National College of Ireland

A Comparative Study of CNN, RNN-LSTM, and Transfer Learning Models for Facial Emotion Recognition in context of gaming

MSc Research Project
MSc in Artificial Intelligence

Ayush Gole
Student ID: x23224100

School of Computing
National College of Ireland

Supervisor: Arundev Vamadevan

National College of Ireland
Project Submission Sheet
School of Computing

National College of Ireland

Student Name:	Ayush Gole
Student ID:	x23224100
Programme:	MSc Artificial Intelligence
Year:	2024-25
Module:	MSc Research Project
Supervisor:	Arundev Vamadevan
Submission Due Date:	12/12/2024
Project Title:	A Comparative Study of CNN, RNN-LSTM, and Transfer Learning Models for Facial Emotion Recognition in context of gaming
Word Count:	6401
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	A.V.Gole
Date:	26th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

A comparative study of CNN, RNN-LSTM, Transfer Learning model for facial emotion recognition in context of gaming

Ayush Gole
x23224100

Abstract

The boomed growth of Esports industry, highlights intense competition within industry and gameplay. This underscores the need of adaptive gaming and deeper understanding of player emotions with gaming context. By analysing gamers emotions with game context, we aim to develop base for framework for tailoring gaming experiences according to gamers. Leveraging different neural network models and methods like CNN , transfer learning VGG16 and RNN-LSTM to compare and find more suitable one for gaming emotion detection is prime goal of this research. This research shows initial preprocessing standards needed, data collection and standardization along side method, models suitable for implementation. This research will contribute to advancement to adaptive gaming by providing insights of research models and implementation.

Keywords: Adaptive gaming, CNN, Emotion analysis, RNN-LSTM, Transfer learning

1 Introduction

There is big booming growth in global Esports industry, mainly because of easy availability of internet service and passionate young generation. (Shrivastav; 2024) This increasing popularity of gaming players and gaming streamers is leading to high interest in game creation, configuration and understanding of how it works. To facilitate more research and for business purpose leading experts proposed to categorize into groups. For instance, Laird and van Lents proposed taxonomies of classification in groups like adventure, horror, and action serve as concrete bases for exploration of diverse landscapes of gaming experience. (Laird and van Lent; 2001)

While gaming industry is under continuous advancement and research, creating games that adapt to number of gamers and platforms remains quite a hard challenge. This is part of adaptive gaming development process. Adaptive parameters such as player's emotions, player's skills, and device capability tailoring them for better performance is primary goal of adaptive gaming. This concept grew exponentially after the usage of serious games for military gaming use cases, where gaming environment showed random adjustments to create numerous unknown situations. (Xu et al.; 2018) While traditionally gaming analysis focused mainly on high-level strategic insights by human roles such as game analyst, team coaches, slowly it has started to integrate with more advanced

digitized methods. A deeper understanding of player emotions and health is necessary for optimizing gameplay and gaming experiences.

This research study delves into enhancing player's experience by leveraging gamer's facial emotions in context of game scenes such as boss fights, 1vs 4 situations, regular driving in gaming world. The study seeks to develop foundation for adaptive gaming which will tailor gameplay experience with gamer emotions by comparing various emotion detection methods. This research builds upon existing studies of facial emotion detection experiments and knowledge of gameplay-associated emotions. These research findings will help bridge gap between traditional method and the emerging possibility of real-time gamer's emotion detection with game context for adaptive gaming.

The subsequent sections deep dive into study in section 2 Related foundation work, section 3 research methodology which is proposed, section 4 design architecture of research methodology and section 5 implementation of research. Following this section 6 explains evaluations of experimented methodology and section 7 concludes with conclusion of this research study and future work.

2 Related Work

Previously explored methods for emotion detection and related gaming papers are the main topics of discussion in this section. Prior research has explored gaming-related emotion analysis on various other parameters and methods such as ECG, EMG, and facial emotion detection is explored with various machine learning models as well. These corresponding methodologies address problems separately through only facial emotion detection or gaming emotion analysis through physical signal parameters respectively. To enhance gaming emotion analysis in real time with simpler implementation is primary aim of this research by comparison of models.

2.1 Prior Research on Game Emotion Analysis

Affective gaming is detecting real-time emotions of players during various gaming stages and enhancing gamers engagement. (Kalansooriya et al.; 2020) Affective gaming is similar to adaptive gaming concept. This paper studies utilizing real-time physiological signals such as ECG, EEG to monitor emotional state of gamers. By analyzing gamer's emotional state and level, the adjustments in game can be done by adjusting parameters such as Background music, difficulty level. This paper is researched Car racing games and relative emotional state, by utilizing EEG signals of player. The paper includes EEG signal modeling and developing machine learning models like BR, CC, Knn, RakEL that have been employed to categorize music excerpts based on gamer's emotional impact, enabling game to tailor soundtrack in terms of frequency to player's current mood. RakEL produced highest accuracy of 58% out of all this. While study serves significant foundation for adaptive gaming progression but lacks in wider range of parameters.For example number of emotions considered in this study, number and categories of games considered for study, machine learning models, emotion detection techniques.

The second study case delves into integration of affective computing with gaming context and challenges with same.(Yang et al.; 2018) The study deep dives into physiological signals like EMG, ECG, EDA, respiration and body moment with help of accelerometer to find relationship with gaming emotional state. Study highlights the difficulty of emotion control in various gaming scenarios, which showcase need of more refined stimulation

techniques. The study utilized various physiological signal to capture emotional state in gaming events by employing machine learning techniques to model emotion recognition. The study also discusses peripheral physiological signal struggle to detect and imprint small changes in short period, particularly capturing the subtle changes in emotions. This shows it might miss more complex feelings. It is important to address cognitive biases in subjective evaluation, as it will impact accuracy of emotion detection model. Furthermore, it shows importance of feature selection, and segmentation on model performance which drives optimal results. At the end author concluded with importance of personalized models that will account for individual's physiological responses and emotional responses. Study furthermore talks about need for more physiological signals and different gaming emotion categories consideration and underlying issues with employed methodology.

Previous studies use EEG for detection of single emotion and gaming categories, whereas it's important to study it for wider range of games and emotions. By analyzing EEG signals from input of players researchers aimed to classify emotions into 4 categories: Bore, Stress, Happy, Relax. (Khan and Rasool; 2022) The study employed various ML models like SVM, GBM, Random Forest to extract features from EEG data for emotion classification. GBM classifier worked best in this study with 82% overall accuracy. Hybrid approach of combining wavelet domain feature, frequency, time period for analysis yields better results. However, this hybrid combination analysis highlights limitations like Data acquisition from EEG signal, individuality of gamers data, EEG based emotion recognition and method implementation. The author describes need of future research exploring more advanced ML models like deep learning, to improve accuracy. Also EEG can be combined with other modalities such as facial expression or more physiological signals for robust emotion recognition. This study gives concrete base for Adaptive gaming research and direction for research.

