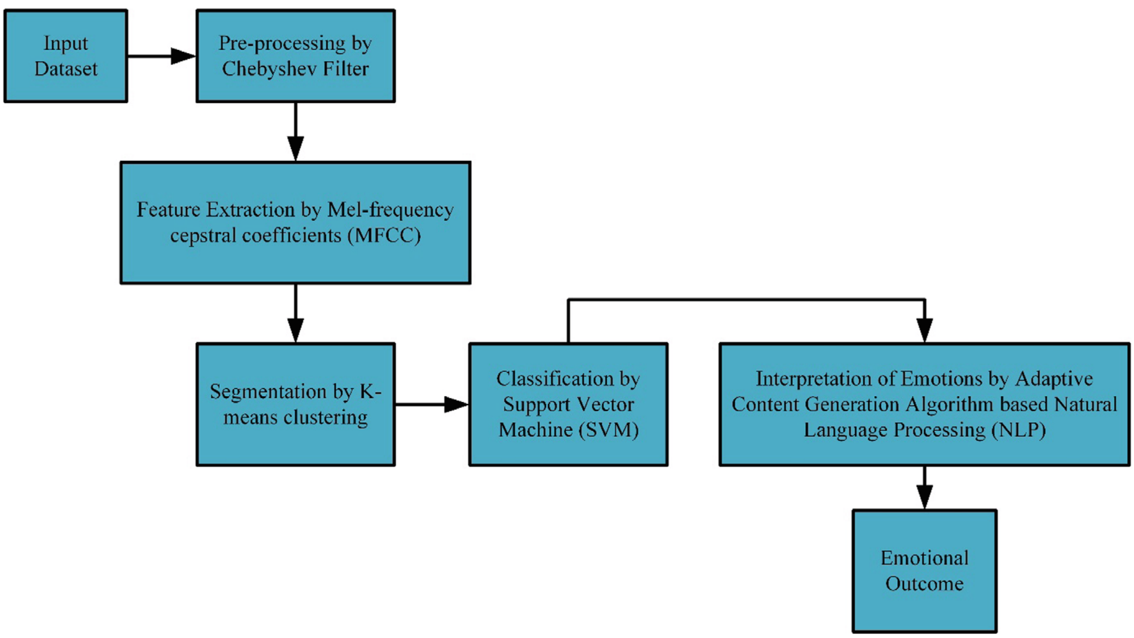

FIGURE 1 Architectural diagram of proposed SVM-NLP framework.

3.1 | Data collection

The dataset utilized here is the Affect Net dataset.[24] In this study, the Affect Net dataset serves as the primary source for emotion data, comprising approximately 40,000 datasets. To ensure a robust and comprehensive investigation, a balanced approach is adopted, with 50% of the datasets allocated for the training process and the remaining 50% for testing. This careful partitioning of the dataset is crucial for evaluating the proposed neuro-fuzzy-based emotion recognition and adaptive content generation algorithms. The utilization of a substantial number of datasets allows for a diverse and representative sample, enhancing the model's ability to generalize and accurately recognize emotions across various scenarios within the Metaverse context. The balanced distribution between training and testing datasets ensures a thorough assessment of the model's performance, validating its efficacy in real-world applications and contributing to the reliability of the study's findings.

3.2 | Preprocessing

Beforehand separating the characteristics, the samples are subjected to a method called preprocessing.[25] Some samples will have noise and other incorrect data, as well as certain contextual differences. By applying filters, it might get removed at this stage. The samples in this research are cleaned up using a Chebyshev filter. The Chebyshev filter is chosen for preprocessing due to its ability to efficiently remove high-frequency noise in physiological data, such as biometric signals. Its sharp roll-off characteristics help maintain the integrity of emotional data extracted from physiological reactions, enhancing the accuracy of emotion recognition in the Metaverse.

3.3 | Feature extraction

Feature extraction is the process of taking the most important and instructional characteristics out of raw data. The data must be transformed into a more compact, relevant space for features while keeping crucial patterns and traits.[26] By distinguishing these qualities, the focus is placed on acquiring the essential elements of the data that are required for evaluation, classification, or other operations, which reduces extra or redundant details from the data. The aim of the method of feature extraction is to produce a concise and thorough representation that renders the data analysis and simulation more effective and efficient. This will vary depending on the data source and issue region.

3.4 | Mel frequency cepstral coefficients

The idea of Mel reciprocity was created by Davies and Mermelstein to explain how the human auditory system reacts to vibrations of various wavelengths. As shown in Equation (1), the Mel frequency and linear frequency have a quantitative relation.

$$f_{\text{mel}} = 2696 \times \log_{10}\left(1 + \frac{f}{800}\right) \tag{1}$$

The MFCC operates on the idea of Mel frequency, that is one of the most commonly employed and effective characteristic variable representations and may account for the phenomenon of multilayer channel imperfection. Normalized processing, framing, windowing, fast Fourier transform (FFT), Mayer filter group filtering, logarithmic, and discrete cosine transform (DCT) are the different steps in the MFCC extraction procedure. The technique of separating MFCC with emotions like a joyful tone is demonstrated in the sample that follows.

3.4.1 | Normalized processing

In order to produce a signal with sound $y(n)$, the speech is normalized.

3.4.2 | Framing

Framing is the procedure of converting the time series into a sequence of frame with partially overlapped portions. If the frame's values are too large to be separated, the final zero is appended to the full frame.

3.4.3 | Windowing

Windowing is used to reduce the gap among frames that develops after framing to a particular degree. The Henning window, as well as rectangular and triangular windows, is the standard. The Henning window is used by convention. The particular equation is displayed in Equation (2).

$$u(m) = 0.5 \times \left(1 - \cos\left(\frac{2\pi m}{(M-1)}\right)\right), 0 \leq m \leq M-1 \tag{2}$$

3.4.4 | Fast Fourier transform

In (3), the speech information is subjected to a FFT.

$$y_b(l) = \sum_{m=0}^{M-1} y(m)e^{-i\omega l}, 0 \leq l \leq M \tag{3}$$

Here, $\omega_{l=}\frac{2\pi}{M}l$, M is represented as the length and $y(m)$ is represented as the input signal.

