KeAi
CHINESE ROOTS
GLOBAL IMPACT

International Journal of Innovation Studies

ISSN:2096-2487 | E-ISSN:2589-2975

EMOTION-AWARE REINFORCEMENT LEARNING FOR ADAPTIVE VIDEO GAMING THERAPY

J.Usha
Research scholar, Department of Computer Applications, Dr.M.G.R.Educational and Research Institute, Chennai, Tamil Nadu,ushajay11@gmail.com

Dr.Viji Vinod
Head of the department, Computer Applications, Dr.M.G.R.Educational and Research Institute, Chennai, Tamil Nadu,hod-mca@drmgrdu.ac.in

ABSTRACT

Reinforcement Learning (RL) in Video Gaming Therapy, combined with emotion detection, dynamically adapts game environments to optimize therapeutic outcomes by personalizing experiences based on real-time emotional responses.A key challenge is integrating accurate emotion detection with reinforcement learning to personalize video game therapy effectively while ensuring real-time responsiveness and user engagement.This model presents a novel framework for an Inverted Residual Adaptive Convo-Depth NeuroNet that integrates Emotion Detection and Reinforcement Learning (RL) to create a therapy system for an adaptive video gaming therapy system. The Emotion Detection Module uses a pre-trained MobileNetV2 to analyzethe data and classify the emotional state, such as Happy, Fear, Angry, Disgust, Neutral, Sad, and Surprise. The Reinforcement Learning Module leverages Deep Q-Network (DQN)algorithms, where the state combines the player's emotional state and game performance. The RL model adjusts the game difficulty and provides personalized rewards. This methodpurposes to create a more immersive and effective video game therapy environment, where the game evolves in response to both emotional and performance rewards, ensuring a personalized and engaging experience.The Python-based simulation achieved performance metrics ranging from 83% to 85%, including F1-score, precision, recall, and accuracy, validating the effectiveness of the emotion-aware adaptive therapy system.

Keywords: Reinforcement Learning (RL),Video Gaming Therapy, Emotion Detection, Convolutional Neural Network (CNN), Python, Deep Q-Network (DQN).

1. Introduction

The advancement of AI has progressively affected various fields, including health care, education, and entertainment (Baker and Xiang, 2023). According to Bakerand Xiang(2023), one of the most promising applications for AI is now being introduced to video games for health-related applications. The use of video game therapy is an exciting new type of modality designed to relieve stress or treat psychological issues or any neurocognitive issues while considering one's emotional wellbeing (Vindigni, 2023). Regular therapeutic video games have the disadvantage that they cannot consider the user's emotional state, which limits their effectiveness, as the gaming experience can diminish the user's emotional state. Emotions are pivotal to human behavior, cognition, and decision-making (Mehra et al., 2023). In gaming therapy, understanding and responding to the players' feelings can greatly increase therapeutic benefits (Parvaria and Schofield, 2025). Traditional therapeutic games often follow predetermined rules and stable difficulty levels, which may not be aligned with individual users' emotional and cognitive needs. Some players may experience despair, boredom, or anxiety, which may obstruct the effectiveness of therapy (Bocci et al., 2023). An emotion-aware system to detect emotional states uses advanced AI techniques, including facial expression analysis, physiological (heart rate, skin conductance, brain activity), and voice recognition (Pal et al., 2021). This allows real-time emotional input to reinforcement learning models to create a personal and responsible medical experience to adapt to the mechanics, difficulty, challenges, and storyline of the game.

Possible applications of emotion-aware reinforcement learning in gaming therapy have various domains, including mental health, rehabilitation, and neurological disorders. For individuals suffering from anxiety, depression, or PTSD, adaptive gaming therapy can provide a controlled and immersive environment where they can gradually face and manage their emotions (Yadav et al., 2024). For example, CBT -The comfort-based game can use emotion-aware reinforcement to adapt the user's relaxation level and tailor exposure scenarios, which can make therapy more effective. Similarly, patients recovering from stroke or brain injuries can benefit from adaptive games that adjust the difficulty and reaction mechanisms based on their emotional and cognitive progress (Spytska, 2024). Children with ASD can also benefit from personalized gaming interventions that teach social and emotional skills attractively and supportively. To develop an adaptable video game therapy system, this model offers a unique framework for an Inverted Residual adaptable Convo-Depth NeuroNet that combines Reinforcement Learning (RL) and Emotion Detection.

- **Dataset Collection:** The FER-2013 dataset is collected, containing various facial expressions for emotion identification.
- **Emotion Detection Training:** Inverted Residual Adaptive Convo Depth NeuroNet, integrating CNN and MobileNetV2, is used to analyze and classify emotions.
- **Reinforcement Learning:** The player's emotional state and game performance are combined to define the state in DQN algorithms.
- **Evaluation:** The proposed model achieved high accuracy, F1-score, precision, and recall in dynamically adapting gameplay based on players' emotional and performance feedback.

This research is structured as follows: Part 2 examines the conceptual work on recognizing facial expressions. Part 3 underlines the functioning, covers the dataset details, emotion detection, and classifies the emotional states. Part 4 discusses the results and the effectiveness of the model. Finally, part 5 ends with the scalability of the framework and insight into future research directions.

2. Related works

This section discusses recent developments in emotion-aware systems used in media, gaming, mobile networks, and user experience platforms.

EAIG and optimal signal features were used to create aSERJ in Lin et al., (2023). During games, the SERJ recognized shifts in a player's emotional state. The EAIG and SERJ were tested on ten participants and were shown to be effective. By evaluating unique occurrences brought on by a player's emotions, the game adjusts itself to improve the player's experience.