Other than physiological signals like EEG,ECG next study investigates the feasibility of using pen-based input data and in-game performances of player's emotional states. (Frommel et al.; 2018) The researchers employed machine learning models SVM, Random Forest, dummy classifier, RF(LNO) for classification of emotion of players in specific gaming intensity levels of valence, arousal, dominance. Random forest classifier showed best F1 score of 0.577 here. The study done on serious games showed promising results but also highlighted limitations in research method. This includes biased subjective evaluation, the limitations of physiological signals, the limitations of controlling emotional responses in various gaming scenarios, and similar to previous studies, impact of feature selection and segmentation for optimized performance.

The study suggests machine learning models like random forest, SVM can be employed for custom-made serious game to capture and predict gamer's emotions. The results show further scope of refining feature extraction for better results and accuracy. This study underlines use of newer method of using pen-based input as physiological signal for analysis of player's emotions but it lacks in many areas. Such as it only includes one category- serious game and also has limited range of emotional levels considered. Also, method only considered different models but not different types of games. This method shows promising ease but needs further development as all games cannot be incorporated with pen-based input.

Multimodal Game Frustration Database can be employed for research in adaptive gaming and human-computer interactions.(Song et al.; 2019) All previous research shows particular game categories and selected machine learning model deployment in order

with physiological signals, whereas this database research captures and uses audio, video data of individuals experiencing frustration. The study enables the development and evaluation of Automatic frustration emotion recognition system with MGFD.

The author proposes more prominent method of combining audio and visual features in order to utilize method like SVM, LSTM-based models, and feature selection. The study shows best results with LSTM model (60%) and Audio-Video data, but also highlights challenges of capturing and recognizing subtle range of emotions associated with MGFD, particularly with fast changing gaming environment. This study serves as promising base for development of adaptive gaming systems. It also underlines future research direction including wider range of emotions, wider range of game datasets. This study's results show need of explore of more sophisticated feature extraction methods, refine models and methods. The research only delved into frustration emotion. Future directions of expanding database can include wider range of emotions and gaming contexts. These researches show different methods for particular emotion detection in gaming context and need of development of system with features that should be considered for same.

2.2 Prior Research on Facial Emotion Analysis

There are numerous methods of facial emotion recognition. The study by author investigates facial emotion recognition using CNN -ResNet50v2 architecture on large dataset. (Bayunanda et al.; 2022) The study is done on 35887 images dataset mainly for impact of epochs on accuracy of overall model. Initial trained model on 25 epochs showed poor result and high loss , as numbers of epochs gradually increased accuracy improved and loss reduced significantly. The study suggested higher number of iterations resulted in better classification of new images and identified anger as dominant emotion due to dataset. However initial high loss value indicated potential issue in dataset or model that requires further investigation. This study gives us concrete base in this research for method selection as it shows prominent results with FER2013 dataset.

There are multiple pre-trained models so its important to choose one for emotion recognition. The study by authors deep dives into various ResNet pre-trained models for facial emotion recognition. ResNet101V2 models showed best result among all tested pre-trained models by accuracy of 93%. (Gondkar et al.; 2021) Other models like ResNet152V2, ResNet50V2, and mobilenet also shows prominent results which is around same accuracy of ResNet101v2 model. Mobilenet which is comparatively smaller model than others also perform well for real time applications, as showed by its Android deployment application example. This study showcases potential of using transfer learning for facial emotion recognition model tasks, whereas it also shows challenges with same. Based on training dataset and application it is important to choose potential model, in this study author chose Mobilenet even though its accuracy is less compared to others. The study doesn't show prominent and distinctive conclusion about model selection but necessities in parameter tuning.

The similar model of deep learning like RNN can be employed with more advanced techniques for facial recognition too. On this idea research paper investigated how recurrent neural network (RNN) model can be used, as moreover Long short-term memory network (LSTM) for facial emotion recognition system. (Mostafa et al.; 2018) The model shows promising results for emotions like anger, disgust by employing LSTM model trained over sequence of facial frames. The model effectively captured dynamic nature of facial emotion from dataset leading improved accuracy between distinguishing differ-

ent facial emotions. This method application shows prominent path for facial emotion recognition of neural networks that overperformed traditional models.(82%) The study successfully concluded with suggestion of various use case applications, while the study findings is limited to sequence of frames and perform well, particularly for anger, disgust only. Also, it requires video data and suggests applications under specific frames which limits its usage. For this research, it shows prominent methods that can be applied.

There are multiple models that can be employed for facial emotion recognition and comparing them is important. This research deep dives into the application of various machine learning model like Support Vector machine (SVM), Random Forest, CNN, RNN, and Long Short-Term memory network for facial emotion detection. (Negi et al.; 2024) The study showed RNNs employed together have superior performance than other models for facial emotion detection. The integration of deep learning models like CNN, LSTM in this experiment showed promising results but also showed critical challenges for facial emotion detection applications. The implementation result shows challenges like bias in dataset, Real-time performance, cross-cultural recognition, and subtle emotions recognition. Also it underlines biggest challenge of using deep learning model computation expense with real-time integration. This research gives us broader base for this research for model selection and base learnings.

2.3 Summary

Above extensive study deep dive into emotion recognition within gaming context and facial emotions. Prior research methods for gaming emotion detection focused on physiological signals and particular gaming categories. These studies isolate to specific game genre , modality or range of emotion. While this study shows physiological signals offering granular insights, they require particular equipment and are efficient for subtle emotion nuances detection only. Studies further suggest optimizing detection methods by employing various methods, which leads to base of this research of using facial emotion detection method.

The literature review on facial emotion recognition methods focuses on deep learning methods. They highlight effectiveness of CNN particularly RESNET architecture for classification. The study highlighted use of transfer learning models for model improvement. Additionally, LSTM is explored for pattern recognition. The study serves base of research by highlighting various pros, and cons of these methods like dataset bias, real-time performance, hyper parameter tunings needed for various applications. The study highlights potential use case of deep learning model on game scenes. Both these case studies help to explore application of deep learning models on facial emotion recognition in context of gaming.