3.4.5 | Mayer filter group filtering

Build a Mayer filter bank and the electromagnetic spectrum is put through the dot-product operations in order to accomplish a dot conversion from the power wavelength to a Mayer frequency that is nearer to the human ear's function. By utilizing (4), the Eigen frequencies of the filter are scattered equally across the Mel frequency range.

$$g(k) = \frac{M}{G_t} f^{-1}\left(f(K_k) + k \times \frac{f(V_k) - f(K_k)}{N+1}\right) \tag{4}$$

Here, G_t is represented as the frequency of actual acquisition; M is represented as the length of FFT; N is represented as the number of filters; $f(K_k)$ is represented as the lower limit frequency of single filter; f^{-1} is represented as the inverse function and this can be calculated from (1).

3.4.6 | Logarithmic

To imitate the logarithmic characteristics of the human ear, the major goal of collecting logarithms. To get the range of energy, the logarithmic conversion is required. In (5) and (6), the exponential transformation is shown.

$$R(n) = \ln\left(\sum_{l=0}^{N-1} |Y_b(l)|^2 I_n(L)\right), 0 \leq n \leq N \tag{5}$$

$$\sum_{n=0}^{N-1} I_n(L) = 1 \tag{6}$$

Here, $I_n(l)$ is represented as the frequency of triangular filter; N is represented as the triangular filter.

The discrete cosine transformation is followed by generation of the MFCC constants through (7).

$$D(m) = \sum_{n=0}^{M-1} r(n) \cos\left(\frac{\pi_m(n-0.5)}{N}\right), m = 1, 2, \ldots K \tag{7}$$

Here, K is represented as the number of MFCC coefficients and N is represented as the triangular filter.

3.5 | Segmentation by K means clustering

A dataset or an image is segmented into unique and significant subsets or areas based on shared attributes. It makes it simpler to understand and obtain pertinent information from complicated databases or images through making it simpler to discover patterns, group related components, and perform statistical analysis or image manipulation activities. Data are divided into K separate clusters using the unsupervised machine learning technique K-means clustering, where K is a user-defined value. The information points are repeatedly assigned to the closest clustered median utilizing the method, and the centroids are subsequently updated using the average of the data points inside every cluster. The cluster centroids continue to move through this procedure until convergence.[27] The objective of the algorithm is to reduce the sum of square within-cluster lengths, efficiently combining similar information and forming different clusters. K-means clustering is a popular technique for organizing and comprehending patterns of data in huge datasets. It is employed for a variety of applications, including segmentation of customers, compression of images, and identifying anomalies. The K-means clustering procedures have some of the following techniques and are given below in steps.

Step 1: It is necessary to identify the core and the L-cluster.

Step 2: For each pixel, the Euclidean distance E_d is provided and is stated in (8).

$$E_d = ||h(i,j) - Pl|| \tag{8}$$

Step 3: The logic for allocating all of the pixels to the middle is E_d.

Step 4: By ensuing the task assigned to all the pixels, the new center positions are re-computed by Equation (9)

$$Pl = \frac{1}{L} \sum_{j \in Pl} \sum_{i \in Pl} h(i,j) \tag{9}$$

Step 5: Continue doing this until the requirement is met.

3.6 | Classification

An essential task in machine learning and pattern recognition is the classification or labeling of data into specified groups or divisions in accordance with its traits or features. The initial step in the classification process is to build an algorithm utilizing machine learning on a set of labeled information that is connected to recognized classification identification. The algorithm for machine learning modifies from the data used for training to look for connections and trends between the characteristics and the appropriate class labels. After been trained, the model with predictions can be utilized for predicting the labeling of categories of new, unidentified information. The accuracy and effectiveness of the system of classification are evaluated based on the degree to which it can classify data into relevant subcategories.

3.6.1 | Support vector machine

Supervised machine learning technique that may be used for both regression and classification problems is the SVM. SVMs are often used in NLP for sentiment analysis, text categorization, and other language-related tasks. An effective supervised machine learning technique employed in both regression and classification applications are the SVM. In spaces with many dimensions, SVM functions well. It is useful for applications like text categorization and gene expression analysis since it performs well even in cases when there are more dimensions than samples. SVMs seek to maximize the margin, or the separation between the decision border and the points of various classes that are closest to each other. Better generalization and a decreased danger of over fitting are frequently the results of this margin-maximizing mindset. SVM's main objective is to identify the optimum hyper plane for classifying the data space's categories. SVM determines the hyper plane for a binary classification that minimizes the border, which is the separation among the hyper plane and the nearest data points (support vectors) for each class. Because of its strength toward over fitting and effectiveness in high-dimensional spaces. Additionally, SVM can deal with both linearly separable and non-linearly separable data by employing a variety of kernel algorithms that raise the source information's spatial scale, allowing for the usage of a linear hyper plane for categorization. SVMs work effectively on datasets that range in size from small to larger. They can handle modest quantities of data effectively with the right optimization and the application of kernel functions, even though their training times might be longer on bigger datasets. Because of its adaptability and solid conceptual underpinning, SVM maintains to be a preferred option for task classification in a number of fields, notably categorization of texts, image recognition, and biotechnology. Finding the global optimum can be more difficult with neural networks or other techniques, but SVMs' convex optimization problem guarantees convergence to the global minimum, offering a unique approach.

3.6.2 | Adaptive content generation based natural language processing algorithm

A complex computational method called adaptive content generation, which is centered on NLP, constantly develops material that is adapted to certain settings and preferences of the user.[28]

The system analyses input data, including interactions between users or domain-specific data, using cutting-edge NLP methods to generate highly pertinent and customized content. The developed material is kept current and accurate in terms of context by adapting its results to accommodate evolving input. The use of this potent technology extends to chatbots to communicate, content suggestion systems, and language generating jobs, all of which improve the user experience by presenting relevant and interesting information in context. The algorithm for the proposed SVM-NLP is given below and the flowchart of SVM-NLP framework was shown in Figure 2 (Algorithm 1).

