University of Kerbala
College of Computer Science & Information Technology
Computer Science Department

GROUP EMOTION DETECTION IN A SMART SIMULATION SYSTEM

A Thesis

Submitted to the Council of the College of Computer Science & Information Technology / University of Kerbala in Partial Fulfillment of the Requirements for the Master Degree in Computer Science

Written by

Ali Hussein Mousa Hamza

Supervised by

Prof.Dr. Asia Mahdi Naser Alzubaidi

2024 A.D. 1445 A.H.

بسم الله الرحمن الرحيم

الآية (11)

صدق الله العلي العظيم

Supervisor Certification

I certify that the thesis entitled (**GROUP EMTION DETECTION IN A SMART SIMULATION SYSTEM**) was prepared under my supervision at the department of Computer Science/College of Computer Science & Information Technology/ University of Kerbala as partial fulfillment of the requirements of the degree of Master in Computer Science.

Signature:

Supervisor Name: prof. Dr. Asia Mahdi Naser Alzubaidi

Date: / /2024

The Head of the Department Certification

In view of the available recommendations, I forward the thesis entitled "GROUP EMOTION DETECTION IN A SMART SIMULATION SYSTEM" for debate by the examination committee.

Signature:

Assist. Prof. Dr. Muhannad Kamil Abdulhameed

Head of Computer Science Department

Date: / /2024

Certification of the Examination Committee

We hereby certify that we have studied the dissertation entitled (**GROUP EMOTION DETECTION IN A SMART SIMULATION SYSTEM**) presented by the student (**Ali Hussein Mousa Hamza Al-khafaji**) and examined him in its content and what is related to it, and that, in our opinion, it is adequate with (**Very Good**) standing as a thesis for the degree of Master in Computer Science.

Signature:
Name: Baheeja Kudhair Shukur
Title: Prof. Dr.
Date: / / 2024
(Chairman)

Signature:
Name: Saif Ali Abd Alradha Alsaidi
Title: Assist. Prof. Dr.
Date: / / 2024
(Member)

Signature:
Name: Ali Retha Hasson Khayeat
Title: Assist. Prof. Dr.
Date: / / 2024
(Member)

Signature:
Name: Asia Mahdi Naser Alzubaidi
Title: Prof. Dr.
Date: / / 2024
(Member and Supervisor)

Approved by the Dean of the College of Computer Science & Information Technology, University of Kerbala.

Signature:
Assist. Prof. Dr. Mowafak khadom Mohsen
Date: / / 2024
(Dean of Collage of Computer Science & Information Technology)

Dedication

I’d like to dedicate this work:

To The Prophet Mohammad and his Ahl al-Bayt and especially to Al-Imam Al-Mahdi; Peace & blessings be upon them. To my parents, wonderful wife, brothers, sisters, supervisor and friends. To everyone who supported me.

Acknowledgement

First of all, I thank Allah, my creator and Ahl al-Bayt and especially to my guider Al-Imam Al-Mahdi (Peace & blessings be upon them); for aiding me to finish what I started and presenting this work in the best way; all my thanks and gratitude to them.

I want to express my appreciation to my supervisor: Prof. Dr. Asia Mahdi Naser Alzubaidi; For her dedication, guidance and encouragement during this work. Special thanks go to my Mother and a lot of thanks to my brother, sisters and my friends for encouraging and supporting me. Finally, I'd like to state that no amount of words could adequately convey how grateful I am to My Lovely Wife “Hadil” for her faithful efforts to give me the maximum endless support in everything.

Ali Hussein Mousa Hamza

Abstract

In the field of Human-Robot Interaction (HRI), the ability to recognize and understand human emotions plays an essential role in facilitating contextually appropriate and socially sensitive interactions. Explore within the details of how get benefit of Group Emotion Recognition (GER) regards as a foundation for improving decision-making processes within the context of HRI.

To fix the problem of how to make the suitable decision, the methodology of this thesis consists from many steps. Begin by looking for using a datasets encompassing a wide array of emotional expressions and group scenarios. These datasets, derived from both simulated and real-world environments, serve as the basis for following analyses. A two-tiered approach to face detection, integrating Haar Cascade and Histogram of Oriented Gradients (HOG) Descriptors, ensures robust and efficient face detection. Feature extraction relies on Dlib's facial landmark detector, enriched by the integration of Multi-task Cascaded Convolutional Networks (MTCNN) to enhance accuracy.

The contribution of this research lies in the enhancement of a Convolutional Neural Network model, designed for accurate multi-class image classification of facial expressions. This Attention CNN model, trained on the prepared dataset, enables precise emotion recognition with an impressive accuracy rate. Then this model has been tested on two types: essential and secondary datasets. The ROS/Gazebo Generated regarded as a main Dataset of 23,222 images was achieved an accuracy of 98.52% across all emotion classes while the RAF dataset regarded as a supplementary dataset contain 3068 images achieved around 66.69% of accuracy.

Furthermore, a two videos dataset represent two different scenarios has been well preprocessed, including frame extraction, multi-face detection, and GER classification, follows has been applied. This process includes affective

steps such as face localization, emotion labeling, entropy calculation, and GER classification, providing valuable insights into the spatial dynamics of group emotions.

Finally, the action will be taking depending on two aspects: positive or negative labels of the emotions of the individual’s group. Also the evaluation process demonstrates the model's effectiveness, achieving high accuracy in differentiating emotions and making informed decisions.

Declaration Associated with this Thesis

[1] A. H. M. Alkhafaaji and A. M. N. Alzubaidi: “A review of Emotions Recognition via Facial Expressions for Human–Robot Interaction” in 5th International Conference on Information Technology, ICITAMS 2023, Qadisyah University, Iraq, IEEE Explore

[2] Ali Hussein Mousa Hamza Alkhafaaji and Aisa Mahdi Naser Alzubaidi: “ENHANCING HUMAN-ROBOT INTERACTION THROUGH GROUP EMOTION RECOGNITION” in Iraqi Journal for Computers and Informatics (IJCI), Online ISSN: 2520-4912, institutional journal issued by the University of Information Technology and Communications (UoITC) in Baghdad- Iraq

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

ABBREVIATION	DESCRIPTION
CNN	**C**onvolutional **N**eural **N**etwork
Dlib	**D**avis E. King **lib**rary
DM	**D**ecision **M**aking
FD	**F**ace **D**etection
FN	**F**alse **N**egative
FP	**F**alse **P**ositive
GER	**G**roup **E**motion **R**ecognition
Haar	**Haar** cascade classifier algorithm
Hog	**H**istogram of **O**riented **G**radients
HRI	**H**uman **R**obot **I**nteraction
IER	**I**ndividual **E**motion **R**ecognition
MTCNN	**M**ulti-task **C**ascaded **C**onvolutional **N**etworks
RAF	**R**eal-world **A**ffective **F**aces
RNN	**R**ecurrent **N**eural **N**etwork
ROI	**R**egion **O**f **I**nterst
ROS	**R**obot **O**perating **S**ystem
FPS	**F**rame **P**er **S**econd
TN	**T**rue-**N**egative
TP	**T**rue-**P**ositive

CHAPTER ONE

INTRODUCTION

1.1 Introduction

In recent years, the field of robotics has witnessed remarkable advancements, especially in human-robot interaction (HRI). The ability of robots to understand and respond the human emotions has become a crucial aspect of creating more natural and meaningful interactions with robots. Emotion recognition within the robotics field has demonstrated its transformative potential as a technology, gives robots the capability to distinguish and interpret human emotions [**1**]. Thus, this progression supports interactions that are filled by heightened empathy and personalized engagement. Thus, the emotion recognition has emerged as a pivotal aspect within this field, empowering robots to not only discern but also interpret human emotional states [**2**]. This transformative potential has lent robots the capability to engage in interactions that are characterized by heightened empathy and personalized engagement, resulting in more natural and meaningful exchanges [**3**]. Therefore, to enable this profound leap in emotional understanding, the research community has fervently pursued the development and deployment of a range of emotion recognition techniques [**4**]. These techniques are applied across diverse applications, each harnessing the power of robots' newfound emotional insight. For instance, within the healthcare sector, robots are being equipped with emotion recognition capabilities to offer empathetic support to patients, promoting a sense of companionship and emotional well-being [**5**]. Similarly, in educational settings, emotion-aware robots adapt their teaching methodologies based on students' emotional responses, creating an adaptive and dynamic learning environment. Customer service is another realm where robots are being employed to enhance customer experiences, offering empathetic interactions that cater to individual emotional states [**6**]. The entertainment industry benefits from robots capable of engaging audiences through emotionally rich interactions, leading to more immersive and captivating experiences. Finally, Facial expression analysis involves deducing emotional states, often utilizing techniques such as facial landmark detection and machine learning algorithms [**7**]. Speech analysis deciphers patterns in speech tones and prosody to infer emotions, making use of speech recognition and sentiment analysis techniques. Monitoring physiological signals, like heart rate and skin conductance, adds an additional layer of emotional insight. Text analysis delves into written or typed text to extract emotional cues, with natural language processing and sentiment analysis playing key roles insights from written or typed content [**8**].

1.2 Research Problem

The current researches on emotion recognition in robotics mainly focus on collecting the emotions of individuals to determine the suitable decision-making situations. This research problem presents a large gap in the field, as the dynamics of group emotions can substantially influence the outcome of HRI [**1**]. By addressing the challenge of group emotion recognition, aim of studying is to provide an Attention CNN which is an advanced variant of a CNN that combines attention mechanisms. These mechanisms allow the network to dynamically emphasize specific parts of input data that are most relevant to the emotion recognition task [**5**]. Therefore, an Attention CNN would enable the network to allocate more attention to crucial aspects such as the distribution of emotions among the group and the sizes of individual faces in the frame. By doing so, the model can effectively capture and analyze key features of the emotion recognition system. This approach aligns well with the project's goal of deriving insights from multiple data resources to understand group emotions more comprehensively [**9**].

1.3 Objectives

The specific objectives of this research are as follows:

1. Developing an Attention CNN model for emotion recognition: To build a robust and accurate emotion recognition system capable of detecting emotions from individual faces in a given scene. Implements the Attention CNN model on video data, could allow continuous emotion analysis in dynamic human-robot interaction scenarios.
2. Proposing a method for determining group emotion in a scene: To introduce an approach that determine the emotions of a group of individuals based on the entropy of distributed emotions and the size of each face appearing in the frame.
3. Explore the influence of positive and negative emotions on HRI and how emotions can impact the dynamics of HRI. To establish a robust framework for harmonious HRI. The frame work has strong emphasis on sensitivity, adaptability, and ethical considerations, enabling robots interpret and respond the emotional states of human groups in diverse scenarios.

1.4 Scope and Limitations

The scope of this research involves recognizing a range of emotions, focusing on social robots designed for interaction in diverse environments. However, several limitations and constraints must be considered.

- The effectiveness of the proposed emotion recognition system can vary based on input video data quality and the accuracy of the trained Attention CNN model. Generalizing this method to highly dynamic and crowded settings may present challenges.
- Also, hardware limitations, including processing power and memory capacity, could affect speed and accuracy.

- Similarly, sensing limitations, calibration, and maintenance are critical factors, as inadequate hardware maintenance might lead to inaccurate results.
- Likewise, software complexities, such as bugs and compatibility issues, pose risks to the system's functionality.

1.5 Thesis Structure

This thesis is organized into five chapters to present an analysis of the proposed method for group emotion recognition in HRI.

Chapter Two: Theoretical Background

This chapter provides an overview of HRI and the significance of emotion recognition. It reviews existing techniques used and highlighting the gaps in the current literature that this research aims to address.

Chapter Three: Proposed Methodology

In this chapter, the development and implementation of the Attention CNN model for individual facial emotion recognition are discussed. The process of training the model and its validation on independent datasets is explained; also present the proposed method for determining group emotions and how decision making specified.

Chapter Four: Results

Show, discuss and analysis the result of the proposed model.

Chapter Five: Conclusion

Provides study conclusions and states future works.

CHAPTER TWO

THEORETICAL BACKGROUND

2.1 Overview

In recent years, the field of Human-Robot Interaction (HRI) has advanced significantly due to the integration of cutting-edge technologies and a deeper understanding of human psychology and behavior. Emotions play a pivotal role in human communication and interaction, influencing social dynamics and conveying intentions. Robots need to recognize and interpret these emotional signals to engage effectively with humans and respond appropriately [**10**]. Decision-making in robots extends beyond algorithms, requiring a nuanced understanding of human emotions, preferences, and social cues to tailor responses and adapt to dynamic environments. The theoretical chapter aims to bridge emotion recognition and decision-making within HRI, leading to the practical implementation of a new approach: Human-Robot Interaction decision-making based on Group Emotion Recognition. It explores individual emotion recognition, challenges in accurate emotion recognition, and transitions to group emotions, delving into emotional dynamics within collective settings [**11**]. GER, introduced as an innovative paradigm, extends emotion recognition to multiple individuals within a scene. Concepts such as entropy are explored as powerful metrics for quantifying emotional diversity and uncertainty, enhancing understanding and shaping adaptive decision-making in HRI scenarios. The theoretical framework as in Figure 2-1 involves determining GER, emotion classification, decision-making, and integrating a chatbot, forming the foundation for recognizing emotions in group settings and triggering appropriate robot behaviors. The interconnected approach enhances human-robot communication and lays the groundwork for subsequent chapters on practical implementation, experimental validation, and real-world implications in HRI [**12**].

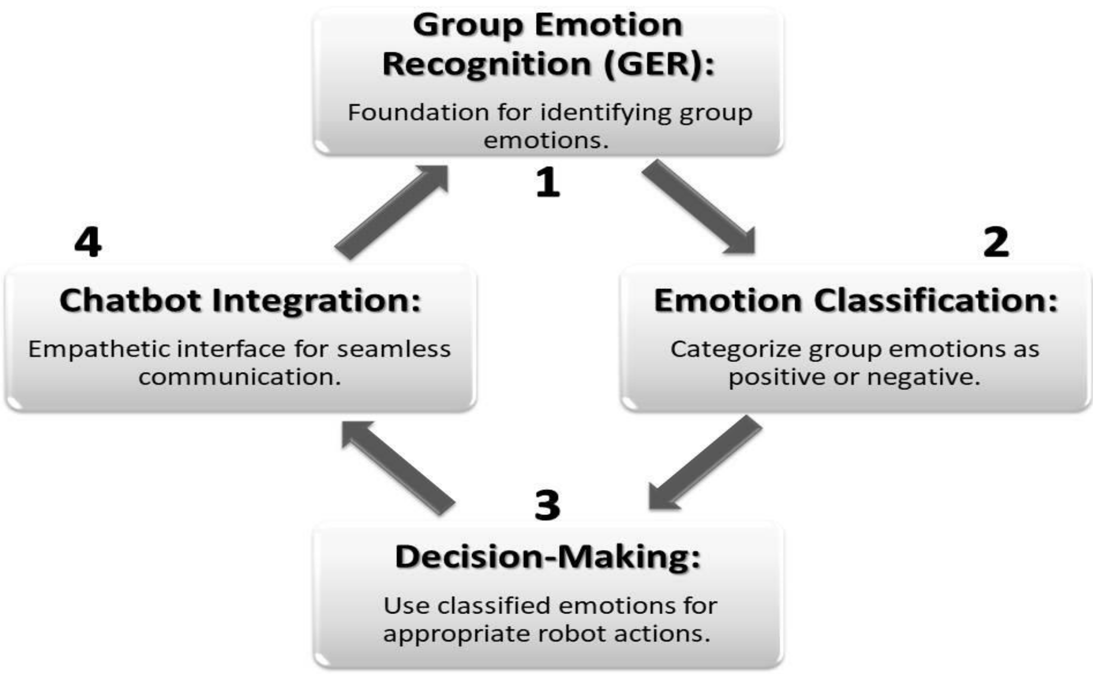


Figure 2-1 decision action based on GER.

2.2 Emotion Recognition: Individual and Collective

Understanding and interpreting human emotions is a complex yet essential task in the field of HRI. Individual Emotion Recognition (IER) is the foundational step in this process, where the objective is to accurately detect and classify emotions expressed by an individual through their facial expressions. [**13**]. This recognition serves as a fundamental building block for extending emotion analysis to group settings and ultimately influencing robot behaviors and interactions. At the core of IER is the analysis of facial expressions, it has a wealth of emotional information, ranging from subtle microexpressions to overt macroexpressions. In the field of IER, it is important to distinguish between microexpressions and macroexpressions Microexpressions are fleeting, involuntary facial movements that occur in response to genuine emotions [**14**]. They often provide authentic cues about a person's emotional state but are challenging to capture due to their rapid nature. Macroexpressions, on the other hand, are more sustained facial expressions that align with the underlying emotional state as shown in Figure 2-2. Here are some examples of microexpressions and macroexpressions for each of the six basic emotions, based on real-life scenarios:

1. **Fear** [15]**:**

* Microexpression: A person briefly widens their eyes and raises their eyebrows when they hear a sudden loud noise.
* Macroexpression: Someone screams and jumps back when they encounter a snake unexpectedly on a hiking trail.

2. **Surprise** [16]**:**

* Microexpression: A split-second widening of the eyes when someone opens a gift and finds an unexpected item inside.
* Macroexpression: A person exclaims loudly and jumps out of their chair when they walk into a surprise birthday party thrown for them.

3. **Anger** [17]**:**

* Microexpression: A subtle clenching of the jaw and a brief narrowing of the eyes when someone receives frustrating news via.
* Macroexpression: A person shouts and slams a door in anger when they discover their car has been towed without notice.

4. **Disgust** [18]**:**

* Microexpression: A momentary wrinkling of the nose when someone smells a foul odor while passing a garbage dumpster.
* Macroexpression: Someone visibly gags and pushes away a plate of spoiled food after taking a bite.

5. **Sadness** [19]**:**

* Microexpression: A brief downturn of the corners of the mouth and a slight trembling of the lower lip when someone receives a sad text messageemail.
* Macroexpression: Tears well up in a person's eyes, and they sob uncontrollably when they hear about the loss of a loved one.

6. **Happiness** [**20**]:

* Microexpression: A quick, genuine smile that appears when someone receives a compliment from their.
* Macroexpression: Someone bursts into laughter, hugs their friends, and dances around with joy after learning they've won a major lottery jackpot.

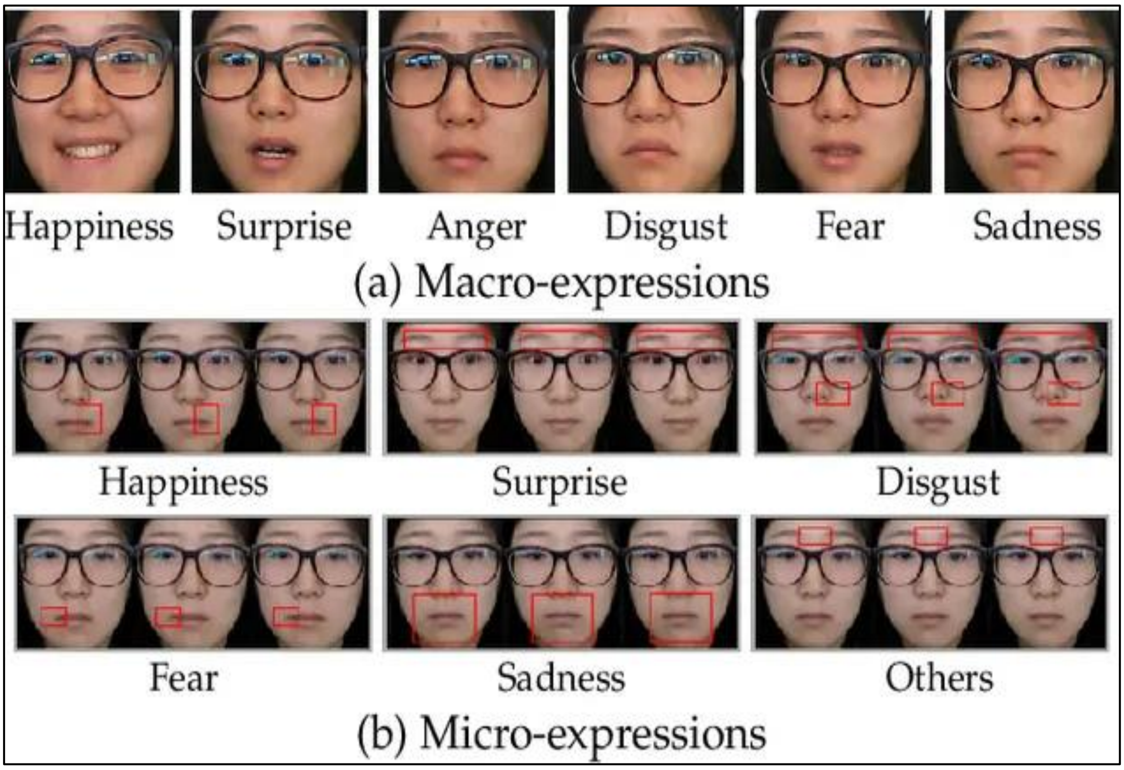


Figure 2-2 Macro_expression and Micro_expression [**21**].

In the domain of IER, numerous challenges and complexities exist that can significantly impact the accuracy and reliability of emotion recognition systems crush. These challenges encompass a range of issues, including occlusions that obscure facial features, variations in lighting conditions, diverse and nuanced facial expressions, rapid temporal dynamics of emotions, subjectivity influenced by individual and cultural differences, the need for cross-domain generalization, data imbalance in biased datasets, ambiguity in emotion categorization, contextual dependency of emotions, and ethical considerations related to privacy and potential misuse [**21**].