3 Research Methodology

The following sections describe the corresponding structure and the methodology used for comparing various ML models for facial emotion detection with gaming context. The accompanying section's specific major objective is to understand detailed description of dataset and go into details about various types of methods that has been utilized throughout study. The research process employed in this research study involves several distinct stages. The research study starts with finding idea and filed of research of interest. It further goes into initial literature review to establish foundation for relative

research domain and deeper topic. This initial review provides comprehensive methods and results of existing research, identifies future directions, gaps, and direction for future methodologies and questions that can be addressed or studied. Following initial literature study process outlined in figures and boundaries for study, the next step formulating the research question. This was followed by in depth and more relevant technical literature review. This in-depth and technical literature helps for preliminary findings which can be addressed while further design and implementations. Building upon these concrete foundational stages, research progresses with designing methodologies and acquisition of all relevant datasets and libraries. The subsequent phase is conducting experiments means implementation and execution of the designed experiments. This is phase is directly influenced by previous stages of literature review, research question formulation, and methodology designing.

The final stage of process is detailed analysis of results and the subsequent writing of thesis. This stage entails thorough analysis of data collection, finding throughout process and experiments which then serve as the foundation for thesis composition.

3.1 Proposed methodology

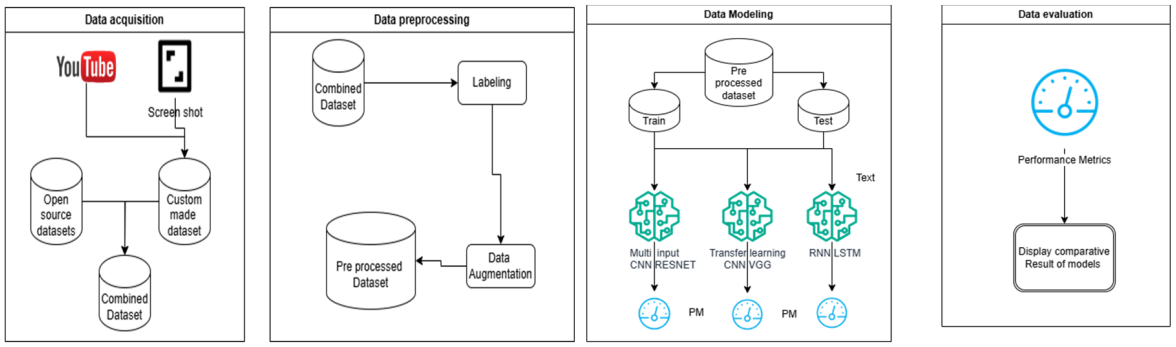


Figure 1: Proposed Methodology

The Figure 1 shows proposed methodology for research study. Methodology mainly works in four stages as shown in the figure. The initial phase is very important for research which is data collection and acquisition. Obtaining two publicly available Face Image Dataset : CK+ and FER2013 is critical step, particularly selecting this dataset is result of previous literature reviews. Alongside this custom dataset is created, which contains Game scene images and Gamer's face images. Custom made dataset is created by author of this research. Particular frames from YouTube streams of famous gaming streamers is selected, and then categorized into different emotions associated with them. This process has captured frames that showcase particular emotions. This colored image dataset is stored similar to CK+ and FER2013 folder structure. Once this custom-made dataset is created, custom-made face image dataset is combined with the other two datasets for future application for combinational study purposes.

The second phase includes Data preprocessing of collected data. There are now two combined data sets one is a custom-made face image dataset + CKplus dataset and other one is custom-made face image + FER2013. The custom-made face image dataset has RGB and irregular sizes from range 400x300 to 500x380. Whereas the publicly available datasets CK+ and FER2013 have standardized size of 48x48 and is Gray scaled for better

modeling. So, custom made face image dataset is pre-processed and augmented in this phase, so it is ready for next phase as Pre-processed dataset.

Once Second phase of data preprocessing is completed, third phase of data modeling can be initiated. Pre-processed dataset from previous stage is applied to three selected methods. From previous literature reviews future scope, this study explores three methods of Multi-input CNN model, Transfer Learning CNN VGG model, and RNN LSTM model. Both pre-processed datasets are applied to all three models for analysis. Once models are ready performance metrics will be applied to understand and compare performances of models.

At the final stage, all performance of all models is collected together. Then these performances are compared with each other. Model performance is evaluated with Accuracy metrics and computational efficiency. Along with these trained model's metrics, more constraints are considered which are tuned while training and modeling. These all findings and result comparisons then can be added as research final findings.

3.2 Dataset Description

As mentioned in the above section, the "CK+ dataset" and "FER2013 dataset" datasets available for the public from Kaggle open repository are used as open-source data sources for one of primary objectives of face emotion detection. A custom-made dataset is created with the help of open-source tool You tube platform for collection. This dataset contains both face image and game scene image datasets, which addresses both objective of emotion detection for gamer's face and game scene

3.2.1 Cohn-Kanade Dataset

The Cohn-Kanade 48 dataset is widely used for face emotion detection studies in previous papers and was created mainly for same purpose.(Lucey et al.; 2010) The proposed paper showcase use-case of CK dataset and all the insights of dataset for emotion detection applications. There are different versions available like extended too but 48 is easily publicly available and subset of CK+. The dataset contains 981 images with 7 emotions annotated: anger fear, disgust, happened, sad, neutral. Previous studies showed great results with usage of this dataset.

3.2.2 FER2013 dataset

After reviewing multiple paper and dataset availabilities FER2013 dataset is most suitable for this research.(Zahara et al.; 2020) As proposed in paper FER2013 is more ideal for micro expressions emotion detection applications. FER2013 dataset provides foundational resources for establishing baseline for facial emotion recognition. This data set have images standardized to size 150x150 and grayscale images. Total 35,887 images in dataset each annotated with seven emotions: anger, fear, disgust, happened, sad, neutral. This dataset is widely used for initial modeling and training.

3.2.3 Custom dataset

The custom dataset is created particular for this research. There are number of images captured from open-source platform YouTube. As the data collection is not done by any standardized tool image sizes vary from 400x300 to 500x400. The dataset contains total

game scene images 193 for all emotions: anger, fear, disgust, happened, sad, neutral. It also contains 52 images of gamer's face.

