Paris Saclay University

Master 2 Internship presented by:

Clara MARTINEZ

at Ecole Technologique Supérieure (ETS)

in Montréal, Canada

Deep Learning Framework leveraging Multi-Teacher Knowledge Distillation using privileged information with Physiological Signals.

Internship supervised by:

Dr. GRANGER Eric — **FRSQ Double Chair in AI and Digital Health**

Director of LIVIA - Systems Engineering Department

In collaboration with:

Dr. ETEMAD Ali — **Queen's University – Faculty, Ingenuity Labs**

Research Institute, Electrical and Computer Engineering

Internship advised by:

Dr. ZEHRAOUI Farida — **Deep Learning and Personalized Medicine Teacher**

IBISC Laboratory

Memoire delivered on June 19th, 2024 – **2 months internship**

From 1st April 2024 to 31st September, 2024

Acknowledgments

I am inspired and guided by numerous people who are helping me for two months to start this internship. First and foremost, I would like to sincerely thank my internship supervisor, Dr. Eric Granger, who has provided me with generous support throughout my stay at Ecole Technologique Supérieure (ETS). I am very grateful to him for allowing me to work on such an interesting project. I have learned immensely through my interactions in the team laboratory on emotion expression. Furthermore, I want to thank Dr. Ali Atemad, who collaborated to supervised throughout the research and development of the project. I am very appreciative to have participated in various discussions in the collaborative and interdisciplinary environment that represent the LIVIA laboratory.

I would like to thank particularly a few members of the group with whom I had the chance to connect: PhD candidate Muhammad Haseeb Aslam, Post-Dr. Soufiane Belharbi, PhD Muhammad Osama Zeeshan.

Besides, I would like to acknowledge my internship adviser Dr. Farida Zehraoui, who supported me during this two-months internship, and even before by opening my mind to deep learning applied to e-health application.

Abbreviations

ACC: Accuracy

AI: Artificial Intelligence

BVP: Blood Volume Pulse

CV: Cross Validation

DNN: Deep Neural Network

EDA: Electrodermal Activity

ECG: Electrocardiogram

EEG: Electroencephalogram

EMG: Electromyogram

ETS: École de Technologie Supérieure

HEOG: Electrooculogram

ICCV: International Conference on Computer Vision

KD: Knowledge Distillation

LIVIA: Laboratory of Imagery Vision and Artificial Intelligence

MEG: Magnetoencephalogram

MER: Multimodal Emotion Recognition

NIR: Near-Infrared

R&D: Research & Development

RESP: Respiration

sEMG: Surface Electromyography

SSL: Self-Supervised Learning

TEMP: Skin Temperature

tEMG: Trapezius-Electromyogram

Glossary

Artificial Intelligence (AI): The simulation of human intelligence in machines programmed to think and learn like humans.

Convolutional Neural Network (CNN): A type of deep learning algorithm designed to process structured grid data such as images, using convolutional layers to capture spatial hierarchies.

Domain Adaptation: A technique in machine learning where a model trained in one domain is adapted to work in another domain with different but related data.

Electrocardiogram (ECG): A medical test that measures the electrical activity of the heart over a period of time, often used in physiological signal processing.

Electromyography (EMG): A technique for evaluating and recording the electrical activity produced by skeletal muscles, used to assess muscle function and detect neuromuscular abnormalities.

Galvanic Skin Response (GSR): A method of measuring the electrical conductance of the skin, which varies with its moisture level, indicating psychological or physiological arousal.

Knowledge Distillation (KD): A process in machine learning where a smaller, simpler model (student) is trained to replicate the performance of a larger, more complex model (teacher).

Machine Learning (ML): A branch of artificial intelligence that involves the development of algorithms that enable computers to learn from and make predictions or decisions based on data.

Multimodal Data Fusion: The process of integrating multiple types of data (e.g., text, audio, and visual) to improve the performance of a machine learning model.

Physiological Signals: Biological signals such as ECG, EMG, and GSR that provide information about the body's physiological state.

Recurrent Neural Network (RNN): A type of neural network designed for processing sequential data, capable of maintaining context over time through its architecture.

Self-Supervised Learning (SSL): A type of machine learning where the model learns to predict part of the input from other parts, using unlabeled data to create labels for training.

Transformer Encoder: A deep learning model component that uses self-attention mechanisms to process and encode sequential input data, widely used in natural language processing and other applications.

List of Figures

Figure 1: Six pillars of LIVIA's research and development activities.

Figure 2: Organigramme of LIVIA.

Figure 3: The limbic system.

Figure 4: Emotion models.

Figure 5: Biosensors position and typical waveforms.

Figure 6: Classification of deep learning techniques.

Figure 7: Multimodality databases on emotion recognition.

Figure 8: Distribution of dataset usage in emotion recognition using EEG signals.

Figure 9: Classification of deep learning techniques.

Figure 10: Schematic representation of the self-supervised Learning paradigm.

Figure 11: Overview of knowledge distillation methods.

Figure 12: Hierarchically structured taxonomy of knowledge distillation with teacher-student learning.

Figure 13: Graphical illustration for KD with multiple teachers.

Figure 14: Schematic framework of learning privileged information through teacher-student distillation.

Figure 15: Table presenting the models selected as a baseline for the study

Figure 16: Pseudo-algorithm of Baseline PainAttnNet.

Figure 17: Schematic representation of cross-validation. (Sklearn, 2022)

Figure 18: Schematic representation of a cross-validation procedure. (Sklearn, 2022)

Figure 19: Leave-One-Out Cross-Validation methodology.

Figure 20: Pseudo-algorithm presenting the LASO methodology.

Figure 21: Architecture of 1D-CNN with knowledge distillation form the logits.

Figure 22: Architecture of 1D-CNN with knowledge distillation with a cosine loss minimization run.

Figure 23: Architecture of 1D-CNN with knowledge distillation with an intermediate regressor run.

Figure 24: Architecture of the baseline PainAttnNet.

Figure 25: Architecture of PainAttnNet with knowledge distillation from the logits.

Figure 26: Architecture of PainAttnNet with knowledge distillation from the intermediate features.

Figure 27: Hyper-parameters tuning approach.

Figure 28: Table presenting the validation accuracy (%) cross-entropy runs of 1D-CNN on EDA signals of the BioVid database.

Figure 29: Table presenting the test accuracy (%) of cross-entropy runs of 1D-CNN on EDA signals of the BioVid database.

Figure 30: Table presenting the performance results of the distillation loss from the logits of 1D-CNN on EDA signals of the BioVid database.

Figure 31: Table presenting the performance results of cosine loss minimization run of 1D-CNN on EDA signals of the BioVid database.

Figure 32: Table presenting the performance results of intermediate regressor run of 1D-CNN on EDA signals of the BioVid database.

Figure 33: Table presenting the performance results of PainAttnNet.

Figure 34: Training loss and validation curve of fold 77 of PainAttnNet.

Figure 35: Training and validation accuracy of PainAttnNet with logits knowledge distillation.

Figure 36: Training and validation loss of PainAttnNet with logits knowledge distillation.

Figure 37: PainAttnNet's performance across five tasks on the BioVid dataset. (Lu, 2023)

Figure 38: Performance comparison between PainAttnNet and other SOTA approaches.

List of Appendix

Appendix 1: XKD framework.

Appendix 2: HKD-MER framework.

Appendix 3: SDT framework.

Appendix 4: Performance results of knowledge distillation with PainAttnNet Baseline.

Table of contents

Introduction 11

Part I: Ecole Technologique Supérieure **12**

A. LIVIA: a multidisciplinary artificial intelligence laboratory 15

B. Theorical and Bibliographic Analysis **15**

1. Emotion Recognition 15

2. Physiological signals 18

3. Multimodality 21

4. Deep Learning 25

5. Knowledge Distillation 28

C. Internship's missions **33**

Part II: Materials and Methods 34

A. Preparation of the data **34**

B. Protocol 35

1. General approach and softwares 35

2. Data 37

a. Physiological Signals 37
b. Data Preprocessing 39

3. Validation procedure 39

a. Cross validation 39
b. Leave One-Cluster-Out Cross-Validation 40

4. Models Construction 41

a. 1-D CNN 41
b. PainAttnNet 42
c. Hyper-parameters Tuning 44

5. Evaluation of the models 47

Part III: Presentation and Analysis of the Results **48**

1. Model development and validation 48

a. 1-D CNN 49
c. PainAttnNet 50

2. Generalization of the models 51

3. Knowledge distillation for pain classification using physiological signals 52

Conclusion ***53***

Summary ***54***

Bibliography ***56***

Appendices ***62***

Introduction

Emotion recognition is an interdisciplinary field in computer science, psychology, and neuroscience, which aims to understand human emotions using facial expressions, vocal intonations, and physiological signals. Emotion recognition plays a crucial role in many real-world applications ranging from healthcare and human-computer interaction to educational technologies.

This study focuses on advancing emotion recognition using deep learning techniques, with an emphasis on integrating multimodal data to improve the robustness and accuracy of these systems. The primary aim of this research is to develop and evaluate deep learning models capable of recognizing expressions linked to ambivalence, hesitancy and pain in healthcare applications. Indeed, these models are particularly important in clinical settings where accurate emotion detection can lead to better patient care and outcomes. The study leverages domain

adaptation, knowledge distillation and multimodalities (vocal, physiological signals, video) fusion techniques.

Furthermore, existing research in emotion recognition has demonstrated the potential of deep learning models in various applications. Numerous studies have shown that convolutional neural networks (CNNs) and recurrent neural networks (RNNs) can effectively process visual and sequential data, respectively (LeCun et al., 2015). However, these models often require large, annotated datasets, which are challenging to obtain, especially in the medical field (Wodzinski et al., 2020). Additionally, the integration of multiple data modalities has been explored to enhance the performance of emotion recognition systems (Thiam et al., 2021). Despite these advancements, there remains a significant gap in the literature regarding the effective fusion of multimodal data and the application of these models in real-world healthcare settings.

To addresses these gaps, we propose a novel deep learning architectures that incorporate self-supervised learning (SSL) and knowledge distillation (KD) techniques. SSL allows the models to learn robust representations from vast amounts of unlabeled data, while knowledge distillation enables the transfer of knowledge from larger, complex models to smaller, more efficient ones (Hinton et al., 2015; Jing & Tian, 2021). The combination of these techniques is expected to enhance the generalization and accuracy of emotion recognition models, making them more suitable for practical applications.

The research is carried out at the Laboratory for Imaging, Vision, and Artificial Intelligence (LIVIA) at the École de Technologie Supérieure—a leading research unit in Canada dedicated to advancing AI through visual perception and environment modeling. LIVIA is a multidisciplinary approach, and its contributions to AI research offer a solid basis to this research. This paper addresses the research problem of developing efficient and accurate deep learning models for identifying ambivalence and hesitancy expressions in healthcare data using multi-modal data.

The study aims to show the effectiveness of the SSL and KD techniques in improving the performance of these models when the available annotated data is limited. The novelty and contributions of this study are diverse: the design of new deep learning architectures with SSL and KD for emotion recognition; the complete evaluation of these models on multimodal data in combination with physiological signals; and the possibility of actual application of the model in healthcare settings, pointing toward its use in enhancing patient care.

Part I: Ecole Technologique supérieure (ETS)

A. LIVIA: a multidisciplinary artificial intelligence laboratory

The Imaging, Vision, and Artificial Intelligence Laboratory (LIVIA) at the École de Technologie Supérieure (ÉTS) is a cutting-edge research unit dedicated to advancing the field of artificial intelligence (AI) through the visual perception of 2D and 3D scenes and the static and dynamic modeling of environments. Established nearly 30 years ago, LIVIA has played a pivotal role in the development and dissemination of innovative AI technologies. Thus, ÉTS,

with the contributions of LIVIA, ranks 6th in Canada for computer vision according to CSRankings, reflecting the laboratory's prominent role in advancing AI research.