Figure 2 depicts the overall flow diagram of the proposed SVM-NLP Framework. Chebyshev filter is used in pre-processing and MFCC and K means clustering is used for feature extraction and segmentation. SVM-NLP is employed for classification.

4 | RESULTS AND DISCUSSIONS

The suggested technique has been examined by means of some datasets. Here, the neuro fuzzy SVM-NLP adaptive content generation framework is secondhand in this investigation for the enhancement of VR in the Metaverse for the detection of emotion recognition. The depiction of the suggested method is deliberated by certain factors such as Accuracy, Recall, Precision, F1-score.

4.1 | Performance metrics

Accuracy, which represents as a percentage of every instance in which a data set has been accurately categorized, is the metric that is the easiest to comprehend. It offers a thorough indication of general accuracy. The accuracy equation is presented in Equation (10) below

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} \tag{10}$$

Precision is the proportion of the occurrences that were precisely foretold as positive to all other cases. The precision calculation is offered in Equation (11) below

$$\text{Precision} = \frac{\text{TP}}{(\text{TP} + \text{FP})} \tag{11}$$

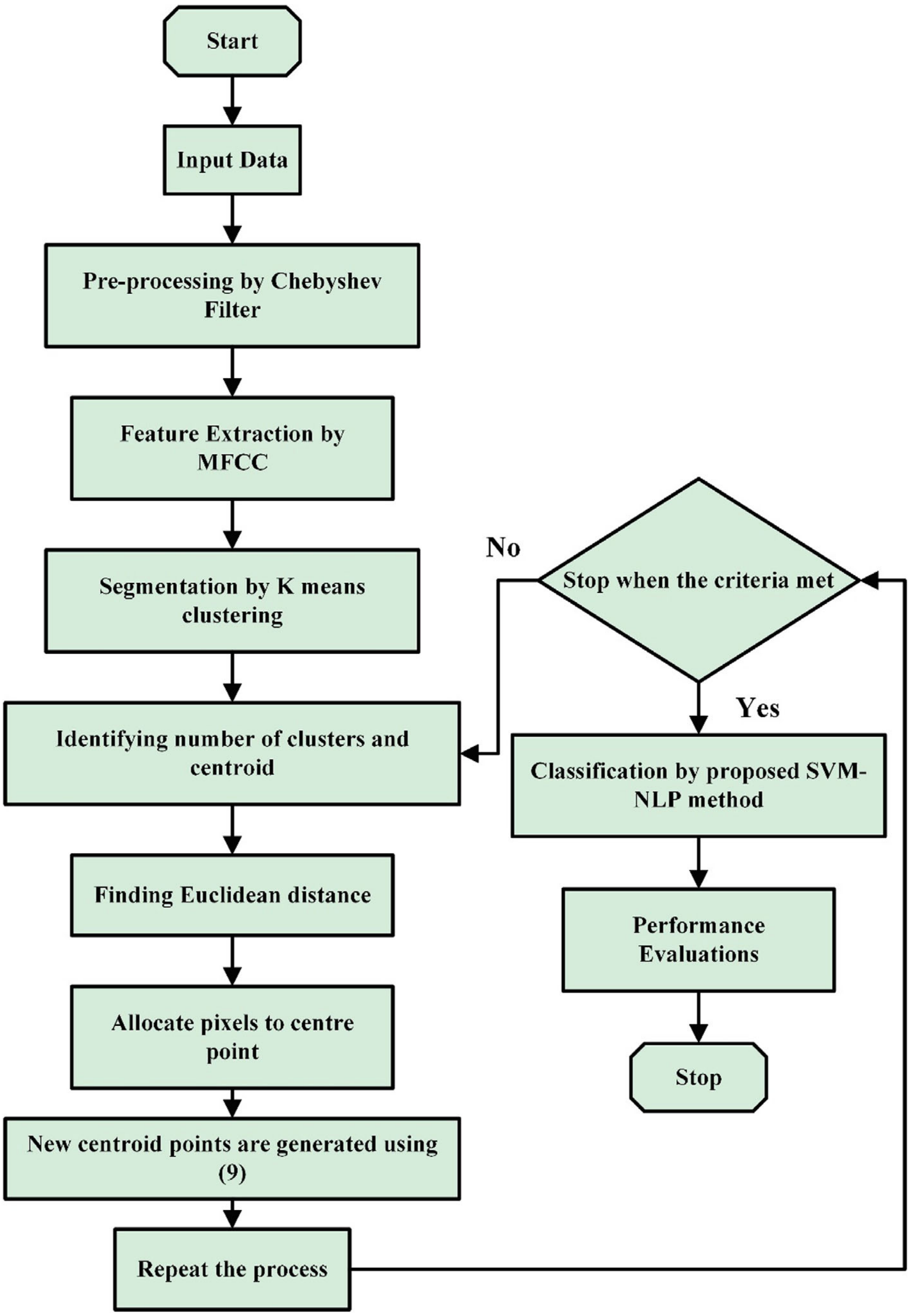

FIGURE 2 Flowchart of proposed SVM-NLP framework.