According to the results, SERJ, based on optimal signal characteristics, outperforms traditional machine learning-based techniques.
The MFGD voice dataset was used by Song et al., (2024) to examine how players communicate their emotions while playing games. It created a multi-task learning structure and included descriptions for continuous degrees of valence and arousal. It was recommended to evaluate a Temporally Oriented ResNet. Outperforming the traditional ResNet18 results, the suggested framework accurately predicted players' arousal and valence from speech. It assisted voice interface designers in modifying and personalizing player experiences.
To increase accuracy, Patel and Mehta, (2024) evaluated many models for identifying emotions from facial expressions and presented a novel CNN. To increase user engagement on music streaming platforms, the technology delivered personalized music recommendations based on facial expressions recorded by webcams or images. It illustrated that adding facial expression recognition to recommendation systems could improve the user experience by offering musical recommendations that were sensitive to user emotions.
Awwad (2021) proposed a visual emotion-aware cloud localization user experience architecture based on mobile location services. Emotion-aware data was captured by the framework, processed on a cloud server, and then analyzed for real-time localization. After establishing a relationship between the user's emotional response and the default language of the app, the framework loads the resources and modifies screen elements in response to location and real-time feelings.
ML was used in Ok et al., (2024) to identify emotional stages and improve the dynamically modified information to improve e-reading experiences. ML algorithms can modify textual complexities, suggest materials, and provide measures by analyzing behavior patterns, biometric data, and facial expressions. It emphasized the ability of emotionally intelligent e-reading platforms while discussing important machine learning models and problems.
Duan et al., (2024) presented a sophisticated feeling identity system to improve emotional response and create more organic, spontaneous, and psychologically hypnotic experiences.Duan et al., (2024) presented a sophisticated feeling identity system. Technology explains complex emotional indicators in real time by integrating text data, speech, and facial expressions. It improved traditional techniques and promoted strong connections between users and the digital environment, demonstratingeffective emotion identification and a notable improvement in the F1 score.
Stoynov, (2023) proposed a framework called ENPto integrate emotion-aware networking in practice in 5G mobile networks. It attempts to understand the user's feelings and modify services accordingly. The ENM framework includes user profile management, providing EASUPMF and individual emotional services. The performance of the ENM was evaluated using a new KPI. According to findings, the ENM could bring a revolution to the mobile network industry.

3. Methodology

The FER-2013 dataset is initially collected, having varied facial expressions for emotion identification. The emotion detection phase applies the Inverted Residual Adaptive Convo Depth NeuroNet, which combines CNN and MobileNetV2, to classify the identified emotions and classify them. The player's emotional state and game performance are combined to create the state in DQN algorithms. This adaptive strategy enables real-time emotional evaluation, providing customized game difficulty adjustment for an enjoyable therapy session. Figure 1 shows the overall process of recognizing facial expressions for video game therapy.

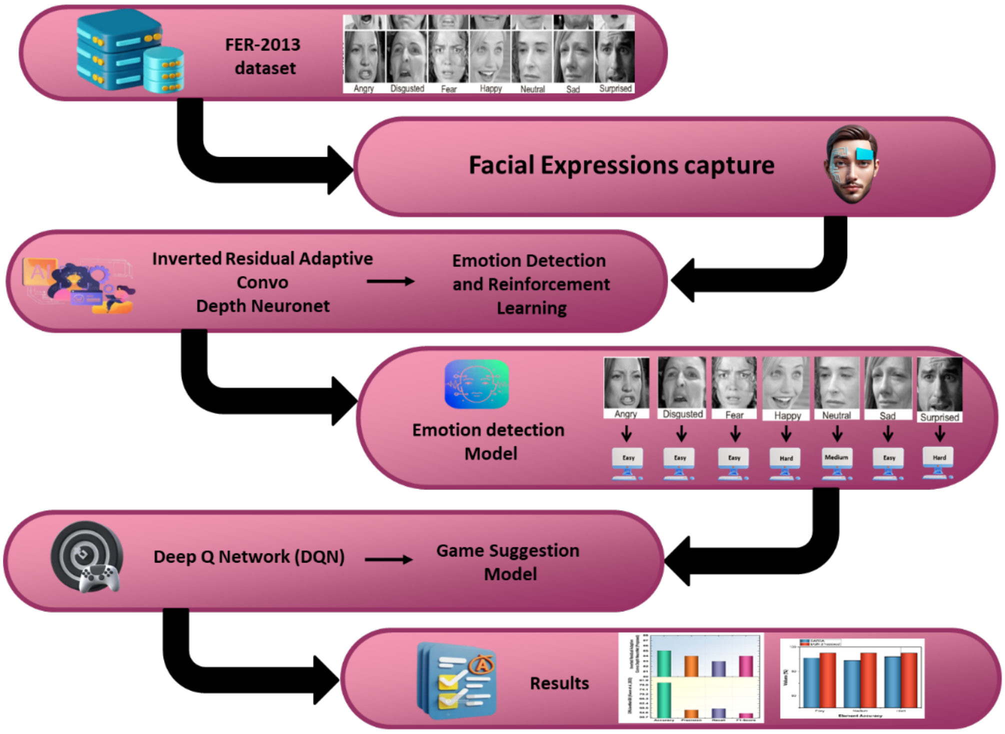

Figure 1: Overall process of recognizing facial expressions for video game therapy.