LIVIA's research activities are oriented around several core areas: machine learning, computer vision, pattern recognition, adaptive and intelligent systems, information fusion, and the optimization of complex systems. The laboratory is renowned for its excellence in AI engineering, particularly in developing complex models for deep learning that can handle massive amounts of data with limited annotations. This focus is crucial for applications that require precise and accurate interpretations of vast datasets, such as medical imaging and surveillance systems.

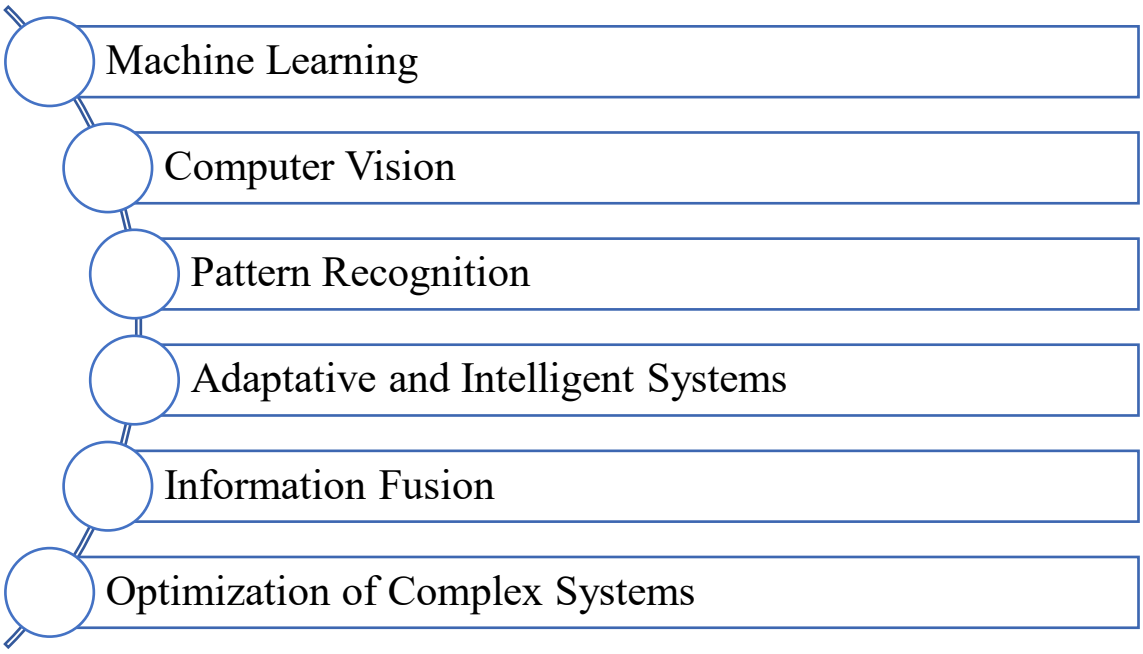


Figure 1: Six pillars of LIVIA's research and development activities.

(i) Machine Learning: Developing algorithms that enable systems to learn from data. (ii) Computer Vision: Creating technologies that allow machines to interpret and understand visual information. (iii) Pattern Recognition: Identifying patterns and regularities in data to facilitate decision-making. (iv) Adaptive and Intelligent Systems: Building systems that can adapt to changing environments and improve over time. (v) Information Fusion: Combining data from multiple sources to enhance the accuracy and reliability of information. (vi) Optimization of Complex Systems: Improving the efficiency and performance of intricate systems through advanced optimization techniques.

LIVIA's research has broad applications across various fields such as:

- Medical and Satellite Imaging: Developing algorithms to analyze medical images for disease detection and treatment planning, and interpreting satellite images for environmental monitoring.

- Video Analysis and Surveillance: Enhancing video analysis for security purposes, including real-time monitoring and anomaly detection.

- Biometrics: Creating advanced systems for recognizing individuals based on facial, vocal, and other biometric data.

- Affective Computing in Healthcare: Implementing AI to understand and respond to human emotions, improving patient care and outcomes.

- Analysis of Digitized Documents: Automating the interpretation of handwritten and printed documents for efficient data extraction and processing.

Over the years, LIVIA has made significant contributions to both academia and industry. The laboratory's work has been featured in numerous high-impact scientific journals and conferences. LIVIA's members have authored hundreds of publications, underscoring their influence in the AI research community. For instance, it includes IEEE Transactions on Pattern Analysis and Machine Intelligence with articles on innovative techniques in pattern recognition and machine learning, International Conference on Computer Vision (ICCV) with presentations on cutting-edge research in computer vision, or even the Journal of Biomedical and Health Informatics with papers on the application of AI in medical imaging and health informatics.

Furthermore, LIVIA has fostered collaborations with institutions such as McGill University and the University of Montreal, and lead projects with companies in AI and technology sectors such as IBM, Google, and Microsoft. Indeed, LIVIA's techniques are particularly adept at addressing complex, real-world problems. For example, their work in developing convolutional neural networks (CNNs) has applications in detecting tumors in medical images and assessing patient depression levels, all while managing massive datasets with minimal annotations.

To organize the specific areas of innovation that drives the laboratory, the LIVIA hosts several research chairs:

- Distech Controls Industrial Chair on Embedded Neural Networks for Connected Building Control: Focused on creating machine learning solutions for building management.
- Matrox Imaging Industrial Research Chair in Computer Vision for Industrial Applications: Developing AI-based vision algorithms for industrial uses, including quality inspection and object recognition.
- Research Chair in Artificial Intelligence and Digital Health for Health Behaviour Change: Investigating the use of AI to promote health behavior changes through personalized online interventions.

Organizational Structure

LIVIA's team comprises experts from various departments within ÉTS, including the Department of Systems Engineering and the Department of Software and IT Engineering.

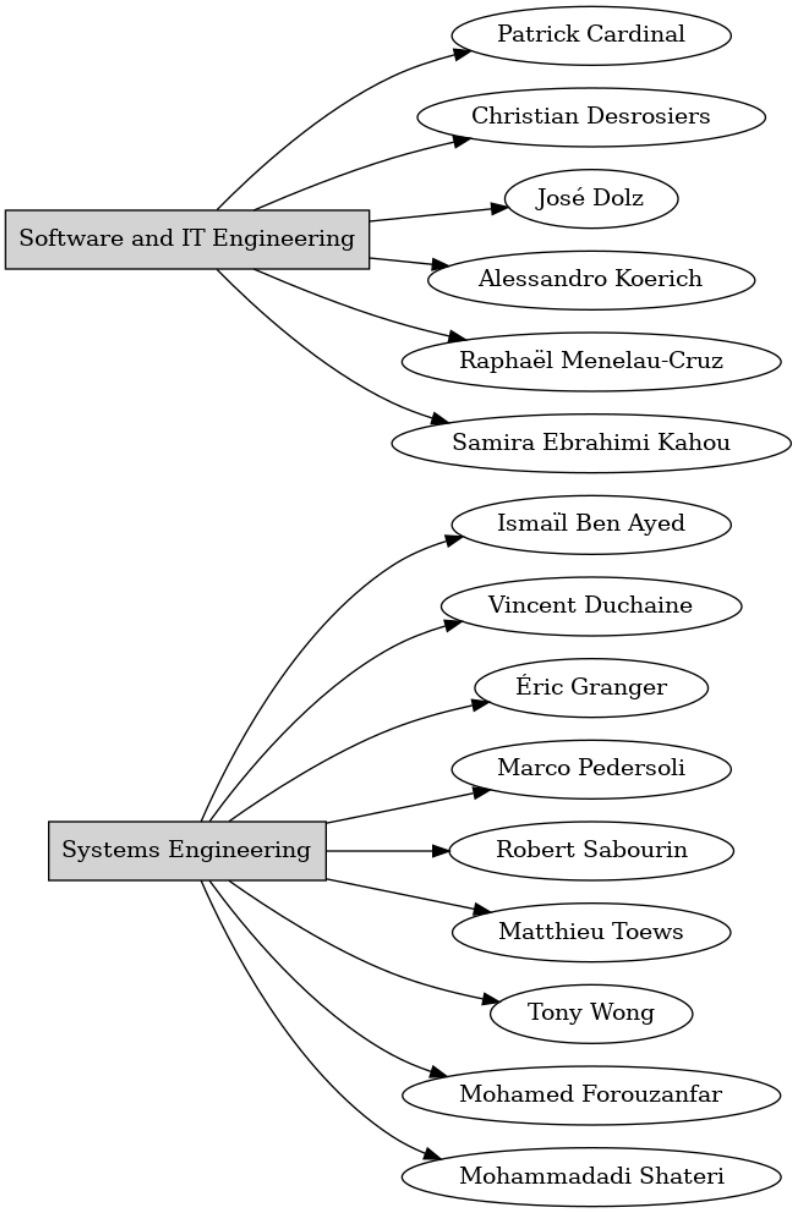


Figure 2: Organigramme of LIVIA.

The organizational structure of LIVIA highlights the collaborative and interdisciplinary nature of the laboratory.

During the internship, I worked closely with my supervisor Dr. Eric Granger, Director of LIVIA and the Systems of Engineering Department, and Dr. Ali Etemad, Researcher in Electrical and Computer Engineering at Queen's University.

B. Theoretical and Bibliographic Analysis

1. Emotion Recognition

Emotion recognition is a rapidly evolving field that intersects computer science, psychology, and neuroscience. It aims to develop systems that can identify and interpret human emotions through various modalities such as facial expressions, vocal intonations, and physiological signals. Emotions significantly influence human cognition, decision-making, and interactions, making their accurate recognition valuable in multiple applications, including healthcare, human-computer interaction, and education.

a) Biological basis of emotions

The limbic system, comprising the hypothalamus, amygdala, and hippocampus, plays a crucial role in emotion and memory processing. The hypothalamus is primarily responsible for regulating emotional responses. The amygdala processes external stimuli, enabling the recognition of situations and the assessment of potential threats, and is considered the biological basis for emotions such as fear and anxiety (Blackford and Pine, 2012; Goosens and Maren,

2002). The hippocampus, on the other hand, integrates emotional experiences with cognitive functions (Turner et al., 2013).