2.3 Group Emotion Dynamics and Recognition

GER extends conventional emotion recognition approaches from individual emotions to the complexities of group emotions, offering potential benefits in enhancing user engagement, personalizing interactions, improving group dynamics, and fostering social acceptance of robots in HRI scenarios [**22**].

2.3.1 The Complexity of Group Emotions

To explore deeper into the field of GER, a rich and complicated mixture of emotional dynamics has been faced that emerge within a social environment. The transition from understanding individual emotions to comprehending the complexities of group emotions unveils a fascinating interplay of psychological and social phenomena. This section aims to clarify the multifaceted nature of group emotions, exploring concepts such as emotional contagion, emotional convergence, and social influence, which collectively contribute to the formation of collective emotional states. By unraveling these dynamics, this thesis gain essential insights into the intricacy and depth of emotions within groups, understanding that is essential for effectively capturing the nuanced landscape of emotions in this context [**23**].

1. Emotional Contagion: Emotional contagion refers to the phenomenon where the emotions of one individual within a group can spread and "infect" other individuals, leading to a shared emotional experience. This contagion occurs through nonverbal cues, facial expressions, body language, and vocal intonations. When a person expresses a strong emotion, such as excitement, happiness, or anxiety, those around them may unconsciously mirror and adopt similar emotional states. In a group setting, emotional contagion can create a ripple effect, amplifying and propagating emotions across individuals. For instance, the infectious laughter of one person in a group can trigger laughter in others, creating a collective atmosphere of joy [**24**].

2. <u>Emotional Convergence:</u> Emotional convergence explore into the tendency of individuals within a group to synchronize their emotional states over time. As group members interact and share experiences, their emotions tend to align, resulting in a convergence of emotional expressions. This convergence is driven by the human inclination to seek emotional harmony and maintain social cohesion. Emotional convergence can lead to the amplification of shared emotions and the emergence of a dominant emotional theme within the group. In collaborative tasks or social interactions, emotional convergence can shape the collective mood and influence decision-making processes [**25**].

3. <u>Social Influence and Emotional Framing:</u> The concept of social influence plays an essential role in shaping group emotions. Individuals within a group are not only influenced by their internal emotional states but also by the emotional expressions and behaviors of others. Social influence can stem from perceived social norms, group dynamics, and the desire for social approval. Emotional framing, where individuals interpret and label emotions based on social cues and context, further contributes to the fluidity of group emotions. For instance, if a leader within a group displays enthusiasm and optimism, others may adopt similar emotions, thereby contributing to a positive group atmosphere [**26**].

4. <u>Emergence of Collective Emotional States</u>: As emotional contagion, emotional convergence, and social influence intertwine, collective emotional states emerge as a result of complex interactions. These collective emotional states transcend the sum of individual emotions, giving rise to a unique emotional climate that characterizes the group as a whole. The emergent emotional state can influence group cohesion, decision-making, and overall group dynamics. Whether it's a crowded at a sporting event, a team working on a project, or an audience at a concert, the collective emotions that emerge define the shared experience and interactions within the group, In Figure 2-3 a Ellipse A represents individual emotions, including one subtype of individual emotions, namely group-based emotions. Ellipse B represents collective emotions. Where collective emotions regarded as many individuals emotions (represented by the smaller circles). Collective emotions unfold as a result of emotional interactions among individuals. These interactions can involve either non-group-based or group-based individual emotions [27].

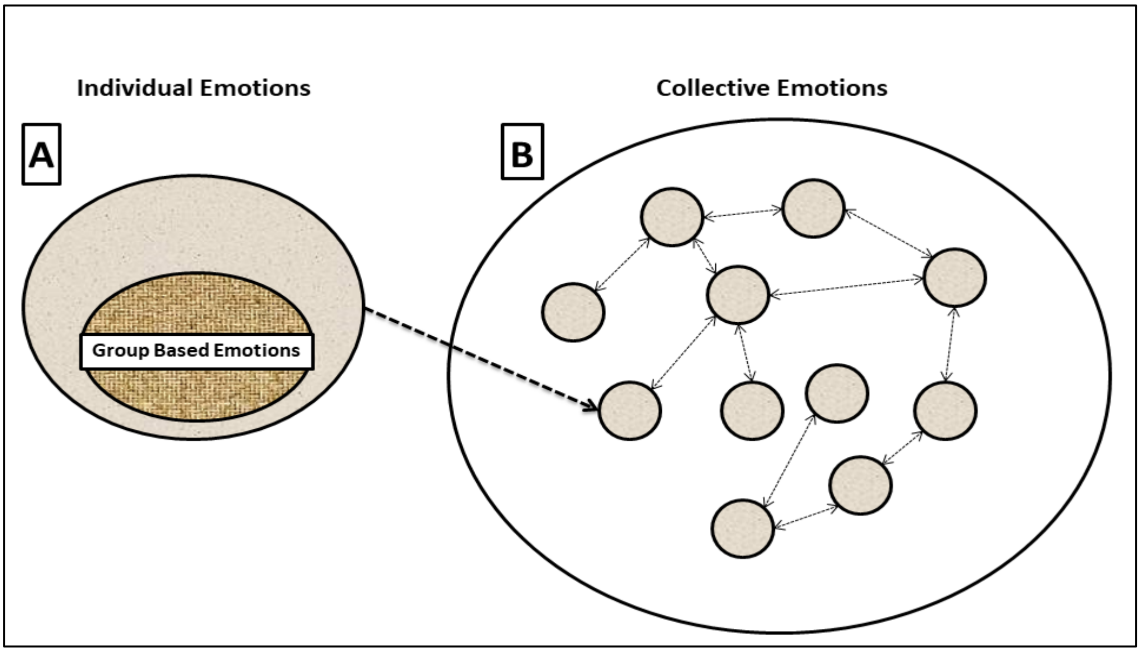


Figure 2-3 : Collective Emotions [28]

Understanding the intricacies of emotional contagion, emotional convergence, and social influence provides a comprehensive lens through which to view the complexity of emotions within groups. These phenomena underscore the interdependence of individual emotional experiences and the profound impact of social interactions on the emotional landscape of a collective setting. By unraveling these dynamics, the thesis covers the way for effective Group Emotion Recognition, enabling robots to navigate and respond to the intricate tapestry of emotions that shape human interactions within groups [**29**] .

2.3.2 Group Emotion Recognition:

GER represents a new paradigm in the field of HRI as it extends the conventional emotion recognition approaches from individual emotions to emotions within a collective group setting. By recognizing and understanding the emotional climate of a group, GER opens up new possibilities and potential benefits for HRI. Now, let's explore these potential benefits and their impact on various applications [**30**]:

1. Enhancing User Engagement Explanation: GER enhances user engagement in HRI by accurately recognizing and responding to group emotions. Emotional interactions foster a strong connection between humans and robots, resulting in more enjoyable and emotionally meaningful interactions [**31**].

2. Personalizing Interactions Explanation: GER enables personalized responses based on individual emotional states within a group. Robots adapt their behaviors to cater to unique emotional needs, creating more individualized interactions. Users feel valued and understood, strengthening the bond between humans and robots [**32**].

3. Improving Group Dynamics Explanation: GER captures emotions of group members, identifying trends and dynamics. This understanding fosters harmonious interactions, resolves conflicts, and promotes positive emotional states within the group [**33**].

4. Supporting Emotional Regulation Explanation: GER detects emotional fluctuations within a group, aiding emotional regulation. Robots adjust behaviors to diffuse tension or provide support during conflicts, enhancing group emotional well-being [**34**].

5. Enhancing Social Acceptance of Robots Explanation: Integrating GER into robot behaviors increases social awareness and emotional sensitivity. Robots demonstrate a deeper understanding of human emotions, potentially leading to greater acceptance of robots in various contexts [**35**].

6. Facilitating Human-Robot Collaboration Explanation: In collaborative scenarios, GER comprehends emotions among team members. Robots fulfill roles that support group emotional needs, enhancing teamwork and cooperation [**36**].

7. Enabling Emotional Support in Group Contexts Explanation: In therapy or support groups, GER identifies and empathizes with emotional states of multiple individuals simultaneously. Robots provide emotional support and comfort to the entire group, contributing to a supportive atmosphere [**37**].

8. Supporting Psychosocial Well-Being Explanation: GER-equipped robots support emotional well-being in groups. In therapy, robots adapt interactions based on emotional states [**38**].

2.3.3 GER Calculations

In the GER determination process, the initial step involves localizing faces within each frame and calculating proximity metrics to evaluate the significance of these faces. This includes determining the number of faces in a frame and computing metrics such as the largest face and nearest face using Euclidean distance calculations. Formally, for a face 'i' in frame 't' with coordinates (x_i, y_i) relative to the group's center (x_c , y_c) ,the distance 'd_i' between face 'i' and the group's center is calculated as follows [**39**]:

$$\boldsymbol{d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}} \qquad \textbf{(2-1)}$$

These proximity metrics provide crucial insights into the spatial dynamics of the group, which are essential for understanding group emotions. Then, in the GER procedure, the pre-trained Attention CNN model is applied to recognize the emotions expressed by each individual detected face in every frame. This recognition process produces emotion labels for each face, which are a serious component for further analysis. Once emotions labels are assigned for each face within a frame, then the determination of the largest and nearest face calculated, in order to replace the determined face with the corresponding emotion label. Then calculate the probability distribution of all obtained emotions within the group. This distribution helps assess the diversity of emotions expressed [**40**]. To quantify this diversity, the concept of entropy, drawn from information theory, is introduced. Entropy measures the uncertainty or diversity of a distribution, which, in this context, reflects the diversity of emotions within the group. The following steps are taken to calculate the entropy of the emotion labels:

- Let 'E' be the set of emotion labels ,
  E={*happy*, *surprise* , *sad*, *fear*, *disgust*, *angry*}.
- 'N' represents the total number of faces detected in the frame.
- 'n_e' signifies the number of faces with emotion label 'e', where 'e' is an element of the set 'E'.
- Calculate the probability distribution of emotion labels as [**41**]:

  $$\boldsymbol{P_e = \frac{n_e}{N}} \qquad \textbf{(2-2)} \qquad , \quad \text{for each 'e' in 'E'.}$$

- Calculate the entropy of the emotion labels distribution as [**42**]:

$$\boldsymbol{Entropy\_Value = -\sum P_e * log_2(P_e)} \quad \textbf{(2-3)}$$

- The entropy value increases with greater diversity or uncertainty in the distribution of emotion labels within the group, while it decreases when a dominant or focused emotion is expressed by the group. These steps collectively form a comprehensive approach to understanding and quantifying the emotions within a group context. Later to calculate the mean (average) GER value for the scene by summing all GER values for the all frames within that scene and dividing by the number of frames in the scene [**43**]:

$$\mathbf{Mean_{GER}} = \frac{\sum \textbf{GER values for all frames}}{\textbf{Number of frames in the scene}} \quad \textbf{(2-4)}$$

In summary, GER introduces a range of potential benefits for HRI scenarios. By recognizing and understanding the emotions within a group setting, robots can enhance user engagement, personalize interactions, improve group dynamics, support emotional regulation, and foster social acceptance. Additionally, GER enables robots to play more active and adaptive role in collaborative settings and provide emotional support in various group contexts. Overall, GER holds promise for enriching the quality of human-robot interactions and creating more emotionally intelligent and socially aware robots [**44**].

2.4 Decision-Making and Human-Robot Interaction

Within the context of Human-Robot Interaction (HRI), decision-making assumes a critical role, particularly in scenarios involving groups of individuals. As humans engage with robots within a collective setting, a multitude of choices and actions are required to navigate the complex dynamics of group interactions. This section explore into the field of decision-making within group contexts, unraveling the underlying cognitive processes, models, and frameworks that shape choices and behaviors. By comprehending the intricacies of decision-making, this thesis lay the foundation for the development of robots with adaptive, contextually appropriate, and socially sensitive behaviors [**45**] .

2.4.1 Complexity of Group Decision-Making

The decision-making process in group contexts is inherently intricate due to the interplay of multiple perspectives, preferences, and objectives. Individuals within a group often possess diverse viewpoints, priorities, and emotional states, leading to a dynamic and sometimes conflicting decision landscape. Group decisions are influenced by social dynamics, power structures, and the desire for consensus. Understanding this complexity is paramount for robots aiming to facilitate and participate in group decision-making processes; the following aspects are outlines complexities in group decision-making process [**46**]:

1. **Nature-of Decision Process**: Involves multiple perspectives, preferences, objectives, leading to a dynamic and sometimes conflicting decision landscape [**47**].

2. **Diverse Viewpoints**: Individuals within a group possess diverse viewpoints, priorities, and emotional states, further contributing to complexity [**48**].

3. **Social Dynamics**: Group decisions are influenced by social interactions, power structures, and the quest for consensus, adding layers of complexity [**49**].

4. Dynamic and Conflicting: The interplay of various factors makes the decision-making process inherently intricate and sometimes marked by conflicting choices [**50**].

5. Importance-of Understanding: Complete understanding of this complexity is essential for robots facilitating or participating in group decision-making [**51**].

2.4.2 Cognitive Processes in Group Decision-Making

Decision-making within groups is characterized by a series of cognitive processes that guide individuals in evaluating options, assessing outcomes, and arriving at choices. As shown in Figure 2-4; these processes include information gathering, problem framing, consideration of alternatives, and the weighing of potential consequences. Group members engage in information sharing, persuasion, negotiation, and compromise as they collectively work towards a decision. Recognizing and emulating these cognitive processes is essential for robots to contribute meaningfully to group decision-making [**52**].

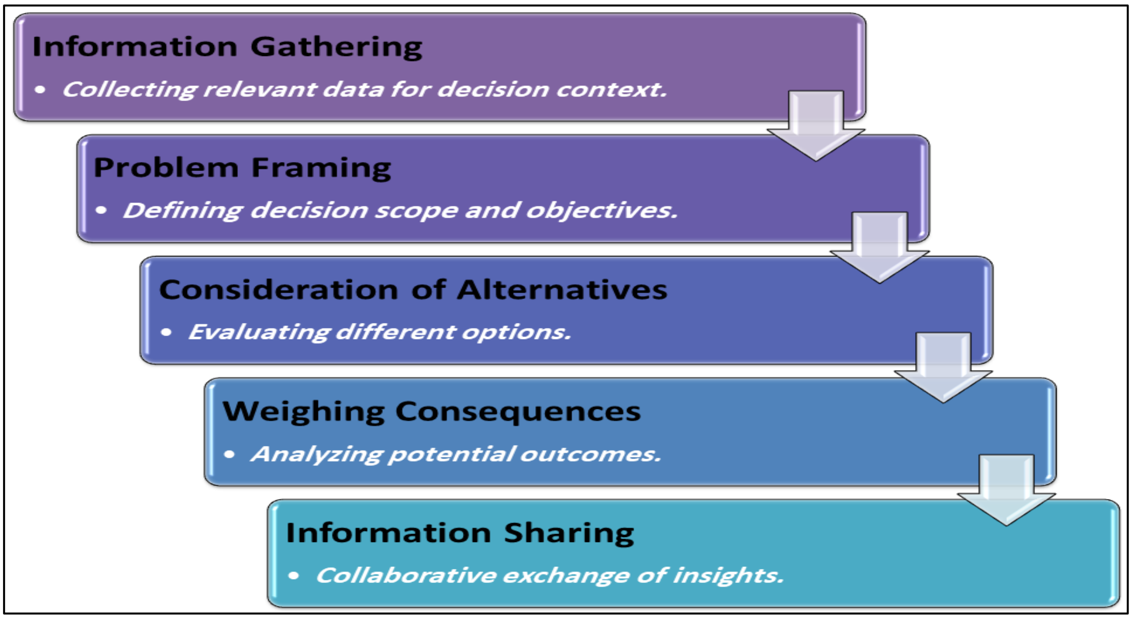


Figure 2-4 Cognitive Processes in Group Decision-Making [**53**].

2.4.3 Adaptive and Contextually Appropriate Behaviors

In order to facilitate effective HRI within group environment, it is imperative that robots have the capability to make decisions that are both adaptive and contextually appropriate. This needs the integration of cognitive processes, decision models, and real-time analysis of group dynamics. Robots must evaluate the emotional climate, interpret social indications, and take into account individual preferences in order to determine the most suitable course of action. Whether these robots are assisting in collaborative tasks, mediating conflicts, or engaging in social interactions, their ability to make informed and adaptive decisions significantly enhances the overall group experience [**54**].

2.4.4 Enhancing Group Dynamics and Interaction Quality

Robots equipped with decision-making capabilities play an essential role in enhancing group dynamics and interaction quality. They facilitate efficient and complete decision-making processes, thus modifying conflicts, promoting active participation, and fostering a sense of shared ownership among group members. Furthermore, robots that align their behaviors with group decisions and emotional states create a harmonious and engaging environment that resonates with human users, making them valuable collaborators and contributors within group environment [**55**].

2.5 Emotion-Informed Decision-Making

Emotions play an essential role in shaping decision-making processes, influencing both individual and group choices. This section explore into the profound impact of emotions on decision-making strategies, emphasizing the importance of adopting emotion-informed approaches in HRI. By leveraging emotional cues, robots can tailor their behaviors and responses, leading to enhanced engagement, effective communication, and more meaningful interactions with humans [**56**].

2.5.1 Emotional Influence on Decision-Making

Research in psychology and neuroscience has demonstrated the profound connection between emotions and decision-making. Emotions act as influential signals, guiding individuals and groups through options and choices. These emotional responses shape how can perceive various alternatives and assess potential outcomes in terms of risk. Positive emotions can drive optimism and a willingness to embrace risks, while negative emotions often instill caution and aversion to risk [**57**]. Understanding these emotional dynamics is essential for robots to align their decisions with human users' emotional states and preferences. The sequence depicted in Figure 2-5 illustrates the significant impact of emotions on decision-making. Research in psychology and neuroscience reveals that emotions play an essential role in guiding the decision-making process. Emotions influence the perceptions, particularly in risk assessment, when this work considers various choices [**58**]. Positive emotions tend to encourage optimism and a readiness to take risks. Conversely, negative emotions tend to promote cautiousness and risk aversion. In the pursuit of effective HRI, robots strategically align their decisions with users' emotions and preferences. This alignment ensures optimal outcomes that resonate emotionally with users, enhancing meaningful and empathetic interactions within group settings. The robot's ability to integrate emotional understanding with rational evaluation leads to decisions that consider both aspects, contributing to more nuanced and profound interactions [**59**].

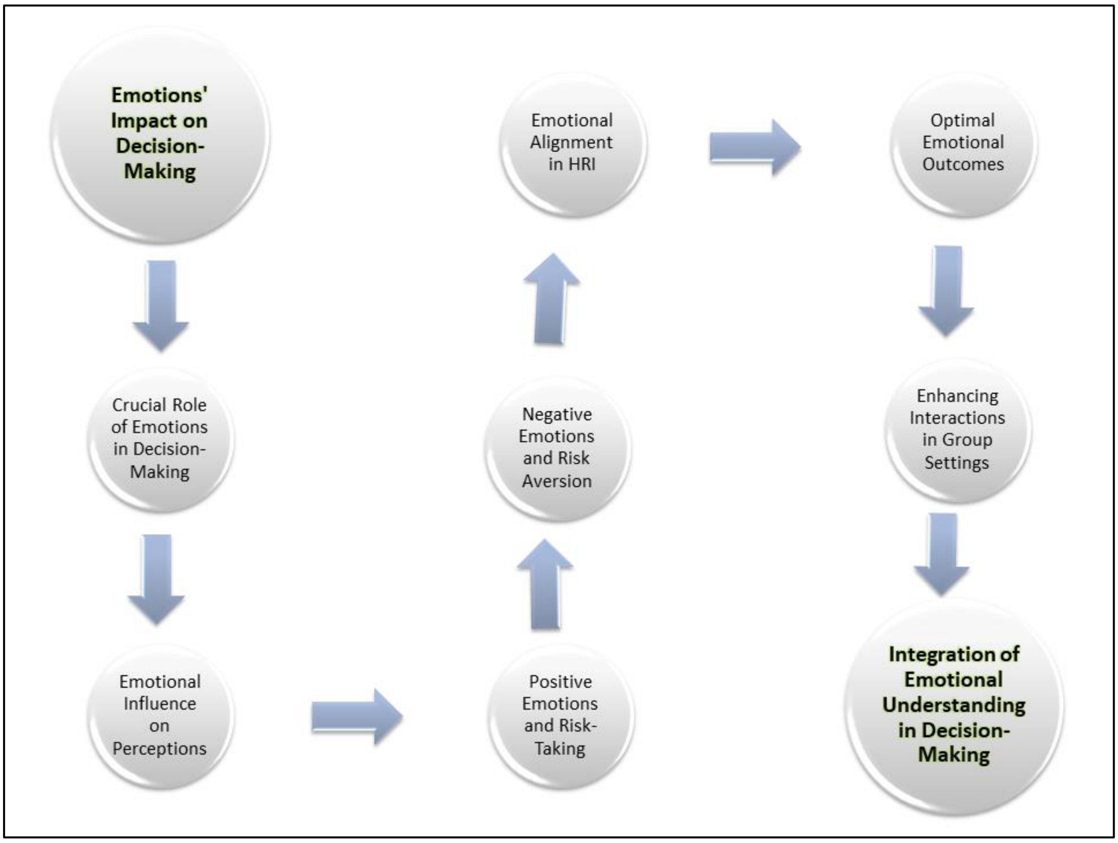


Figure 2-5 Emotional Influence on Decision-Making [**59**]

2.5.2 Modifying Behaviors and Responses

Emotion-informed decision-making empowers robots to modify their behaviors and responses to match the emotional context of the interaction. Robots can adjust their tone of voice, language, gestures, and overall demeanor based on the emotional states of the individuals they are interacting with. For instance, when detecting signs of happiness or excitement, a robot may respond with enthusiasm and encouragement, contributing to a positive and enjoyable experience. On another hand, when detect signs of sadness or distress, a robot may adopt a more comforting and supportive demeanor to offer solace. This ability to adapt emotionally allows robots to establish a stronger emotional bond with human users, fostering trust and understanding [**60**].