3.1 Dataset and Data Pre-processing

The collection consists of 48x48 pixel grayscale facial images. Each face's expression needs to be categorized according to the emotion into one of the seven groups. There are 28,709 samples in the training set and 3,589 examples in the public test set. The ImageDataGenerator is used for data augmentation, rescaling pixel values, and randomly transforming images during training, including flipping, zooming, and rotating. A 20% validation split is specified for the training data, which is loaded from directories and batched for training. After resizing the images to 224x224 pixels and batching them for training, the training and validation data are loaded from the folders.

Source: https://www.kaggle.com/datasets/msambare/fer2013

3.2 Emotion detection using Inverted Residual Adaptive Convo Depth NeuroNet

The proposed emotion detection model takes advantage of a CNN architecture designed to efficiently classify emotions from facial expressions. The CNN consists of convolutional, pooling, and FCL layers,where the convolutional layers extract spatial features, and the pooling layers reduce dimensionality. The model includes Mobilenetv2, which increases efficiency through Depthwise Separable Convolutions while reducing computational complexity, while maintaining accuracy. Linear bottlenecks are employed to prevent information loss in non-linear layers, ensuring better feature representation. In addition, IR supports good gradient propagation and enhances network efficiency with minimal memory resources. The lightweight yet resilient structure guarantees real-time emotion classification and thus can find applications in adaptive gaming.Figure 2 shows the overall process of emotion detection using MobileNetV2-based CNN.

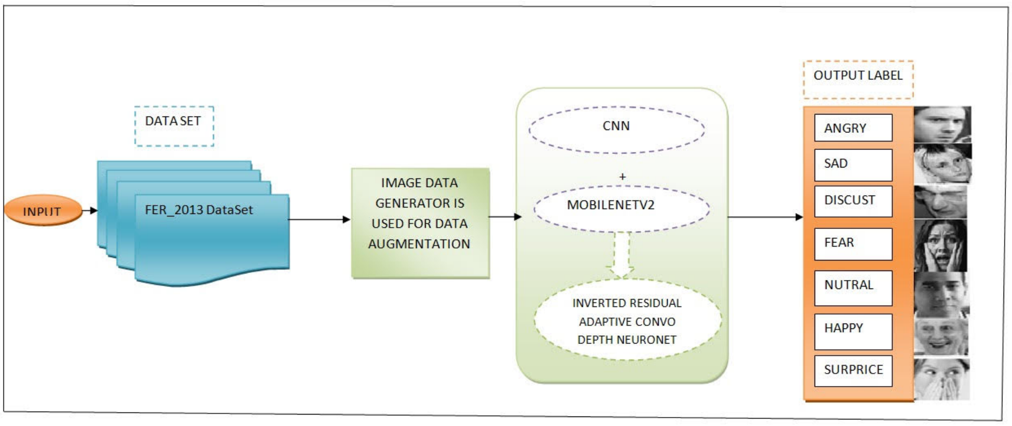

Figure 2: The overall process of emotion classification using MobileNetV2-based CNN

3.2.1MobileNetV2

MobileNetV2 is a light DL model specifically designed for mobile and real-time applications. This design enables efficient emotion classification using minimal resources, however, it retains the accuracy required. The model is ideally suited to adaptive gaming therapy because it has actual and precise facial expression recognition. It focuses on mobility and uses the MobileNetV2 model for image classification. MobileNetV2 uses the DSC method to make it compatible.

A unique structure called inverted residuals was created to solve the problem of data loss for irregular layers in convolutional blocks. This structure not only uses linear bottlenecks but also preserves data.

Depthwise Separable Convolutions (DSC): The DSC that was used in MobileNetV1 is also used in MobileNetV2. Combining the Depthwise and Pointwise convergence can reduce the overall number of parameters and computing cost to about 18 of the normal convolution.

Linear Bottlenecks (LB): The presence of LB is supported by two features. If the layer's output necessitates a type of ReLU(Aw) and its outcome is still not 0, then the relevant portion of input space (w) is positively restricted to an explosive transformation (Aw).

Deep networks can only perform linear classifiers on the non-zero size part of the output domain. However, ReLU is probably likely to lose information in a channel when it compresses data. Information can be retained in other channels in various combinations, provided that the set of activation functions has a structure. A layer can include the input into a lower-dimensional subset of the activation space, which is modeled by this type of structure. If the input area can be consolidated into a low dimension, then using $ReLU$ as an activation method only requires a linear translation to retain all the pertinent information. It can be achieved by adding linear bottleneck layers to the convolution blocks.

Inverted Residuals (IR): In residual blocks, each bottleneck block begins with an input, develops, and then passes through several bottlenecks. To increase the gradients' communication capabilities in numerous layers, shortcuts can be placed between various bottlenecks for two primary reasons:

- The bottlenecks comprise the majority of the information.
- The expansion layer can be considered an application feature with a non-linear transformation of the tensor.

Additionally, the IR consumes less memory than the traditional architecture.

Model Architecture: It classifies 160 x 160 RGB images using the depthwiseseparable convolution model MobileNetV2. The framework uses a lightweight depthwise convolution to eliminate the reinforcement input after increasing its dimensions to high levels. A linear combination is used to shift features back to a low-dimensional form. This framework addresses the firm filter count of MobileNetV1, preserves information, and keeps the block lightweight. It builds layers for certain categorization problems using transfer learning,as shown in Figure 3.