Algorithm 1. SVM-NLP

Input: Image data
Output: Faces with different types of emotions
Collection of data //Affect Net Dataset
Pre-processing //Chebyshev Filter
Feature Extraction //MFCC
f_{mel} equation is given in Eq. (1)
Windowing equation is given in Eq. (2)
Fast Fourier Transform equation is given in Eq. (3)
MFGR equation is given in Eq. (4)

Logarithmic equation is given in Eqs. (5) and (6)
Discrete Cosine Transform equation is given in Eq. (7)
Segmentation //K means clustering
Euclidean Distance equation is given in Eq. (8)
Centroid equation is given in Eq. (9)
Classification using SVM NLP
Initialize the transformer and NLP Classifier
Training the classifier
Initialize SVM Classifier //SVM-NLP
Comprise modifiers; //Add the utilization of transformers to the imports.
Comprise NLP
Load the data that is utilized for the training process
Train x_data = read_data (samples_ training);
Train y_data = read_data (result_ training);
Specify the transformer set to be used
Set models = Group of transformers;
List the embedded trained dataset [];
Train the weak learner linked to every transformer
for each transformer modules do
NLP classifier;
Model generate embedding (x_train); Array embedding
Classifier training;
Store the data obtained by the poor learner
Stop

Recall processes the percentage of properly projected helpful measures that really arisen among all satisfactory circumstances. It scales how effectively the model can classify each positive occurrence. In Equation (12), the recall formulation is offered.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{12}$$

The F1-score is produced by efficiently summing precision and recall. By mixing the two dimensions, it creates a sole number that delivers an exact assessment of a representation's efficiency. In Equation (13), the F1-Score calculation is delivered.

$$\text{F1} - \text{Score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})} \tag{13}$$

Table 1 gives the accuracy values of the proposed method with the existing methods like FER using CNN- and EEG-based bio signal method.

Figure 3 shows the comparison Accuracy graph of the proposed method and the existing methods like FER using CNN- and EEG-based bio signal method. Moreover, the proposed method produces the higher accuracy of about 90% than the other two methods. Table 2 gives the performance comparison with the other methods.

TABLE 1 Comparison table of accuracy.

Method	Accuracy
FER using CNN	78
EEG based bio-signal method	82
Neuro fuzzy	92

Abbreviations: CNN, convolutional neural network; FER, facial emotion recognition; EEG, electroencephalogram.

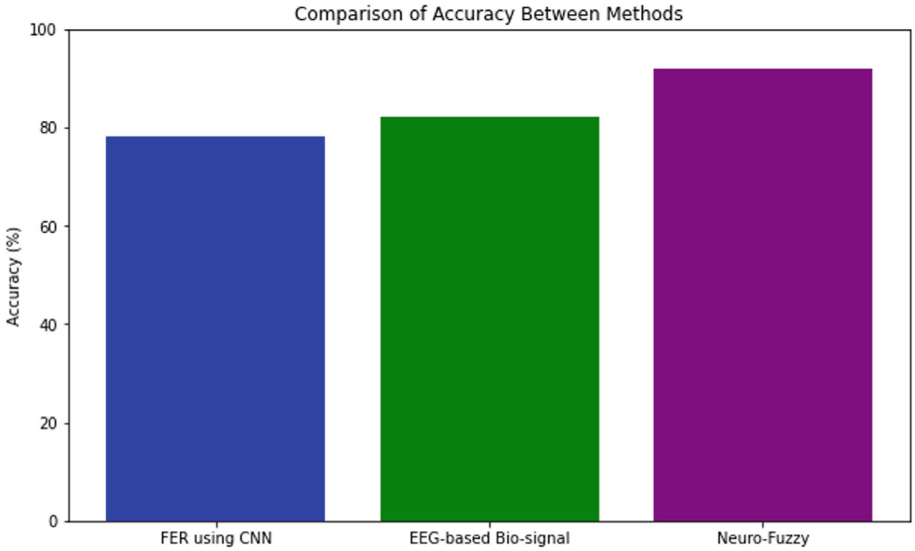

FIGURE 3 Comparison graph of accuracy.

TABLE 2 Performance comparison with existing methods.

Methods	Precision	Recall	F1-score
MARBERT[29]	56.9	56.7	54.8
Bi-LSTM[29]	53.1	56.9	50.56
Bi-GRU[29]	60.23	54.89	60.56
Proposed method	80.5	85	87

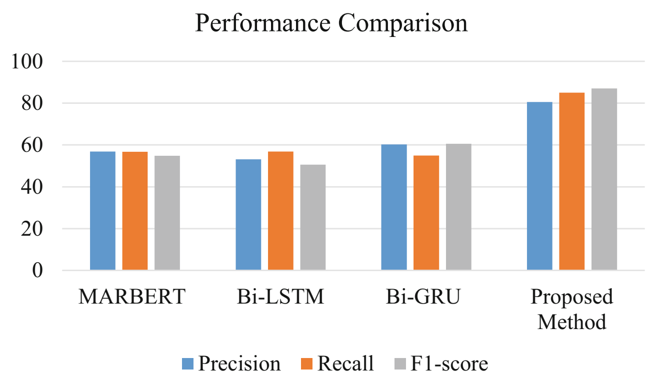

FIGURE 4 Comparison graph with exiting methods.

Figure 4 shows the comparison accuracy graph of the proposed method and the existing methods like MARBERT using BI-LSTM and Bi GRU-based bio signal method. Moreover, the proposed method produces the higher scores than the other methods.

4.2 | Confusion matrix and heat map

The confusion matrix offers a thorough assessment of the effectiveness of emotion recognition by showing the proportion of correct and incorrect forecasts for every emotional category and its pictorial representation is given in Figure 5. Scientists and users can understand the algorithm's advantages and disadvantages in identifying different feelings in