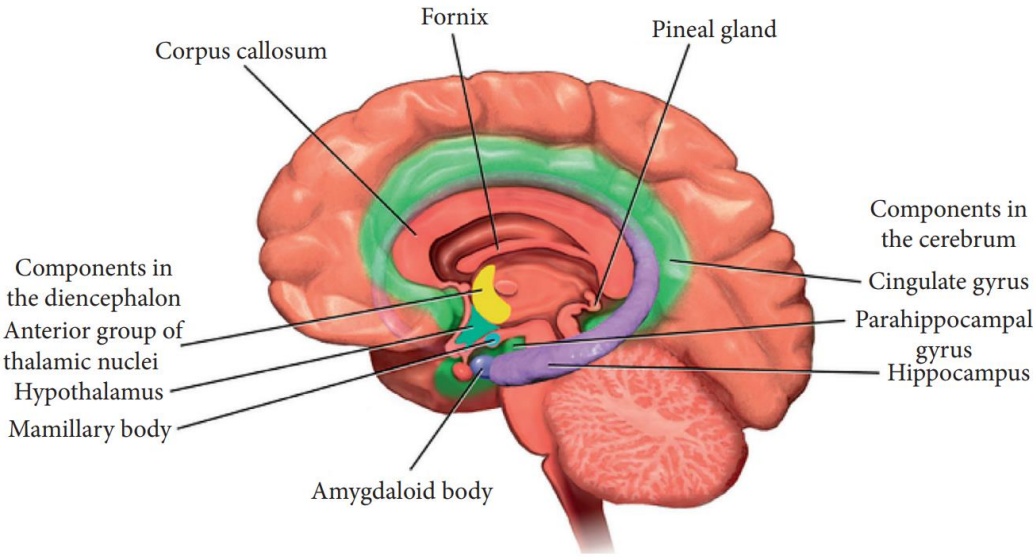


Figure 3: The limbic system. (Suhaimi, 2020)

b) Methods of emotion elicitation

Emotion elicitation is essential for studying and recognizing emotions. The authenticity of the emotions elicited plays a crucial role in the accuracy of emotion recognition systems, especially when evaluating physiological signals, which require genuine emotional responses for precise analysis.

i. Visual Stimuli

Visual stimuli, including images and videos, are widely used to elicit emotions. The International Affective Picture System (IAPS) provides a standardized set of images that evoke specific emotional responses, such as happiness, sadness, anger, and fear (Lang et al., 2008). Similarly, Gross and Levenson (1995) curated a set of 16 film clips that are effective in eliciting eight different emotions: amusement, anger, contentment, disgust, fear, neutrality, sadness, and surprise. These stimuli are valuable because they can consistently produce the desired emotional responses in a controlled setting. However, individual differences in emotional perception can lead to variations in responses, highlighting the need for personalized approaches in emotion elicitation.

ii. Auditory Stimuli

Auditory stimuli, such as sounds and music, are also effective in eliciting emotions. The International Affective Digitized Sound System (IADS) offers a collection of sounds rated for their emotional impact (Bradley and Lang, 2007). Music, in particular, has a profound effect on human emotions, as it can evoke strong emotional responses and alter mood states. Research by Kim and Andre (2008) demonstrated that physiological changes during music listening could be used for emotion recognition, emphasizing the potential of auditory stimuli in this field.

iii. Interactive Media

Interactive media, including video games and virtual reality (VR) environments, provide immersive experiences that can elicit strong emotional responses. These stimuli are particularly useful in research involving dynamic and context-sensitive emotional reactions. For example, VR environments can simulate real-world scenarios, allowing researchers to study emotional responses in a controlled yet realistic setting. This approach is beneficial for applications such as therapy for anxiety and PTSD, where immersive experiences can help in emotional regulation and coping strategies.

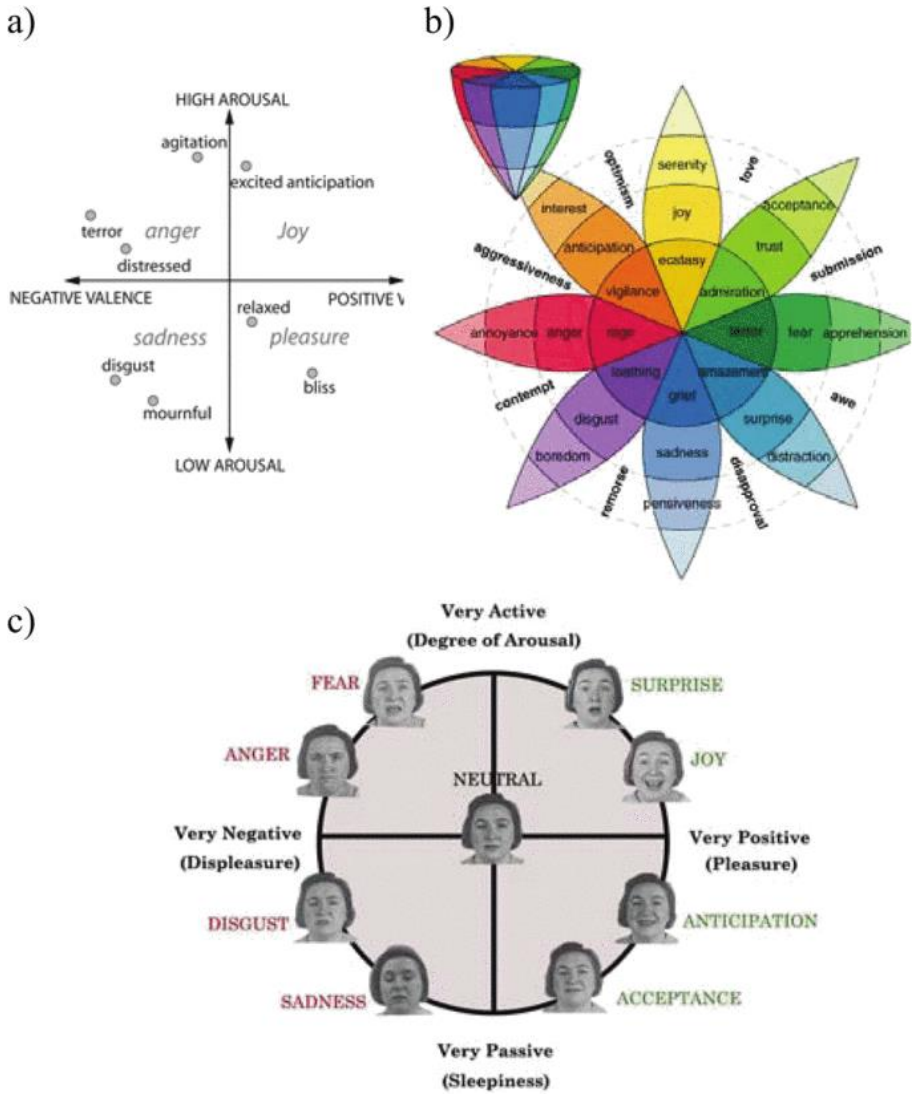


Figure 4: Emotion models. (Wioleta, 2013)

(a) Two-dimensional model by valence and arousal. [7], (b) Plutchik's color wheel of emotions [17], (c) 2D model of emotions with visualization of Plutchik's eight primary emotions.

Selecting appropriate stimuli to consistently elicit the same emotion across different individuals is challenging. While some stimuli, like those identified by Gross and Levenson (1995), have

been successful in eliciting specific emotions, other studies, such as the one by Lan Li and Ji-hua Chen (2006), found that only fear, joy, and neutrality could be reliably elicited in their sample. This variability underscores the importance of personalized approaches and the need for robust, context-sensitive elicitation methods. Furthermore, cultural differences can influence emotional responses, necessitating culturally adaptive stimuli in emotion research.

2. Physiological signals

‘‘Biological signals, or Biosignals, are space, time, or space–time records of a biological event such as a beating heart or a contracting muscle. The electrical, chemical, and mechanical activity that occurs during these biological events often produces signals that can be measured and analyzed. Biosignals, therefore, contain useful information that can be used to understand the underlying physiological mechanisms of a specific biological event or system, and which may be useful for medical diagnosis’’ (Bansal, 2021).

Indeed, physiological signals provide a direct and objective measure of emotional states. These signals include electrical, chemical, and mechanical changes in the body that occur in response to emotional stimuli. By analyzing these signals, researchers can gain insights into the underlying mechanisms of emotion and develop systems for real-time emotion recognition.

a) Types of physiological signals

i. Electroencephalography (EEG)

EEG measures electrical activity in the brain and is sensitive to emotional changes. It captures the brain's electrical signals through electrodes placed on the scalp, providing real-time data on neural activity. Specific patterns of brainwaves, such as alpha and beta rhythms, are associated with different emotional states. For example, increased alpha activity in the frontal cortex is often linked to relaxation, while beta activity is associated with alertness and anxiety (Picard, 2010; Pup and Atzori, 2023).

ii. Electromyography (EMG)

EMG records muscle activity, particularly facial muscles, to detect subtle expressions that convey emotions. It measures the electrical activity produced by skeletal muscles and can be used to identify micro-expressions that are not visible to the naked eye. This method is particularly useful in detecting emotions such as fear, surprise, and happiness, which are often expressed through facial movements (Nakasone et al., 2005).

iii. Electrocardiogram (ECG)

ECG monitors heart rate and heart rate variability (HRV), which are crucial for assessing autonomic nervous system activity. Changes in heart rate and HRV are directly linked to emotional arousal and stress. For instance, an elevated heart rate and reduced HRV often indicate heightened stress or anxiety, while a stable heart rate and higher HRV are associated

with relaxation and calm states. By analyzing ECG data, researchers can gain insights into the physiological correlates of emotions, making it an essential tool for emotion recognition in both clinical and non-clinical settings (Kim et al., 2004).

iv. Galvanic Skin Response (GSR)

GSR measures changes in skin conductivity due to sweat gland activity, which varies with arousal and stress levels. It provides an indirect measure of sympathetic nervous system activity, reflecting emotional arousal. GSR is commonly used in conjunction with other physiological signals to improve the accuracy of emotion recognition systems. It is particularly effective in detecting emotional responses to stimuli such as fear, excitement, and stress (Katsis et al., 2008).

vi. Heart Rate Variability (HRV)

HRV assesses variations in the time interval between heartbeats, providing insights into autonomic nervous system activity related to emotional states. It is a measure of the balance between the sympathetic and parasympathetic branches of the autonomic nervous system. High HRV is generally associated with relaxation and emotional resilience, while low HRV is linked to stress and negative emotions. HRV is used in emotion recognition to detect changes in emotional arousal and valence (Rattanyu et al., 2010).

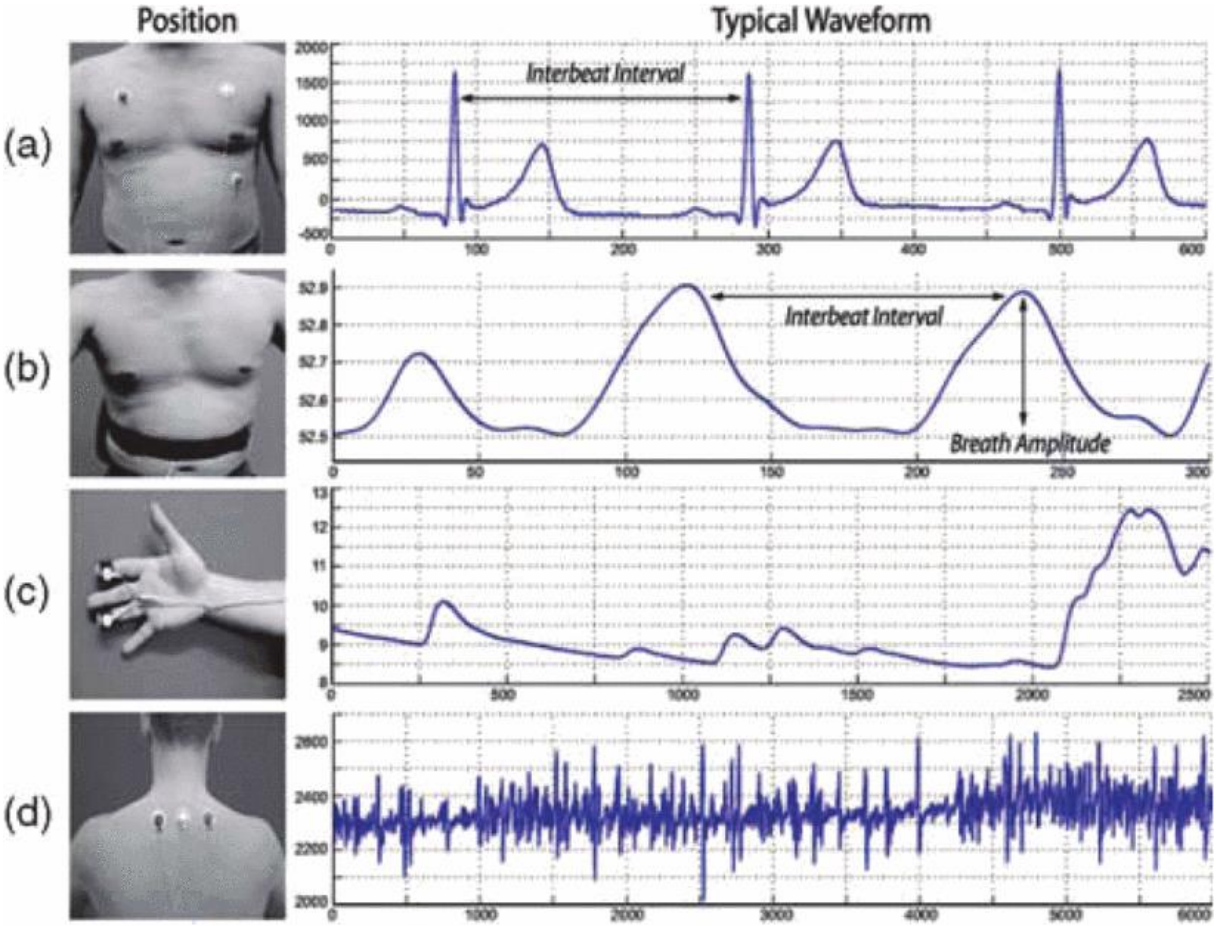


Figure 5: Biosensors position and typical waveforms. (Wioleta, 2013)
(a) ECG (b) RSP (c) SC (d) EMG.