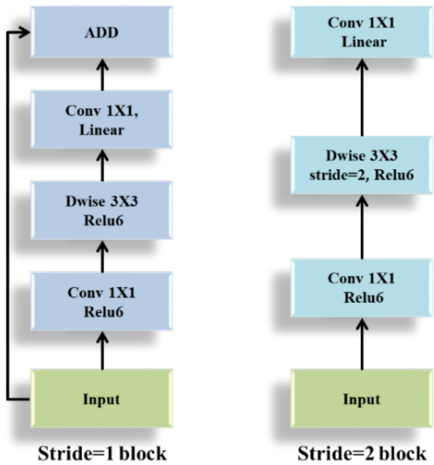

Figure 3: The convolutional blocks in MobileNetV2

3.2.2 CNN

CNN is a DL architecture designed for image classification, which makes it ideal for facial expression recognition. It consists of a CL that identifies spatial patterns, pooling layers that reduce the dimensionality, and an FCL for classification. Activation functions such as $ReLU$ improve non-linearity, ensuring better performance in emotion recognition functions.

CNNs are a type of deep FF-ANN that are employed to yieldcorrect results in tasks involving AI,including image recognition and categorization. Compared to a typical neural network, CNNs contain more layers. A non-linear activation function is used to determine weights, biases, and outcomes. Three dimensions, height, width, and depth, are used to structure the neurons in the CNN.

Figure 4 shows the CNN architecture, which consists of the FCL, pooling layer, and CL.Usually, the pooling and CL are alternated, and each filter's length grows from one side to the other while the output size decreases. The final stage, known as the FCL, is analogous to CNN's final layer.

Pixel values in the image are used as input data. Its three dimensions are height, width, and depth (RGB channels), as shown in $[50\ x\ 50\ x\ 3]$. The CL computes the outputs of neurons at particular locations in the input. A set of attainable filters, called kernels, comprises the parameters of the layer. To determine the dot product across inputs and filter entries, these filters employ convolution to move over each level of the volume of input and span its length and width. When the system detects a certain kind of feature at a particular spatial point in the input, it acquires filters that activate.As a result, the activation map of that filter is two-dimensional. The element-wise activation function is performedby the ReLU layer function. Equation (1) defines theReLU function.

$$e(w) = \max(0, w) \quad (1)$$

For positive values, this function grows linearly, while for negative values, it is zero. The volume size does not change as a result. The maximal activation within an area is output by the pooling layer. Spatial metrics like length and width are down-sampled using this method. The FCL is the neural network's last result layer. This layer uses the widely used softmax activation to produce a probability rating for the number of output categories.

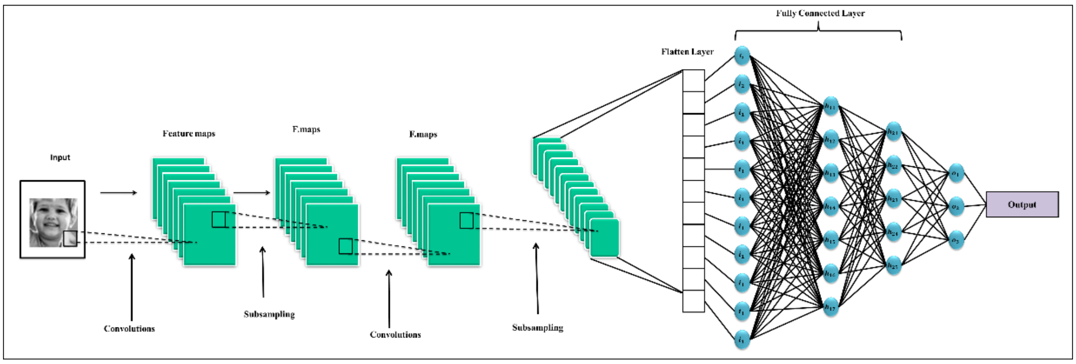

Figure 4: CNN architecture

3.3 Deep Q-Network (DQN)

In this research, DQN (Deep Q-Network) is used solely for facial expression recognition. Facial expressions are converted into emotion labels. These labels are given as input to the DQN model. The DQN then accurately identifies the facial expressions based on the learned emotional patterns. Figure 5 shows the overall process for DQN-based facial expression recognition.

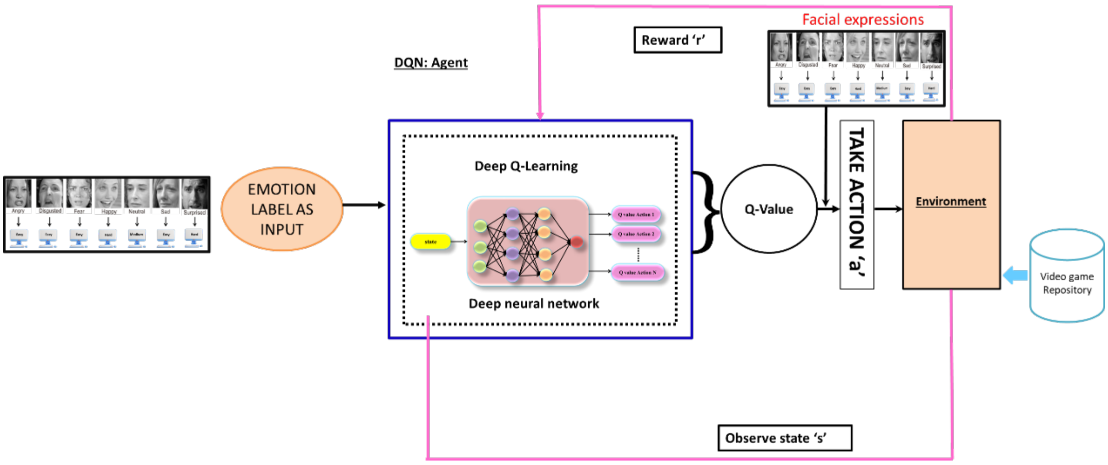

Figure 5: DQN-based real time facial expression recognition

DQN has been employed to dynamically adjust the video game parameters based on the emotional status and the player's performance. The DQN algorithm estimates the Q function using a DNN, enabling the system to learn complex decision policies. By leveraging DRL, the system adjusts individuals' gaming experiences, the level of difficulty, and reward structures to maintain emotional stability and maximize therapeutic effectiveness.