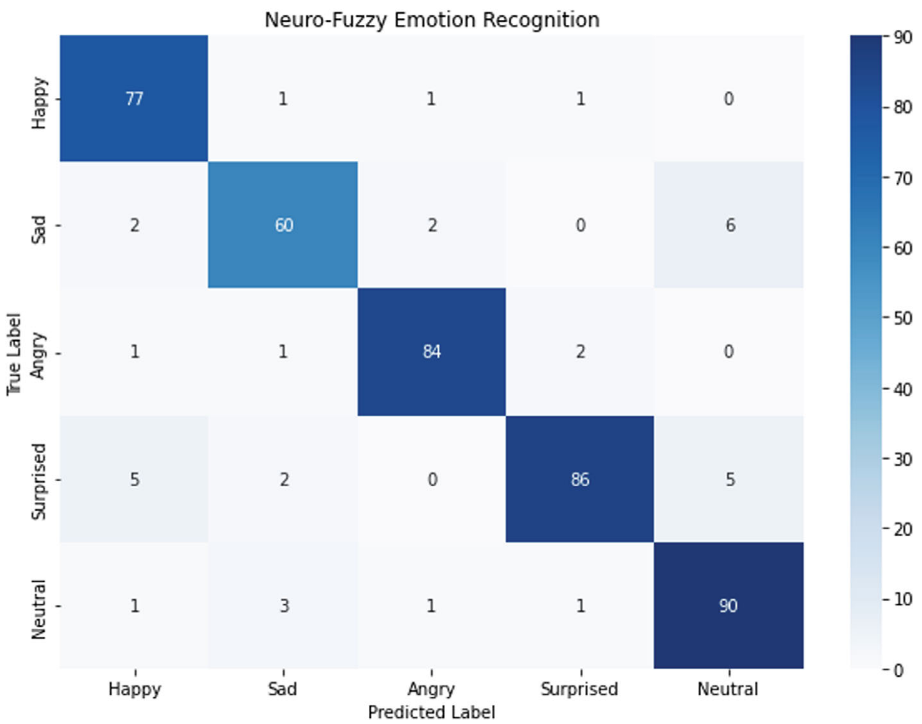

FIGURE 5 Confusion matrix of emotions.

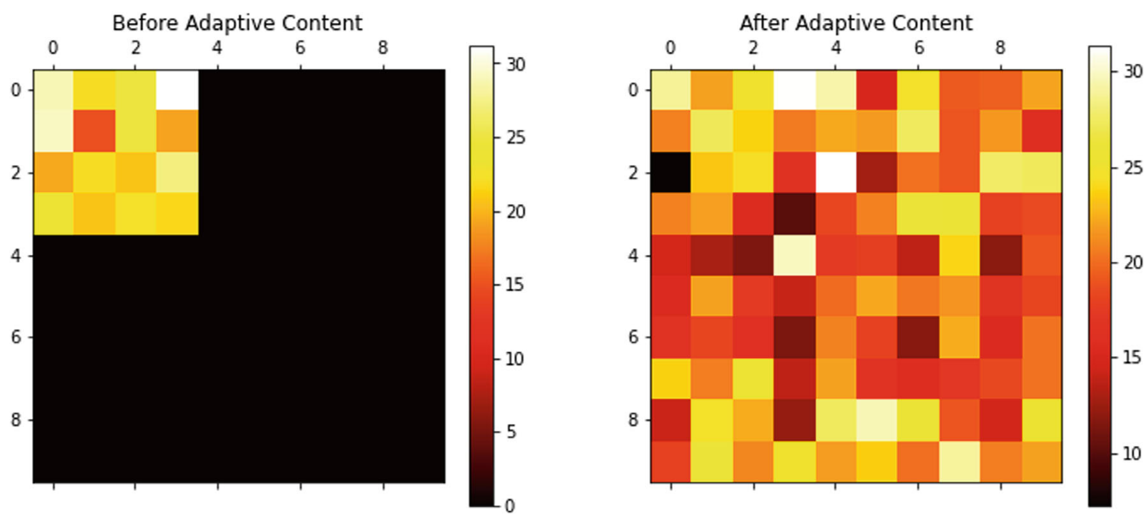

FIGURE 6 Heat map matrix before and after adaptive content.

the virtual Metaverse space by using the heatmap that comes with the study[30] and its pictorial representation is given in Figure 6. The use of these analytical instruments provides useful information about the functionality and possible enhancements of the newly developed neuro-fuzzy-based SVM-NLP emotion recognition for future developments in VR and their wide use in the Metaverse.

4.3 | Average frequency graph

The average frequency graph can indicate how feelings are distributed among different classifications, such as joy, sorrow, rage, astonished, and neutral, and can also demonstrate which feelings are more or less prevalent in the data that has been gathered.[31] These data can help in evaluating the method's efficiency, especially if it excels at over representing some

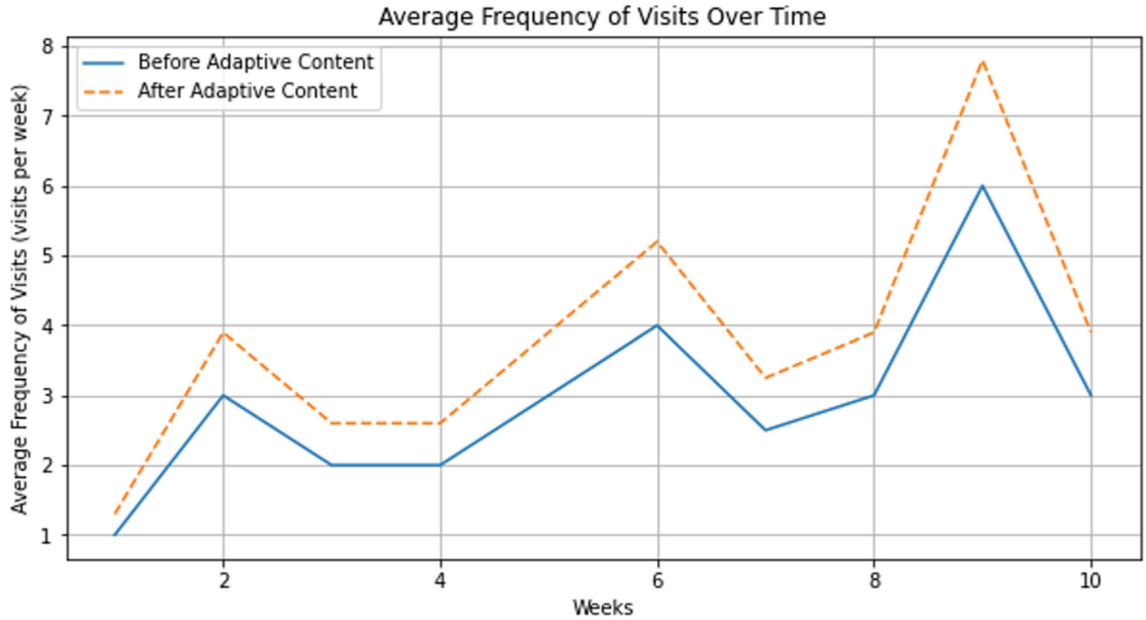

FIGURE 7 Adaptive frequency graph.