2.6 Measure the Emotion Diversity through Entropy

Following subsections, explore the concept of entropy and its groundbreaking application in understanding and quantifying emotions within group dynamics. Entropy, derived from information theory, serves as a powerful metric to measure the diversity and uncertainty within probability distributions, offering a numerical representation of data disorder and randomness [**61**]. In the context of group emotions, entropy becomes a valuable tool for measuring the richness and variability of emotional states showed by individuals. This approach joins entropy as a means to GER and its implications for interaction design, providing insights into the complicated dynamics of emotions within groups and enhancing decision-making and engagement HRI [**49**].

2.6.1 Entropy Calculations

Entropy, a concept borrowed from information theory, emerges as a potent metric for quantifying the diversity and uncertainty within a probability distribution. In essence, entropy provides a numerical representation of the level of disorder or randomness present in a set of data. In the context of emotions within a group, entropy serves as a powerful tool to gauge the richness and variability of emotional states exhibited by individuals. Entropy can be calculated using the following formula [**54**]:

$$\boldsymbol{H(X) = -\sum_{i=1}^{n} p(x_i) log_2(p(x_i))} \qquad \textbf{(2-5)}$$

Where:

- $H(X)$ represents the entropy of the data distribution.
- n is the number of different emotional states in the group.
- $P(xi)$ is the probability of occurrence of emotional state xi.

The formula captures the concept that entropy measures the uncertainty or randomness in a set of data. In the context of emotions within a group, H(X) calculates the entropy based on the probabilities of occurrence of each emotional state x_i . If all emotional states are equally likely, the entropy will be high, indicating a diverse and unpredictable emotional landscape. On the other hand, if one emotional state dominates, the entropy will be low, suggesting a more predictable emotional scenario [**62**]. The Figure 2-6 illustrates the process of calculating entropy using a given formula. It begins with the user providing emotion probabilities for each category to the system. The system then initiates the entropy calculation through the "EntropyCalculator" component. The calculator sequentially calculates the surprise value for each event based on probability, applies the formula involving negative logarithm and multiplication with probabilities, and sums up the results. Finally, the calculated entropy is returned to the user. This diagram visually represents the sequential steps involved in quantifying the diversity and uncertainty within a probability distribution, capturing the essence of entropy calculation and its significance in evaluating emotional states within a group [**53**].

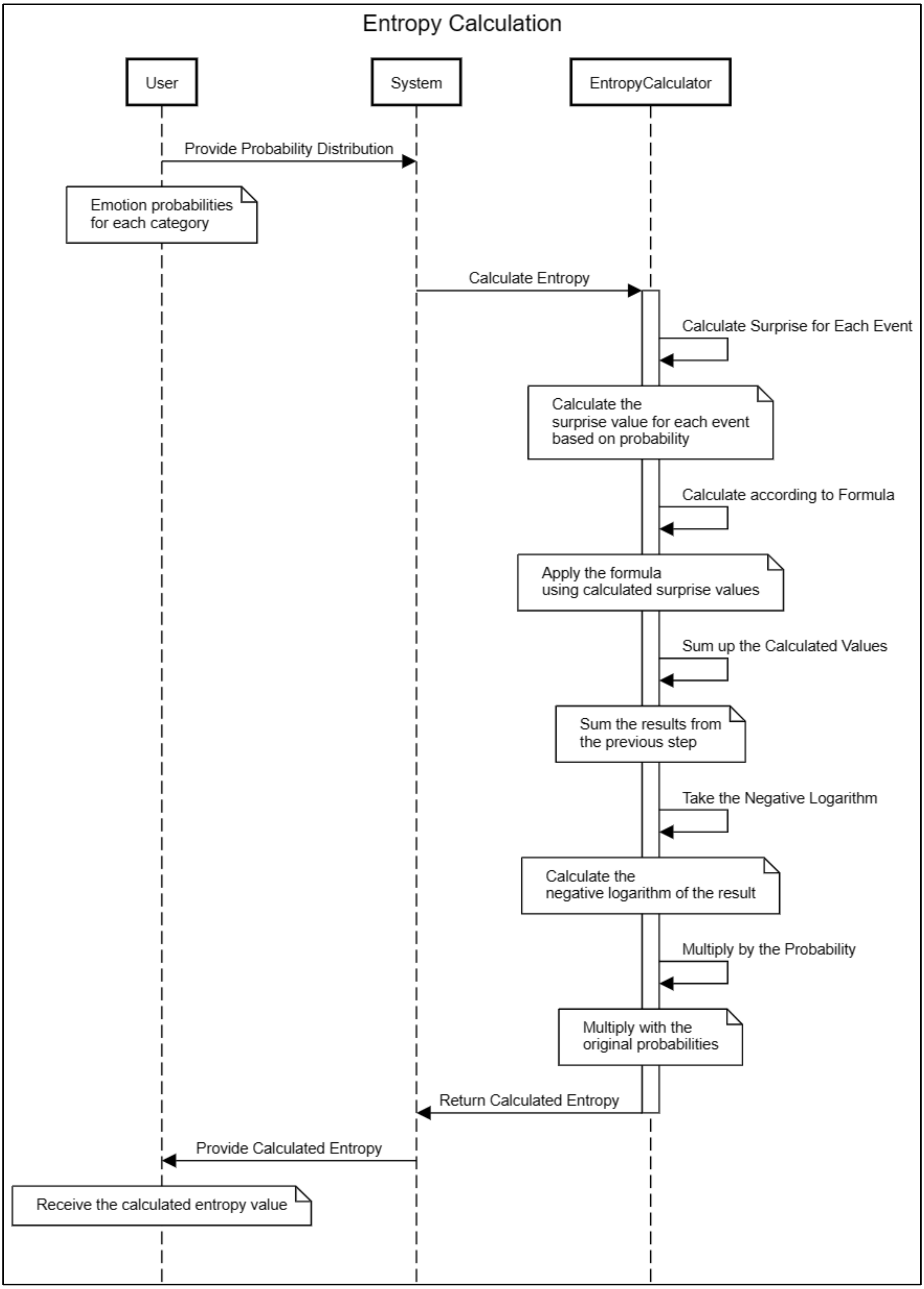


Figure 2-6 (the process of calculating entropy)

2.6.2 Comprehensive Entropy-Based Analysis for Improved GER and Interaction Design

This innovative approach harnesses entropy as a potent tool to GER and its implications for interaction design. By applying entropy-based analysis, this thesis explore into the intricate dynamics of emotions within groups, transcending traditional recognition methods [**63**]. This analysis quantifies the diversity of emotional states among group members, uncovering hidden patterns and trends that conventional approaches might overloo k. As a result, the comprehension of collective emotions deepens, revealing the coexistence of multiple emotional states and the overall emotional climate. Moreover, the insights gained through entropy-based analysis provide profound benefits for decision-making and interaction design in HRI. This holistic understanding of emotional dispersion empowers robots to tailor their behaviors appropriately, ensuring inclusivity and engagement across a range of emotional landscapes [**64**].

2.7 Combining Entropy-Based Analysis with Face Size

In order to enhance HRI through GER, a combined multiple dimensions of emotional indications has been used. This complete approach not only refines the determination of group emotions but also empowers robots with a heightened sensitivity to the emotional intricacies of human interactions [**65**]. By merging the insights derived from entropy-based analysis with the parameters of face size and proximity, that establish a robust foundation for precise emotion labeling and informed decision-making within the framework of HRI [**66**].

Figure 2-7 illustrates the integration of multiple dimensions in enhancing HRI through GER. The diagram depicts how entropy-based analysis provides insights into emotional diversity, which, when combined with face size assessment and proximity evaluation, forms a robust foundation for precise emotion labeling [**67**]. This comprehensive approach empowers robots with heightened sensitivity to emotional intricacies, refining the determination of group emotions. The combination of these insights feeds into informed decision-making within the framework of HRI, ultimately contributing to a more empathetic and context-aware interaction between humans and robots [**68**].

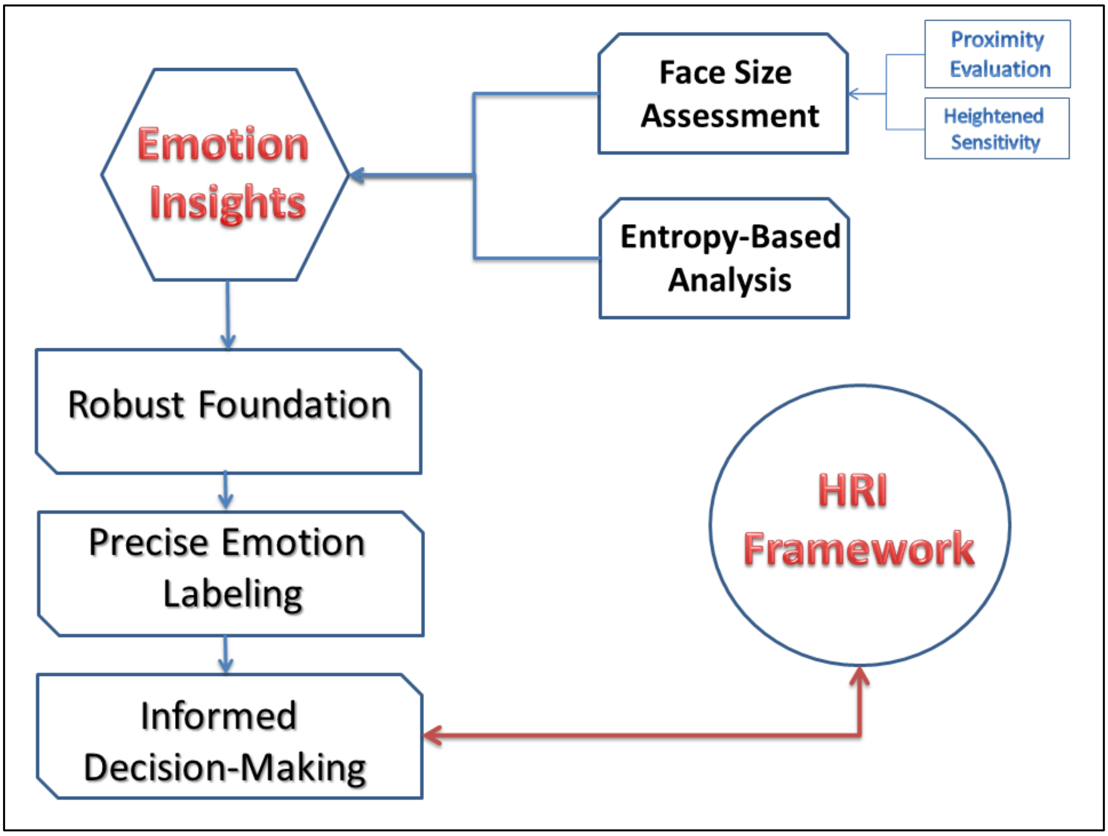


Figure 2-7 Integration of emotional cues for enhanced HRI

2.8 Enhancing HRI through Emotion-Informed Chatbot

Within the context of HRI, the new approach to GER forms the foundation for improved user engagement, tailored interactions, and improved overall user experiences. Drawing from a deep understanding of group emotions, introduce an emotion-informed decision-making process seamlessly integrated with an advanced chatbot. This fusion harmonizes emotion recognition, context-sensitive decision-making, and natural language processing, redefining the landscape of HRI within social environments [**69**]. After entropy, largest face, and nearest face calculations, the group's emotion label is classified into either "positive" (happy, surprise) or "negative" (angry, disgust, fear, sad) emotions. This classification lays the groundwork for subsequent decision-making. The derived emotion label becomes a critical trigger for decisions. In scenarios with a "negative" emotion label, the robot engages in a locomotion sequence to move away, providing space to prevent potential escalation of negative emotions. Conversely, when the emotion label is "positive," the robot initiates a brief and friendly greeting conversation, fostering positive interactions. Figure 2-8 illustrates the process of integrating emotions into a HRI scenario. Three users, User1, User2, and User3, express their emotions to the Social Robot. The Social Robot interacts with the System by first requesting face detection and then emotion analysis. The System triggers the Entropy Calculator to calculate entropy and face size metrics based on the recognized emotions. The Emotion Classifier then classifies the emotion label, which is sent back to the Social Robot [**58**]. Depending on the emotion label, an "Alternative" path is followed. If the emotion label is "Negative," the Social Robot initiates a "Move Away Sequence." On the other hand, if the emotion label is "Positive," the Social Robot engages in a friendly greeting conversation with the Chatbot. The diagram demonstrates the dynamic interaction between users, the Social Robot, and the various components of the system, highlighting the adaptation of robot behavior based on detected emotions [**68**].

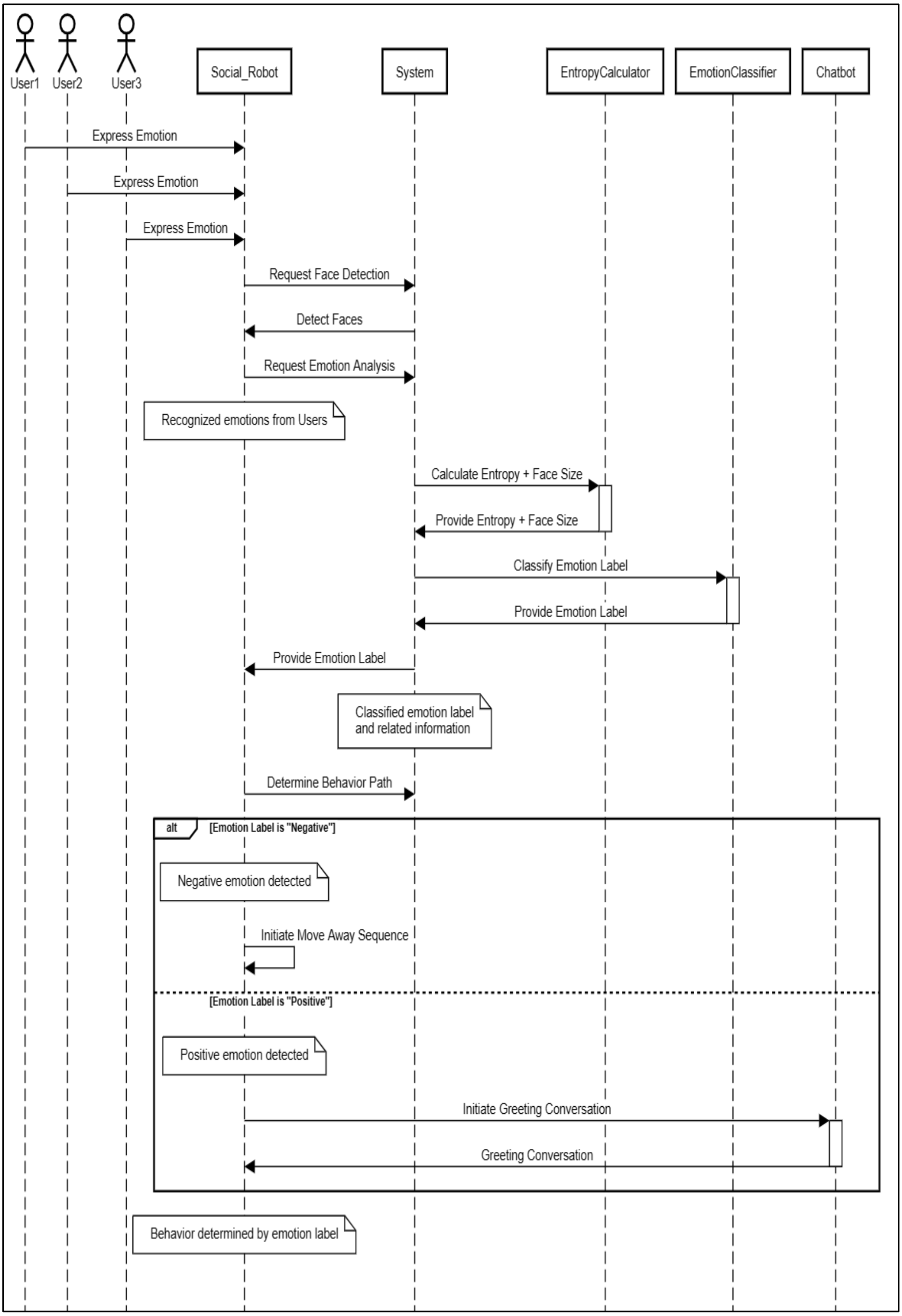


Figure 2-8 integrating emotions into HRI scenario.

2.9 Related Works

In the field of human-robot interaction, a multitude of research efforts have been devoted to advancing the field of facial expression recognition (FER).

As in Table 2-1 the beginning in 2008, where Matthias Wimmer and colleagues introduced a prototype for FER utilizing a social robot model (B21r) equipped with 1/3in Sony cameras and an LCD panel displaying a robot face. Their approach focused on recognizing six universal human facial expressions (happiness, anger, disgust, sadness, fear, and surprise) using Support Vector Machines (SVM) [**70**].

Subsequently, in 2017, Suhaila N. Mohammed and Loay E. George introduced an approach based on texture analysis of facial expressions to extract the face region determination and recognize seven basic emotional states and achieving an accuracy of 90.05% [**71**].

Shih-Chung Hsu et al, in 2017 proposed a hybrid FER system employing Gabor features and SVM classifiers to recognize emotions including anger, disgust, fear, happiness, sadness, and surprise. Their dual-layer hierarchical SVM system achieved an impressive 95% accuracy [**72**].

In 2018, Zhiqin Zhang delved into deep face emotion recognition, utilizing DenseNet-like structures to identify emotions like anger, disgust, fear, happy, neutral, sad, and surprise. The study demonstrated an accuracy of 90.2% for training and 86% for validation [**73**].

In 2019, and team introduced a cGAN-based approach for facial expression recognition, utilizing the AffectNet and RAF-DB datasets and achieving an average recognition accuracy of 81.83% [**74**].

Concurrently, Alia K. Hassan and Suhaila N. Mohammed introduced a facial emotion recognition scheme based on graph mining, achieving an accuracy of 90.00% [**75**].

Further contributions emerged in 2019 with Alireza Esfandbod et al. developing a facial expression imitation system on the RASA robot, achieving an accuracy of 90.10% [**76**].

In 2020, Mikel Val-Calvo and colleagues explored real-time emotion estimation using facial expressions and physiological signals, achieving an accuracy of 72.47% for facial expression recognition [**77**].

Daniel Octavian Melinte and Luige Vladareanu extended the field in 2020 by introducing a facial expression recognition system based on deep convolutional neural networks (CNNs), with accuracies exceeding 97% [**78**].

Silvia and their team explored facial expression evaluation with a social robot in 2020, achieving an accuracy of 43% using a convolutional neural network (CNN) [**79**].

In 2021, Chiara et al. improved human-robot interaction by enhancing NAO robot awareness of human facial expressions, achieving varying accuracy levels for different expressions [**7**].

Also in 2021, Jianmin Wang and colleagues conducted an experimental study on abstract expression in human-robot emotional communication, emphasizing the efficiency and recognition accuracy improvements of such design [**80**].

Advances persisted in 2022 as Marco Quiroz et al. developed a group emotion detection system using a Pepper robot, achieving an accuracy of 90.84% [**81**].

Finally, in 2022, Grazia D’Onofrio and their team introduced the EMOTIVE Project, integrating CNN-based FER models into the Pepper robot platform and achieving accuracy levels of up to 98% [**82**].