The term "DQN" specificallyrefers to the use of DNNs in conjunction with the Q-Learning method. Because of their strong capacity to learn multiple levels of abstraction from data, DNNs are a good option for generating generalizations about the Q-function from these abstractions. These abstractions might even outperform human-designed features as representations if there is sufficient information.

It employs a parameterized Q-function $Q(t,b;\theta) \approx Q(t,b)$due to DQNemploying a DNN as a function approximator rather than storing an action-value table. In this case, θ stands for the DNN's parameters. It uses gradient descent minimization of the loss function instead of the Q-Learning iterative update approach for training in Equations (2) and (3):

$$K_s(\theta_s) = \left(z_s - Q(t_s, b_s; \theta_s)\right)^2 \quad (2)$$

Where $z_s$is

$$z_s = q_s + \gamma \max_b Q(t_{s+1}, b; \theta_s) \quad (3)$$

This loss function is derived from the idea that the framework should forecast the Q-value at timestep s as the sum of the greatest expected reward from the updated state as predicted by the model and the reward received from the action chosen in that state. Since the model parameters from the previous time stepθ_s are utilized to produce the new model parameters θ_{s+1}in the subsequent timestep, this maintains the iterative updating process employed in Q-learning.

It assumes that the agent's optimal behavior of action throughout training is to consistently choose the course of action with the highest value Q-value estimate. These Q-value estimates, however, are the result of the model's initialization rather than of acquired knowledge during the early training stages. Therefore, relyingsolely on the model's predicted actions might not yield beneficial training examples that could lead to improvements. The Q −learning algorithm uses a ϵ-greedy strategy to deal with this, where the model's predictive action is determined with frequency $1-\epsilon$ and a random action is selected with probability ϵ. At the beginning of training, the value of ϵ is set close to 1, and as training progresses, it decays to a small number.

4. Results

4.1 Simulation setup

The simulation setup utilizes a Python-based deep learning environment, employing libraries such as TensorFlow and Keras for model development and training. These tools provide the necessary flexibility and efficiency for implementing the Inverted Residual Adaptive Convo-Depth NeuroNet. Python's extensive ecosystem ensures smooth integration of emotion detection and reinforcement learning modules. The environment is optimized for real-time processing and performance evaluation of the adaptive video gaming therapy system.

4.2 Performance evaluation of the proposed method

The primary objective is to develop a real-time, emotion-aware adaptive gaming therapy system that dynamically adjusts game difficulty based on the user's emotional states. This system takes advantage of the advanced emotion-detecting techniques to create a personalized and engaging gaming experience,as well as learning techniques that optimize therapeutic results. The approach focuses on the continuous monitoring of users' facial expressions in real-time, allowing the game to dynamically respond to their emotional changes.

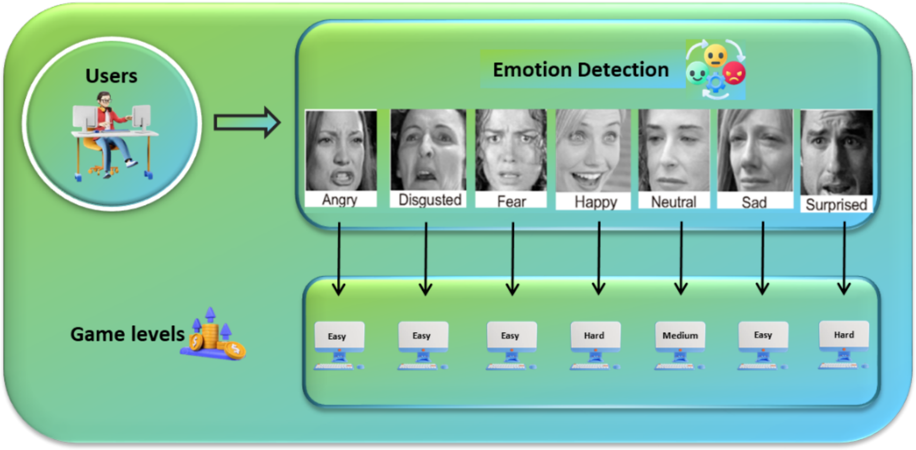

Figure 6: Emotion Detection-Based Game Level Model

Captures real-time facial expressions of the user. These high-frequency frames are analyzed to accurately detect and evaluate the user's emotional state in real time. These facial images are then analyzed using a deep learning-based emotion detection model, their emotional state, such as Happy, Fear, Sad, Angry, Disgust, Neutral, and Surprise. However, in this research, we have focused on analyzing only seven emotions: Happy, Fear, Sad, Angry, Disgust, Neutral, and Surprise, as shown in Figure 6. This real-time analysis is essential in ensuring that the gaming system remains responsive and adapts accordingly to the user's emotional state. To maintain engagement, enhance emotional well-being, and maximize the therapeutic benefits of gaming by providing an interactive and adaptive environment.