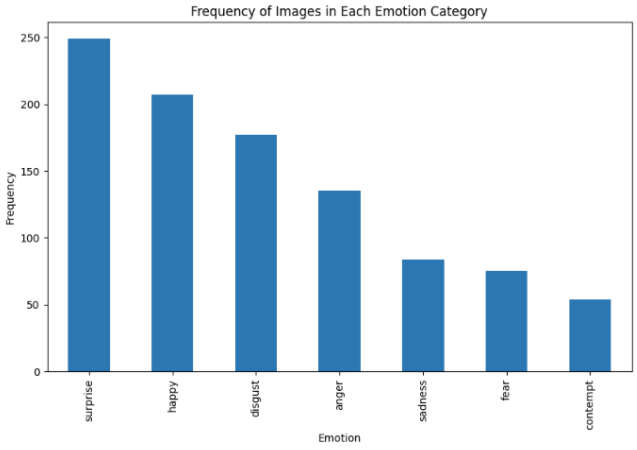
```
Games:
                        Horror: Kamala
                        multiplayer, adventure ,action : chained, PUBG, valorant
                        ROLEPLAY: GTA
                        comdey, intelligence: scribble, among us
Streamers:
                       Male: Mortal, shreeman, Dynamo,carry minati, snax,Binks, jonathan
                       Female: sharkshee, kash plays,payal gaming, dobby is live, ankita
Situations:
                       tricky,surprise,challenging: 1VS4 in PUBG,valorant,Police chase in GTA
                       surprise,fear : Kamala bhoot entry
                       laugh, comedy: scribble, among us
                       lie, tricking: among us meeting room defend
                       co-ordination, frustration: chained falling someone, spelling issue in scribble
                       anger: granade by teammate, or no support less calls in PUBG, teammate only looting not reviving
                       anxiety, surprise: kamala bhoot coming sound
                       happy ,joy ,excitement: killing squad in PUBG, winning round in Valorant
                       boredom, tired: long stream for pubg, roleplay in GTA
```

Figure 2: Custom dataset collection information

Figure 2 shows how which games, streamers and situations are considered while creating dataset. Data collected is based on various streamers and games information as follows:

3.3 Dataset EDA, Preprocessing and Validation

While working with open-source datasets, thorough exploratory data analysis is crucial step for researching and modelling. By understanding data insights, informed decision for preprocessing and modelling can be made.




```
Null values in the dataset:
image      0
emotion    0
dtype: int64
```

Figure 3: CK dataset exploration

CK dataset and FER2013 contain images for seven types of emotions and is important to explore how many each emotion has datapoints. Figure 3 shows the distribution of data points according to each emotion for CK dataset and Figure 4 shows for FER2013 dataset. Missing data can reduce overall effectiveness of ML models, which results in biased model prediction and unreliable insights. To ensure dataset is well-integrated dataset has been verified if there are any data points missing or contain any null values. Analysis confirmed there is no absence of any such issue.

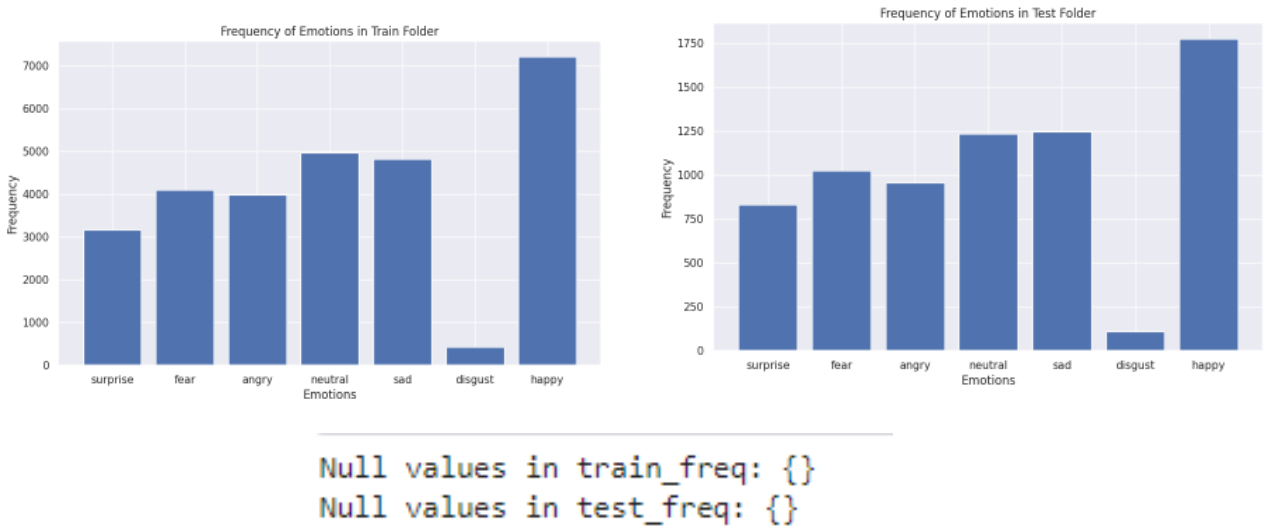


Figure 4: FER2013 dataset exploration

Understanding properties of data points is also crucial for the Preprocessing step. For model selection and reproducibility, consistency of research understanding properties of data points like Dimensions, dtype, min and max value is important. Figure 5 shows properties of data point in happy folder of CK dataset and Figure 6 shows datapoint of happy folder of FER2013 dataset

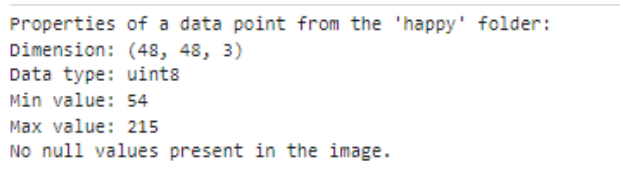


Figure 5: CK properties

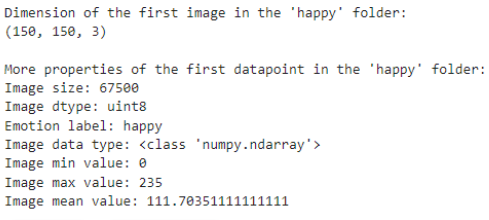


Figure 6: FER2013 properties

Preprocessing and Data augmentation: Based on previous EDA and to make dataset ready for modeling various preprocessing data augmentation techniques have been employed as follows:

Rescaling: Pixel values of images are rescaled normalized to range from 0 to 1to ensure consistency and prevent numerical instability.

Grayscale conversion: Images that were in RGB format are converted to Grayscale for better training purposes.

Horizontal Flipping: Images are randomly flipped horizontally to account for variation in dataset.