Table 2-1: Summary of Related Works

Research	Recognition	Emotions	Algorithm	Methodology	Dataset	Metrics Accuracy	Results
Wimmer et al. (2008)	FER	Six universal facial expressions	SVM	Image/Video -> Face detection -> Feature extraction -> Classification	CKDB, real-world scenes	Confusion matrix 70% (CKDB), 67% (robot)	Prototype suitable for FER by robots
Mohammed and George (2017)	Face region extraction	Seven basic emotions	Haar wavelets transform	Preprocessing -> ROI extraction -> Feature extraction -> Classification	JAFFE dataset	Optimal ANN configuration 90.05%	Neural network's effect on emotion recognition
Hsu et al. (2017)	Facial expressions	Anger, Disgust, Fear, Happy, Sadness, Surprise	Gabor features, SVM classifiers	Gabor extraction -> SVM training -> AU-based random forest training	Extended CK+ Dataset	The Accuracy about 95%	Effective facial expression recognition
Zhang (2018)	Facial Expression	Seven emotions	DenseNet-like structure	Feature fusion -> CNN	SFEW, real-time Wild Challenge	Accuracy 90.2% (training), 86% (validation)	Beyond RGB input for CNNs
Deng et al. (2019)	FER	Seven facial expressions	cGAN	Detect, crop, normalize -> Conditional GAN	AffectNet, RAF-DB datasets	Features worth investigating 81.83%	Simultaneous generative and discriminative representations
Hassan and Mohammed	Facial emotion recognition	Happy, Surprise, Sad, Fear,	gSpan algorithm	Graph theory, gSpan algorithm	SAVEE database	Binary classification 90.00%	Frequent sub-graphs in

(2019)	based on graph mining	Disgust, Angry					emotional classes
Esfandbod et al. (2019)	Facial expression imitation system	Five artificial expressions	MicroExpNet CNN	FER algorithm on robotic platform -> Imitation through display screen	CK+ facial expression database	Evaluation by participants 90.10%	Real-time facial expression imitation
MIKEL VAL-CALVO et al. (2020)	FER and Physiological signals Preprocessing	Seven emotions	CNN, SVM, KNN, etc.	Analyzing physiological responses -> Real-time emotion estimation	FER-2013, RAVDESS, real-time video	Macro-average F1-score 72.47% (FER), >80% (Preprocessing)	Causal emotion generation mechanism
Melinte and Vladareanu (2020)	Face recognition (FR) and Facial expression recognition (FER)	Seven emotions	Deep CNN	Object detection (SSD, RPN) -> CNN models (VGG, ResNet50, Inc)	COCO, Google OpenImage database	PASCAL VOCs evaluation metrics 97.8% (Faster R-CNN), 97.42% (SSD Inception)	Improved accuracy using serialized CNN
Ramis et al. (2020)	Facial expression recognition system	Seven emotions	CNN	Robot interaction simulation with user	CK+, BU4DFE, JAFFE, etc.	Accuracy, TP ratio 43%	Social robot for evaluating facial expressions
Filippini et al. (2021)	Facial expressions	Five emotions	CNN-based FER model	CNN integration into NAO -> Real-world human-robot interaction	Real-life situations, FER2013	Precision, recall, F1 score Varied for each	Enhancing human-robot interaction

						emotion	
Wang et al. (2021)	Facial expressions	PAD model	OCC model	Avatars' facial expression design using PAD model	Virtual image	Non-verbal self-assessment manikin N/A	Improved recognition through abstract expression
Quiroz et al. (2022)	Group Emotion Recognition	Happy, Surprise, Sad, Fear, Disgust, Angry	VGG neural network	Feature extraction -> Frame motion determination	Simulated environment	ROS/Gazebo tests 90.84%	Efficient system applicable in robotics
D'Onofrio et al. (2022)	Emotion Recognizing by a Robotic Solution Initiative	Three types of attitude	KNN, RF	Video capture -> Feature extraction -> Classification	IAPS database	Accuracy, precision, recall, etc. RF (98%), KNN (86%)	Enhancing robot's emotion recognition

CHAPTER THREE

PROPOSED METHOLODGY

3.1 Overview

The chapter focuses on practical implementation, face detection using Haar Cascade and HOG Descriptor, and feature extraction with Dlib. Integration of MTCNN enhances the accuracy of face detection. The Attention CNN model, developed using TensorFlow and Keras, enables accurate facial expression classification. A comprehensive video dataset is prepared for GER, including face localization, emotion labeling, entropy calculation, and classification. Decision-making based on GER labels facilitates adaptive HRI. Evaluation metrics like confusion matrix, precision, recall, and F1-score highlight the model's effectiveness in emotion differentiation, emphasizing its potential in HRI and group emotion recognition.

3.2 Face Detection Framework

In the field of face detection, a robust and effective strategy often involves the sequential utilization of both the Haar Cascade and the HOG descriptor. Each of these classifiers contributes distinct advantages to the process, creating a more reliable and accurate face detection system.

3.2.1 Framework Strategy Description

The rationale behind employing the Haar Cascade classifier as the initial step in face detection lies in its efficiency in swiftly identifying potential regions of interest within an image. However, this efficiency can sometimes come at the cost of a higher false positive rate. Haar Cascade's primary role is to serve as a quick filter, rapidly flagging areas in the image that might contain faces. While it may yield some false positives, it significantly reduces the computational load by pinpointing regions that warrant further scrutiny. On the other hand, the HOG Descriptor is typically used as a confirmation step. Although HOG is computationally more intensive, it excels in providing a more detailed and accurate analysis of regions identified by the Haar Cascade. By applying the HOG Descriptor to the potential face regions highlighted by Haar Cascade, false positives can be reduced, and the detection process gains a higher level of confidence. This two-tiered approach, with Haar Cascade as the initial filter and HOG as the confirmation mechanism, strikes a balance between speed and accuracy in face detection.

3.2.2 Framework Implementation Details

The framework implementation steps showed in Figure 3-1that outlines the process of combining Haar Cascade classifier and HOG descriptor for face detection. The user begins by loading an input image, followed by loading both the Haar Cascade and HOG Descriptor. To prepare the image for Haar Cascade and for more processing facility the image color is converted to grayscale. Haar Cascade is then applied to detect potential face regions in the grayscale image. For each face detected by Haar Cascade, a region of interest (ROI) is cropped from the original image and converted to grayscale. HOG is applied within these ROIs to further detect faces, and rectangles are drawn around the detected faces. The final step involves displaying the image with rectangles, showing the detected faces. This sequence illustrated how Haar Cascade acts as an initial filter, and HOG confirms face detection within the potential regions of interest, resulting in accurate face detection and visualization.

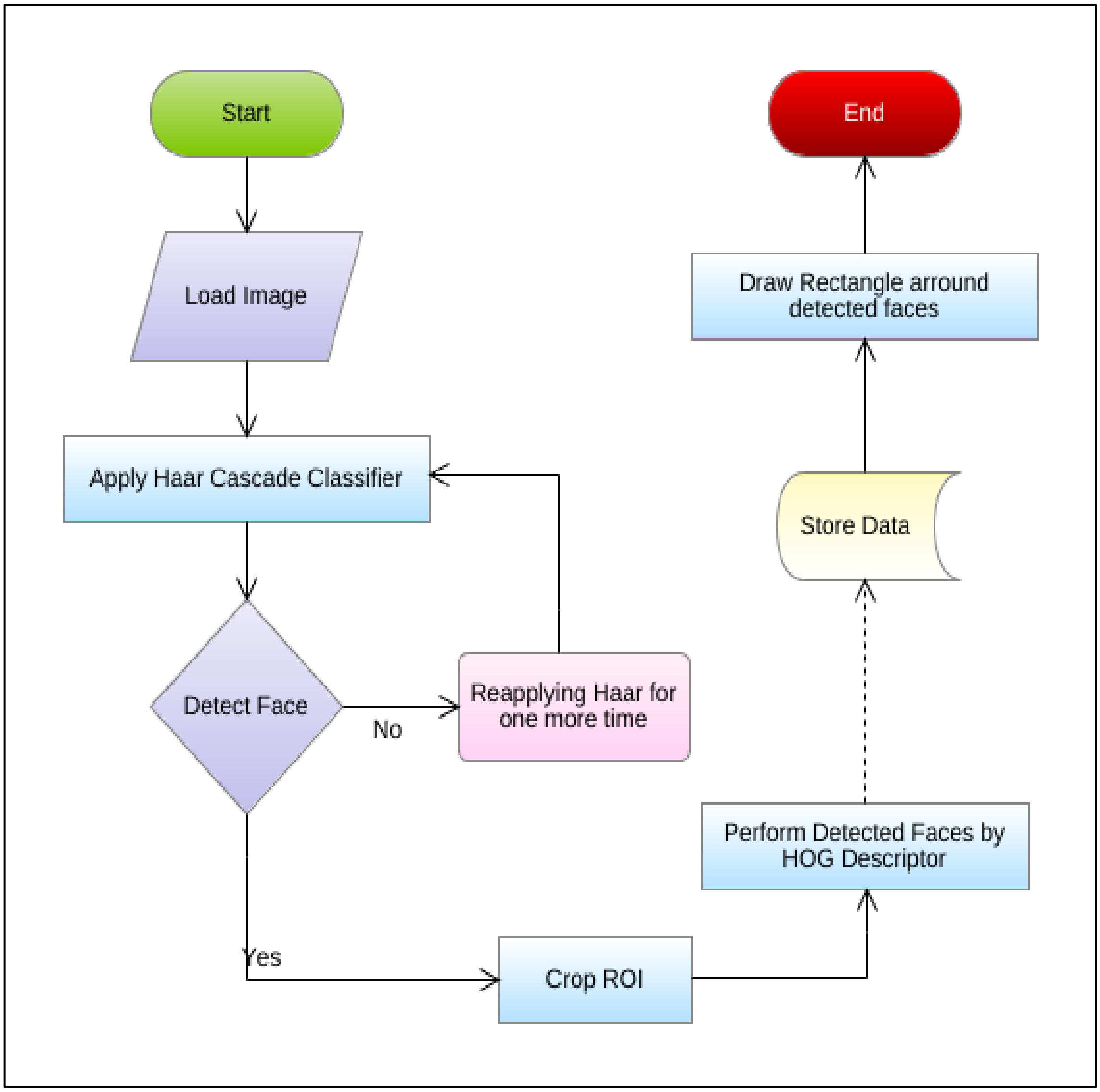


Figure 3-1 Step-by-step process of combining Haar and HOG for face detection.

3.3 Feature Extraction and Selection

The extraction and selection of facial features play a pivotal role in the study, as they are fundamental for both face detection and emotion recognition. In this section, need to explore into the significance of facial landmarks, the use of Dlib, the facial landmark detection process, and the approach to feature selection.

3.3.1 Facial Landmarks

Facial landmarks are critical in the study due to their capacity to provide detailed spatial information about facial features as in Figure 3-2. These landmarks serve as key reference points on the face, such as the corners of the eyes, the tip of the nose, and the edges of the mouth. The significance of facial landmarks lies in their ability to capture the structural nuances of facial expressions, which are vital for both face detection and emotion recognition.The methodology for extracting these landmarks involves leveraging Dlib's pre-trained facial landmark detector, specifically the "shape_predictor_68_face_landmarks.dat" model. This detector is adept at accurately locating and identifying 68 key facial landmarks in input images. However, to streamline the process and reduce computational complexity, choosing to work with a subset of 37 landmarks those are most relevant to the study.

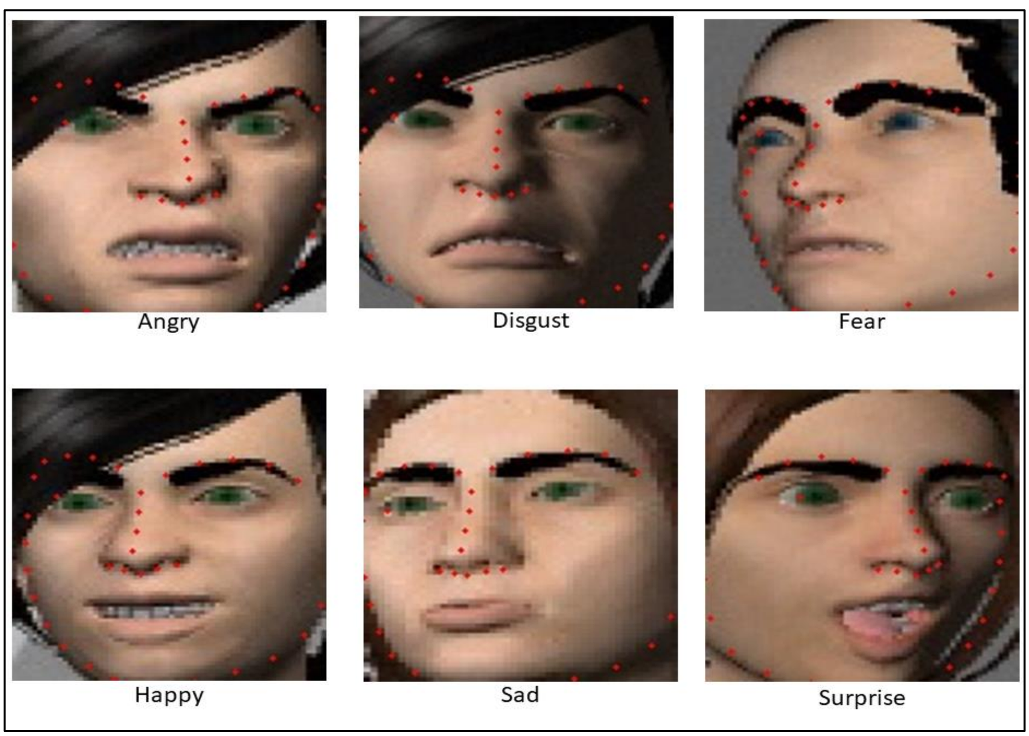


Figure 3-2 Samples of facial landmarks

3.3.2 Dlib: Enhancing Facial Feature Extraction

Dlib, or the Davis E. King Library, is an instrumental tool in the feature extraction pipeline. It offers robust solutions for facial feature extraction and facial landmark detection. Dlib's capabilities empower us to accurately and efficiently locate facial landmarks in the input images. Thus for the “Facial Landmark Detection” the feature extraction process begins with the utilization of Dlib's pre-trained facial landmark detector. This detector is designed to locate facial landmarks with remarkable precision. Then apply this detector to input images, enabling us to pinpoint the positions of these critical facial reference points. Furthermore the “Feature Selection” is a strategic step in the study, primarily aimed at dimensionality reduction. By working with a subset of the 68 detected landmarks, specifically the most relevant 37 features, several advantages achieved. Firstly, dimensionality reduction helps optimize the Attention CNN model, making it more efficient and less computationally intensive. This is particularly beneficial for real-time applications where computational resources may be limited. Secondly, the chosen 37 landmarks encompass the essential structural information required for both face detection and emotion recognition. They capture the key facial characteristics that convey emotional expressions, allowing us to focus the model on the most discriminative features. The following pseudo-code outlines the essential steps of a Python program for facial landmark detection and visualization. It first initializes a face detector using the Dlib library and loads a pre-trained model for facial landmark prediction. Then, it iterates through a folder containing input images; for each image, it detects faces using the face detector and, for each detected face, identifies 37 specific facial landmarks using the predictor model. These landmarks are represented as red circles drawn on the image. Finally, it displays the image with the marked facial landmarks as shown in (Algorithm 3-1) below.

Algorithm 3-1 : Detect Facial Landmarks in Images

Input: Path to input_images folder, Pre-trained face detector, Pre-trained facial landmark predictor model

Output: Display the image with Facial Landmarks

Begin

-Initialize face_detector using provided Dlib model

-Load pre-trained facial landmark predictor model

For each image in input_images folder:

Load the image

-Detect faces in the image using face_detector

For each detected face:

-Detect 37 facial landmarks using predictor model

For each landmark:

Get x and y coordinates of the landmark

Draw a red circle at (x, y) on the image

-Output

-Close all image display windows

End

3.4 Attention Mechanism

In order for more attention on the important features an integration of MTCNN employed within the facial analysis pipeline, after initial face detection with Haar Cascade and HOG confirmation, and subsequent feature extraction with Dlib, resulted in significant improvements in system performance. MTCNN, known for its accuracy and robustness, enhanced face detection accuracy by identifying faces of various sizes, orientations, and poses, and provided crucial facial feature localization. Its confirmation step reduced false positives, ensuring high detection precision, and enhanced system robustness against challenges like occlusions and lighting variations. Despite increased computational demands, MTCNN maintained reasonable processing times suitable for real-time applications. Its seamless integration with Dlib facilitated comprehensive spatial information gathering, enabling precise emotion analysis of facial expressions. The incorporation of pre-trained models expanded research and application possibilities across domains like human-computer interaction and sentiment analysis. The output of this pipeline served as the input dataset for training and evaluating the subsequent Attention CNN model, enhancing the overall effectiveness and accuracy of the facial analysis system.

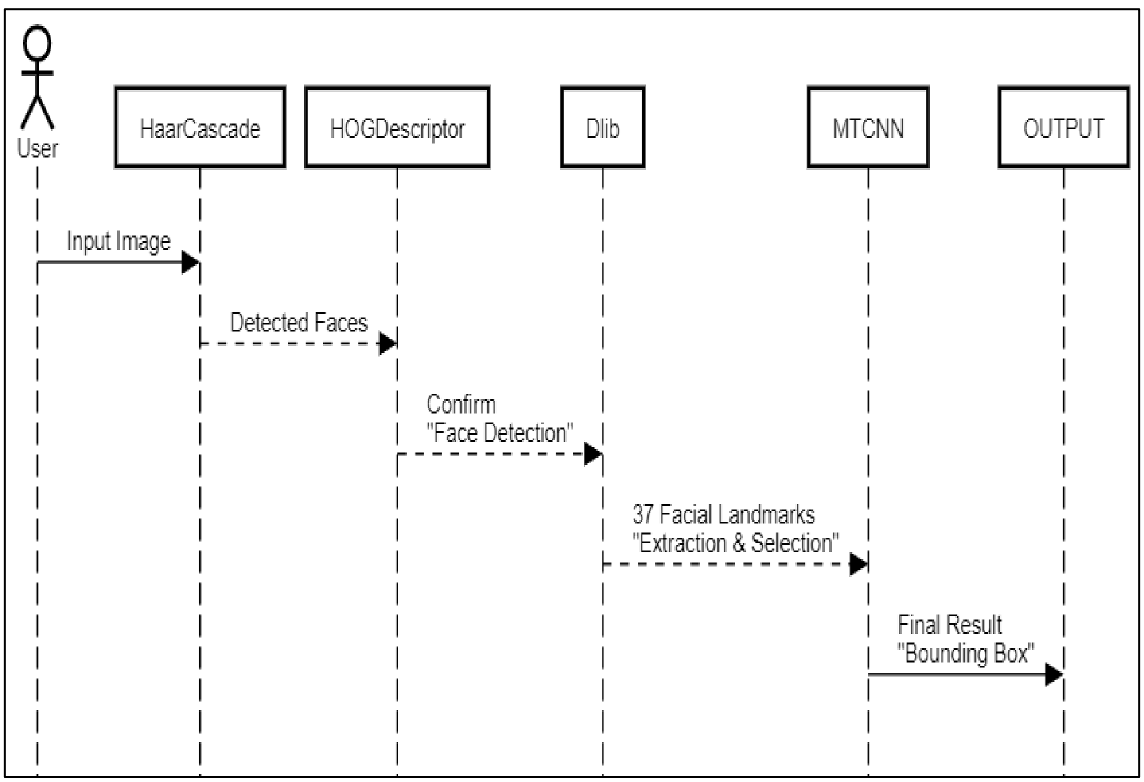


Figure 3-3 Pipeline of Face Detection

3.5 Attention CNN Model

Designing Attention CNN architecture, regarded as an essential component of the overall facial analysis system. This Attention CNN model is instrumental in achieving the wider objectives of HRI and GER. Attention CNN model's procedures begins with the well-prepared dataset, meticulously processed through a pipeline that involves the integration of various components, including Haar Cascade, HOG, Dlib, and MTCNN. This logically arranged pipeline extracts valuable features from facial images, which serve as the input dataset for Attention CNN model. By taking advantages of TensorFlow and Keras, Attention CNN architecture is formulated, designed, and trained to classify facial expressions into six distinct categories. Also there are many alternative ways to design Attention CNN model as shown in Figure 3-4.

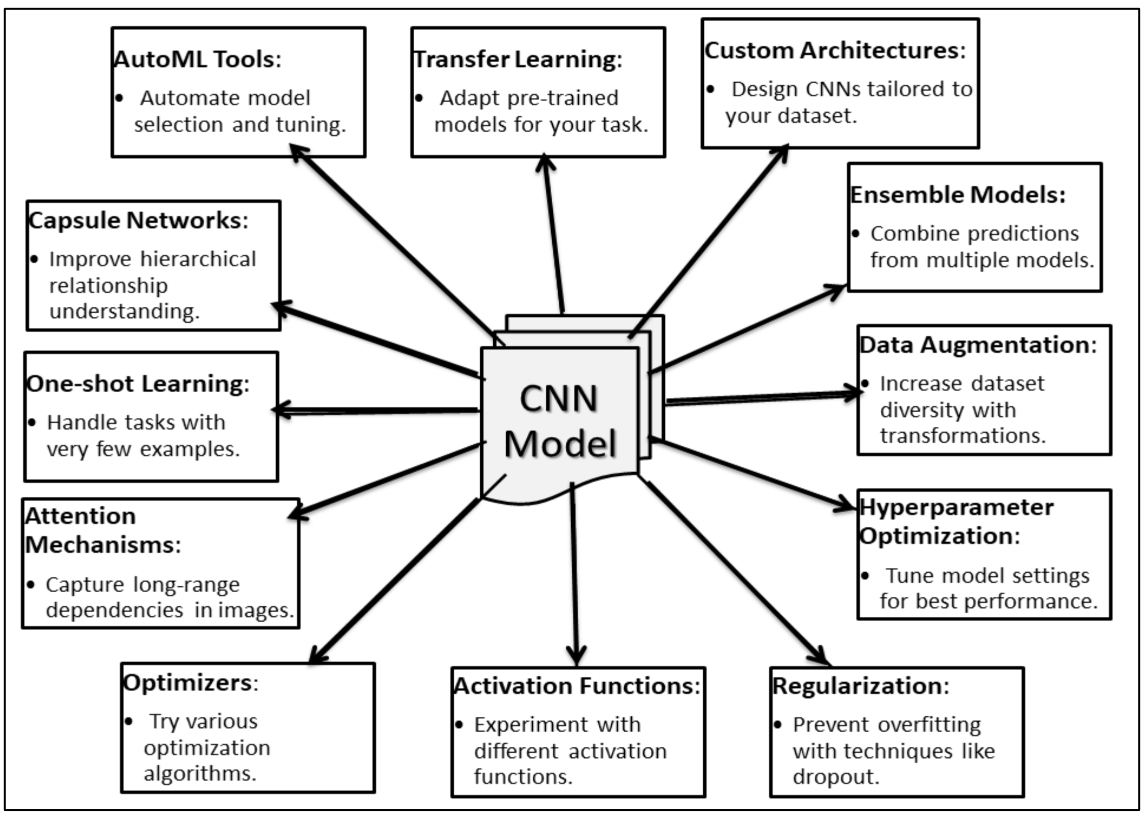


Figure 3-4 Alternative methods and approaches for creating Attention CNN models

3.5.1 Attention CNN Model Description

The Attention CNN model serves as the computational core of a facial analysis system as shown in Figure 3-5, adept at multi-class image classification, particularly in deciphering facial expressions across six distinct emotion categories. It operates by processing preprocessed facial images through a pipeline involving Haar Cascade, HOG, Dlib, and MTCNN. Developed using TensorFlow and Keras, it facilitates feature extraction and recognizes complex patterns within facial expressions. The model's versatility extends to interpreting diverse emotions, enabling Human-Robot Interaction systems to respond to human emotions accurately and sensitively.