Once the emotion is detected, the reinforcement learning model determines the most suitable game difficulty level for the user. The system follows a structured approach to modify game difficulty based on the seven primary emotional states:

- **Happy:** If the user expresses happiness, the system raises a challenge by allocating a hard-level game. The reasoning behind this modification is that difficult sports can act as an outlet for frustration, and the user can expend their energy. By involving the user in a challenging activity, the system diverts their attention away from negative feelings and encourages focus, flexibility, and emotional control.
- **Anger:** When the user becomes angry, the system offers an easy level game. This ensures that the user continues to experience pleasure and relaxation, reinforces positive emotions, and maintains motivation to play. Easy-level game acts as a reward mechanism, which encourages continuous engagement without facing excessive challenges that can disrupt the positive emotional state.
- **Fear:** If the user is seen to be fearful, the system delivers an easy-level game. This provides an average amount of challenge that does not hold the user without overwhelming them. A balanced difficulty level allows confidence-building, as the user is motivated to advance without causing undue stress or anxiety. This strategic compensation enables the user to progressively dispel fear and gain a feeling of accomplishment through gaming.
- **Disgust:** The system offers an easy-level game. Disgust often leads to avoidance behavior; presenting a challenging activity can redirect attention, foster engagement, and motivate the user to channel their focus into overcoming the difficulty. This helps diminish the intensity of disgust by shifting the emotional response toward achievement.

- **Neutral:** The system delivers a medium-level game. A balanced difficulty ensures the user remains engaged without becoming bored or overly challenged. This strategy maintains a neutral emotional state while promoting focus and a sense of flow.
- **Sad:** The system provides an easy-level game. Sadness can sap energy and reduce motivation, so a low-difficulty task allows the user to experience small, incremental successes. This reinforces positive emotions and encourages continued engagement.
- **Surprise:** The system opts for a hard-level game. Surprise can be stimulating and heighten attention. By presenting a challenging task, the system leverages this heightened state to channel the user's focus and encourage problem-solving, helping them capitalize on their alertness.

The system's adaptive mechanism enables real-time monitoring and adjustment, ensuring a personalized gaming experience. It integrates real-time detection with reinforcement learning, enhancing user engagement and emotional welfare, resulting in an immersive, therapeutic gaming experience.

4.3 Metrics for evaluating the effectiveness of the proposed model

- **Accuracy** measures the system's overall effectiveness by calculating the percentage of correct emotional state predictions and game adjustments.
- **Precision** evaluates how accurately the system identifies positive emotional states without misclassifying neutral or negative ones.
- **Recall** assesses the system's ability to detect and respond to all relevant emotional signals during gameplay.
- **F1-score** balances precision and recall, reflecting the system's overall ability to make accurate and timely therapeutic adjustments.

The proposed system, Inverted Residual Adaptive Convo-Depth NeuroNet with DQN, provides excellent performance in adaptive video gaming therapy with an accuracy rate of 85.04%, meaning predictions of emotional states and video game adaptations are highly reliable. The precision of the model was 84% with a recall of 83%, indicating the model strongly detected some critical emotion cues and responded appropriately. Finally, the F1-score of 84% indicates that the system performed consistently well and was robust, allowing it to be employed in real-time personalized emotion-aware reinforcement learning therapy.

4.4 Comparative results analysis

This section discusses the suggested method's performance outcomes. The results are compared to the existing methods, EfficientNet-B0 (Kwon et al., 2022). The four most important performance metrics, F1-score, accuracy, recall, and precision, were used to evaluate the proposed strategy. Subsequently, the state-action-reward-state-action (SARSA) method will be employed to evaluate the accuracy of recommended game levels in comparison to the proposed DQN approach.

In balancing the accuracy and computing cost of EfficientNet-B0 (Kwon et al., 2022), it tends to fail real-time inference when managing a wide range of highly dynamic emotional states in a low-latency therapy gaming context. Itis complicated scaling, and a static input size will produce delays or incorrect classifications due to the processing of rapidly shifting emotional cues. However, the SARSA method does not have the depth or flexibility in understanding complex emotional patterns or contextual gameplay behavior, and this generalization of potential boundaries prevents personalization and adaptation, which is a key aspect of adaptive video game therapy systems.

According to Table 1 and Figure 7, the proposed Inverted Residual Adaptive Convo-Depth NeuroNet model outperforms the existing EfficientNet-B0 model. With an accuracy of 85.04% the suggested method outperforms EfficientNet-B0, which achieves 80% accuracy, and offers more precise emotion-based adaptations. Additionally, compared to the existing model's 57% and 58% precision and recall, it improves both to84% and 83%, respectively. The significant development is the F1-score, which rises from 54% to 84%, suggesting a robust and balanced characterization of affective states. These enhancements improvethe model's capacity to recognize subtle emotional states and offer tailored therapeutic gaming experiences.