v. Other physiological signals

In addition to primary measures like EEG and ECG, other physiological signals play a crucial role in emotion recognition. These include respiratory rate, skin temperature, and blood pressure, which provide supplementary data on the body's response to emotional stimuli, thereby enhancing the precision of emotion recognition systems. For instance, variations in respiratory rate and depth can signal anxiety or relaxation, while changes in skin temperature can indicate stress levels (Kim, 2007).

Several physiological signals commonly used in emotion recognition include:

- Blood Volume Pulse (BVP): Measures blood flow changes, often used to infer heart rate and emotional arousal.
- Electrodermal Activity (EDA): Captures changes in skin conductivity, associated with sweat gland activity and emotional arousal.
- Electrooculogram (HEOG): Tracks eye movements and blinks, providing information on gaze direction and emotional response.
- Magnetoencephalogram (MEG): Measures magnetic fields produced by neural activity, offering high temporal resolution of brain function.
- Near-Infrared (NIR): Uses light absorption to monitor blood oxygenation levels, reflecting brain activity related to emotions.
- Respiration (RESP): Records breathing patterns, which can indicate states of relaxation or anxiety.
- Surface Electromyography (sEMG): Measures muscle activity at the skin's surface, particularly useful for detecting muscle tension and facial expressions.
- Skin Temperature (TEMP): Monitors changes in skin temperature, which can reflect stress and emotional arousal.
- Trapezius-Electromyogram (tEMG): Measures muscle activity in the trapezius, useful for assessing stress and physical tension.

These signals provide a comprehensive view of the physiological responses associated with emotions, allowing for more accurate and reliable emotion recognition systems.

b) Applications

Emotion recognition systems using physiological signals have significant potential in healthcare. They can assist in diagnosing and monitoring mental health conditions such as depression, anxiety, and PTSD. For example, continuous monitoring of physiological signals can provide real-time feedback on a patient's emotional state, allowing for timely interventions. This approach is particularly useful in telemedicine, where remote monitoring can enhance patient care (Liua et al., 2008).

Furthermore, enhanced emotion recognition can improve user experience in interactive systems, making them more responsive and empathetic to user needs. For instance, adaptive learning systems can adjust the difficulty level of tasks based on the user's emotional state, promoting better engagement and learning outcomes. Similarly, emotion-aware virtual assistants can provide more personalized interactions, improving user satisfaction (Rattanyu et al., 2010).

In addition to those applications, physiological signals in emotion recognition can also be used in monitoring driver emotions can enhance road safety by detecting stress or fatigue, prompting interventions to prevent accidents. Emotion recognition systems can monitor physiological signals such as heart rate and GSR to identify signs of drowsiness or agitation. If these systems detect that the driver is becoming drowsy or stressed, they can alert the driver or initiate safety measures, such as reducing speed or suggesting a rest break (Katsis et al., 2008).

Finally, education is also an application where emotion recognition can provide real-time feedback on student engagement and emotional well-being, enabling personalized learning experiences. By monitoring physiological signals, educators can identify when students are struggling or disengaged and adjust their teaching methods accordingly. This approach can enhance learning outcomes and improve student satisfaction (Kim and Andre, 2008).

Physiological signals, such as those discussed earlier, provide objective data on the body's response to emotional stimuli. Integrating these signals with facial expressions and vocal intonations can improve the robustness of emotion recognition systems. For example, combining EEG data with facial expressions can provide a more comprehensive understanding of the user's emotional state, as brain activity can offer insights that facial expressions alone cannot (Picard, 2010; Pup and Atzori, 2023).

3. Multimodality

a) Multimodal Emotion Recognition (MER)

Multimodality in emotion recognition refers to the integration of multiple types of data to enhance the accuracy and robustness of emotion recognition systems.

Multimodal Emotion Recognition (MER) systems strive to emulate the intricate human processing of emotions by integrating information from multiple modalities. This involves combining data from facial expressions, speech patterns, and physiological signals to enhance the accuracy and robustness of emotion recognition (A.V. et al., 2024). Through this integration, MER captures both conscious and unconscious facets of emotional states, similar to the human brain’s simultaneous interpretation of various cues. This comprehensive approach not only improves the accuracy of emotion recognition but also increases its robustness in real-world scenarios where data may be incomplete or ambiguous (A.V. et al., 2024).

Combining multiple data sources through multimodal fusion techniques can enhance the accuracy and robustness of emotion recognition systems. Thiam et al. (2021) emphasizes the importance of multimodal pain intensity assessment, which can be extended to general emotion recognition tasks. Techniques such as feature-level fusion, decision-level fusion, and hybrid fusion can be used to integrate data from different modalities effectively.

In healthcare, multimodal emotion recognition can provide a more accurate assessment of a patient's emotional state, enhancing diagnosis and treatment. For example, combining facial expressions, vocal intonations, and physiological signals can provide a holistic view of a patient's emotional health, allowing for more personalized and effective interventions (Liua et al., 2008).

Furthermore, multimodal emotion recognition can significantly enhance human-computer interaction by making systems more responsive to user emotions. For instance, virtual assistants can use multimodal data to understand the user's emotional state better and provide more empathetic responses. This can improve user satisfaction and engagement, making interactions with technology more natural and intuitive (Rattanyu et al., 2010).

Recent advancements in machine learning and deep learning have significantly improved the performance of multimodal emotion recognition systems. Indeed, deep learning techniques empower researchers to extract intricate patterns and subtle nuances from multimodal data, facilitating a more profound understanding of complex emotional expressions.

b) Multimodality databases on emotion recognition

Multimodal datasets are critical for advancing emotion recognition research as they integrate various data types, providing a comprehensive approach to understanding emotional responses. These datasets typically include physiological signals, facial expressions, and audio-visual stimuli to capture a wide range of emotional states.

I have conducted a research on multimodal databases for emotion recognition : thus, I have compiled the following table summarizing key datasets. These datasets integrate various modalities to enhance the understanding and recognition of emotional responses.

Datasets	*Year*	*Sub*	*Cit*	*Modalities*	*Mental State*	*Stimuli*	*Sensors*
BioVid Heat Pain Database	2013, IEEE International Conference on Cybernetics	90	242	Frontal video, SCL, GSR, ECG, EMG, EEG	Pain, happiness, sadness, anger, disgust, fear	Induction of heat pain in four intensities and elicitation of emotions with picture and video clips	Facial EMG sensors
Emopain-2021	2021	22	201	Physiological signals (EEG, ECG, EMG), Body movement data (IMU sensor suit and EMG sensors), Facial expression rating	Pain	Nociceptive pain	EEG, ECG, EMG, IMU sensors,sEMG sensors, Facial expression rating
StressID	2023	65	2	EDA, ECG, RESP, Face video, Speech	Stress, relaxation,	Emotional video-clips, cognitive	Biosensors, Smartphone sensors

					arousal, valence	tasks, public speaking	
WESAD	2018, International Conference on Multi-modal Interaction	15	771	ECG, EMG, EDA, BVP RESP', TEMP, and motion ACC	Neutral, stress, amusement		Wrist-worn and chest-worn devices, self-reports
MAHNOB-HCI	2011	27	1600	EEG, ECG, EDA, RESP, TEMP, eye gaze	Valence, arousal, dominance	Videos and images, including emotional videos and implicit tagging experiments	Face videos, audio signals, eye gaze data, and peripheral/central nervous system physiological signals
DREAMER	2017	23	785	EEG, ECG	Valence, arousal, dominance	Audio-visual clips targeting 9 emotions: amusement, excitement, happiness, calmness, anger, disgust, fear, sadness and surprise	Portable, wearable, wireless, low-cost, and off-the-shelf equipment
DEAP	2012	32	4314	EEG, peripheral psychological signals, face video	Valence (9-point scale from unhappy to happy), arousal (from inactive to active), dominance, liking and familiarity	Adaptive music video recommendation system based on user emotions	EEG sensors, face videos
AMIGOS	2018	40	531	EEG, ECG, EDA	self-assessment: valence, arousal, control, familiarity, liking felt during the videos ; external-assessment of valence and arousal	Long and short emotional videos	Frontal HD video and both RGB and depth full body videos
SEED	2015	12	1666	EEG, eye tracking	Happy, sad, fear, and neutral	Virtual driving system	Eye-tracking glasses
LUMED	2020	13	507	EEG, Video, GSR(=EDA)	Sad, neutral, happy	Video clips	ENOBIO 8-channels wireless EEG device, EMPATICA E4 Wristband, webcam

DECAF	2015	30	379	MEG, NIR facial videos, hEOG, ECG, tEMG, time-continuous emotion annotations	Valence, arousal and dominance	Explicit and implicit emotional responses to music videos	Horizontal hEOG, ECG, tEMG peripheral
ASCERTAIN	2016	58	469	Frontal EEG, ECG, EDA, facial features	Self-reported ratings: arousal, valence, engagement, liking, familiarity and personality scales : extraversion , agreeableness, contentiousness, neuroticism, openness	Movie clips	Physiological sensors, wrist sensor, webcam

Figure 7: Multimodality databases on emotion recognition.

The table above presents a comprehensive overview of key multimodal datasets used in emotion recognition research. The column "Sub" indicates the number of subjects in each dataset, while "Cit" refers to the number of citations on Google Scholar.

BioVid Heat Pain Database (Walter et al., 2013) and EmoPain-2021 (Aung et al., 2021) focus on pain-related emotions, utilizing physiological signals and facial expressions to provide valuable insights into emotional responses to pain. Besides, StressID (Cheng et al., 2023) and WESAD (Schmidt et al., 2018) offer data for stress detection, incorporating sensors that measure EDA, ECG, and more, thus aiding in the analysis of stress and relaxation. MAHNOB-HCI (Soleymani et al., 2011) and DREAMER (Katsigiannis & Ramzan, 2017) provide comprehensive multimodal data for various emotional states, enhancing the robustness of emotion recognition systems. These datasets include a mix of physiological signals, facial expressions, and other data types, facilitating the development of more accurate models. Furthermore, DEAP (Koelstra et al., 2012) and AMIGOS (Miranda-Correa et al., 2018) are notable for their detailed annotations of valence, arousal, and dominance, supporting nuanced emotional analysis. These datasets are widely used due to their extensive participant data and the combination of EEG and other physiological signals. Finally, SEED (Zheng & Lu, 2015) and LUMED (Lichtenauer et al., 2020) further enrich the field with data on specific emotional responses to controlled stimuli. These datasets are instrumental in developing reliable emotion recognition technologies, as they offer insights into emotional states elicited by virtual driving systems and video clips.