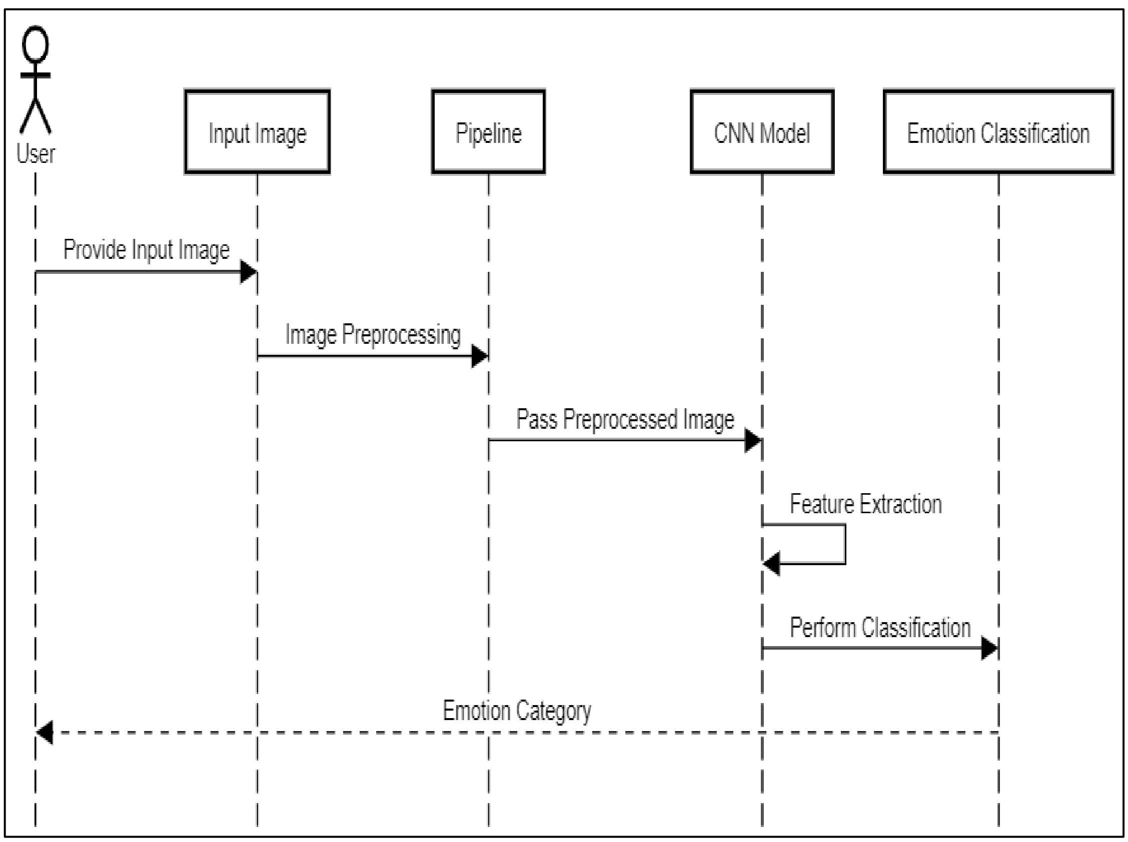


Figure 3-5 Attention CNN Description

3.5.2 Attention CNN Model Architecture

The Attention CNN model's architecture is a meticulously constructed group of layers designed to perform multi-class image classification effectively. At its core, this architecture excels in analyzing facial expressions and discerning emotions. As shown in Figure 3-8 each layer plays a distinct role in the feature extraction and classification process:

- **Input Layer**: procedures starts with the input layer, accommodating images of 150x150 pixels with three color channels (RGB).
- **Convolutional Layers**: The Attention CNN architecture comprises three convolutional layers, each with an increasing number of filters and a 3x3 kernel size. These layers serve as feature extractors, detecting patterns and intricate details within the images. ReLU activation functions introduce non-linearity, enhancing the model's ability to capture complex relationships.
- **Max-Pooling Layers**: After each convolutional layer, a max-pooling layer follows, compression the spatial dimensions by a factor of 2x2. This process reduces computation and helps keep translation constant
- **Flatten Layer**: Following the convolutional and max-pooling layers, a flatten layer converts the 2D feature maps into a 1D vector. This transformation prepares the data for the fully connected layers.
- **Fully Connected Layers**: Two fully connected dense layers comprise the final stretch of the architecture. The first dense layer contains 512 units, employing ReLU activation to process the flattened features. A dropout layer with a rate of 0.5 is strategically inserted to mitigate overfitting. The second dense layer, the output layer, consists of six units, corresponding to the six emotion

categories in the dataset. A softmax activation function is applied to generate class probabilities as shown below in (Algorithm 3-2). Also, the Figure 3-6 shows the following Attention CNN model layers:

Algorithm 3-2 : Convolutional Neural Network (CNN) for Emotion Classification

Input: Resized input images (150x150 pixels, 3 color channels)

Output: Predicted emotion label

Begin

- Preprocessing sub-image to 150x150 pixels with 3 color channels (RGB)
- Apply 32 filters of size 3x3 for feature extraction
    - Use ReLU activation to detect features in the input images
- Apply max-pooling with a 2x2 pool size to reduce spatial dimensions
- Apply 64 filters of size 3x3 for more feature extraction
    - Use ReLU activation for feature enhancement
- Apply max-pooling with a 2x2 pool size for further dimension reduction
- Apply 128 filters of size 3x3 for complex feature extraction
    - Use ReLU activation for advanced feature representation
- Apply max-pooling with a 2x2 pool size for feature reduction
- Flatten the data into a 1D vector for input to fully connected layers
- Implement a fully connected layer with 512 neurons for feature combination
    - Use ReLU activation for enhanced feature mapping
- Apply dropout with a rate of 0.5 to prevent overfitting
- Create the final layer with 6 neurons for multi-class classification (6 emotions)
    - Use softmax activation for emotion prediction probabilities
- Predict the emotion label based on the highest probability in the output layer

End

This Attention CNN is compiled with the Adam optimizer and categorical cross-entropy loss function, which is appropriate for multiclass classification tasks. It designed for providing dataset for 10 epochs.

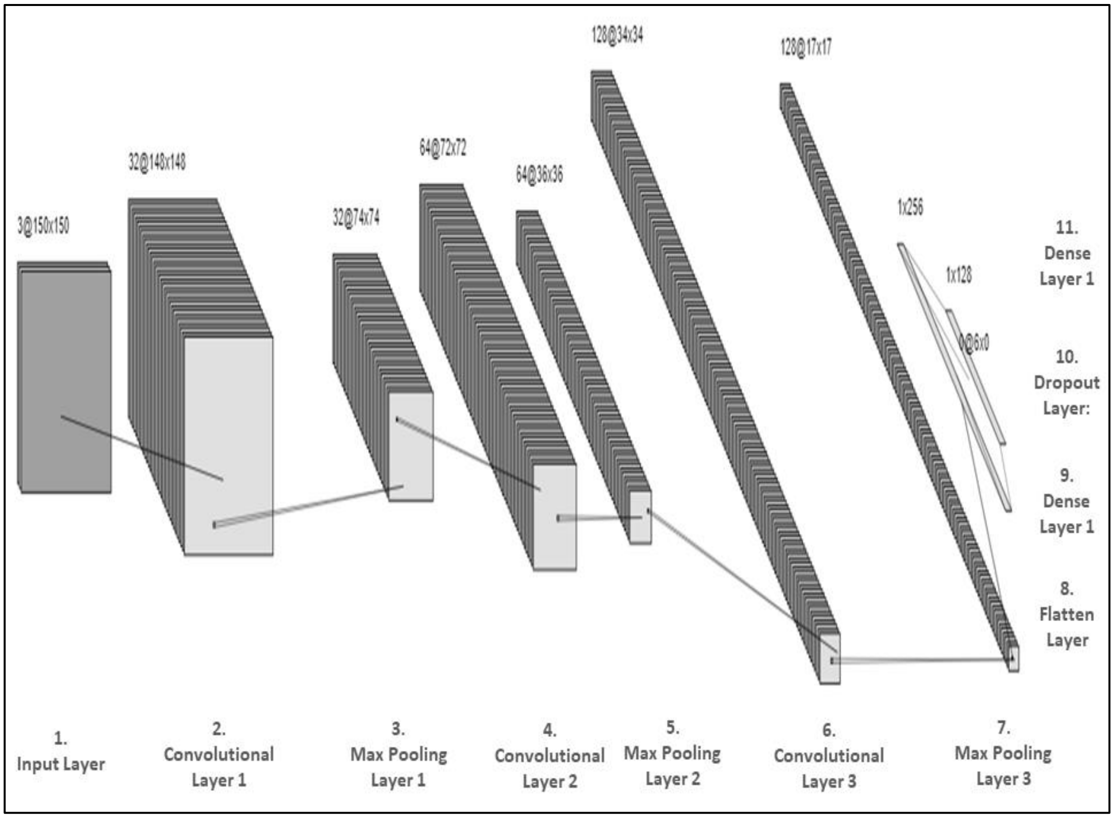


Figure 3-6 Attention CNN Model Architecture

3.6 Video Preparation: Extract Frames and Detect Multi-Face

This section exploring the essential process of preparing the video dataset for subsequent analysis, specifically aimed at GER. The dataset under consideration comprises two distinct scenarios, "Cafeteria" and "Museum," each consisting of 12 videos. The objective is to transform these videos into a format suitable for manipulation by Attention CNN model. To achieve this, a Python program has been exactly created to accomplish several tasks, including calculating the video duration, determining FPS, extracting frames from each video, and counting the total number of frames in each video. Subsequently, each extracted frame is saved as an individual image, with the file name encoding information about its source folder and the corresponding frame number.

3.6.1 Multi-Face Detection in a Frame

In the proceeding of multi-face detection within each frame, need for deploying a classy pipeline leveraging state-of-the-art techniques. The pipeline kicks off with initial face detection using the Haar Cascade classifier, known for its speed in identifying potential regions of interest. Followed by confirm using HOG Descriptor, which provides a detailed analysis, thereby reducing false positives. Then, proceed to feature extraction with Dlib, an integral component for facial landmark detection. Finally the multi-faces detection completed by using MTCNN, a powerful tool for precisely localizing and identifying multiple faces within a frame. The combination of these methodologies yields accurate and efficient multi-face detection, a fundamental requirement for the GER system.

3.6.2 Scene Determining

To simplify the processing and facilitate efficient analysis, need to introduce a concept of scene determination. After the frames have been exactly extracted from each video, now collect every ten repeated individual frames into a single scene. This grouping not only condenses the data but also aligns with the practical notion of scenes in video analysis, allowing for more meaningful analysis of emotions in context. In cases where the number of remaining frames is less than ten, practically treat them as a single scene, ensuring that no frames are left unaccounted for. This scene determination process is an essential step in preparing the dataset for subsequent Group Emotion Recognition, as it enables the model to capture emotional dynamics within meaningful segments of the videos.

3.7 Group Emotion Recognition (GER):

Analyzing emotional dynamics in Group Scenarios shape the mainstay of research to enable human-robot interactions and decision-making based on group emotions. To accomplish this, need to employ a systematic approach, starting with the creation of a repository designed to store critical information required for emotion classification. This information includes the frame file name, the number of detected faces in each frame, the emotions associated with each individual detected face, the probability distribution of emotions, entropy calculations, and information regarding the largest face, nearest face, group scenes, as Figure 3-7 shown.

3.7.1 Faces Localization: Calculating Proximity Metrics

The determination of GER initiates with the specific localization of faces within each frame, accompanied by essential calculations to estimate the significance of these faces. This involves determining the number of faces present in a frame and subsequently calculating metrics such as the largest face and nearest face. Accomplishing this using Euclidean distance calculation, which allow us to measure the spatial relationship between detected faces and the center of the group as shown in as shown previously in the section 2.3.3. These proximity metrics provide valuable insights into the spatial dynamics of the group, and they play an essential role in determining group emotions.

3.7.2 Labeling Faces: Recognizing Individual Emotions

The next stage in the GER procedures involves the application of the pre-trained Attention CNN model to recognize the emotions expressed by each individual detected face in every frame. This process yields emotion labels for each face, a critical component for subsequent analysis.

3.7.3 Entropy Calculation: Measuring Emotion Diversity

With the emotion labels determined for each face within a frame, need for proceeding to calculate the probability distribution of emotions. This distribution allows us to understand the diffusion of different emotions within the group. To measure the collective emotion of a group, need to introduce the concept of entropy, a fundamental concept from information theory as shown in section 2.3.3. The entropy value increases with greater diversity or uncertainty in the distribution of emotion labels, and decreases when there is a dominant or focused emotion showed by the group. Also based on the median value among resulted entropies the threshold metric is calculated to be used as a benchmark for compression in this specific dataset that used.

3.7.4 GER Classification:

The bellow (Algorithm 3-3) provides a structured approach to classify emotions within group scenarios based on facial expressions and spatial indications. It can be implemented in HRI system to make informed decisions and adapt to the emotional dynamics of the group:

Algorithm 3-3 : Group Emotion Recognition (GER)

Input: Detected faces in a frame, Pre-trained Attention CNN model

Output: Group Emotion Recognition (GER) score

Begin

- Utilize a pre-trained Attention CNN model for emotion recognition
  - For each detected face in a frame, obtain emotion labels:
    (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprised)
- Compute the Euclidean distance between each detected face and the center of the group based on bounding boxes $d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$
  - Calculate proximity metrics to understand spatial relationships
- Replace emotion labels for the largest and nearest faces in the frame
  - Each detected face is associated with an emotion label
- Compute the probability distribution P_e of emotion labels n_e within the group N

$$P_e = \frac{n_e}{N}$$

  - Use this distribution to calculate entropy, a measure of diversity in

emotions $H(X) = -\sum_{i=1}^{n} p(x_i) log_2(p(x_i))$

- Determine a suitable threshold for entropy based on the median value among the distributed values of entropy obtained from 1458 frames
- This threshold categorizes scenes based on frame entropies of detected face emotions
- Compare the computed entropy value with the predefined threshold
  - Determine if the frame exhibits clear and consistent emotions or diverse and uncertain emotions
- Calculate the GER score by combining proximity metrics and entropy $Mean_{GER} = \frac{\sum GER\ values\ for\ all\ frames}{Number\ of\ frames\ in\ the\ scene}$
  - Assign the corresponding emotion label to the largest and nearest face
    - Calculate the probability distribution for these emotion labels
      - Compute the entropy based on this distribution
- The GER score reflects the overall emotional dynamics of the group in the frame

End

3.7.5 Identifying GER_Label:

Final step involves identifying the GER_Label, which characterizes the overall emotional state of a scene consisting of multiple frames (typically 10 frames per scene). The process describes in (Algorithm 3-4):

Algorithm 3-4 : Calculate and Categorize GER for Scenes

Input: List of GER values for frames, Predefined threshold

Output: Categorized GER values for scenes (Positive or Negative)

Begin

- Utilize the algorithm mentioned earlier to calculate the GER for each frame, considering both weighted proximity and entropy
  - Append each calculated GER to the list of GER values for frames
- For each scene consisting of every 10 frames or fewer:
  - Calculate a summary statistic to assess the variability or dispersion of emotions within that scene
    - Keep in mind to gather the last remaining frames in the repository to form one scene, even if they are less than ten frames
      - Calculate the average GER value for the scene by summing all GER values for the frames within that scene and dividing by the number of frames in the scene
- Compare aggregated GER value for the scene with a predefined threshold value
  - If the aggregated GER <= 0.5645:
    - Categorize GER value for the scene as Positive
      - If the aggregated GER > 0.5645:
        - Categorize GER value for the scene as Negative

End

This structured approach provides an evaluation of emotional dynamics within scenes to improve calculations of GER during HRI. It provides robot with the capability to estimate the emotional state of groups and make context-aware decisions effectively as shown in Figure 3-7 bellow.

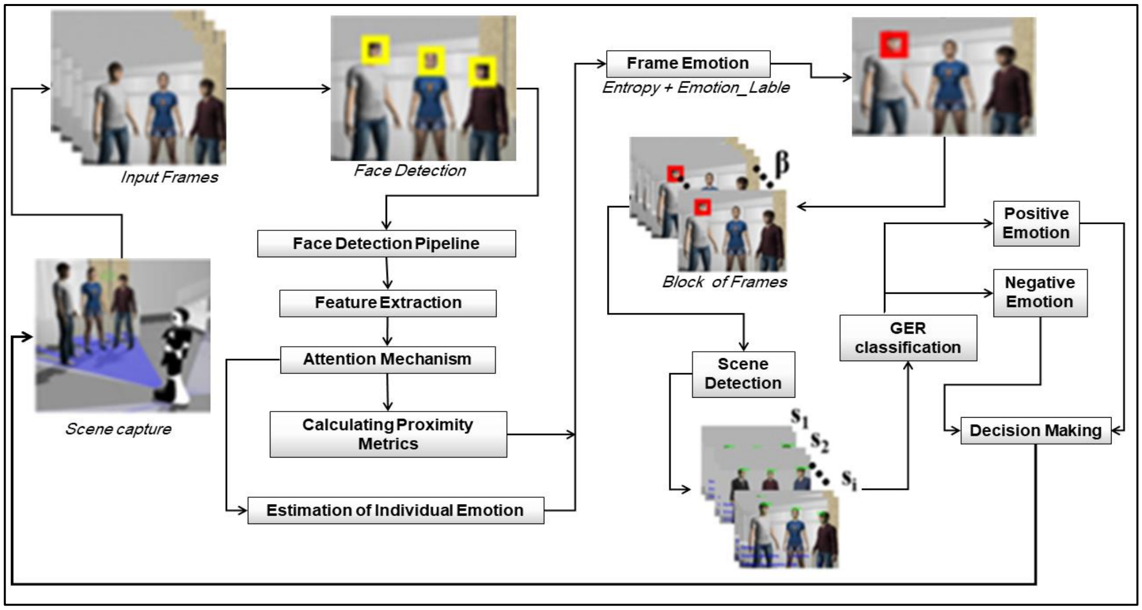


Figure 3-7 GER Estimation

3.8 Decision Making

Creating a decision-making procedure for a social robot based on the Positive or Negative labeling of GER is a critical step to ensure appropriate human-robot interactions. The decision-making process should be implemented with sensitivity and adaptability. Emotions can be complex, and the robot's responses should reflect an understanding of human emotions and social cues for various scenarios and user groups. Also, observe to ethical guidelines to ensure respectful and responsible interactions with humans. Thus, the robot firstly extracts GER information, which is processed by the GER component to determine emotion labels. These labels are then classified by a separate Classifier component. The GER component communicates the emotion classification back to the robot. Depending on whether the GER result is positive or not, as shown in Figure 3-8 the robot either initiates a chatbot interaction with the group or moves away. This diagram illustrates a basic flow of interactions in this context, recognizing that actual implementations may involve more complexity and conditions.