Zooming and Rotation: Datapoint images are randomly zoomed by range 80 to 120 percent and rotated with range of -10 to 10 degrees to simulate data augmentation techniques for different viewing perspectives. Channel shift: The color channels in data points are randomly shifted by 10 percent to introduce variation in datapoints brightness.

Data validation: To ensure the accuracy of dataset which can replicated further and be more robust of facial emotion detection model, datapoints must be rigorously tested so models can be easily implemented into real-world applications. In previous step

after completion of data augmentation in which images are flipped to increase diversity in dataset resulted in increased model generalization and training. For validation purposes mapping facial points and checking with further steps can help rigorous testing. Facial landmarks like nose, eyelids, and mouth the main 5 key points of face are plotted on some random face images using dlib library.

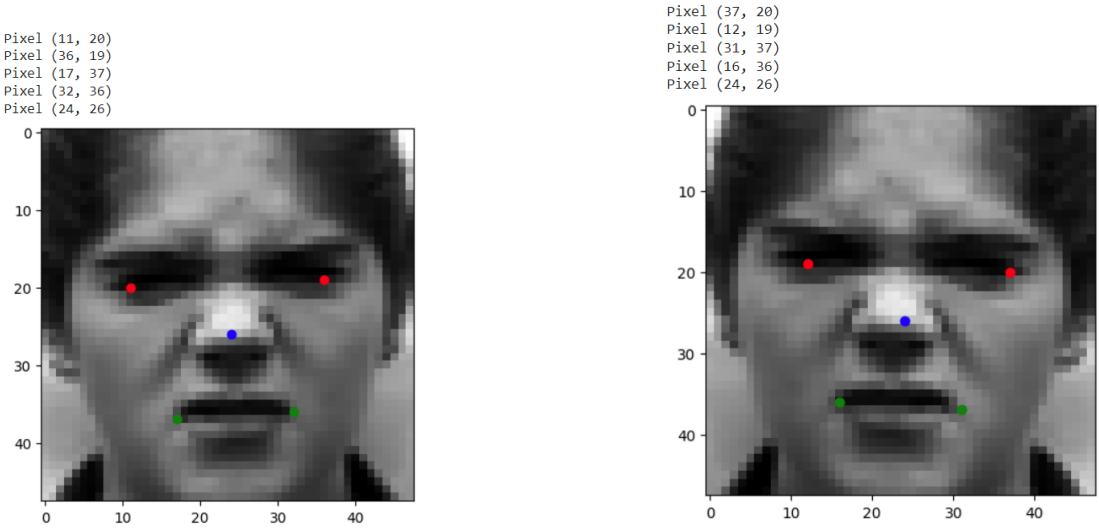


Figure 7: Data validation: face mapping original image and flipped image

In figure 7 first image shows facial landmarks mapped on original photo and second shows landmarks mapped on flipped image. To test datapoints same datapoint is flipped horizontally and again pretrained shape predictor was employed to plot and map facial landmarks. The coordinates of facial landmarks of original image and flipped image were appropriately adjusted. This data verification process showed reliability of dataset and facial features extracted from them, which will contribute to overall accuracy of further model deployments.

3.4 Methodology

For first machine learning model in this research, multi-input CNN ML model is trained. In real-world applications, data comes from multiple different sources or modalities. For instance, in this research, data is acquired from multiple sources two are publicly available datasets and one is custom-made from YouTube. Multi-input CNN is a powerful ML Model for handling such cases. (Sánchez-Cauce et al.; 2021) In this research for first method multi-input CNN model is used multiple times. Multi-input CNN model is first employed for face image dataset combinational learning. Then it is employed for multi-input models like face image model and game scene CNN model for creating combined model. Key advantage of using this method is it enhances feature extraction from multiple data sources/ datasets which is case in this experiment. Along with multi-input CNN, pre-trained model ResNet50 is used for face models for better results.

Second machine learning model employed in this research is Transfer learning with CNN-VGG16 pretrained model. Transfer learning is method of using pre-trained models knowledge gains from large datasets and applying it to new tasks. In this study CNN-VGG16 ImageNet pretrained model has been employed on research datasets. (Habibullah et al.; 2023) The pre-trained model is applied to both combined face dataset and

game scene dataset for emotion detection. Then these trained models are combined further for final combinational model creation. The transfer learning model is created by freezing base layers and excluding top layers, which helps fine-tune model and train more concerning the new assigned task dataset.

Last machine learning model employed for Gaming facial emotion detection research is RNN LSTM. RNN- Long short-term memory model is beneficial for sequential image datasets or video datasets. (Khademi et al.; 2022) For real-time application of this research, it is very important to experiment task of facial and gaming emotion detection with the datasets. The dataset used in this experiment is not sequential or video dataset, so transfer learnings is used at initial stage of experimenting third method. The initial models are trained with CNN-ResNet50 transfer learning for both face, game scene datasets. Then sequential model is constructed over both model learnings, comprising of two LSTM layers having a dropout layer for regularization. This model of transfer learning along with RNN-LSTM gives base insights for future real-time applications.

4 Design Specification

The following sections describes corresponding system architecture of each method employed in research process.

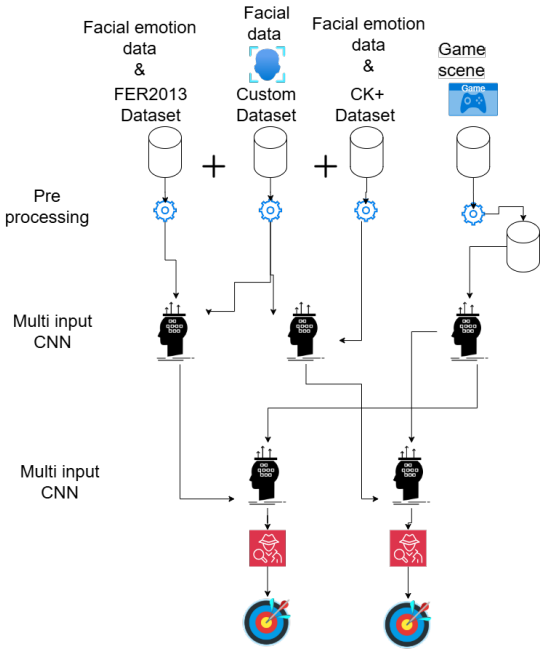


Figure 8: Method 1 Multi-input CNN

The Figure 8 shows design architecture shows multilevel architecture of multi-input CNN model. Design architecture begins with data collection and preprocessing at initial level. Followed by on level two multiple CNN models trained over datasets, separately for game scene, and multiple face image datasets. In the final level, combined models are constructed from previous level models, and evaluated created models.