Table 1: Numerical outcome of the video gaming therapy system

Parameters	EfficientNet-B0 (Kwon et al., 2022) (%)	Inverted Residual Adaptive Convo-Depth NeuroNet (Proposed) (%)
Accuracy	80	85.04
Precision	57	84
Recall	58	83
F1-Score	54	84

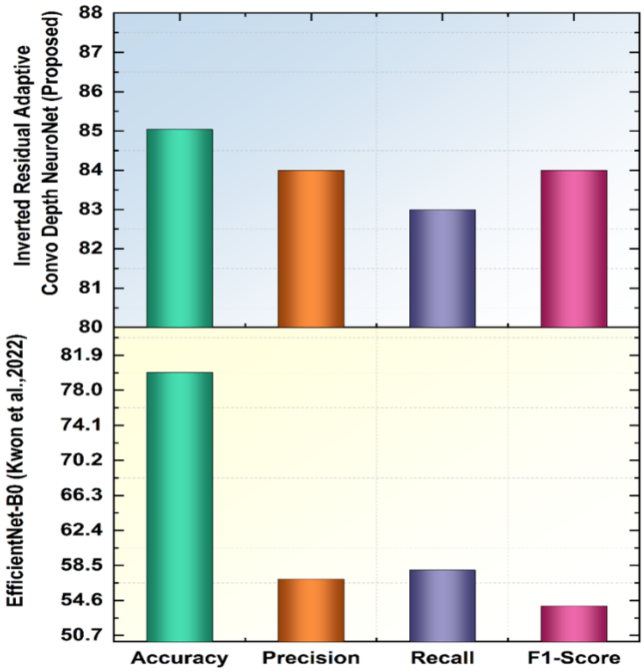

Figure 7: The overall outcome of the video gaming therapy system

The proposed DQN outperforms the SARSA model at each difficulty level, as shown in Table 2 and Figure 8. For the Easy level, the proposed model achieves 95% accuracy when compared to the current method. For Medium and Hard levels, the proposed model maintains 95% accuracy, while the SARSA has 85% accuracy. The overall accuracy of the proposed model improves, highlighting its effectiveness in capturing diverse emotional nuances and varying levels of play complexity. Assuming its consistent performance, the proposed model is confirmed to be flexible and effective for delivering personalized and emotion-appropriate video game therapy.

Table 2: The numerical outcome of each level and combined accuracy.

Element Accuracy	SARSA(%)	DQN(Proposed)(%)
Easy	91	95
Medium	89	95
Hard	92	95
Combined Accuracy	85	95

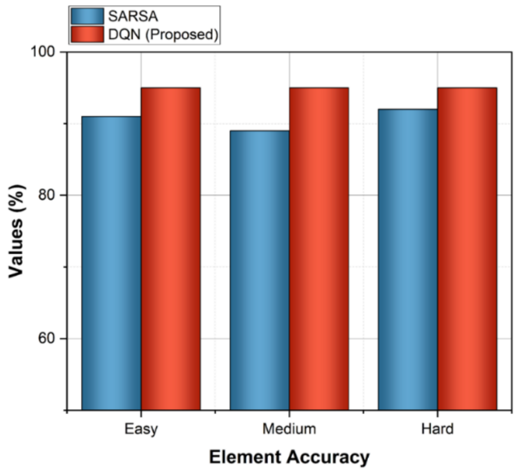

Figure 8: The outcome of each level and combined accuracy.

To address these limitations, the Inverted Residual Adaptive Convo-Depth NeuroNet presents a flexible, tailored approach to an adaptive video gaming therapy system for mental health. Each inverted residual block is a lightweight neural network block, resulting in efficiency and providing a capacity for real-time computing while tracking emotional fluctuations with precision during gameplay. Overall, DQN promotes intelligent, online, continuous, dynamic changes in game difficulty (and reward systems) based on emotional feedback and performance by the user, simultaneously integrating the two systems to enhance user engagement and therapeutic value, and builds a highly personalized gaming experience for each user's emotions and performance measures.

5. Conclusion

Reinforcement learning in video gaming therapy, using emotion detection, can adapt the game environment in response to an athlete's emotional reactions in real time. The system utilizes a unique blend of emotion detection and reinforcement learning via an Inverted Residual Adaptive Convo Depth NeuroNet to create an adaptive video gaming therapy. The emotion detection module uses a pre-trained MobileNetV2 model to classify emotions as Happy, Fear, Sad, Angry, Disgust, Neutral, and Surprise. The DQN reinforcement learning experience utilizes algorithms to modify the complexity of the game, adjust intrinsic rewards based on the emotional state of the athlete, and assign rewards based on performance. The system creates a truly immersive and responsive video gaming therapy experience that adapts dynamically to emotional and performance-driven responses in real time. The platform is built using Python-based tools and libraries.The system exhibited performance metrics ranging from 83% to 85% of F1-score, accuracy, recall, and precision for detecting emotions, and game-level suggestion (DQN) accuracy had 95% accuracy,confirming the effectiveness of the emotion-aware adaptive therapy system.Facial expression recognition under diverse lighting and diffraction conditions presents a challenge for the Inverted Residual Adaptive Convo Depth NeuroNet. The Deep Q-Network (DQN), however, successfully suggested suitable emotions to guide game-level recommendations.

Future scope

Future research will expand on multimodal emotion sensing, incorporate transfer learning for scalable deployment, and establish robust privacy safeguards, paving the way for truly personalized, engaging, and effective digital therapy experiences. This approach holds significant potential in developing therapeutic tools for mental health conditions such as anxiety, depression, and Post-Traumatic Stress Disorder (PTSD). It can be integrated into telehealth platforms, virtual reality (VR) therapy, and rehabilitation programs. Further advancements may include real-world adaptive learning systems and enhanced multi-modal emotion detection, broadening its impact across healthcare, education, and assistive technologies.