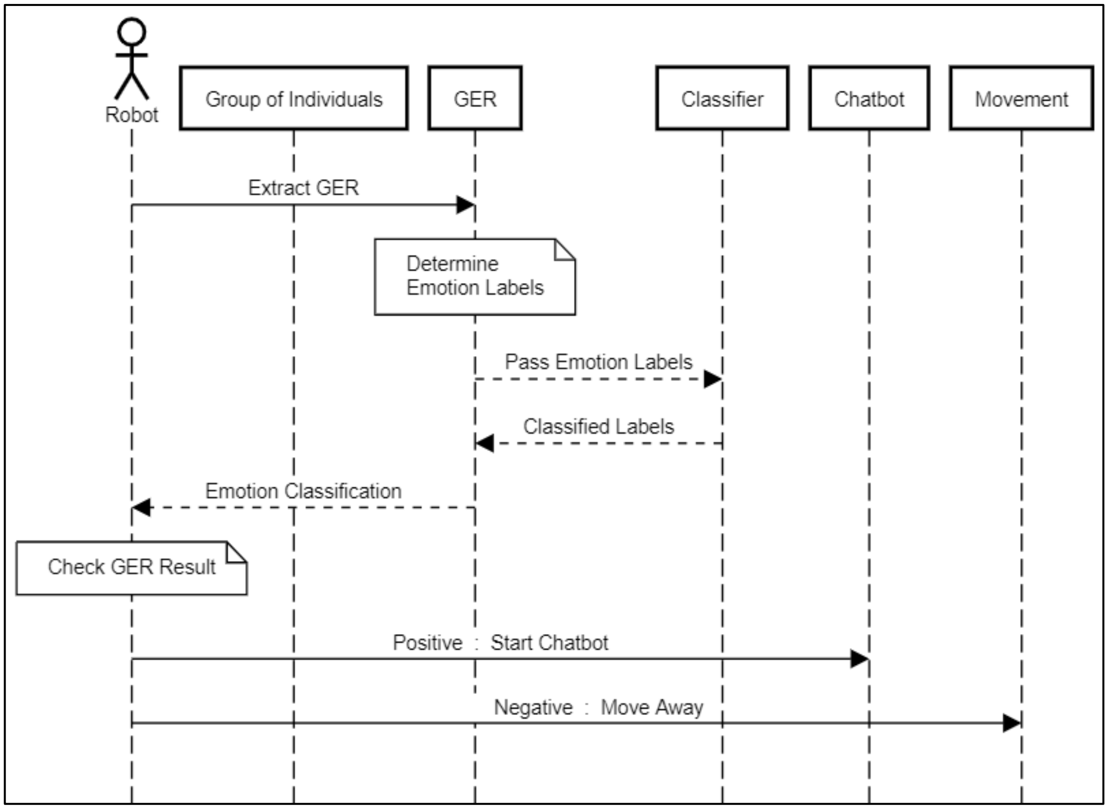


Figure 3-8 Robot Decision Making

3.8.1 Positive GER Label:

Initiating the Chat Box If the GER label for a scene is Positive, it indicates that the group's emotions are relatively clear and positive. The robot can initiate a chatbot or engage in a conversation with the group.

3.8.2 Negative GER Label:

Moving Away If the GER label for a scene is Negative, it suggests that the group's emotions are diverse or less clear, possibly indicating negative emotions. In this case, the robot should consider moving away to avoid potentially negative interactions.

CHAPTER FOUR

RESULTS AND DISCUSSION

4.1 Overview

This chapter, provide overview of GER's critical elements. Introduce Dataset and its relevance in HRI. Also, Present the Attention CNN model architecture and it's evaluation performance using various metrics. Finally, outline a decision-making framework for a social robot based on GER results.

4.2 Technical Specifications of the System:

➔ **Hardware:**

Computer: A personal computer with the following specifications used:

Processor (CPU): Equipped with a processor running at Intel(R) Core(TM) i5-2430M CPU @ 2.40GHz for high computational power.

Random Access Memory (RAM): The computer was configured with 8 gigabytes of RAM to enhance application performance.

Hard Drive: 1-terabyte hard drive was used for data and file storage.

Graphics Card: A 2 gigabyte shared graphics card was employed to support applications and games with high requirements.

➔ **Software:**

Operating System: The Windows 10 Pro operating system was installed to ensure system stability and application compatibility.

Programming Language: Python was used as the primary programming language for application development and algorithm implementation.

➔ **Platforms:**

Colab: https://colab.google.com/

Google Drive: https://drive.google.com

Diagram Drawing: https://sequencediagram.org/

Flowchart Drawing : https://app.genmymodel.com/personal/

CNN model Drawing: https://alexlenail.me/NN-SVG/LeNet.html & https://netron.app/

4.3 Dataset

The datasets utilized are essential resources containing images and videos with diverse emotional expressions. They are methodically prepared, standardized, and augmented to support Attention CNN model training and testing in the context of HRI and Group GER. The ROS/Gazebo Generation Dataset, created in a simulated environment, refill the lack of group emotion recognition resources in HRI, while the Real-world Affective Faces (RAF) Database provides a realistic collection of facial images from real-world contexts. These datasets are fundamental to exploring the role of emotion recognition in decision-making processes within HRI and GER scenarios.

4.3.1 The primary dataset

The primary dataset, known as the ROS/Gazebo Generated Dataset, plays a fundamental role in exploring emotions within HRI. It consists of constructed image and video datasets, of 23,222 images categorized into six classes of emotions (happy, sad, angry, surprise, disgust, fear) as shown in Table 4-1.

Table 4-1: ROS/Gazebo Generated Images Dataset

Emotions	**Happy**	**Sad**	**Angry**	**Disgust**	**Surprise**	**Fear**
Training Data	3371	3145	3363	2872	3179	2976
Test Data	750	750	750	566	750	750
Total	4121	3895	4113	3438	3929	3726

Remarkably, each separate emotion class contains an average of around 3,000 images. This dataset is mainly valuable due to its lack in catering to group emotion recognition and scene emotion detection in the context of HRI. Moreover, In the Videos dataset, and after the preprocessing actions two distinct sets are explored as shown in Table 4-2.

Table 4-2: ROS/Gazebo Generated Videos Dataset

Subfolder Group	**G1**	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**	**G9**	**G10**	**G11**	**G12**
Cafeteria	68	61	56	53	61	58	66	64	65	63	63	68
Museum	61	64	59	59	59	56	60	61	59	56	59	58

The Cafeteria subfolder comprises 12 videos, with varying frame counts for each group, summing up to a total of 746 frames. In contrast, the Museum subfolder also contains 12 videos, with a total of 711 frames across different groups. Videos dataset complement the image dataset, providing a comprehensive foundation for experimental validation. The two distinct simulated settings in ROS/Gazebo museum and a cafeteria are the platforms for experimentation as shown in Figure 4-1.

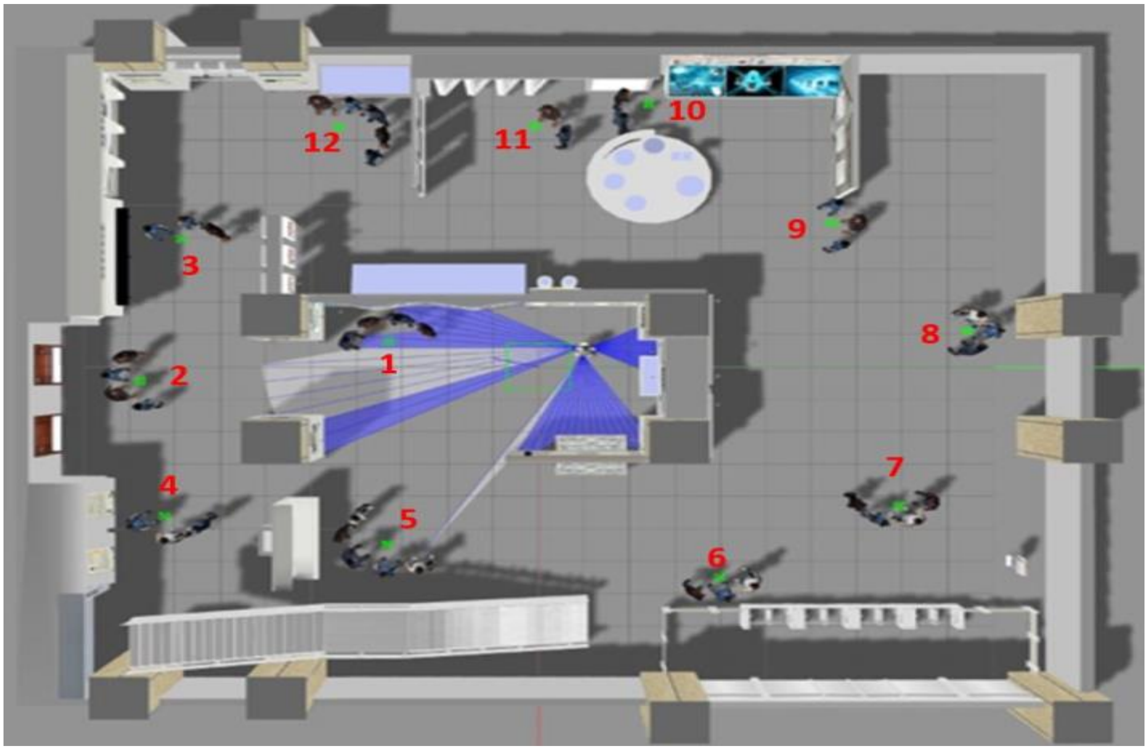


Figure 4-1- (A) Museum DATASET environment [**81**]

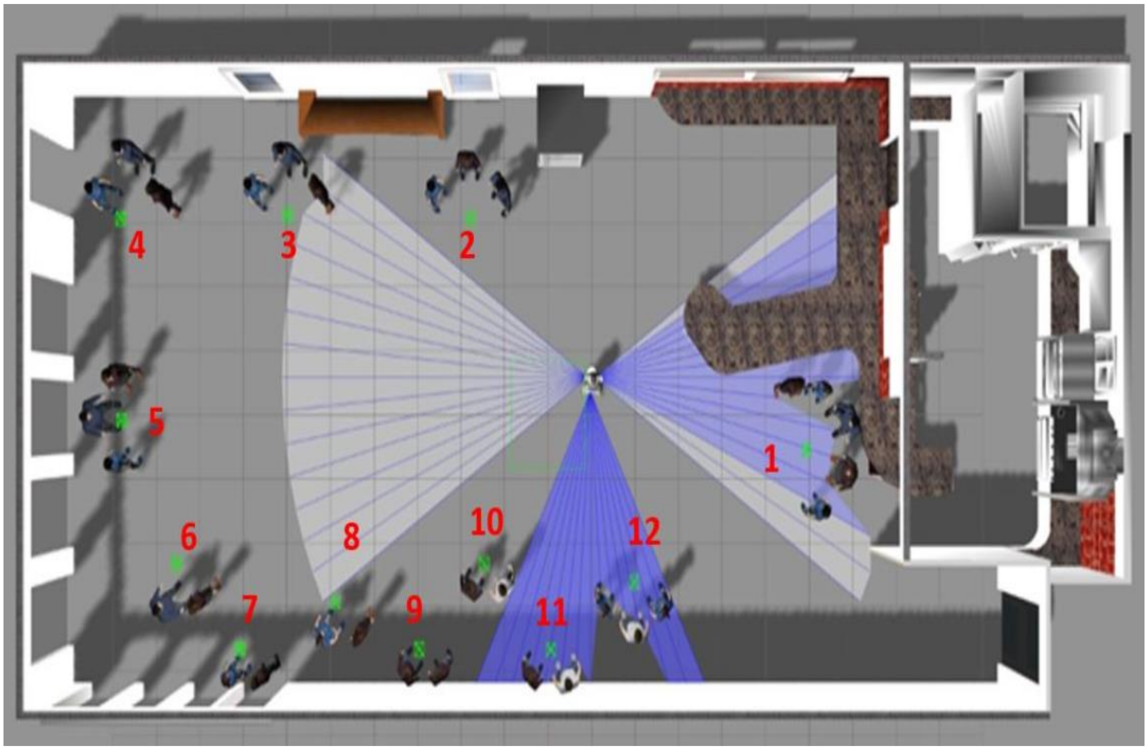


Figure 4-1- (B) Cafeteria DATASET environment [**81**]

4.3.2 The supplementary dataset

The RAF Dataset which regarded as a secondary dataset assumes a critical role in evaluating the adaptability of the Attention CNN model to real-world scenarios. Its test folder consists of 3068 images, complemented by individual text files containing emotion labels corresponding to each image. This supplementary dataset serves to evaluate the Attention CNN model and provides valuable insights into its performance beyond the controlled environment of the primary dataset. RAF regarded as a repository tailored for facial expression recognition research as show in Figure 4-2, it encompasses diverse images sourced from real-world contexts.

Figure 4-2 samples of RAF dataset

4.4 Attention CNN Model Architecture

The Attention CNN model, an essential component of the facial analysis system in this thesis, plays a central role in achieving the main objectives of HRI and GER. This Attention CNN architecture is thoughtfully designed to classify facial expressions into six distinct emotion categories, thereby enabling accurate emotion recognition within the context of HRI. The Attention CNN model's architecture includes several key layers, each contributing to the feature extraction and classification process. As shown in Figure 4-3 the input layer accommodates images resized to 150x150 pixels with RGB color channels. Three convolutional layers with increasing filter numbers and 3x3 kernels extract intricate patterns, employing ReLU activation functions for added complexity. Max-pooling layers follow each convolutional layer, reducing spatial dimensions by a 2x2 factor. A flatten layer then transforms the 2D feature maps into a 1D vector, preparing the data for fully connected layers. Also The Attention CNN architecture includes two fully connected dense layers. The first, with 512 units and ReLU activation, processes the flattened features, while a dropout layer with a rate of 0.5 mitigates overfitting. The second dense layer, serving as the output layer, consists of six units corresponding to the six emotion categories in the dataset. A softmax activation function generates class probabilities.

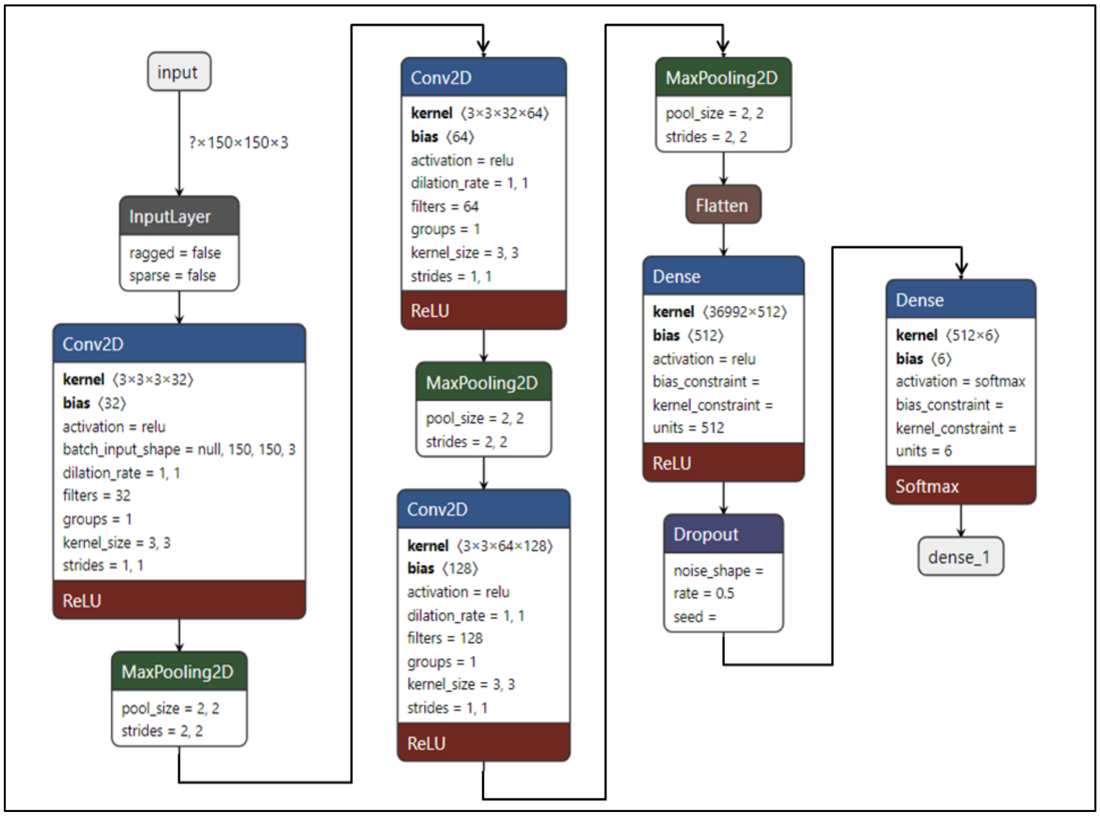


Figure 4-3 Attention CNN Model Architecture

4.5 Training the Attention CNN Model

This Attention CNN model has been training, as shown in Table 4-3:

Table 4-3 (Ten Epochs) for Training Attention CNN Model

Epoch	Duration	Loss	Accuracy
1	5066s	0.3462	0.8778
2	1127s	0.0467	0.9861
3	1124s	0.0312	0.9902
4	1118s	0.0199	0.9932
5	1114s	0.0217	0.9930
6	1109s	0.0147	0.9955
7	1137s	0.0102	0.9968
8	1108s	0.0141	0.9950
9	1118s	0.0100	0.9970
10	1108s	0.0117	0.9958

For 10 epochs, represents a fundamental landmark in the detection of robust emotion recognition and decision-making within the field of HRI, successfully handling the diversity of emotions captured within the dataset.

4.6 Evaluation Metrics for using Attention CNN Model

In this section, completely evaluate the performance of Attention CNN model for emotion recognition. The Attention CNN model represents a decisive section of overall facial analysis system, playing an essential role in achieving the wider objectives of HRI and GER.

4.6.1 Test Accuracy

Test accuracy serves as a fundamental metric to assess the model's overall performance. It measures the proportion of correctly classified samples in relation to the total number of samples in the test dataset. By calculate the ratio of correctly classified samples to the total number of samples in the test dataset.

➔ Applying Attention CNN model on ROS/Gazebo Generated Dataset, obtain the following Table 4-4 of accuracy values for each emotion class:

Table 4-4 Classes Accuracy (ROS/Gazebo Generated Dataset)

Emotion Class	Measuring	Accuracy Result
Angry	$^{750}/_{750}$	100%
Disgust	$^{564}/_{566}$	99.65%
Fear	$^{747}/_{750}$	99.60%
Happy	$^{744}/_{750}$	99.20%
Sad	$^{745}/_{750}$	99.33%
Surprise	$^{702}/_{750}$	93.60%

These high accuracy values underscore the model's capability to accurately classify emotions across all classes.

➔ While applying Attention CNN model on RAF Dataset, obtain the following Table 4-5 of accuracy values for each emotion class:

Table 4-5 Classes Accuracy (RAF Dataset)

Emotion Class	Measuring	Accuracy Result
Angry	$316/483$	0.6542
Disgust	$330/481$	0.6860
Fear	$329/562$	0.5854
Happy	$397/560$	0.7089
Sad	$293/467$	0.6274
Surprise	$381/515$	0.7389

4.6.2 Precision

Precision assesses the model's ability to correctly classify instances of a specific emotion class among all instances predicted as that class as described in Table 4-6.

Table 4-6 Classes Predictions

Predictions	Description
True Positive (TP)	The model correctly predicted the positive class.
True Negative (TN):	The model correctly predicted the negative class.
False Positive (FP):	The model incorrectly predicted the positive class (Type I error).
False Negative (FN):	The model incorrectly predicted the negative class (Type II error).

It reveals the number of TP, TN, FP, and FN predictions for each class. Thus, precision for each class measures the ability of the model to correctly classify that class among all instances predicted as that class.

➔ Applying Attention CNN model on ROS/Gazebo Generated Dataset, obtain the following Table 4-7 impressive precision values for each class:

Table 4-7 Classes precision (ROS/Gazebo Generated Dataset)

Emotion Class	Measuring	Accuracy Result
Angry	$^{750}/_{750}$	100%
Disgust	$^{564}/_{566}$	99.65%
Fear	$^{747}/_{750}$	99.60%
Happy	$^{744}/_{750}$	99.20%
Sad	$^{745}/_{750}$	99.33%
Surprise	$^{702}/_{750}$	93.60%

➔ While applying Attention CNN model on RAF Dataset, obtain the following Table 4-8 precision values for each class:

Table 4-8 Classes precision (RAF Dataset)

Emotion Class	Measuring	Accuracy Result
Angry	$^{316}/_{483}$	0.6542
Disgust	$^{330}/_{481}$	0.6860
Fear	$^{329}/_{562}$	0.5854
Happy	$^{397}/_{560}$	0.7089
Sad	$^{293}/_{467}$	0.6274
Surprise	$^{381}/_{515}$	0.7389

4.6.3 Recall (Sensitivity)

Recall, also known as sensitivity, measures the model's ability to correctly identify all instances of a particular emotion class by measuring the ratio of true positives for an emotion class to the summation of true positives for that emotion class with the false positives for the same emotion class.

➔ Applying Attention CNN model on ROS/Gazebo Generation Dataset, obtain the excellent recall values for each class with the same results as found previously in the section **4.4.2 Precision**. These recall values emphasize the model's efficiency in capturing all instances of emotions within the dataset.

➔ While applying Attention CNN model on RAF Dataset, obtain the following Table 4-9 of recall values for each emotion class:

Table 4-9 Classes recall (RAF Dataset)

Emotion Class	Measuring recall	Accuracy Result
Angry	316 / (316 + 56 + 316 + 23 + 27 + 18)	0.418
Disgust	330 / (330 + 23 + 17 + 27 + 47 + 330)	0.4268
Fear	329 / (329 + 67 + 37 + 21 + 329 + 49)	0.3959
Happy	397 / (397 + 51 + 29 + 21 + 37 + 25)	0.7107
Sad	293 / (293 + 64 + 37 + 19 + 51 + 33)	0.5895
Surprise	381 / (381 + 34 + 29 + 381 + 19 + 23)	0.4395

4.6.4 F1-Score

The F1-score, a harmonic mean of precision and recall, provides a balanced measure of the model's performance.