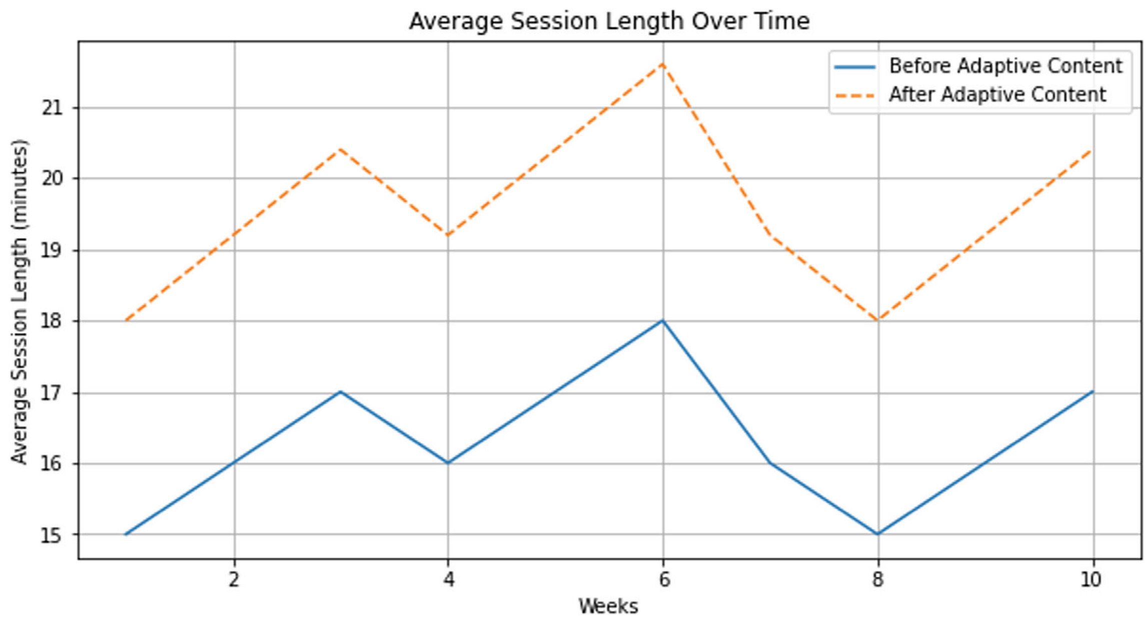

FIGURE 8 Average session graph.

emotions while having trouble with others. Figure 7 shows the average frequency length over a period of 10 weeks before and after adaptive content visit.

4.4 | Average session graph

The average session graph is a time-based illustration that enables researchers to examine how the software performs as users engage with it repeatedly or over an extended period of time.[32] Researchers can identify trends and patterns in the platform's behavior by graphing metrics like emotion detection accuracy, content relevancy, user happiness, or any other pertinent indicators of performance. Figure 8 shows the average session length over a period of 10 weeks before and after adaptive content visit.

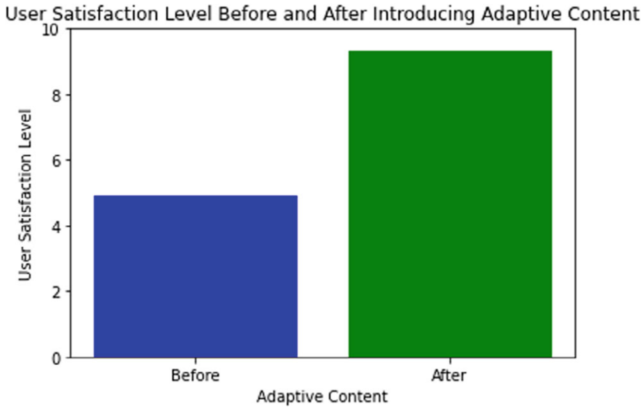

FIGURE 9 User satisfaction graph before and after using adaptive content.

TABLE 3 Precision, Recall, and F1-score of different types of emotions.

Classes	Precision	Recall	F1-score
Happy	0.90	0.96	0.93
Sad	0.90	0.86	0.88
Angry	0.95	0.95	0.95
Surprised	0.96	0.88	0.91
Neutral	0.89	0.94	0.91

The level of comfort and acceptance that users show concerning the ability of the system to correctly identify and react to their respective emotional states is referred to as user satisfaction. High satisfaction among users shows that the emotion detection technology successfully records and translates users' feelings, resulting in a more individualized and comprehensive experience.[33] Users who are happy with the system are more inclined to think of it to be responsive and sympathetic, which encourages confidence and emotional investment. In order to improve user interactions, optimize user interfaces, and ensure that the system has practical applicability in a variety of real-world circumstances, it is essential to comprehend user satisfaction in emotion recognition. Figure 9 shows the diagrammatic representation of the user's satisfaction before and after introducing the adaptive content.

Table 3 shows the performance metrics of precision, recall and F1-score value of different types of emotions like Happy, Sad, Angry, Surprised, and Neutral. Figure 10 shows the performance metrics graph of Precision, Recall, and F1-score of different kinds of emotional faces of peoples.

4.5 | Discussion

In order to improve VR experiences in the Metaverse, the research investigates neuro-fuzzy-based SVM-NLP emotion recognition adaptive content generation methods. Outcomes are used to assess the effectiveness of emotion recognition, comprising an accuracy graph and confusion matrix. The average frequency graph examines the dataset's distribution of classes and identifies any potential abnormalities. The typical session graph analyses effectiveness over a period of time and identifies patterns and incremental advancements. The focus on user pleasure draws attention to the value of emotional interactions in fostering favorable attitudes and user engagement. In general, the investigation improves VR technology, offering profoundly emotional and comprehensive metaphysical experiences for leisure, learning, and mental wellness. The proposed paper introduces a methodology for real-time emotion recognition within the Metaverse using neuro-fuzzy-Based Emotion Recognition and Adaptive Content Generation Algorithms. However, a critical concern that remains unaddressed is the issue of latency. The paper should meticulously investigate and quantify the time required to obtain emotion detection results using the proposed methodology. Understanding the latency associated with

FIGURE 10 Performance metrics graph of different types of emotional faces.

the emotion recognition process is essential for assessing the feasibility of real-time applications, particularly in scenarios where swift responses, such as modifying avatar expressions or adapting content, are crucial for enhancing VR experiences within the Metaverse. Addressing this fundamental issue would significantly contribute to the practical applicability and effectiveness of the proposed methodology.