The designed architecture is shown in Figure 9 leverages multi-level transfer learning CNN-VGG16 architecture model. The initial stage involves data collection, and dataset gathering followed by data preprocessing. Subsequently, in the next stage, pre-trained

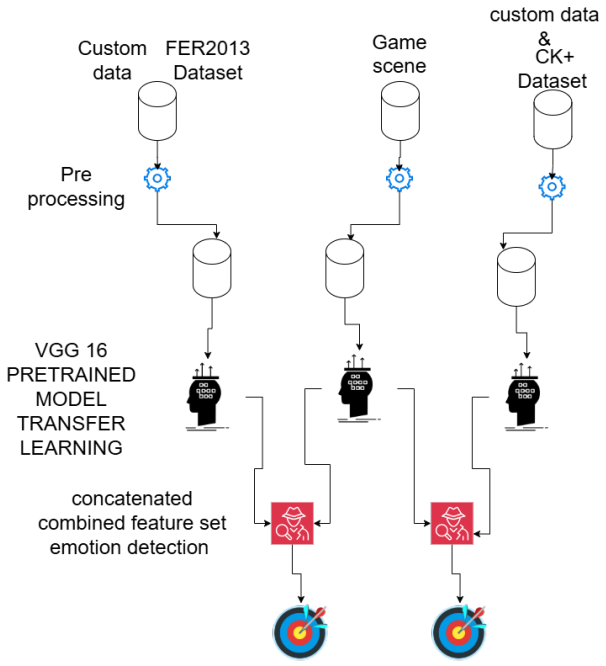


Figure 9: Method 2 Transfer learning CNN VGG16

model CNN- VGG16 which is trained over vast datasets applied on research datasets, and specialized models are created tailored for specific tasks. In the final stage, these models are combined and integrated together to develop final models. Once combined models are ready, overall performance of the models is evaluated. This multilevel transfer learning design approach leverages pretrain models' knowledge gains and adapts the unique characteristics of research datasets.

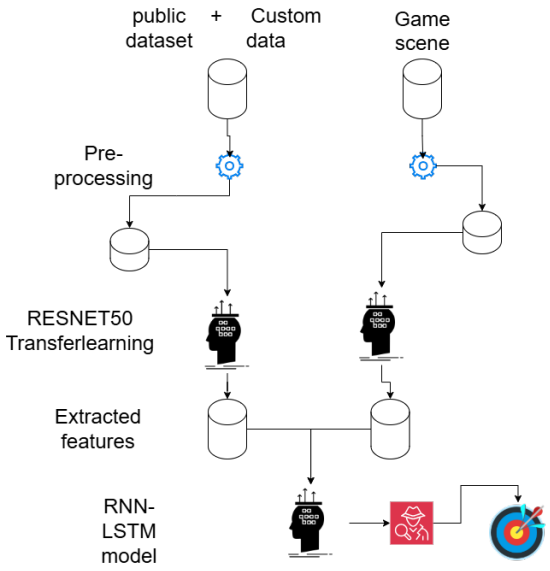


Figure 10: Model 3 RNN LSTM

Figure 10 shows the designed architecture of method 3 which utilizes a multilevel combination of transfer learning & RNN-LSTM. In the starting level of architecture dataset is collected and pre-processed for further application. Subsequent level is created model by utilizing transfer learning with pre-trained model CNN ResNet50.Then

features from models trained over research datasets are saved. On the next level, these features are combined and fed to the RNN-LSTM network, which will inherent sequential dependencies between features from previous models. The final models are evaluated using evaluation metrics. This design architecture aims to effectively capture temporal patterns and spatial within the dataset.

5 Implementation

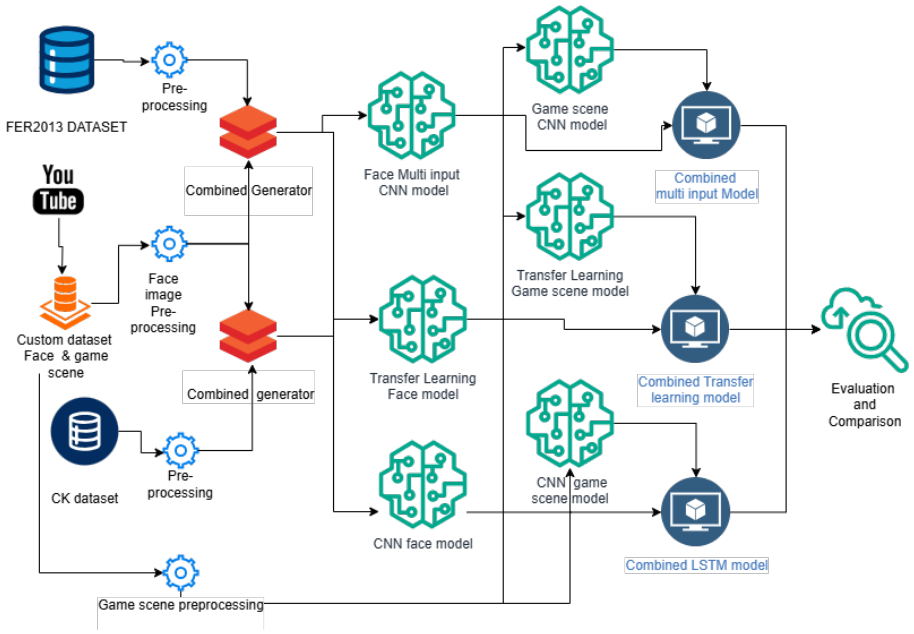


Figure 11: Implementation

The associated study's implementation is depicted in Figure 11. The whole implementation of this research study is carried out by using Python coding language which is more suitable and easier for machine learning and coding practices. For initial Dataset collection Kaggle dataset library is used to find and download CK and FER2013 datasets. The open-source streaming platform YouTube is used for creation of Custom datasets. Open live-streamed video of gamers with high quality is used for creation by capturing frames according to requirements. These images as data points are stored similarly to other dataset structures. The exploratory data analysis is done on these datasets to understand data insights using various inbuilt libraries like NumPy and pandas. Following this Preprocessing step is done on dataset for further application. TensorFlow library's preprocessing function provides numerous functions for the same. ImageDataGenerator function from preprocessing library is employed for preprocessing step implementation. For facial image models there is more than one dataset is employed so a local function combined generator is created. For modeling step implementations TensorFlow library keras provides all required libraries. During implementation, various libraries like models, layers, applications, and utils are imported from Keras for machine learning model training.