➔ Applying Attention CNN model on ROS/Gazebo Generation Dataset, obtain the following Table 4-10 of (F1-scores) for each emotion class:

Table 4-10 Classes F1-scores (ROS/Gazebo Generated Dataset)

Emotion Class	Measuring F1	Accuracy Result
Angry	$\mathbf{2X1.00X1.00}/_{1.00+1.00}$	100%
Disgust	$\mathbf{2 \cdot 0.9965X0.9965}/_{\mathbf{0.9965+0.9965}}$	99.65%
Fear	$\mathbf{2X0.996X0.996}/_{\mathbf{0.996+0.996}}$	99.60%
Happy	$\mathbf{2X0.992X0.992}/_{\mathbf{0.992+0.992}}$	99.20%
Sad	$\mathbf{2X0.9933X0.9933}/_{\mathbf{0.9933+0.9933}}$	99.33%
Surprise	$\mathbf{2X0.936X0.936}/_{\mathbf{0.936+0.936}}$	93.60%
Average		98.56

These F1-scores reinforce the model's ability to maintain a harmonious balance between precision and recall, indicative of its robust performance.

➔ While applying Attention CNN model on RAF Dataset, obtain the following Table 4-11 of (F1-scores) for each emotion class:

Table 4-11 Classes F1-scores (RAF Dataset)

Emotion Class	**Measuring**	**Accuracy Result**
Angry	(2 * 0.6804 * 0.418) / (0.6804 + 0.418)	0.5175
Disgust	(2 * 0.6912 * 0.4268) / (0.6912 + 0.4268)	0.5271
Fear	(2 * 0.6431 * 0.3959) / (0.6431 + 0.3959)	0.4872
Happy	(2 * 0.6628 * 0.7107) / (0.6628 + 0.7107)	0.6856
Sad	(2 * 0.558 * 0.5895) / (0.558 + 0.5895)	0.5734
Surprise	(2 * 0.7744 * 0.4395) / (0.7744 + 0.4395)	0.5669
Average		0.5596

4.6.5 Classification Report

The classification report provides a comprehensive summary of the model's performance, including precision, recall, and F1-score for each emotion class. The weighted averages of these metrics, as well as macro and micro averages, offer an overall assessment of the model's effectiveness.

➔ Applying Attention CNN model on ROS/Gazebo Generation Dataset, obtain the following classification report for the Attention CNN model as in Table 4-12.

Table 4-12 The classification report of Main Dataset

Emotion	**Precision**	**Recall**	**F1-Score**	**Support**
Angry	1.00	1.00	1.00	750
Disgust	0.99	0.99	0.99	566
Fear	1.00	1.00	1.00	750
Happy	0.99	0.99	0.99	750
Sad	0.99	0.99	0.99	750
Surprise	0.94	0.94	0.94	750
Statics				
Accuracy			0.99	4316
Macro Avg	0.99	0.99	0.99	4316
Weighted Avg	0.99	0.99	0.99	4316

This table provides a brief overview of the precision, recall, and F1-score for each emotion class, along with support and the macro and weighted averages for the entire classification report. Both macro and weighted averages provide insights into the overall performance of classification model. In this case, both averages have the same values (0.99) because the dataset seems balanced, with an equal number of instances for each class.

➔ While applying Attention CNN model on RAF Dataset, obtain the following classification report for the Attention CNN model as in the below Table 4-13:

Table 4-13 the classification report of Secondary Dataset

Emotion	Precision	Recall	F1-score	Support
Happy	0.7089	0.6628	0.6851	560
Sad	0.6274	0.5592	0.5913	467
Angry	0.6542	0.6796	0.6667	483
Surprised	0.7398	0.7744	0.7567	515
Fearful	0.5854	0.6451	0.6138	562
Disgusted	0.6861	0.6904	0.6882	481
Statics				
Accuracy			0.6669	3068
Macro Avg			0.6670	3068
Weighted Avg			0.6664	3068

4.6.6 Confusion Matrix

A confusion matrix is typically presented as a table, and it helps to understand where the model is making correct predictions and where it's making errors. From this matrix, it can calculate various performance metrics like precision, recall, and the F1-score, which provide insights into the model's accuracy, sensitivity, and overall effectiveness. The bellow Figures 4-5 and 4-6 shows the obtained results of confusion matrix when applying the Attention CNN model on both mainly and secondary datasets.

➔ Applying Attention CNN model on ROS/Gazebo Generation Dataset, obtain the following confusion matrix:

Applying CNN Model @ ROS/Gazebo Dataset >>> Show the Following *RESULTS*

TARGET / OUTPUT	Angry	Disgust	Fear	Happy	Sad	Surprise	SUM
Angry	750 17.37%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	0 0.00%	750 100.00% 0.00%
Disgust	0 0.00%	566 13.11%	0 0.00%	0 0.00%	0 0.00%	2 0.05%	568 99.65% 0.35%
Fear	1 0.02%	1 0.02%	747 17.30%	0 0.00%	1 0.02%	0 0.00%	750 99.60% 0.40%
Happy	0 0.00%	6 0.14%	0 0.00%	744 17.23%	0 0.00%	0 0.00%	750 99.20% 0.80%
Sad	1 0.02%	4 0.09%	0 0.00%	0 0.00%	745 17.25%	0 0.00%	750 99.33% 0.67%
Surprise	0 0.00%	0 0.00%	0 0.00%	47 1.09%	1 0.02%	702 16.26%	750 93.60% 6.40%
SUM	752 99.73% 0.27%	577 99.65% 1.91%	747 100.00% 0.00%	791 94.06% 5.94%	747 99.73% 0.27%	704 99.72% 0.28%	4254 / 4318 98.52% 1.48%

Figure 4-5 The Confusion Matrix of Main Dataset

➔ While applying Attention CNN model on RAF Dataset, obtain the following:

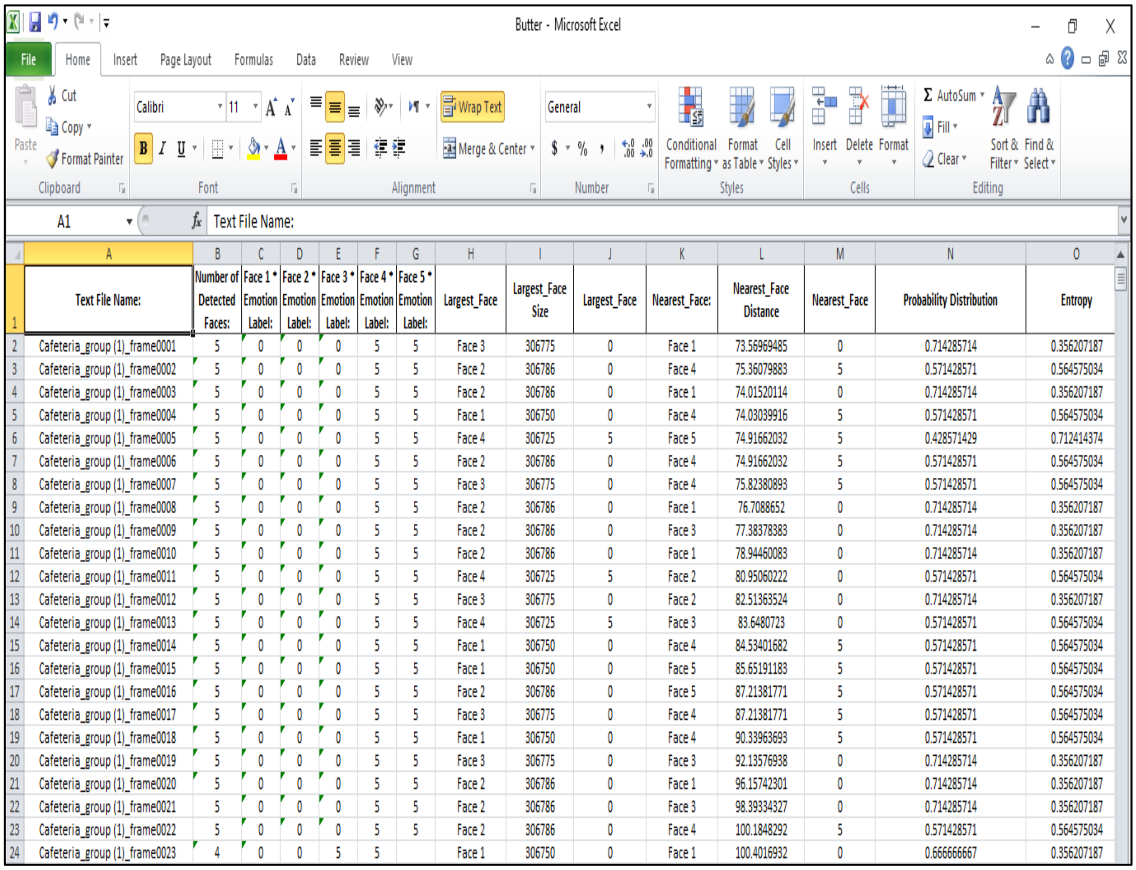
Apply CNN Model @ Test - RAF_DataSet

TARGET / OUTPUT	Happy	Sad	Angry	Surprised	Fearful	Disgusted	SUM
Happy	397 12.94%	51 1.66%	29 0.95%	21 0.68%	37 1.21%	25 0.81%	560 70.89% 29.11%
Sad	34 1.11%	293 9.55%	37 1.21%	19 0.62%	51 1.66%	33 1.08%	467 62.74% 37.26%
Angry	43 1.40%	56 1.83%	316 10.30%	23 0.75%	27 0.88%	18 0.59%	483 65.42% 34.58%
Surprised	29 0.95%	34 1.11%	29 0.95%	381 12.42%	19 0.62%	23 0.75%	515 73.98% 26.02%
Fearful	59 1.92%	67 2.18%	37 1.21%	21 0.68%	329 10.72%	49 1.60%	562 58.54% 41.46%
Disgusted	37 1.21%	23 0.75%	17 0.55%	27 0.88%	47 1.53%	330 10.76%	481 68.61% 31.39%
SUM	599 66.28% 33.72%	524 55.92% 44.08%	465 67.96% 32.04%	492 77.44% 22.56%	510 64.51% 35.49%	478 69.04% 30.96%	2046 / 3068 66.69% 33.31%

Figure 4-6 The Confusion Matrix of supplementary dataset

4.7 Group Emotion Recognition (GER):

In the field of GER, the objective is to analyze emotional dynamics in group scenarios to enable better human-robot interactions and decision-making based on group emotions. This process involves several key steps. First, as shown in Figure 4-7 a repository is created to store critical information required for emotion classification within each frame, including the number of detected faces in each frame, individual emotions, probability distributions, entropy calculations, and information about the largest and nearest faces. Then, the process goes through faces localization and proximity metrics calculations, determining the number of faces, the largest face, and the nearest face within each frame using Euclidean distance calculations. Following this, emotion labels are assigned to each detected face using the pre-trained Attention CNN model. The probability distribution of emotions within the group is computed, and entropy is used to measure diversity in emotions. The threshold for entropy is determined based on a median value, allowing frames to be categorized as exhibiting either clear and consistent emotions or diverse and uncertain emotions. The GER score is calculated by combining proximity metrics and entropy, reflecting the overall emotional dynamics of the group in each frame. Finally, the GER_Label is identified for each scene, summarizing the emotional state of multiple frames based on aggregated GER values and predefined thresholds. This structured approach enhances the understanding of group emotions during human-robot interactions, enabling robots to make context-aware decisions effectively.



Text File Name:	Number of Detected Faces:	Face 1 * Emotion Label:	Face 2 * Emotion Label:	Face 3 * Emotion Label:	Face 4 * Emotion Label:	Face 5 * Emotion Label:	Largest_Face	Largest_Face Size	Largest_Face	Nearest_Face:	Nearest_Face Distance	Nearest_Face	Probability Distribution	Entropy
Cafeteria_group (1)_frame0001	5	0	0	0	5	5	Face 3	306775	0	Face 1	73.56969485	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0002	5	0	0	0	5	5	Face 2	306786	0	Face 4	75.36079883	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0003	5	0	0	0	5	5	Face 2	306786	0	Face 1	74.01520114	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0004	5	0	0	0	5	5	Face 1	306750	0	Face 4	74.03039916	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0005	5	0	0	0	5	5	Face 4	306725	5	Face 5	74.91662032	5	0.428571429	0.712414374
Cafeteria_group (1)_frame0006	5	0	0	0	5	5	Face 2	306786	0	Face 4	74.91662032	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0007	5	0	0	0	5	5	Face 3	306775	0	Face 4	75.82380893	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0008	5	0	0	0	5	5	Face 2	306786	0	Face 1	76.7088652	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0009	5	0	0	0	5	5	Face 2	306786	0	Face 3	77.38378383	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0010	5	0	0	0	5	5	Face 2	306786	0	Face 1	78.94460083	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0011	5	0	0	0	5	5	Face 4	306725	5	Face 2	80.95060222	0	0.571428571	0.564575034
Cafeteria_group (1)_frame0012	5	0	0	0	5	5	Face 3	306775	0	Face 2	82.51363524	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0013	5	0	0	0	5	5	Face 4	306725	5	Face 3	83.6480723	0	0.571428571	0.564575034
Cafeteria_group (1)_frame0014	5	0	0	0	5	5	Face 1	306750	0	Face 4	84.53401682	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0015	5	0	0	0	5	5	Face 1	306750	0	Face 5	85.65191183	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0016	5	0	0	0	5	5	Face 2	306786	0	Face 5	87.21381771	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0017	5	0	0	0	5	5	Face 3	306775	0	Face 4	87.21381771	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0018	5	0	0	0	5	5	Face 1	306750	0	Face 4	90.33963693	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0019	5	0	0	0	5	5	Face 3	306775	0	Face 3	92.13576938	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0020	5	0	0	0	5	5	Face 2	306786	0	Face 1	96.15742301	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0021	5	0	0	0	5	5	Face 2	306786	0	Face 3	98.39334327	0	0.714285714	0.356207187
Cafeteria_group (1)_frame0022	5	0	0	0	5	5	Face 2	306786	0	Face 4	100.1848292	5	0.571428571	0.564575034
Cafeteria_group (1)_frame0023	4	0	0	5	5		Face 1	306750	0	Face 1	100.4016932	0	0.666666667	0.356207187

Figure 4-7 a view of samples from the main dataset repository

Furthermore, the Figure 4-8 show the Emotions_Label's classification of all the GER scenes that obtained within ROS dataset of the both cafeteria and museum environments.

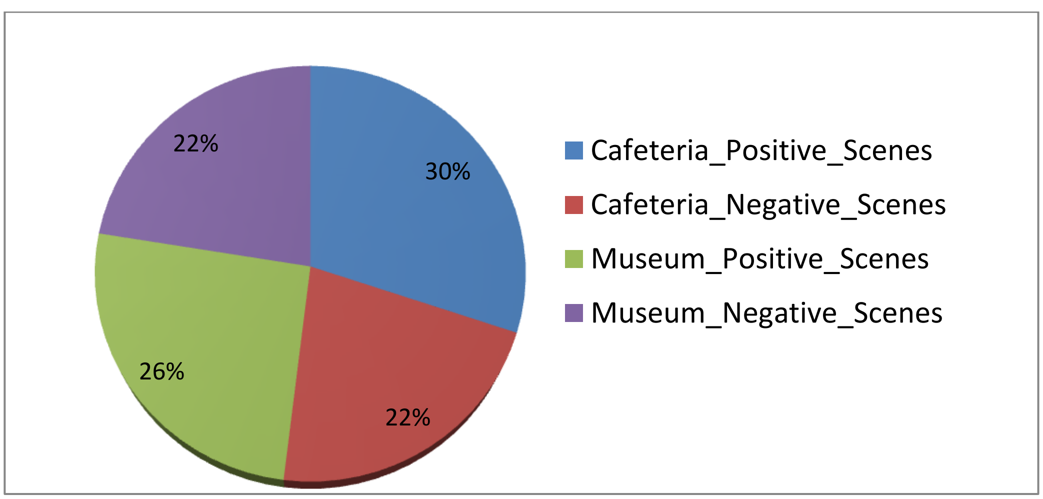


Figure 4-8 ROS dataset Emotions_Label's classification

4.8 Decision making

Establishing a robust decision-making framework for a social robot, depending on the Positive or Negative labeling generated by the GER system, establishes an essential phase for ensuring harmonious HRI. This process must be characterized by sensitivity, adaptability, and a deep understanding of the intricacies of human emotions and social indications across diverse scenarios and user's natures. It is also imperative to follow to ethical rules to foster respectful and responsible interactions with humans. Thus, the decision-making procedure for a social robot, based on the Positive or Negative labeling of GER, is an essential step in enabling suitable HRI. In this procedure, the robot initially gathers GER information, which is then processed by a dedicated GER component to ascertain emotion labels. These emotion labels are then classified by a distinct Classifier component. The GER component communicates the emotion classification back to the robot. Depending on whether the GER result is positive or not, as previously illustrated in Figure 3-11, the robot will be enable to make the right suitable decision making according to the detailed situations bellow.

4.8.1 Positive GER Label:

When GER labels a scene as Positive, signifying a collective emotional state that is clear and positive, the robot is empowered to engage with the group through a chatbot interface.

The following Table 4-6:

Table 4-14 Robot actions for Positive GER

Action	Description
Initiate Chat Box	Start a chat box or conversation interface on the robot's display or communication device.
Greetings	Begin the interaction with a friendly greeting or introduction, like "Hello, how can I assist you today?"
Engage Actively	Actively engage with the group, asking open-ended questions, and attentively listening to their responses.
Emotionally Appropriate Responses	Tailor responses based on the group's emotions. Maintain a cheerful tone for a happy group or respond empathetically to concerns.
Offer Assistance	Be ready to offer assistance or information based on the context of the interaction, answering questions, or providing relevant information.
Monitor Emotions	Continuously monitor the group's emotions during the conversation and adapt the conversation style if emotions change.
Maintain Ethical Conduct	Ensure the robot's behavior remains ethical and respectful throughout the interaction, avoiding actions that may offend or upset the group.
Contextual Awareness	Consider the context of the interaction, including the purpose of the conversation and the preferences of the group members.

This table above outlines the key actions undertaken by the robot in response to this previous scenario.

4.8.2 Negative GER Label:

Conversely, when GER designates a scene as Negative, indicating emotions that are diverse or less clear, possibly leaning towards negativity, the robot's decision-making process diverges. Here, the robot opts for disengagement and potential relocation to avoid potentially negative interactions. The following Table 4-7 outlines the decision-making actions taken in response to this previous scenario:

Table 4-15 Robot actions for Negative GER

Action	Description
Stop Interaction	Politely inform the group that the robot needs to attend to other tasks, ceasing the ongoing interaction.
Dis-engage	Stop the current conversation or interaction in a friendly and non-intrusive manner, leaving the option for future questions
Maintain Distance	If applicable, move the robot to a predefined safe distance or location to avoid disrupting the group.
Monitor	Continuously monitor the group's emotions from a distance to assess whether emotions become more positive or stable.
Alert Human Supervisor	Depending on the context, consider notifying a human supervisor or operator about the situation for further actions.
Safety Considerations	Prioritize safety during the robot's movement, ensuring it doesn't pose physical risks to the group or itself.
Record Data	Record data related to the interaction for later analysis and improvements in the robot's decision-making algorithms.
Review	Periodically review the group's emotions after moving away. If emotions improve or the situation changes, the robot decide whether to re-engage.

4.9 Discussion:

Evaluating the thesis components of the offered system is essential in estimating the performance and effectiveness of the proposed GER system within the context of HRI. These components are consisting of; the datasets, including the ROS/Gazebo Generation Dataset and the RAF Dataset, are detailed, highlighting their significance and contributions to the research. Attention CNN model's architecture, training process, and evaluation metrics demonstrate the system's robustness in recognizing emotions. The decision-making process for social robots based on GER results is outlined with sensitivity and ethical considerations. However, limitations, such as dataset size, emotion diversity, potential biases, overfitting risks, image quality dependency, hardware constraints, contextual understanding, lack of external validation, and real-time processing challenges, are emphasizing the need for cautious interpretation of the system's outcomes in real-world HRI scenarios.

Finally, as a compression of what has been reached between what this thesis and a related research [**81**]; Table 4-9 show that:

Table 4-17 : comparison between thesis and [**81**]

Aspect	**This Thesis**	[**81**]
Methodology and Approach	Two-tiered face detection, MTCNN integration, Attention CNN model in TensorFlow and Keras.	Viola–Jones classifier for face detection, VGGFace-based feature extraction, emotion classification into six categories.
Dataset Utilized	ROS/Gazebo Generated Dataset 23,000 images and 12 videos, and use RAF Dataset	ROS/Gazebo Generated Dataset 23,000 images and 12 videos

Aspect	This Thesis	[81]
Model Architecture	Attention CNN model with convolutional layers, max-pooling, fully connected layers, and softmax output.	Modified VGGFace model for individual emotion classification with dropout layers.
Scene Determining	Grouping frames into scenes every ten repeated frames for group emotion analysis.	Using Blocks of Frames (BOF) to determine scenes based on face size changes between frames.
Group Emotion Aggregation	Group Emotion Aggregating (GER) based on face localization, proximity metrics, entropy calculations, and CNN-based emotion recognition.	Aggregating individual emotions by defining scenes based on face size in Blocks of Frames (BOF).
Results and Performance Metrics	High accuracy, precision, recall, and F1-score values ranging from 93.60% to 100% for all emotion classes.	Average accuracy of 97.19% for individual emotion detection and 89.78% to 90.84% for emotion detection in videos

The table 4-17 provided a concise comparison between two research efforts, represented as "This Thesis" and " " in the field of Emotion Recognition in Human-Robot Interaction. It highlights key aspects of their research, including their research, methodologies; datasets utilized, and model architectures for emotion recognition, group emotion aggregation techniques, and results with performance metrics.