5 | CONCLUSIONS AND FUTURE WORK

The final result of the study highlights the breakthrough impacts of the Metaverse on social interactions and the interaction with digital information while emphasizing the need for modern technology that can understand users' emotions and offer adaptable, customized solutions. The researchers suggest a unique method, the neuro-fuzzy-based SVM-NLP Adaptive Content Generation Algorithms, to solve this problem by generating highly engaging and personalized experiences within the Metaverse. Employing biometric information, the neuro-fuzzy SVM algorithm precisely determines individuals' state of emotions, and NLP content generating algorithm improve the accuracy of emotion classification. NLP incorporation makes it possible to alter individuals, tales, and interactive elements in real-time, improving user pleasure and immersion in VR. The acquired results reveal the SVM-NLP method's excellent accuracy in recognizing emotions, highlighting its ability to increase the impressiveness and mental engagement of the Metaverse. As a consequence, this combination opens the door for many applications that improve human–computer interactions, including virtual rehabilitation, education, entertainment, and social networking. This method's accomplishment, which achieved an astonishing accuracy of about 92%, provides the groundwork for future developments in emotion-aware methods, offering more engrossing and emotionally stimulating moments inside the changing Metaverse environment.

Future study might investigate feasible uses of the suggested strategy, enhance computational effectiveness, and carry out user research to assess its effects on diverse Metaverse applications and populations of users. In the future, further advancements in enhancing VR experiences within the Metaverse can be achieved through the integration of cutting-edge technologies such as Neuro-Fuzzy-Based Emotion Recognition and Adaptive Content Generation Algorithms. Incorporating neuro-fuzzy systems into emotion recognition can lead to more nuanced and accurate understanding of users' emotional states, allowing for personalized and responsive content generation. Efficient real-time processing is imperative in the VR and Metaverse domain, particularly for swiftly detecting and processing emotions to enhance data utilization, such as adapting avatar expressions based on recognized emotions from vocal flow. To ensure a seamless user experience, it is crucial to investigate the feasibility of analyzing the real-time processing capabilities of the proposed Neuro-Fuzzy-Based Emotion Recognition system. Addressing potential limitations is equally vital, necessitating an open acknowledgment and discussion of challenges related to emotion detection and content adaptation speed in the conclusions, along with transparently identifying impediments like bottlenecks, technological constraints, or computational limitations. Proposing future research directions to overcome these challenges contributes to a comprehensive understanding of the proposed system's capabilities and provides valuable insights for enhancing VR experiences within the Metaverse.

AUTHOR CONTRIBUTIONS

Mhd Saeed Sharif: Supervision; resources; writing – review and editing; funding acquisition; conceptualization. **Oshamah Ibrahim Khalaf:** Conceptualization; software; formal analysis; writing – review and editing. **Dhamodharan Srinivasan:** Investigation; methodology; software; data curation. **Sameer Algburi:** Conceptualization; investigation; project administration; writing – review and editing. **Jeevanantham Vellaichamy:** Conceptualization; investigation; validation; visualization; writing – original draft; formal analysis; data curation. **Dhanasekaran Selvaraj:** Conceptualization; investigation; writing – original draft; writing – review and editing; formal analysis; data curation. **Wael Elmedany:** Writing – review and editing; investigation; project administration; resources.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID

Oshamah Ibrahim Khalaf https://orcid.org/0000-0002-4750-8384
Mhd Saeed Sharif https://orcid.org/0000-0002-4008-8049