References

Baker, S. and Xiang, W., 2023. Artificial intelligence of things for smarter healthcare: a survey of advancements, challenges, and opportunities. *IEEE Communications Surveys & Tutorials*, *25*(2), pp.1261-1293. https://doi.org/10.1109/COMST.2023.3256323

Vindigni, G., 2023. Exploring digital therapeutics: game-based and e-health interventions in mental health care: potential, challenges, and policy implications. *British Journal of Healthcare and Medical Research-Vol. 10*(3). https://doi.org/10.14738/bjhmr.103.14804

Mehra, I., Nasir, S. and Prakash, A., 2023. Computer and video games in mental health. *Science Insights*, *42*(4), pp.877-883. https://doi.org/10.15354/si.23.re261

Parvaria, P.R. and Schofieldb, D., 2025. Dementia Therapy: The Role of Gamified AI and Digital Art in Supporting Cognitive and Emotional Well-being. *Procedia Computer Science*, *256*, pp.1391-1398.https://doi.org/10.1016/j.procs.2025.02.253

Bocci, F., Ferrari, A. and Sarini, M., 2023, June. Putting the gaming experience at the center of the therapy—the Video Game Therapy® approach. In *Healthcare* (Vol. 11, No. 12, p. 1767). MDPI.https://doi.org/10.3390/healthcare11121767

Pal, S., Mukhopadhyay, S., and Suryadevara, N., 2021. Development and progress in sensors and technologies for human emotion recognition. *Sensors*, *21*(16), p.5554. https://doi.org/10.3390/s21165554

Yadav, M., Pandey, A., Gaur, N. and Nghiem, X.H., 2024. Mental health and well-being: Metaverse-based therapeutic interventions for anxiety, depression, post-traumatic stress disorder. In *Examining the Metaverse in Healthcare: Opportunities, Challenges, and Future Directions* (pp. 353-380). IGI Global. https://doi.org/10.4018/979-8-3693-1515-6.ch014

Spytska, L., 2024. The use of virtual reality in the treatment of mental disorders such as phobias and post-traumatic stress disorder. *SSM-Mental Health*, *6*, p.100351. https://doi.org/10.1016/j.ssmmh.2024.100351

Lin, W., Li, C., and Zhang, Y., 2023. A system of emotion recognition and judgment, and its application in an adaptive interactive game. *Sensors*, *23*(6), p.3250. https://doi.org/10.3390/s23063250
Song, M., Jing, X., Parada-Cabaleiro, E., Yang, Z., Yamamoto, Y. and Schuller, B.W., 2024, August. Temporal Oriented ResNet for Gaming Dimensional Emotion Prediction. In *2024 32nd European Signal Processing Conference (EUSIPCO)* (pp. 596-600). IEEE. https://doi.org/10.23919/EUSIPCO63174.2024.10715180
Patel, V. and Mehta, K., 2024, November. Emotion-Aware Music Recommendations: Evaluating Custom CNN vs. VGG16 and InceptionV3. In *2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS)* (pp. 499-504). IEEE.https://doi.org/10.1109/FMLDS63805.2024.00092
Awwad, A.M.A., 2021. Visual Emotion-Aware Cloud Localization User Experience Framework Based on Mobile Location Services. *International Journal of Interactive Mobile Technologies*, *15*(14). https://doi.org/10.3991/ijim.v15i14.20061
Ok, E., Noah, N., Mia, L., Lena, A., and Vicente, C., 2024. Enhancing E-Reading Experiences: Leveraging Machine Learning for Emotional Interfaces in Digital Reading Platforms.http://dx.doi.org/10.35940/ijrte.D8391.118419
Duan, S., Wang, Z., Wang, S., Chen, M. and Zhang, R., 2024, November. Emotion-aware interaction design in intelligent user interface using multi-modal deep learning. In *2024, 5th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC)* (pp. 110-114). IEEE. https://doi.org/10.1109/ISCEIC63613.2024.10810240
Stoynov, V., 2023, May. A Novel Emotion-Aware Networking Model for Enhanced User Experience in 5G Networks. In *2023, 33rd Conference of Open Innovations Association (FRUCT)* (pp. 296-308). IEEE.https://doi.org/10.23919/FRUCT58615.2023.10143069
Kwon, S., Ahn, J., Choi, H., Jeon, J., Kim, D., Kim, H., and Kang, S., 2022. Analytical framework for facial expression in game experience test. *IEEE Access*, *10*, pp.104486-104497.https://doi.org/10.1109/access.2022.3210712
Lekha, A., Srikrishna, C.V. &Vinod, V., 2015. *Fuzzy association rule mining*. Journal of Computer Science. https://doi.org/10.3844/jcssp.2015.71.74
RajKumar, N. and Vinod, V., 2015. Integrated educational information systems for disabled schools via a service bus using SOA. https://doi.org/10.17485/ijst/2015/v8i13/55030

APPENDIX

Reinforcement Learning = RL	Emotion adaptive interactive game = EAIG
Deep Q − Network = DQN	system of emotion recognition and judgment = SERJ
Artificial Intelligence = AI	Multimodal Frustration Game Database = MFGD
post − tromatic stress disorder = PTSD	Convolutional Neural Network = CNN
cognitive − behavior therapy = CBT	machine learning = ML
autism spectrum disorder = ASD	Emotion − aware Networking Paradigm = ENP

Emotion − aware Networking Model = ENM	Key Performance Indicators = KPIs
Fully convolutional layer = FCL	feed − forward artificial neural network = FF − ANN
Rectified Linear Unit ReLU	Depthwise Separable Convolutions = DSC
deep neural network = DNN	deep reinforcement learning = DRL
convolutional layers = CL	