CHAPTER FIVE

CONCLUSION

AND

FUTURE WORKS

5.1 Overview

This chapter discusses the conclusion which have verified in this thesis at (section 5.2) and suggests a suitable future works at (section 5.3).

5.2 Conclusions:

The research on HRI decision-making based on GER has reached several significant conclusions:

1. **Dependable Emotion Recognition and Decision-Making:**

The study successfully developed an Attention CNN model for emotion recognition, enabling accurate detection of emotions from individual faces in dynamic human-robot interaction scenarios. This model serves as the foundation for group emotion recognition and subsequent decision-making.

2. **GER Calculations:**

The proposed method for determining group emotion based on the entropy of distributed emotions and the size of each face in the frame demonstrated the feasibility of understanding collective emotions. By combining proximity metrics and entropy calculations, GER provides valuable insights into the emotional dynamics of groups.

3. **Robust Attention CNN Model:**

The Attention CNN model achieved high accuracy across multiple emotions, indicating its robustness in recognizing diverse emotional states. This model is pivotal for interpreting and responding to human emotions in real-world HRI scenarios.

5.3 Future Work Suggestions:

1. **Dataset Enhancement:**

Future research can focus on expanding the diversity and size of the emotion recognition dataset. This would involve capturing a wider range of emotional expressions and considering a more diverse group of participants to address potential biases.

2. **Enhancing Emotional Understanding and Cultural Sensitivity:**

Future work aims to enhance emotional intelligence by recognizing nuanced emotions and diverse emotional states, while also diversifying datasets to improve recognition across different ethnicities and cultural expressions.

3. **Real-Time Processing Optimization:**

Addressing hardware constraints for real-time processing is essential for practical deployment.

4. **Contextual Understanding:**

Future work should aim to enhance the system's contextual understanding of human emotions and social interactions. This might involve incorporating context-aware algorithms that consider situational factors.

REFERENCES

[1] Matteo Spezialetti, "Emotion Recognition for Human-Robot Interaction: Recent Advances and Future Perspectives," *Frontiers in Robotics and AI*, vol. 7, 21 - December 2020.

[2] L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak Barrett, "Corrigendum: Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements," *Psychological Science in thePublic Interest*, 2019.

[3] Eryka Probierz, "On Emotion Detection and Recognition Using a Context Aware Approach by Social Robots– Modification of Faster R-CNN and YOLO v3 Neural Networks," *European Research Studies Journal*, vol. XXVI, no. 1, p. 14, 30/March 2023.

[4] Hang Su, "Recent advancements in multimodal human–robot interaction," *Frontiers in Neurorobotics*, vol. 17, p. 21, 11 - May 2023.

[5] Ruth Stock-Homburg, "Survey of Emotions in Human–Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research," *International Journal of Social Robotics (2022) 14:389–411*, 2021.

[6] Ameer A. Badr, "A Review on Voice-based Interface for Human-Robot Interaction," *raqi Journal for Electrical and Electronic Engineering by College of Engineering, University of Basrah.*, 2020.

[7] Chiara Filippini, "Improving Human–Robot Interaction by Enhancing NAO Robot Awareness of Human Facial Expression," *Sensors MDPI*, 2021.

[8] Muneeb Imtiaz Ahmad, "A Systematic Review of Adaptivity in Human-Robot Interaction," *Multimodal Technologies and Interact.*, vol. 3, no. 14, 20 - July 2017.

[9] Marcos Maroto-Gómez, "A Systematic Literature Review of Decision-Making and Control Systems for Autonomous and Social Robots," *International Journal of Social Robotics (2023) 15:745–789*, 2023.

[10] Sharon Ewa Spisak, "A Study on Social Exclusion in Human-Robot Interaction," *Electronics MDPI*, vol. 12, no. 07, 28 - March 2023.

[11] Kerstin Dautenhahn, "KASPAR – a minimally expressive humanoid robot for human–robot interaction research," *Taylor & Francis / Applied Bionics and Biomechanics*, 2009.

[12] LINN FRIIS-LIBY, "Evaluating the possible applications of social robots in the manufacturing industry," INSTITUTIONEN FÖR INDUSTRI- OCH MATERIALVETENSKAP AVDELNINGEN FÖR PRODUKTIONSSYSTEM, CHALMERS TEKNISKA HÖGSKOLA, Bachelor thesis in Mechanical engineering 2020.

[13] Weria Khaksar*, "Robotics in Elderly Healthcare: A Review of 20 Recent Research Projects," *Department of Informatics, University of Oslo*, 2022.

[14] Yujia Lin, "Human Digital Twin: A Survey," *Springer Nature*, 2021.

[15] Gerald Matthews, "Individual Differences in Trust in Autonomous Robots: Implications for Transparency," *IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS 1*, 2019.

[16] Jo Ann Oravec, "Negative Dimensions of Human-Robot and Human AI Interactions: Frightening Legacies, Emerging Dysfunctions, and Creepiness," in *Good Robot, Bad Robot*.: https://link.springer.com/chapter/10.1007/978-3-031-14013-6_3, 2022, ch. pp 39-89.

[17] ETHAN S. BERNSTEIN, "MAKING TRANSPARENCY TRANSPARENT: THE EVOLUTION OF OBSERVATION IN MANAGEMENT THEORY," *r Academy of Management Annals*, vol. 11, 2017.

[18] Dag Sverre Syrdal, "The Impact of Social Expectations towards Robots on Human-Robot Interactions on Human-Robot Interactions," the University of Hertfordshire, Submitted to the University of Hertfordshire in partial fulfilment of the requirements of the degree of Doctor of Philosophy 2018.

[19] Said Ghani Khan , Guido Herrmann Adam Spiers, *Biologically Inspired Control of Humanoid Robot Arms*. Switzerland, Switzerland: Springer Nature, 2016.

[20] Xingzhi Wang, Dong Zhang, Hong-Zhou Tan, and Dah-Jye Lee, "A Self-Fusion Network Based on Contrastive Learning for Group Emotion Recognition," *IEEE Transactions on Computational Social Systems* , 2023.

[21] Xianye Ben, "Video-based Facial Micro-Expression Analysis: A Survey of Datasets, Features and Algorithms," *EEE T-PAMI*, 2021.

[22] Ali Mollahosseini, "Going Deeper in Facial Expression Recognition using Deep Neural Networks," *IEEE*, 2016.

[23] Maria De Marsico, "Robust face recognition after plastic surgery using region-based approaches," *journal homepage: www.elsevier.com/locate/pr*, 2014.

[24] Brais Martinez, "Automatic Analysis of Facial Actions: A Survey," *IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.*, 2017.

[25] James Z. Wang, "Unlocking the Emotional World of Visual Media: An Overview of the Science, Research, and Impact of Understanding Emotion," *IEEE*, 2023.

[26] Pawel Drozdowski, "Demographic Bias in Biometrics: A Survey on an Emerging Challenge," *IEEE TRANSACTIONS ON TECHNOLOGY AND SOCIETY, VOL. 1, NO. 2, JUNE 2020*, 2020.

[27] Sylvie Serret, "Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new Serious game: a pilot study," *Serret et al. Molecular Autism* , 2014.

[28] Filippo Cavallo, "Emotion Modelling for Social Robotics Applications: A Review," *Journal of Bionic Engineering*, 2018.

[29] Patrik N. Juslin, "Emotional responses to music: The need to consider underlying mechanisms," *BEHAVIORAL AND BRAIN SCIENCES*, 2008.

[30] Sigal G. Barsade, "The Ripple Effect: Emotional Contagion and its Influence on Group Behavior," *sagepublications*, 2002.

[31] Hillary Anger Elfenbein, "The many faces of emotional contagion: An affective process theory of affective linkage," *Organizational Psychology Review*, 2014.

[32] Joost Broekens, "Modeling emotional contagion based on experimental evidence for moderating factors," *https://www.researchgate.net/publication/266021112*, 2012.

[33] Chi-Hua Li & Pei-Ju Chao, "Impact of Emotional Contagion through Social Network Sites on Travel Willingness in the Pandemic," *Journal of Quality Assurance in Hospitality & Tourism* ?, 2022.

[34] Amit Goldenberg, "The Psychology of Collective Emotions," *Stanford University, Department of Psychology*, 2016.

[35] Dacher Keltner and Oliver P. John Cameron Anderson, "Emotional Convergence Between People Over Time," *Journal of Personality and Social Psychology Copyright 2003 by the American Psychological Association, Inc.*, 2003.

[36] Mark Yacoub, "Emotional Framing: How Do Emotions Contribute to Framing Effects?," core.ac.uk, A thesis submitted to the faculty of the University of North Carolina 2012.

[37] Mohammed Issah, "Change Leadership: The Role of Emotional Intelligence," *SAGE Open*, 2018.

[38] Yu Wang, "ConGNN: Context-consistent cross-graph neural network for group emotion recognition in the wild," *https://www.sciencedirect.com/science/article/abs/pii/S0020025522008830*, 2022.

[39] Garima Sharma, Roland Goecke, Tom Gedeon Abhinav Dhall, "EmotiW 2020: Driver Gaze, Group Emotion, Student Engagement and Physiological Signal based Challenges," *ICMI '20: Proceedings of the 2020 International Conference on Multimodal Interaction • October 2020*, 2020.

[40] Xin Guo, "Graph Neural Networks for Image Understanding Based on Multiple Cues: Group Emotion Recognition and Event Recognition as Use Cases," *IEEE*, 2020.

[41] Ahmed Shehab Khan, "Regional Attention Networks with Context-aware Fusion for Group Emotion Recognition," *IEEE*, 2021.

[42] Mo Sun, "Multi-modal Fusion Using Spatio-temporal and Static Features for Group Emotion Recognition," *ICMI '20, October 25–29, 2020, Virtual Event, Netherlands*, 2020.

[43] Juan Aparicio, "On how to properly calculate the Euclidean distance-based measure in DEA," *A Journal of Mathematical Programming and Operations Research*, 2012.

[44] Subarna Shakya, "Human Behavior Prediction using Facial Expression Analysis," *International Conference on Computing, Communication and Automation (ICCCA2016)*, 2016.

[45] Guilherme Deola Borges, "Decision-Making Framework for Implementing Safer Human-Robot Collaboration Workstations: System Dynamics Modeling," *Safty MDPI*, 2021.

[46] Matthew C. Gombolay, "Decision-Making Authority, Team Efficiency and Human Worker Satisfaction in Mixed Human-Robot Teams," *Massachusetts Institute of Technology*, 2014.

[47] Beatriz García-Martínez, "Application of Entropy-Based Metrics to Identify Emotional Distress from Electroencephalographic Recordings," *www.mdpi.com/journal/entropy*, 2016.

[48] Martina Szabóová, "Emotion Analysis in Human–Robot Interaction," *Electronics / www.mdpi.com/journal/electronics*, 2020.

[49] Ya Zheng, "The Influence of Emotion on Fairness-Related Decision Making: A Critical Review of Theories and Evidence," *Frontiers in Psychology | www.frontiersin.org 1 September 2017 | Volume 8 | Article 1592*, 2017.

[50] Jennifer S. Lerner, "Emotion and Decision Making," *Annu. Rev. Psychol.*, 2014.

[51] Hiroshi Ishiguro Takahisa Uchida, "Improving Quality of Life with a Narrative Robot Companion: II – Creating Group Cohesion via Shared Narrative Experience*," *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*

August 31 - September 4, 2020 - Virtual Conference, 2020.

[52] Meredith M Harvey, "Contextually appropriate emotional word use predicts adaptive health behavior: Emotion context sensitivity and treatment adherence," *http://www.sagepublications.com*, 2014.

[53] Atefeh Goshvarpour, "Evaluation of Novel Entropy-Based Complex Wavelet Sub-bands Measures of PPG in an Emotion Recognition System," *Journal of Medical and Biological Engineering*, 2020.

[54] Tom Carter, *An introduction to information theory and entropy*.: Complex Systems Summer School, 2011.

[55] S. R. Ashokkumar, "Emotion identiÒcation by dynamic entropy and ensemble learning from electroencephalogram signals," *International Journal of Imaging Systems and Technology | IMA | Wiley Online Library*, 2021.

[56] Pragati Patel, "EEG-based human emotion recognition using entropy as a feature extraction measure," *https://www.ncbi.nlm.nih.gov/pmc/articles*, 2021.

[57] Alam Lamia, "EXAMINING COGNITIVE EMPATHY ELEMENTS WITHIN AI CHATBOTS FOR HEALTHCARE SYSTEMS," MICHIGAN TECHNOLOGICAL UNIVERSITY, MICHIGAN, DOCTOR OF PHILOSOPHY https://doi.org/10.37099/mtu.dc.etdr/1437, 2022.

[58] Noor Kamal Al-Qazzaz, "Complexity and Entropy Analysis to Improve Gender Identification from Emotional-Based EEGs," *Journal of Healthcare Engineering / Hindawi*, 2021.

[59] Ivo Benke, "Understanding the impact of control levels over emotion aware chatbots," *Computers in Human Behavior - https://www.sciencedirect.com/science/article/abs/pii/S0747563221004453*, 2022.

[60] Shubin Yu, "Designing Emotions for Health Care Chatbots: Text-Based or Icon-Based Approach," *Journal of Medical Internet Researchers* , 2022.

[61] Xin Zuo, "Cross-Subject Emotion Recognition Using Fused Entropy Features of EEG," *https://www.mdpi.com/journal/entropy*, 2022.

[62] Ghazala Bilquise, "Emotionally Intelligent Chatbots: A Systematic Literature Review," *Hindawi - Human Behavior and Emerging Technologies*, 2022.

[63] Xueling Ma, "Group decision-making framework using complex Pythagorean fuzzy information," *Neural Computing and Applications*, 2020.

[64] Susan Mohammed, "Cognitive Diversity and Consensus in Group Decision Making: The Role of Inputs, Processes, and Outcomes," *Organizational Behavior and Human Decision Processes*, 2001.

[65] Bei LUO, "Emotion-regulatory Chatbots for Enhancing Consumer Servicing: An Interpersonal Emotion Management Approach," *Information & Management*, 2023.

[66] Ziang Xiao, "Inform the uninformed: Improving Online Informed Consent Reading with an AI-Powered Chatbot," *CHI '23, April 23–28, 2023, Hamburg, Germany - https://doi.org/10.1145/3544548.3581252*, 2023.

[67] Diana-Cezara Toader, "The Effect of Social Presence and Chatbot Errors on Trust," *www.mdpi.com/journal/sustainability*, 2020.

[68] Muhammad Akram, "Group decision-making based on complex spherical fuzzy VIKOR approach," *journal homepage: www.elsevier.com/locate/knosys*, 2021.

[69] Christian von scheve, *Collective Emotions perspective form psycology philosophy and sociology*., 2012.

[70] Matthias Wimmer, "Facial Expression Recognition for Human-Robot Interaction – A Prototype," *G. Sommer and R. Klette (Eds.): RobVis 2008, LNCS 4931, pp. 139–152, 2008.*, 2008.

[71] Suhaila N. Mohammed, "Emotion Recognition Based on Texture Analysis of Facial Expressions Using Wavelets Transform," *AENSI Publication*, 2017.

[72] Shih - Chuung Hsu, "Facial Expression Recognition for Human-Robot Interaction ," *IEEE International Conference on Robotic Computing*, 2017.

[73] Zhiqin Zhang, "Deep Face Emotion Recognition," *Journal of Physics: Conference Series*, 2018.

[74] Gaoyang Pang, "CGAN Based Facial Expression Recognition for Human-Robot Interaction," *IEEE Access · January 2019*, 2019.

[75] Alia K. Hassan, "A novel facial emotion recognition scheme based on graph mining," *China Ordnance Society. Publishing Services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license*, 2019.

[76] Alireza Esfandbod, "Human-Robot Interaction based on Facial Expression Imitation," *International Conference on Robotics and Mechatronics (ICRoM 2019)*, 2019.

[77] MIKEL VAL-CALVO, "Affective Robot Story-Telling Human-Robot Interaction: Exploratory Real-Time Emotion Estimation Analysis Using Facial Expressions and Physiological Signals," *This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/*, 2020.

[78] Daniel Octavian Melinte and Luige Vladareanu, "Facial Expressions Recognition for Human–Robot Interaction Using Deep Convolutional Neural Networks with Rectified Adam Optimizer," *Sensors MDPI*, 2020.

[79] Jose Maria Buades Silvia Ramis, "Using a Social Robot to Evaluate Facial Expressions in the Wild," *Sensors Mdpi*, 2020.

[80] Jianmin Wang, "Experimental Study on Abstract Expression of Human-Robot Emotional Communication," *Symmetry MDPI*, 2021.

[81] Marco Quiroz, "Group Emotion Detection Based on Social Robot Perception," *MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations*, vol. 22, no. 10, p. 22, 14/May 2022.

[82] Grazia D'Onofrio, "Emotion Recognizing by a Robotic Solution Initiative (EMOTIVE Project)," *Sensors MDPI*, 2022.

الخلاصة

في مجال التفاعل بين الإنسان والروبوت (HRI)، تلعب القدرة على التعرف على المشاعر الإنسانية وفهمها دورًا أساسيًا في تسهيل التفاعل الاجتماعي. حيث إنه يمكن الاستفادة من تفاصيل التعرف على المشاعر الجماعية (GER) كأساس عملي لتحسين عمليات صنع القرار ضمن مبادئ عمل ال(HRI).

تتكون منهجية هذه الأطروحة من عدة خطوات لاتخاذ القرار الأفضل. في بادئ الأمر تم البحث عن مجموعة بيانات تشمل أطياف واسعة من تعابير الوجه العاطفية وسيناريوهات مختلفة لتجمع الأفراد. لذا فقد أعتمدت مجموعة بيانات مستمدة من بيئات محاكاة افتراضية للواقع الحقيقي كأساس رئيسي لتحليل تعابير الوجوه من خلال إستخدام طريقة منهجية لاكتشاف و تحديد الوجوه بشكل أولي ومن ثم أستنباط و تمييز التعابير، وذلك من خلال دمج مصنفات الــ(Haar و HOG)، لضمان تحديد الوجوه بشكل فعال. ثم استخراج العلامات المميزة لتعابير الوجه من خلال استخدام الـ(Dlib)، ولتعزيز الدقة في تحديد الوجه تم إستخدام الــ(MTCNN) .

يكمن جوهر هذا البحث في تطوير نموذج الشبكة العصبية التلافيفية (CNN)، المصمم لتصنيف دقيق للصور متعددة المشاعر طبقا لتعبيرات الوجه. حيث إنه يتيح نموذج الـ(Attention CNN) و الذي تم تدريبه على مجموعة البيانات المهيأة مسبقاً لكي يتمكن من التعرف الدقيق على المشاعر. و من ثم تم اختبار هذا النموذج على نوعين من البيانات الأساسية والثانوية. حيث إنه تم تحقيق دقة عالية عند إستخدامه على بيانات الــ(ROS/Gazebo Generated) الرئيسية محققا نسبة 98.5% لإستكشاف جميع فئات المشاعر الستة المقترحة (الغضب ، الإشمئزاز ، الخوف ، السعادة ، الحزن ، التفاجئ)، بينما حققت مجموعة بيانات الـ(RAF) الثانوية 66% من الدقة.

علاوة على ذلك، تم تطبيق الموديل على مجموعتين من الفيديوات صممت بواسطة برامج محاكاة لبيئتين واقعيتين هما (قاعة متحف و بناية كافيتيريا) ، حيث أنه تم تحويل هذه الفيديوات الى مجموعة إطارات وتم التعامل مع كل إطار على أنه صورة ، ليتم إستخدام نفس العمليات السابقة في تحديد الوجوه و إستنباط المشاعر لأفراد المجاميع ضمن هذه الصور.

بعد تحديد الوجوه و تعابيرها تم إعتماد مبدأ التوزيع المكاني لهذه الوجوه ومبدأ الأنتروبي لتوزيع مشاعر مجموعة من الافراد ضمن الصورة الواحدة ، و بعد دمج كل عشرة أطر لتشكل مشهد واحد ، يتم لاحقا وفق عمليات محددة تصنيف مشاعر المجموعة لكل إطار ضمن إحدى حالتين سلبية أو إيجابية تمهيدا لتمكين الروبوت من إتخاذ القرار المناسب في كيفية التعامل مع مجاميع الافراد ضمن البيئات التفاعلية الموجود فيها ضمن مفاهيم الـ(HRI) ، لتُظهر عملية التقييم فعالية النموذج وتحقيقه دقة عالية في التمييز بين المشاعر المختلفة لكل فرد واتخاذ قرارات مناسبة .

جامعة كربلاء

كلية علوم الحاسوب وتكنولوجيا المعلومات

قسم علوم الحاسوب

إستكشاف المشاعر الجماعية في نظام محاكاة ذكي

رسالة ماجستير

مقدمة الى مجلس كلية علوم الحاسوب وتكنولوجيا المعلومات / جامعة كربلاء وهي جزء من متطلبات نيل درجة الماجستير في علوم الحاسوب

كتبت بواسطة

علي حسين موسى حمزة

بإشـراف

أ.د. اسيا مهدي ناصر

2024 A.D. 1445 A.H.