REFERENCES

1. Abbaschian BJ, Sierra-Sosa D, Elmaghraby A. Deep learning techniques for speech emotion recognition, from databases to models. *Sensors*. 2021;21(4):1249. doi:10.3390/s21041249
2. Imani M, Montazer GA. A survey of emotion recognition methods with emphasis on E-learning environments. *J Netw Comput Appl*. 2019;147:102423. doi:10.1016/j.jnca.2019.102423
3. Issa D, Fatih Demirci M, Yazici A. Speech emotion recognition with deep convolutional neural networks. *Biomed Signal Process Control*. 2020;59:101894. doi:10.1016/j.bspc.2020.101894
4. Spadoni E, Carulli M, Mengoni M, Luciani M, Bordegoni M. Empowering virtual Humans' emotional expression in the metaverse. *Universal Access in Human-Computer Interaction*. Springer; 2023:133-143. doi:10.1007/978-3-031-35897-5_10
5. Pandey P, Seeja KR. Subject independent emotion recognition from EEG using VMD and deep learning. *J King Saud Univ*. 2022;34(5):1730-1738. doi:10.1016/j.jksuci.2019.11.003
6. Mamun MA, Sakib N, Gozal D, et al. The COVID-19 pandemic and serious psychological consequences in Bangladesh: a population-based nationwide study. *J Affect Disord*. 2021;279:462-472. doi:10.1016/j.jad.2020.10.036
7. Zhang Y. The top-of-atmosphere, surface and atmospheric cloud radiative kernels based on ISCCP-H datasets: method and evaluation. *Atmos Sci*. 2021;126(4):1-34. doi:10.1002/essoar.10508753.1
8. Busso C, Bulut M, Narayanan S. Toward effective automatic recognition Systems of Emotion in speech. In: Gratch J, Marsella S, eds. *Social Emotions in Nature and Artifact*. Oxford University Press; 2013:110-127. doi:10.1093/acprof:oso/9780195387643.003.0008
9. Anvarjon T, Mustaqeem, Kwon S. Deep-net: a lightweight CNN-based speech emotion recognition system using deep frequency features. *Sensors*. 2020;20(18):5212. doi:10.3390/s20185212
10. Liu W, Qiu J-L, Zheng W-L, Lu B-L. Comparing recognition performance and robustness of multimodal deep learning models for multimodal emotion recognition. *IEEE Trans Cogn Dev Syst*. 2022;14(2):715-729. doi:10.1109/TCDS.2021.3071170
11. Mustaqeem, Kwon S. A CNN-assisted enhanced audio signal processing for speech emotion recognition. *Sensors*. 2019;20(1):183. doi:10.3390/s20010183
12. Sanguansub N, Kamolrungwarakul P, Poopair S, Techaphonprasit K, Siriborvornratanakul T. Song lyrics recommendation for social media captions using image captioning, image emotion, and caption-lyric matching via universal sentence embedding. *Soc Netw Anal Min*. 2023;13(1):95. doi:10.1007/s13278-023-01097-6
13. Siripattanadilok W, Siriborvornratanakul T. Recognition of partially occluded soft-shell mud crabs using Faster R-CNN and Grad-CAM. *Aquac Int*. 2023. doi:10.1007/s10499-023-01307-0
14. Kittipongdaja P, Siriborvornratanakul T. Automatic kidney segmentation using 2.5D ResUNet and 2.5D DenseUNet for malignant potential analysis in complex renal cyst based on CT images. *J Image Video Proc*. 2022;2022(1):5. doi:10.1186/s13640-022-00581-x
15. Bhavan A, Chauhan P, Shah RR. Bagged support vector machines for emotion recognition from speech. *Knowl-Based Syst*. 2019;184:104886. doi:10.1016/j.knosys.2019.104886
16. Cimtay Y, Ekmekcioglu E. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition. *Sensors*. 2020;20(7):2034. doi:10.3390/s20072034
17. Akhand MAH, Roy S, Siddique N, Kamal MAS, Shimamura T. Facial emotion recognition using transfer learning in the deep CNN. *Electronics*. 2021;10(9):1036. doi:10.3390/electronics10091036
18. Etienne E, Leclercq A-L, Remacle A, Dessart L, Schyns M. Perception of avatars nonverbal behaviors in virtual reality. *Psychol Mark*. 2023;40(11):2464-2481. doi:10.1002/mar.21871

19. Marín-Morales J, Llinares C, Guixeres J, Alcañiz M. Emotion recognition in immersive virtual reality: From statistics to affective computing. *Sensors*. 2020;20(18):5163. doi:10.3390/s20185163
20. Visconti A, Calandra D, Lamberti F. Comparing technologies for conveying emotions through realistic avatars in virtual reality-based metaverse experiences. *Comput Animat Virtual Worlds*. 2023;34(3–4):e2188. doi:10.1002/cav.2188
21. Mehendale N. Facial emotion recognition using convolutional neural networks (FERC). *SN Appl Sci*. 2020;2(3):1-8. doi:10.1007/s42452-020-2234-1
22. Batbaatar E, Li M, Ryu KH. Semantic-emotion neural network for emotion recognition from text. *IEEE Access*. 2019;7:111866-111878. doi:10.1109/ACCESS.2019.2934529
23. Poria S, Majumder N, Mihalcea R, Hovy E. Emotion recognition in conversation: research challenges, datasets, and recent advances. *IEEE Access*. 2019;7:100943-100953. doi:10.1109/ACCESS.2019.2929050
24. Schoneveld L, Othmani A, Abdelkawy H. Leveraging recent advances in deep learning for audio-visual emotion recognition. *Pattern Recogn Lett*. 2021;146:1-7. doi:10.1016/j.patrec.2021.03.007
25. Koduru A, Valiveti HB, Budati AK. Feature extraction algorithms to improve the speech emotion recognition rate. *Int J Speech Technol*. 2020;23(1):45-55. doi:10.1007/s10772-020-09672-4
26. Yang Z, Huang Y. Algorithm for speech emotion recognition classification based on Mel-frequency cepstral coefficients and broad learning system. *Evol Intel*. 2022;15(4):2485-2494. doi:10.1007/s12065-020-00532-3
27. Asghar MA, Khan MJ, Fawad, et al. EEG-based multi-modal emotion recognition using bag of deep features: an optimal feature selection approach. *Sensors*. 2019;19(23):5218. doi:10.3390/s19235218
28. Graterol W, Diaz-Amado J, Cardinale Y, Dongo I, Lopes-Silva E, Santos-Libarino C. Emotion detection for social robots based on NLP transformers and an emotion ontology. *Sensors*. 2021;21(4):1322. doi:10.3390/s21041322
29. Mansy A, Rady S, Gharib T. An ensemble deep learning approach for emotion detection in Arabic tweets. *IJACSA*. 2022;13(4):980-990. doi:10.14569/IJACSA.2022.01304112
30. Bao G, Zhuang N, Tong L, et al. Two-level domain adaptation neural network for EEG-based emotion recognition. *Front Hum Neurosci*. 2021;14:605246. doi:10.3389/fnhum.2020.605246
31. Du Y, Crespo RG, Martínez OS. Human emotion recognition for enhanced performance evaluation in e-learning. *Prog Artif Intell*. 2023;12(2):199-211. doi:10.1007/s13748-022-00278-2.21
32. Mandala V, Senthilnathan T, Suganyadevi S, Gobhinath S, Selvaraj DS, Dhanapal R. An optimized back propagation neural network for automated evaluation of health condition using sensor data. *Measur Sens*. 2023;29:100846.
33. Dhana Sekaran S, Rufus NHA, Anand D, et al. Evolutionary optimization with deep transfer learning for content based image retrieval in cloud environment. In Antona M, Stephanidis C, (eds.), *2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*. IEEE; 2022.

How to cite this article: Khalaf OI, Srinivasan D, Algburi S, et al. Elevating metaverse virtual reality experiences through network-integrated neuro-fuzzy emotion recognition and adaptive content generation algorithms. *Engineering Reports*. 2024;e12894. doi: 10.1002/eng2.12894