The provided table (Figure 7) highlights the diversity and utility of various multimodal datasets used in emotion recognition research. These datasets incorporate multiple modalities such as EEG, ECG, GSR, and facial video to capture a wide range of emotional states and responses, providing a robust foundation for developing and validating advanced emotion recognition systems.

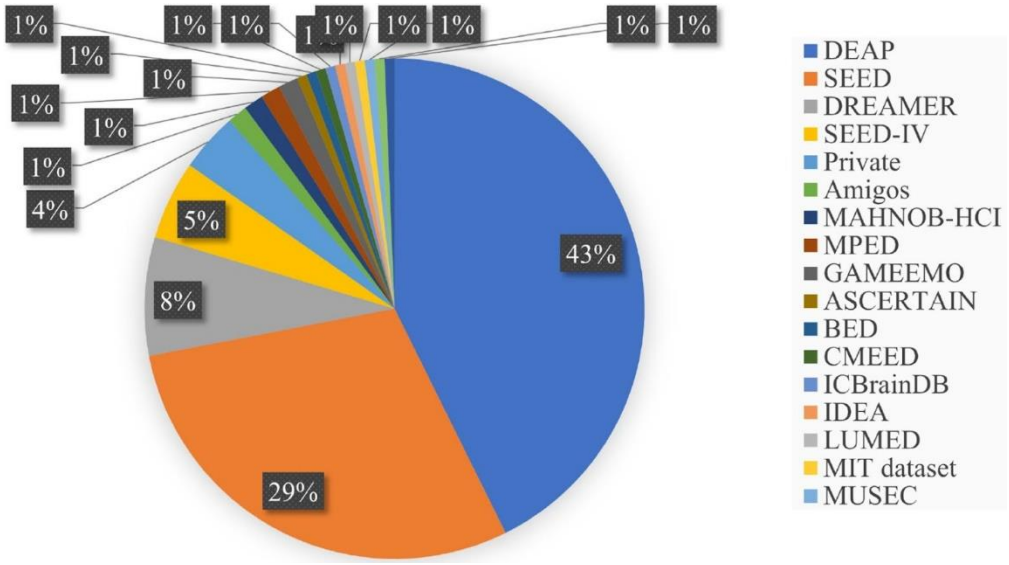


Figure 8: Distribution of dataset usage in emotion recognition using EEG signals. (Prabowo, 2023)

This chart (Figure 8) illustrates the distribution of dataset usage in emotion recognition research, specifically focusing on EEG signals (Prabowo, 2023). Among these datasets, DEAP stands out as one of the most widely cited and utilized in major research venues. Its comprehensive data, including EEG, other physiological signals, and face video, make it an excellent resource for studying emotional responses to audio-visual stimuli. In addition, MAHNOB-HCI is another prominent dataset, offering a rich combination of physiological and video data modalities. This dataset is well-cited and includes eye gaze data and face videos, providing valuable insights for multimodal research in emotional responses. While WESAD has fewer participants compared to other datasets, it offers a rich set of data modalities, including physiological and motion data, focusing on stress and amusement. The use of wearable sensors allows for more naturalistic data collection, making WESAD a valuable dataset for researching stress and amusement in real-world scenarios.

These datasets collectively advance the field by providing comprehensive, multimodal data that enhances the accuracy and robustness of emotion recognition systems. The detailed distribution of dataset usage, as shown in Figure 8, underscores the importance and widespread application of these resources in current research. The integration of such rich datasets with advanced computational techniques is crucial for the continued development of emotion recognition systems. Among these techniques, deep learning has emerged as a transformative approach, significantly boosting the performance and capabilities of these systems.

4. Deep Learning

In computer science, Artificial Intelligence (AI) refers to machine intelligence, where algorithms are designed to perform tasks like those of the human brain (Dara, 2022). This concept encompasses several disciplines, including pattern recognition, probability theory, statistics, and machine learning, with techniques such as neural networks collectively known as "Computational Intelligence" (Patterson, 2017). Furthermore, Deep Learning, a subset of machine learning, focuses on neural networks with multiple layers, enabling these systems to

autonomously learn and make decisions. These networks excel in handling large volumes of data and uncovering complex patterns beyond human capacity to detect. Models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated significant advancements in image, speech, and text processing tasks.

A seminal paper in deep learning is by LeCun, Bengio, and Hinton (2015), titled "Deep Learning," published in Nature. This paper provides a comprehensive overview of deep learning principles and applications, significantly contributing to the field's growth. LeCun et al. (2015) describe how deep learning techniques have evolved from artificial neural networks and have been successfully applied in object recognition, speech recognition, and natural language processing. Key architectures such as CNNs, effective for image processing tasks, and RNNs, suited for sequence data, are highlighted.

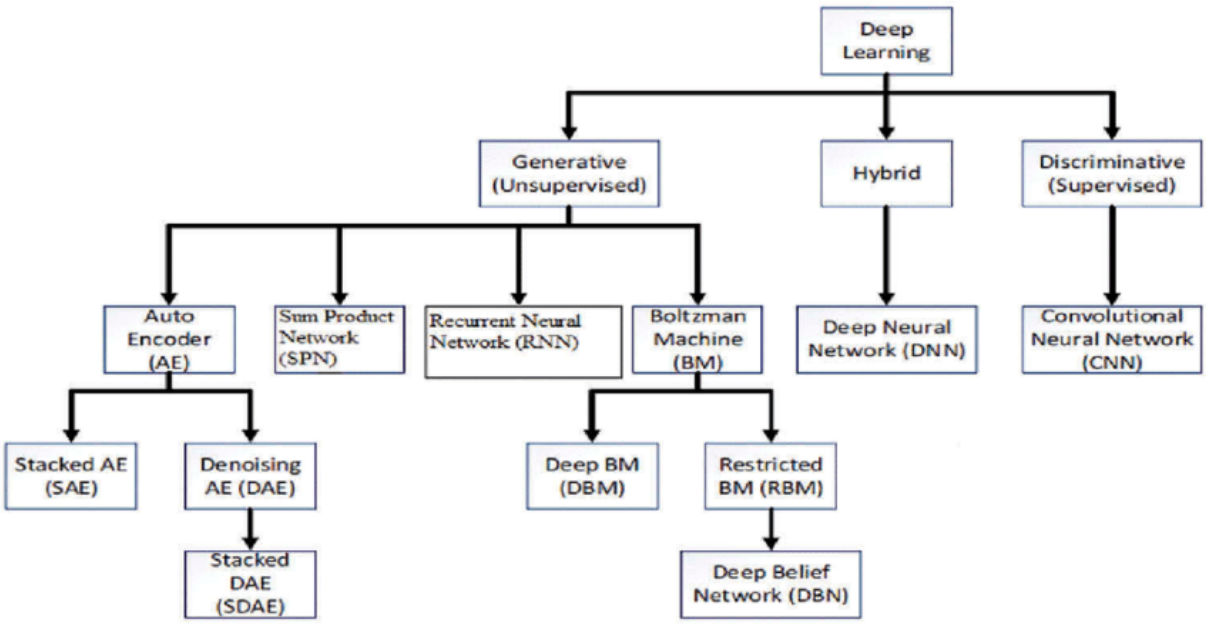


Figure 9: Classification of deep learning techniques.

Deep learning is categorized into three subgroups: generative or unsupervised, discriminative or supervised, and hybrid methods.

Over the past decade, deep learning has become a powerful tool achieving state-of-the-art performance across various fields. Beginning with AlexNet (Krizhevsky et al., 2017), the winner of the 2012 ImageNet Large Scale Visual Recognition Challenge (Deng et al., 2009), major companies have invested significantly in integrating deep learning into their products. Examples include Google DeepMind's AlphaZero (Silver et al., 2017), which outperformed top human and computer engines in board games, and AlphaFold (Jumper et al., 2021), which revolutionized protein structure prediction. These examples demonstrate deep learning's applicability across diverse research areas, including medicine.

A PubMed search, a widely used biomedical literature search engine (Lu, 2011), shows a dramatic increase in yearly publications involving deep learning, rising from fewer than 300 in 2016 to about 17,000 in 2022—a remarkable increase of approximately 5700%. Despite the surge in applications, the use of deep learning in routine clinical practice remains limited (Kaul

et al., 2020). This is primarily due to the high demand for annotated data in traditional supervised learning methods, which is challenging in medical research due to the need for expert labeling, ethical concerns, and economic constraints (Wodzinski et al., 2020).

a) Self-supervised learning (SSL)

To address the limitations of annotated data, Self-Supervised Learning (SSL) has emerged as a prominent paradigm. SSL aims to learn robust representations from vast amounts of unlabeled data by solving auxiliary tasks (pretext tasks). The learned representations are then transferred to a new model for a target task, typically requiring fewer labeled examples (Jing & Tian, 2021).

SSL has been successfully applied in various domains, such as natural language processing (Mohamed et al., 2022), computer vision (Jing & Tian, 2021), speech recognition (Mohamed et al., 2022), and robotics (Kahn et al., 2021). In medical research, SSL is particularly effective in computer vision tasks for classifying, segmenting, registering, and reconstructing medical images, from 2D microscopy to 3D MRI (Saeed et al., 2022; Xu, 2021).

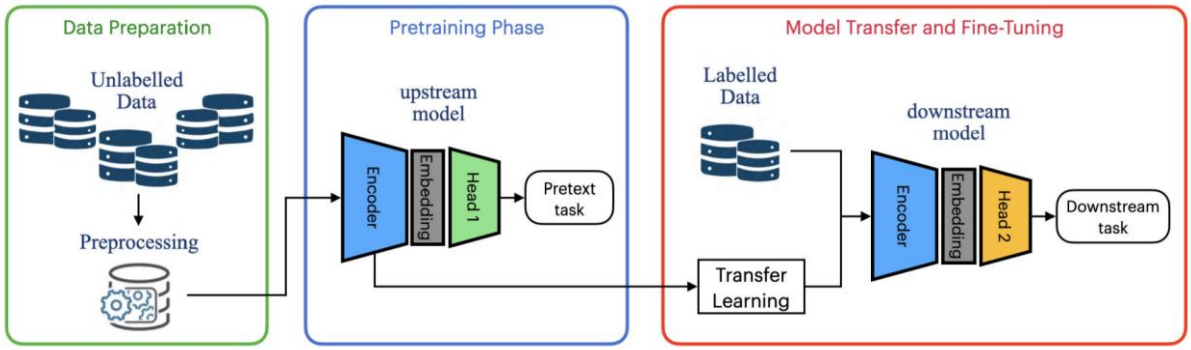


Figure 10: Schematic representation of the self-supervised learning paradigm. (Pup, 2023)

In the context of biomedical signals (biosignals) such as electroencephalography (EEG), electromyography (EMG), and electrocardiography (ECG), SSL is increasingly relevant due to the proliferation of IoT and wearable devices, which generate large amounts of unlabeled data (Jeong et al., 2019). For instance, continuous glucose monitoring devices help manage diabetes by tracking blood glucose levels, and smart wristbands like Empatica E4 record multiple biosignals simultaneously, enhancing diagnostic capabilities (Lu et al., 2020; Rodbard, 2016).