6 Evaluation

Deep learning models have been successfully included in this research study of facial and gaming emotion detection model comparison. This section provides summary of evaluation of models that are employed for research study. This evaluation analysis significantly explores insights during experimentation.

6.1 Experiment Method 1

Implementation of multi-input CNN model on the research datasets. In this implementation, first two models need to be evaluated which are a combination of 3 facial image datasets. The accuracy of CNN model trained over (a) CK & Custom dataset is 90 % and (b) FER2013 & custom dataset is 93.27%.

The combination of FER2013 and custom dataset model shows slightly better results than CK & custom dataset combined model. Even though FER2013 has larger dataset than CK, the accuracy change is small because the CK dataset is created for experiments.

It is very important to validate the computational cost of each model. Table 1 shows the time taken by each model for training. The computational cost for each model is as follows: CK + Custom dataset is trained over T4 GPU hardware accelerator for 31 minutes. FER2013 + Custom dataset is trained over T4 GPU hardware accelerator for 44 minutes.

Table 1: Training time

	CK + Custom dataset	FER2013+ Custom dataset	Game scene dataset
Training Time (in min)	31	34	9

The accuracy of CNN model trained over custom game scene data is 71.795%. The model shows great accuracy but was trained 20 epochs cycle. The computational cost is a model trained for 9 minutes over CPU hardware accelerator.

At the end of the first method implementation the accuracy of combined models, pre-trained models: face image dataset model, and game scene dataset model are combined and accuracy is 73.54 %. The computational cost for this training is 7 minutes of computation over T4 GPU hardware accelerator.

6.2 Experiment Method 2

Implementation of transfer learning CNN model on the research datasets. In this case, two models with transfer learning need to be evaluated. The accuracy of CNN VGG16 transfer learning model trained over (a) CK & Custom dataset is 77.05 % and (b) FER2013 & custom dataset is 23.95%.

The combination of CK and custom dataset model shows a better result than FER2013 & custom dataset combined model. Due to large dataset, FER2013 is trained over less learning rate to find optimized computation cost and accuracy.

Table 2 shows the time taken by each model for training. Table 2 shows importance of the computational cost for training models in research study process. The computational cost for each model is as follows: CK + Custom dataset is trained over TPU v2-8

hardware accelerator for 40.6 minutes.FER2013 + Custom dataset is trained over TPU v2-8 hardware accelerator for 102.5 minutes.

Table 2: Training time

	CK + Custom dataset	FER2013+ Custom dataset	Game scene dataset
Training Time (in min)	40.6	102.5	12

The accuracy of transfer learning CNN VGG16 model trained over custom game scene data is important part of research study. The transfer learning from CNN VGG16 with imagenet weights shows a great accuracy of 82.46%. The model shows great accuracy but was trained for 10 epochs cycle. The computational cost is model trained for 12 minutes over the CPU hardware accelerator.

For accuracy of combined models, pre-trained models: face image dataset model and game scene dataset model are combined, and combined model accuracy for (a) CK & Custom dataset with game scene image is 83.38 % and for (b) FER2013 & custom dataset with game scene image is 34.14%. The computational cost for this training in both these models is 12 minutes and 9 minutes respectively. They are of trained over TPU v2-8 hardware accelerator.

6.3 Experiment Method 3

Implementation of transfer learning with RNN- Long short-term memory model on the research datasets. In this implementation, CNN ResNet50 transfer learning model needs to be evaluated. The accuracy of CNN transfer learning model trained over CK & Custom dataset is 92.32 %. This model is trained for spatial learning for future application so this transfer learning method is evaluated too.

It is very important to validate the computational cost of the model. Following table 3 shows the training time taken by models. The computational cost for each model is CK + Custom dataset is trained over T4 GPU hardware accelerator for 28 minutes.

Table 3: Training time

	Face image dataset	Game scene dataset
Training Time (in min)	28	8

The accuracy of the CNN ResNet50 transfer learning model trained over custom game scene data is 81.62%. The model shows great accuracy but was trained 5 epochs cycle. The training model over the game scene dataset required 8 minutes with the use of CPU computation.

The combination for Long Short-term memory model, the face image transfer learning model, and game scene transfer learning models are combined together create a unified model with the RNN-LSTM method. This combined model achieved an accuracy of 80.66%. The computational cost for this training is 5 minutes of computation over a T4 GPU hardware accelerator.

6.4 Final comparison of all experiments

Accuracy: Accuracy is an important metric for evaluating any machine learning model on a first basis. Accuracy signifies total correct predictions with total prediction with a given dataset by the trained model.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + True\ Negatives + False\ Positives + False\ Negatives} \quad (1)$$

The above formula (1) shows a mathematical representation of calculating accuracy for machine learning models. A higher accuracy percentage shows models are trained well and can classify emotion based on input images well, which is an essential case in this research study. The following Table 4 shows the comparison of accuracies of all experimented methods throughout this research.

Table 4: Accuracy comparison of all methods

	CK + custom face image dataset	FER2013 + custom face image dataset	Game scene dataset	Combined model
Method 1	90 .24	93.27	71.79	73.54
Method 2	77.05	23.95	82.46	83.38 & 34.14
Method 3	92.32	N.A.	81.62	80.66

The following Table 5 shows the comparison of epochs taken to achieve optimized accuracy of all methods throughout this research. The epochs taken to train each model vary and are an important factor influencing model performance and computation time. In this research study, multiple numbers of epochs have been tried to find an optimized number that balances accuracy and training period. Increasing epochs yielded diminishing results like overfitting and increased computational time, whereas decreasing epochs resulted in worsened accuracy. The below table shows a count of epochs for optimized training.

Table 5: Number of epochs

	CK + custom face image dataset	FER2013 + custom face image dataset	Game scene dataset
Method 1	5	5	20
Method 2	10	5	10
Method 3	5	N.A.	5

6.5 Discussion

Concerning the experiment, the outcome showed significant improvement in accuracy and comparison between different methods. Initial experiments showed the importance of preprocessing game scene image dataset, which consisted of varying size images and RGB colored format. To optimize computation cost after multiple testings, training images are

resized to standard size of 150 x 100 and greyscale color format for game scene images. Data is collected specifically for this research purpose but variety in data points is limited and size of the dataset is around 200 only. This small-sized dataset poses challenges to train robust models. The dataset contains various game data points made with less variation, which results in models easily overfitting. Some data point images contain game scenes as well as the small size of gamer's face in corner too, which introduces noise and inconsistency. These limitations and issues in dataset can contribute to overfitting and hinder models' overall performance to generalize unseen data. Future research can focus on expanding datasets, using advanced data augmentation, and investigating more advanced techniques to avoid limitations due to limited datasets.