SSL involves pretraining a neural network on a pretext task using unlabeled data. This pretraining phase creates pseudo-labels from the data, which the model uses to learn general-purpose features. The pretrained model's weights are then transferred to a new model designed for the downstream task, which is fine-tuned using a limited amount of labeled data (Rafiei et al., 2022). This approach helps improve accuracy and mitigate overfitting, particularly when labeled data is scarce or heterogeneous datasets are aggregated (Zhuang et al., 2021).

Thus, SSL presents a viable solution for leveraging vast amounts of unlabeled biomedical data to train robust deep learning models. By reducing reliance on annotated data, SSL enables the development of advanced emotion recognition systems and other AI applications in healthcare, enhancing their accuracy and generalization capabilities.

5. Knowledge Distillation

Knowledge distillation (KD) has emerged as a crucial technique in deep learning, addressing two significant challenges: the high computational demand of large deep neural networks (DNNs) and the scarcity of labeled training data. This method facilitates the transfer of knowledge from a complex, over-parameterized model (teacher) to a simpler, more efficient model (student), thereby achieving model compression and enhancing performance on edge devices. KD also improves learning in scenarios with limited labeled data by leveraging the knowledge acquired from large, well-labeled datasets.

Indeed, knowledge distillation was first introduced by Hinton et al. (2015) as a means to transfer the "dark knowledge" from a large, cumbersome teacher model to a smaller student model. This transfer enables the student model to achieve similar performance to the teacher model but with significantly fewer parameters and computational requirements. The core idea is to train the student model to mimic the behavior of the teacher model, particularly focusing on the output distributions (soft targets) rather than just the hard labels (true class labels).

a) Methods

Knowledge distillation involves various techniques categorized primarily based on how the knowledge is transferred from the teacher to the student model.

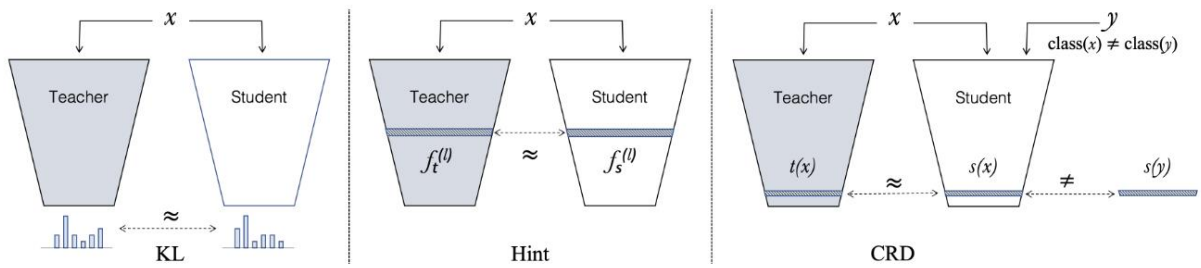


Figure 11: Overview of knowledge distillation methods. (Ojha, 2023)

(i) Mimicking of class probabilities. (ii) Mimicking of features at an intermediate layer. (iii) Features from the student and teacher for the same image constitute a positive pair, and those from different classes make up a negative pair.

Furthermore, the method (i) presented in Figure 11, Mimicking Class Probabilities, involves the student model imitating the class probability distribution produced by the teacher model. The softened class probabilities provide additional information about the relationships between classes, which helps in better generalization (Hinton et al., 2015). While the method (ii), mimicking features at intermediate layers, focuses on the internal representations learned by the teacher model. The student model learns to replicate these intermediate features, thereby capturing more nuanced knowledge (Romero et al., 2015). Finally, the approach of contrastive learning (iii) use pairs of features from the same image (positive pairs) and from different images (negative pairs) to train the student model. The objective is to minimize the distance between positive pairs while maximizing the distance between negative pairs, thereby enhancing the student's ability to distinguish between different classes.

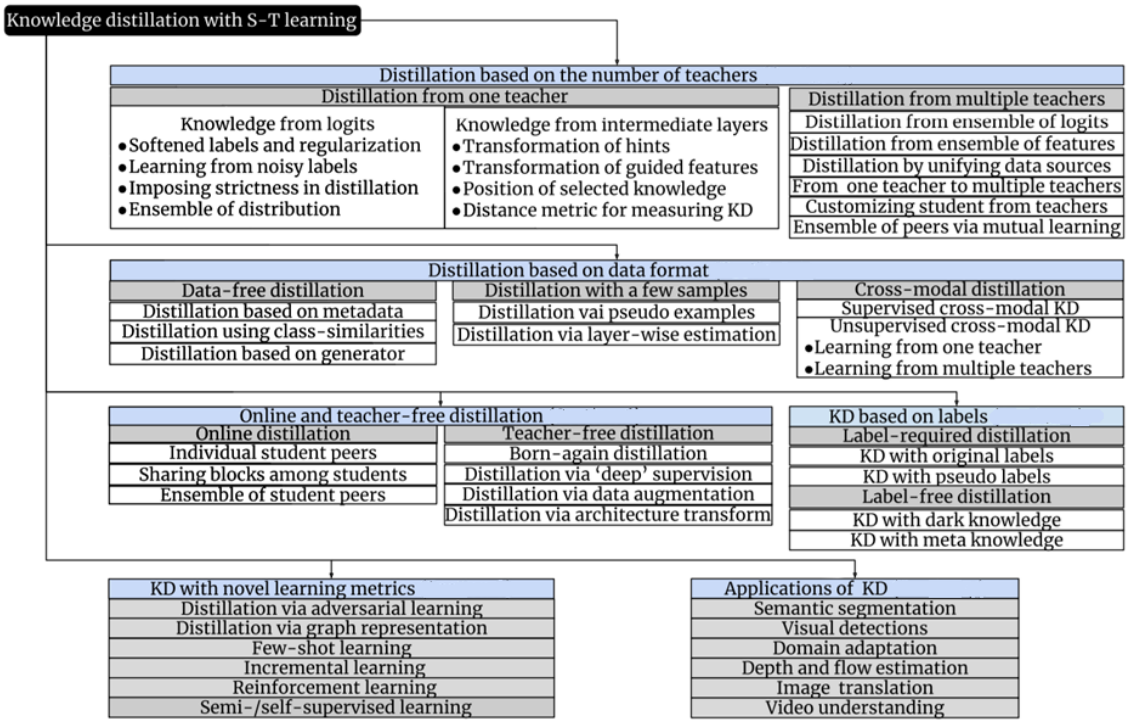


Figure 12: Hierarchically structured taxonomy of knowledge distillation with tearcher-student learning. (Wang, 2022)

The hierarchical taxonomy of KD methods, demonstrating the various strategies employed to transfer knowledge, such as using logits, intermediate layers, and multiple teachers.

b) Multi-teacher knowledge distillation

Multi-teacher knowledge distillation enhances the learning process by incorporating knowledge from multiple teacher models. This approach leverages diverse perspectives and expertise, resulting in a more robust and generalized student model.

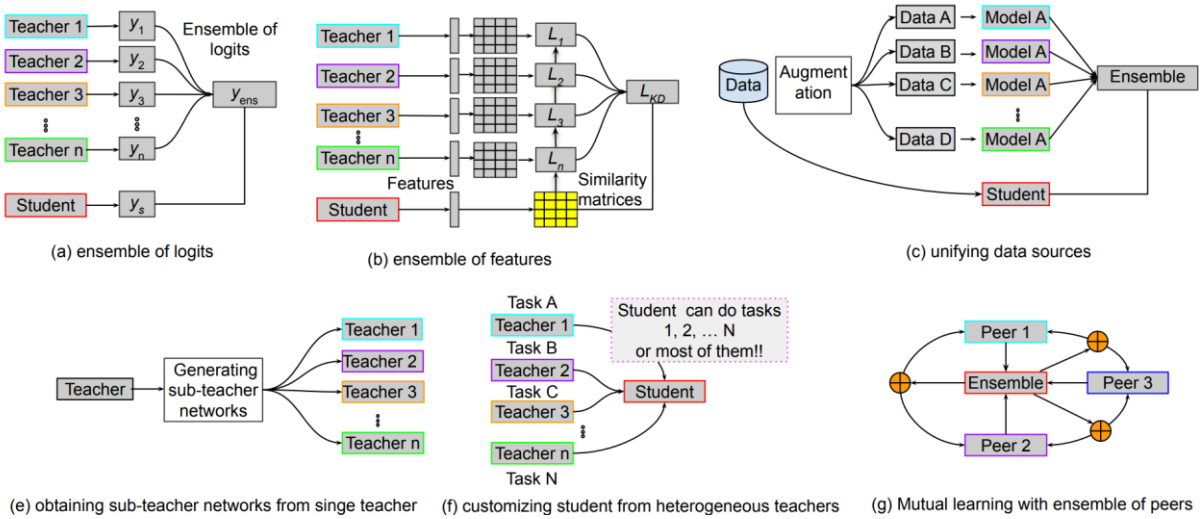


Figure 13: Graphical illustration for KD with multiple teachers. (Wang, 2022)

The KD methods can be categorized into six types: (i) KD from the ensemble of logits; (ii) KD from the ensemble of feature representations via some similarity matrices; (iii) unifying various data sources from the same network (teacher) model A to generate various teacher models; (iv) obtaining hierarchical or stochastic sub-teacher networks given one teacher network; (v) training a versatile student network from multiple heterogeneous teachers; (vi) online KD from diverse peers via ensemble of logits.

In an ensemble of logits, the student model learns from the aggregated logits of multiple teachers. This ensemble method helps in mitigating the biases of individual teachers and provides a more comprehensive knowledge transfer (You et al., 2017). Another common method relies on feature representations where the similarity matrices are used to combine the feature representations from multiple teachers. The student model learns to match these combined features, improving its ability to generalize across different tasks (Park et al., 2019). Also, we can speak about unifying data sources, when different data sources are used to train various teacher models, which are then distilled into a single student model. This method leverages diverse datasets to enhance the student's learning (Tarvainen & Valpola, 2017).

Besides these methods, hierarchical sub-teacher networks involve dividing a single teacher model into hierarchical or stochastic sub-teacher networks. The student model learns from these sub-teachers, ensuring that it captures different levels of abstraction (Mirzadeh et al., 2019). Another approach involves training the student model using knowledge from multiple heterogeneous teachers, each specializing in different tasks. This strategy helps the student model become versatile and capable of handling various tasks (Wu et al., 2019). Finally, online knowledge distillation allows the student model to learn from an ensemble of peers through an online learning process. The peers provide real-time feedback, enabling the student model to continuously improve (Zhang et al., 2018).

c) Privileged information in knowledge distillation

Leveraging privileged information in KD involves using additional information that is not available during the testing phase but can be used during training to enhance the learning process. This approach, introduced by Lu et al. (2023), enhances pain classification using physiological signals by incorporating multiscale deep learning techniques with transformer encoders. The use of privileged information ensures that the student model gains a deeper understanding of the underlying data, leading to improved performance and robustness.