Multi-input Convolutional neural network method proved to be an effective approach to this research study task, showing better results and easy deployment. Whereas creating CNN from scratch takes greater number of epochs which can be seen in game scene image model creation. Models' performance was influenced due to multiple games and the limited size of data points while creating the game scene image dataset. To encounter this data augmentation techniques have been employed in research study. Furthermore, while combining models different sizes and data distributions posed challenges to the overall performance of model.

While implementing Transfer learning model, the necessity of consideration of computation cost for training time with larger datasets can be highlighted. As shown in evaluation section computational cost transfer learning on bigger dataset in this study proved to be computationally intensive, necessitating adjustments of significant parameters like batch size and smaller learning rate for optimal performance. For real-time application selection of proper pre-trained models, fine-tuning parameters is necessary to balance models' accuracy and computational cost. For example, in this study overall computational cost for game scene image dataset training is less due to smaller dataset size and complexity.

RNN-LSTM is well suited for capturing sequential patterns in datasets, whereas in this research study their application to image-based datasets, and limited datasets poses challenges. In this study, we leveraged RNN-LSTM model for a gaming facial emotion detection study. Despite model's strong accuracy on training dataset, its ability to generalize unseen, new data may be limited due to the limited dataset and lack of sequential pattern in the dataset. To encounter and achieve this high accuracy data augmentation technique is employed in this study. Furthermore, investigation and experimenting need to be conducted with extended dataset containing video data points to fully assess the application of RNN-LSTM for gaming facial emotion detection study.

7 Conclusion and Future Work

In conclusion, this research study investigated comparison of different machine learning methods and models for detection of gamer's facial emotion recognition during a specific gaming context. The research study demonstrates various findings from preprocessing, parameter tuning to method selection. Multiple initial experiments indicated need to resize game scene image data to 150 x 100, grey-coloured format for yielding better results. Minimizing size of game scene data less than above makes model overfitting and increasing size results in increased computational cost.

Transfer learning with large pre-trained model CNN VGG16 with ImageNet weights

shows greater results for game scene data models whereas for facial image dataset models CNN ResNet50 shows better accuracy. Transfer learning with larger datasets can increase accuracy on a significant level, but it also tends to increase computational cost. In contrast, models trained over relatively smaller datasets can also achieve substantial performance gain and accuracy. RNN-LSTM model which is traditionally used for sequential datasets or video datasets, was found adaptable after training over image datasets, especially with help of data augmentation performance can be achieved.

Based on conclusion, there are multiple future areas of scope that can be addressed with this research study finding. Future research could focus on the expansion of the game scene image dataset to include diverse games and larger number of data points. Additionally exploring more advanced state of art transfer learning methods that can balance models' accuracy and computational cost will help in real-life use case implementation. The development or usage of advanced standardized techniques or tools for data collection and dataset creation will facilitate future research studies. This will enable more reliable and innovative comparisons between newer more advanced methods and models.

References

Bayunanda, E., Utami, E. and Ariatmanto, D. (2022). Facial expression classification analysis using facial images based on resnet-50v2, *2022 International Conference on Information and Computer Technologies for Intelligent Systems (ICITSEE)*.

Frommel, J., Schrader, C. and Weber, M. (2018). Towards emotion-based adaptive games, *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

Gondkar, A., Gandhi, R. and Jadhav, N. (2021). Facial emotion recognition using transfer learning: A comparative study, *2021 2nd Global Conference for Advancement in Technology (GCAT)*.

Habibullah, M. U., Md.O., K., M., S., Md.S.H., S. and Md.S., R. (2023). Improved convolutional neural network and transfer learning with vgg16 approach for image classification, *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pp. 389–396.

Kalansooriya, P., Ganepola, G. and Thalagala, T. (2020). Affective gaming in real-time emotion detection and smart computing music emotion recognition: Implementation approach with electroencephalogram, *2020 6th International Conference on Smart Computing and Electronic Engineering (ICSCEE)*.

Khademi, Z., Ebrahimi, F. and Kordy, B. H. (2022). A transfer learning-based cnn and lstm hybrid deep learning model to classify motor imagery eeg signals, *Computers in Biology and Medicine* **143**: 105288.

Khan, A. and Rasool, S. (2022). Game induced emotion analysis using electroencephalography, *Computers in Biology and Medicine* **145**: 105441.

Laird, J. E. and van Lent, M. (2001). Human-level ai's killer application: Interactive computer games, *AI Magazine* **22**(2): 15–26.

Lucey, P., Cohn, J. F., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I. (2010). The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression, *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 94–101.

Mostafa, A. A., Khalil, M. and Abbas, H. (2018). Emotion recognition by facial features using recurrent neural networks, *2018 IEEE 15th International Conference on Computer Engineering and Systems (ICCES)* .

Negi, A., Arora, A., Bisht, S., Devliyal, S., Kumar, B. and Kaur, G. (2024). Facial emotion detection using cnn & vgg16 model.

Sánchez-Cauce, R., Pérez-Martín, J. and Luque, M. (2021). Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data, *Computer Methods and Programs in Biomedicine* **204**: 106045.

Shrivastav, Y. (2024). The esports gaming industry has seen exponential growth globally, and india is no exception. with a young population and increasing internet penetration, india has emerged as a promising market for esports.

Song, M., Yang, Z., Baird, A., Parada-Cabaleiro, E., Zhang, Z., Zhao, Z. and Schuller, B. (2019). Audiovisual analysis for recognising frustration during game-play: Introducing the multimodal game frustration database, *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*.

Xu, X., Mei, Y. and Li, G. (2018). Adaptive cgf commander behavior modeling through htn guided monte carlo tree search, *Journal of Systems Science and Systems Engineering* **27**(2): 231–249.

Yang, W., Rifqi, M., Marsala, C. and Pinna, A. (2018). Physiological-based emotion detection and recognition in a video game context, *Proceedings of the IEEE International Joint Conference on Neural Networks*, HAL (Le Centre pour la Communication Scientifique Directe).

Zahara, L., Musa, P., Prasetyo Wibowo, E., Karim, I. and Bahri Musa, S. (2020). The facial emotion recognition (fer-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (cnn) algorithm based raspberry pi, *2020 Fifth International Conference on Informatics and Computing (ICIC)*, pp. 1–9.