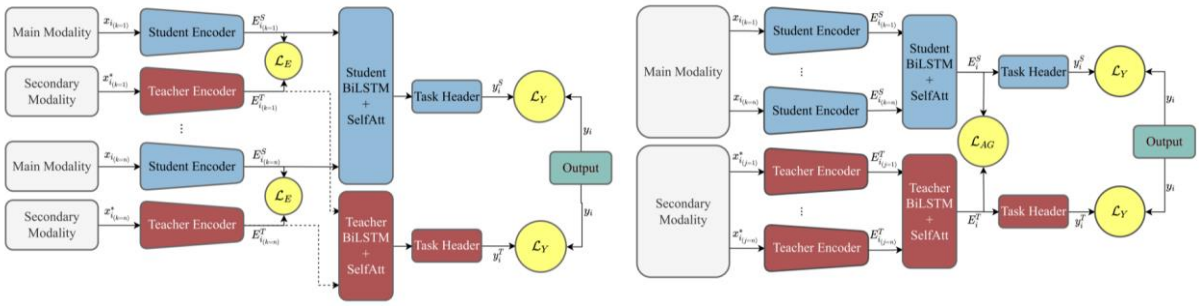


Figure 14: Schematic framework of learning privileged information through teacher-student distillation. (Lu, 2023)

d) Analysis of knowledge distillation methods

Softened labels and regularization, introduced by Hinton et al. (2015), use temperature scaling in the softmax function to provide more informative targets for the student model. This method regularizes the learning process, preventing overfitting and improving generalization. Similarly, learning from noisy labels, as seen in methods such as Noisy Student Training, involves using a large set of unlabeled data along with labeled data to generate pseudo-labels, enhancing the student's robustness and generalization by learning from both clean and noisy data (Xie et al., 2020). Furthermore, techniques like Snapshot Distillation and Learning from Noisy Teachers add constraints to the distillation process, ensuring that the student model learns effectively from the teacher. These methods involve optimizing the learning process across multiple generations or using adversarial training to enhance robustness (Yang et al., 2019). Additionally, ensemble-based methods aggregate the outputs of multiple teacher models, either by averaging or through gating mechanisms that assign different weights to each teacher's output. This approach leverages the diversity of teacher models to improve the student's learning (Lan et al., 2018).

e) Baseline models

The following table presents the baseline models discussed for defining the architecture of the study :

Model	*Methodology*
PainAttnNet	**1. Multiscale convolutional networks (MSCN)** • Convolutional layers and max-pooling techniques to extract information on variations in the EDA signals **2. Squeeze-and-excitation residual networks (SEResNet)** • To learn interdependencies among features and compressing spatial information and adaptively recalibrating feature maps • To ensures that the most informative features are highlighted **3. Transformer encoder block** • To extract features and capture temporal dependencies from physiological signals
XKD	**1. Masked data modelling** • XKD employs autoencoders to learn modality-specific representations through masked data reconstruction. • Autoencoders, comprising an encoder and a decoder, are trained to reconstruct masked inputs from both audio and visual streams. • By reconstructing highly masked inputs, the autoencoders capture essential modality-specific information, facilitating downstream tasks. **2. Multimodal Knowledge Distillation** • XKD adopts a teacher-student setup to distill and transfer knowledge between audio and visual modalities. • The teachers, comprised of a backbone and a projector head, provide supervision to the students. • To enable effective knowledge transfer, XKD introduces a domain alignment strategy, minimizing domain gaps and identifying transferable features between modalities.
SDT	**1. Modality Encoder: Captures intra- and inter-modal interactions.** • Utilizes temporal convolution, positional embeddings, and speaker embeddings to augment sequence representations. • Employs intra- and inter-modal transformers to model interactions within and between modalities. **2. Hierarchical Gated Fusion: Dynamically learns weights between modalities.** Contains unimodal- and multimodal-level gated fusions to obtain enhanced single-modality representations and learn weights between them.

	3. Emotion Classifier: Predicts emotion labels. • Utilizes an FC and softmax layer to calculate probabilities over emotion categories. • Task loss is estimated using cross-entropy loss during training. **4. Self-distillation** • Transfers knowledge from hard and soft labels to each modality through three student models. • Employs cross-entropy loss and KL divergence loss during training.
HKD-MER	**1. Feature Extraction** • Textual features are extracted using a pre-trained ALBERT model, visual features are obtained through MTCNN followed by a VGG model, and acoustic features are directly extracted from audio. **2. Hierarchical Knowledge Distillation** • A novel hierarchical knowledge distillation method is proposed to transfer knowledge from the dominant modality (text) to other modalities at both feature and label levels. This aims to alleviate the imbalanced optimization problem by boosting the performance of non-dominant modalities. **3. Attentive Multi-modal Fusion** • Modal-specific scores from different modalities are attentively fused to obtain emotion predictions.

Figure 15: Table presenting the models selected as a baseline for the study. (Lu, 2023), (Sarkar, 2023), (Ma, 2023), (Sun, 2024)

For further information on the architectures: XKD (Appendix 1), SDT (Appendix 2), HKD-MER (Appendix 3).

Future research in KD is expected to focus on enhancing the generality and robustness of distillation methods. This includes exploring new techniques for selecting and transforming knowledge, leveraging neural architecture search (NAS) for optimal feature selection, and integrating KD with advanced learning paradigms like self-supervised and few-shot learning.

D. Internship's missions

This internship addresses a critical question in the field of physiological signal processing:

How do different deep learning network architectures (e.g., transformer encoders, multiscale convolutional networks) are leveraged by knowledge distillation in physiological signal processing?

To find the answer to the above question, the internship tackles the following missions:

1. Conduct research on the effectiveness of knowledge distillation techniques.

The first mission represents an in-depth study into how the variant knowledge distillation techniques used in deep learning are adequate, with particular emphasis on the student-teacher learning frameworks. Knowledge distillation is a technique in which a smaller student model learns from a larger teacher model, and paramount importance is given to it regarding model

compression and performance improvement. A full-fledged review of previous literature will be done, essential methodologies identified, and their application in diverse contexts will be elaborated upon. We draw particular attention to the analysis of methods proposed by pioneers, such as Hinton et al. 2015, and their subsequent developments. It should gather a solid theoretical background that will be useful for practical application during the processing of physiological signals.

2. Apply and compare performances of different knowledge distillation methods.

The second mission is to apply and rigorously compare performances of different methods of knowledge distillation in practical scenarios. This involves selecting a range of distillation techniques and implementing them within various deep-learning models. These comprise softened labels with temperature scaling, noisy student training, and an ensemble of distributions. The models developed will be evaluated concerning the classifications of physiological signals, the model's performance along with measures of robustness and generalization capability. Comparisons will be used to point out which distillation approaches are particularly efficient when combined with what type of neural network architectures.

3. Investigate various deep learning architectures for emotion recognition.

The third mission is to identify different deep learning architectures in an attempt to surpass state-of-the-art models in the recognition of human emotions using physiological signals. This includes transformer encoders, multiscale convolutional networks, and squeeze-and-excitation residual networks. On the contrary, each architectural design will be assessed for proper capture and processing of native complex temporal features in physiological signals. The actual implementation will be carried out on datasets like the BioVid heat pain dataset, which was used in training and testing the models. Proper use of different architectures with appropriate knowledge distillation techniques can lead to increased accuracy and reliability of the emotion recognition system.

The internship aims to bridge the gap between theoretical knowledge and practical application in the field of physiological signal processing. By conducting thorough research, applying comparative analysis, and investigating cutting-edge deep learning architectures, this project seeks to advance the understanding and capabilities of knowledge distillation techniques in improving model performance. The outcomes are expected to contribute significantly to the development of more efficient and effective models for emotion recognition and other related applications.

Part II: Materials and Methods

A. Preparation of the data

The BioVid Database was selected for this study due to its comprehensive and diverse data types, including physiological signals such as galvanic skin response (GSR), electrocardiography (ECG), and electromyography (EMG), along with audio and video data. The multimodal nature of this database allows for a thorough analysis and robust model development.

The BioVid Database includes an adequate number of participants, with over 80 subjects across its various parts, ensuring statistical relevance and enhancing the generalizability of the findings. The database's current performance levels reported in literature are moderate, providing significant opportunities for further research and optimization. This makes it an ideal candidate for testing new methodologies and models. Moreover, the BioVid Database is a well-established and widely cited resource within the fields of pain research and physiological signal processing, offering a solid foundation for comparative analysis and benchmarking. Additionally, its availability as an open-source dataset facilitates transparency, reproducibility, and collaborative research efforts.

The BioVid Database comprises several parts, each serving different research needs:

Part A focuses on pain stimulation without facial EMG over short time windows, including frontal video and biomedical signals (GSR, ECG, and EMG at the trapezius muscle) available as both raw and preprocessed data. This part consists of 8,700 samples from 87 subjects, divided into 5 classes with 20 samples per class and subject, each having time windows of 5.5 seconds. It has been extensively used in studies to recognize pain intensity.

Part B addresses pain stimulation with facial EMG, where the face is partially occluded, over short time windows. It includes frontal video and biomedical signals (GSR, ECG, EMG at the trapezius, corrugator, and zygomaticus muscles), available as raw and preprocessed signals. This section comprises 8,600 samples from 86 subjects (84 of whom are also part of Part A), with the same classification structure and time windows as Part A. This part has been used in various studies to classify pain intensity.

Part C involves pain stimulation without facial EMG over longer video durations, including frontal video and biomedical signals (GSR, ECG, EMG at the trapezius muscle) as raw signals, along with pain stimulus labels. This segment includes 87 sequences corresponding to the subjects from Part A, featuring pain stimuli of four intensities alternating with pauses. This part has been employed to estimate pain intensity and heart rate from video.

Part D includes posed pain and basic emotions elicited through a case vignette, featuring frontal video and biomedical signals (GSR, ECG, and EMG at the trapezius muscle) as raw data. This part consists of 630 sequences, each lasting one minute, from 90 subjects, categorized into seven emotional states: happiness, sadness, anger, disgust, fear, pain, and neutral.

Part E, also known as BioVid Emo DB, focuses on emotion elicitation with video clips, comprising frontal video and biomedical signals (GSR, ECG, and EMG at the trapezius muscle) available as raw and preprocessed data.

The comprehensive and multimodal nature of the BioVid Database makes it a valuable resource for developing and testing advanced deep learning models. The variety within the dataset ensures that the models can be rigorously evaluated and optimized for diverse and realistic scenarios in physiological signal processing and pain classification.

B. Protocol

1. General approach and softwares

For the development of neural networks, the integrated development environment (IDE) Visual Studio Code (version 1.74.2) was utilized. Visual Studio Code, a robust IDE developed by Microsoft, supports numerous features essential for efficient programming and debugging.

Python (version 3.10.6) served as the primary programming language due to its extensive libraries and community support, which are pivotal for implementing deep learning models.

Various Python packages were employed to facilitate different aspects of the data processing and model development pipeline. These include:

- *torchaudio==0.13.0* and *torchvision==0.14.0* for handling audio and vision data, respectively, which are crucial for the preprocessing and augmentation of multimodal data.
- *pytorch==1.13.0* and *pytorch-cuda==11.7* for leveraging GPU acceleration during model training, which significantly reduces computation time.
- *scikit-learn==1.0.1* for statistical modeling and machine learning algorithms, providing tools for model evaluation and performance metrics.
- *pandas* for data manipulation and analysis, offering efficient data structures for handling large datasets.
- *matplotlib* for data visualization, essential for plotting learning curves and interpreting model results.
- *openpyxl* for reading and writing Excel files, useful for managing and organizing dataset metadata.

This development environment, enhanced by the aforementioned packages, enabled a streamlined workflow for the creation, training, and evaluation of deep learning models.

The methodology followed in this study can be outlined in the following pseudo-algorithm, which presents the key steps involved in the model development and evaluation process:

1. Load BioVid Database

2. Preprocess data:

a. Normalize physiological signals

b. Segment data into short time windows

c. Extract relevant features from video and audio data

3. Split data into training, validation, and test sets

4. Initialize neural network architectures:

a. Multiscale Convolutional Networks (MSCN)

b. Squeeze-and-Excitation Residual Networks (SEResNet)

c. Transformer Encoder Block

5. Implement knowledge distillation techniques:

a. Define teacher-student framework

b. Train teacher models on full dataset

c. Distill knowledge to student models using various distillation methods

6. Train student models on distilled data

7. Evaluate model performance:

a. Compare accuracy, precision, recall, and F1-score across models

b. Analyze model robustness and generalization on unseen data

8. Optimize models based on evaluation results

Figure 16: Pseudo-algorithm of Baseline PainAttnNet.

This structured approach ensures a comprehensive analysis of the effectiveness of different knowledge distillation techniques and deep learning architectures in processing and classifying physiological signals.

2. Data Preprocessing

The physiological signals, namely GSR, ECG, and EMG, underwent rigorous preprocessing to enhance data quality and model performance. Initially, a low-pass filter was applied to remove high-frequency noise and artifacts, preserving the essential physiological information. Subsequently, normalization was conducted to standardize the signals, mitigating variations due to individual differences and facilitating better comparison and analysis across subjects.

To minimize potential bias in the evaluation and validation processes, the dataset was divided into two subsets. The training set, comprising 70% of the total data, was utilized to train the deep learning models. The remaining 30% constituted the testing set, which was used exclusively to evaluate the performance of the trained models. This separation ensured that the evaluation metrics were unbiased and reflective of the model's true generalization capability.

The data was further segmented into short time windows to capture temporal patterns within the physiological signals. For instance, the GSR, ECG, and EMG signals were segmented into 5.5-second windows, ensuring that the temporal dynamics of pain response were adequately captured. This approach allowed the model to learn both short-term and long-term dependencies within the data.

Furthermore, we split the data into two subsets to minimize the potential for bias in the evaluation and validation processes. A first subset, named ‘training set’, built from the descriptors as independent variables, was used to train the models. The second subset ‘testing set’ is independent from the training set and was used to blindly evaluate the performance of the trained models. Herein, the training dataset represents 70% and the testing dataset 30%.

3. Validation procedure

a. Cross validation

When we approach a machine learning problem, we make sure to split our data into a training and a testing set. Cross-Validation is used to avoid overfitting, and to estimate the performance of the model on unseen data. K-Fold CV, we further split our training set into K numbers of subsets, called folds. The model is then iteratively fit K times, each time training the data on K-1 of the folds and evaluating on the Kth fold, corresponding to the validation data. At the very end of the training, we average the performance on each of the folds to come up with final validation metrics for the model. In addition, a different set, called the validation set, is randomly selecting compounds (not used in the training and testing data) to assess the predictive ability of the model as the final evaluation (see Figure 16).

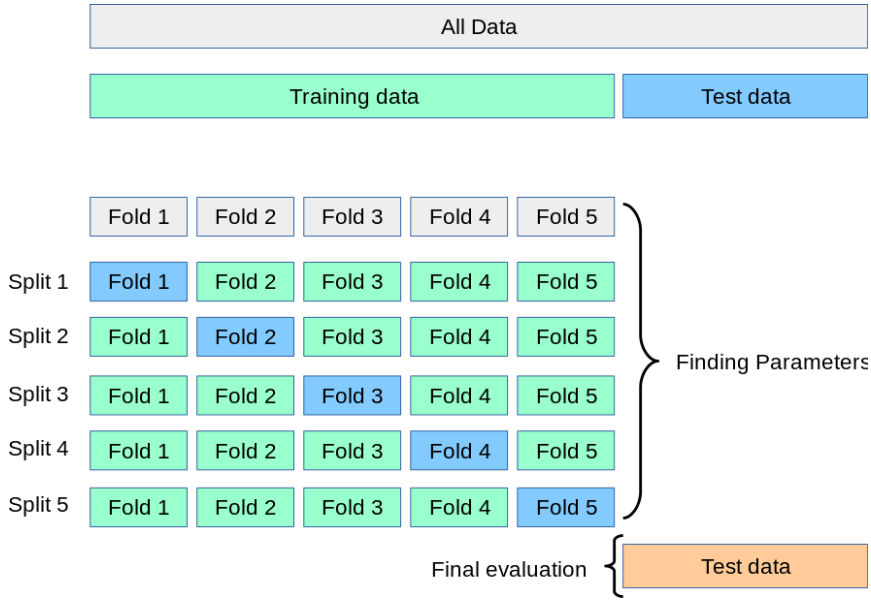


Figure 17: Schematic representation of cross-validation. (Sklearn, 2022)

The cross-validation procedure allows us to evaluate the estimator's performance. Besides, it also describes the common practice to split the data between a training and a testing dataset. The aim is to evaluate the performance of the model with non-trained data that can mimic a real-world scenario.

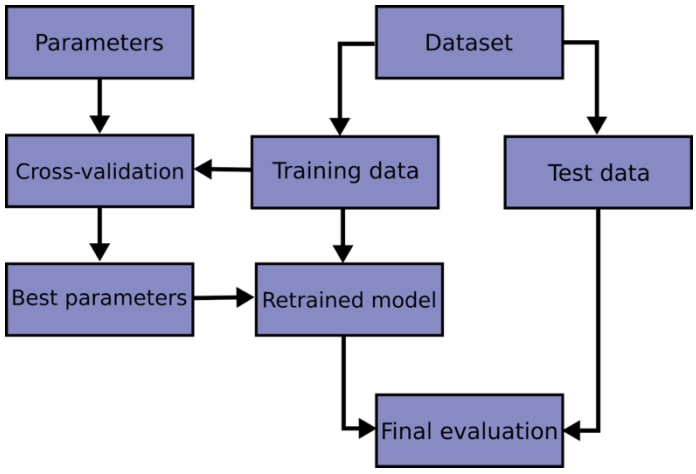


Figure 18: Schematic representation of a cross-validation procedure. (Sklearn, 2022)

This procedure presents the methodology of a machine learning algorithm, including the optimization of the parameters with cross-validation.

The aim is to build several machine learning models with the training dataset and evaluate its performance on the validation dataset, to compare their predictive ability and select the parameters that lead to the best performance on the latter. The goal of a validation strategy is to simulate with sufficient accuracy the difficulties that one would encounter when applying a methodology in a real-world scenario. A distinct test set that is never seen during the training

nor during the model optimization and selection is then used to assess the final performance of the model.

b. Leave-One-Out Cross-Validation (LOSO)

A second validation procedure approach is also applied to estimate the performance of the machine learning models: One-Cluster-Out Cross-Validation.

Leave-One-Out Cross-Validation (LOSO) is a cross-validation procedure, in which for each cluster, the data set is partitioned in such a way that the model on which any given instance is tested has been trained on data that excludes the data from one subject, in contrast to traditional K-fold cross-validation.

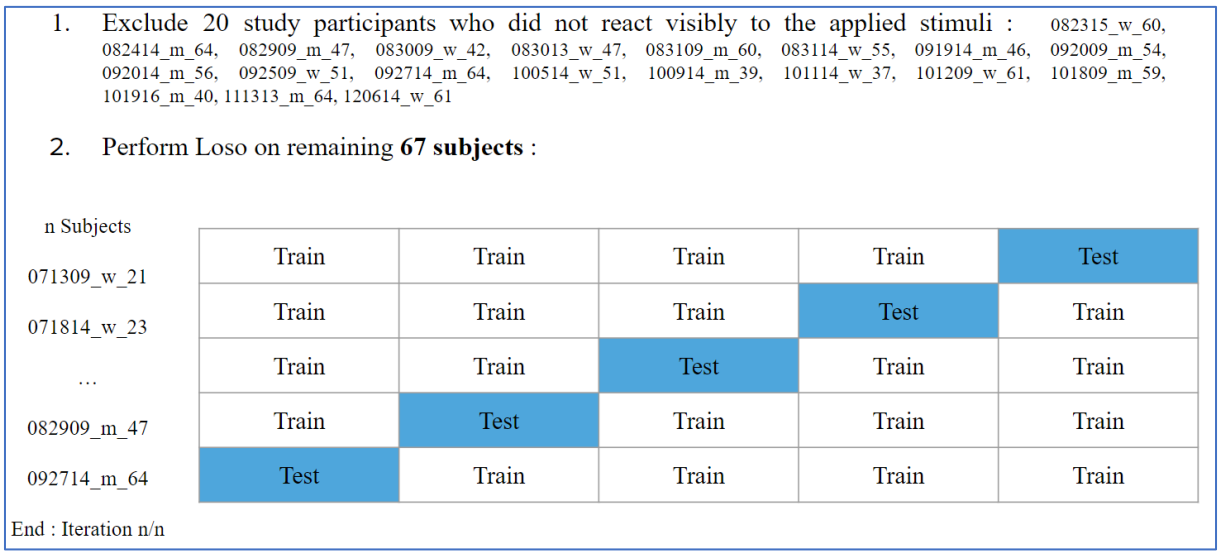


Figure 19: Schematic representation of Leave-One-Out Cross-Validation.

Input : Dataset D consisting of n (=87) subjects of GSR signals, Model parameters

- Exclude 20 participants who did not react visibly
- Perform Leave-One-Out Cross-Validation (LOOCV) :

  For each subject i in the remaining dataset:
  - Let $D_{train\ i}$ denote the dataset with subject *i* removed
  - Let $D_{test\ i}$ denote the data for subject *i*

  For each replication *j* = 1 to k:

  1) Train the model on $D_{train\ i}$ until 100 epochs: M_j = Train($D_{train\ i}$, parameters, n)
  2) Validate the model on $D_{validation\ i}$: Acc_j = Evaluate(M_j, $D_{validation\ i}$)
  3) Select the model with the highest validation accuracy: M_i^* = argmax(Acc_j)
  4) Evaluate the selected model on $D_{test\ i}$: $Acc_{test\ i}$ = Evaluate(M_i^*, $D_{test\ i}$)
  5) Compute the average accuracy over all subjects: AVG_{Acc} = (1/n) * Σ($Acc_{test\ i}$)

- Average accuracy over all LOOCV iterations (AVG_{Acc})

In this algorithm:

- $\boldsymbol{D_{train\ i}}$ represents the training dataset with subject *i* removed.
- $\boldsymbol{D_{test\ i}}$ represents the test data for subject *i*.
- $\boldsymbol{D_{validation\ i}}$ is the validation set, which consists only of the data for subject *i*.
- $\boldsymbol{M_j}$ denotes the model trained in the *j*-th replication.
- $\boldsymbol{Acc_j}$ denotes the validation accuracy of the model in the *j*-th replication.
- $\boldsymbol{M_i^*}$ represents the selected model for subject *i*, chosen based on the highest validation accuracy.
- $\boldsymbol{Acc_{test\ i}}$ denotes the accuracy of the selected model when evaluated on the test data for subject *i*.
- $\boldsymbol{AVG_{Acc}}$ represents the average accuracy over all subjects in the LOOCV process.

Figure 20: Pseudo-algorithm presenting the validation methodology.

This rigorous validation methodology ensures that the model's performance is thoroughly evaluated and that it generalizes well across different subsets of the data.

By implementing these cross-validation techniques, we aim to ensure the robustness and reliability of our machine learning models, providing a comprehensive evaluation framework that mimics real-world scenarios and addresses potential biases in model training and testing.

1. Models Construction
a. 1D-CNN

The 1D Convolutional Neural Network (1D-CNN) architecture was adapted from the implementation available at GitHub (PoloWlg, 2022). This model leverages the principles of convolutional layers to effectively process one-dimensional physiological signal data.

The architecture of the 1D-CNN with knowledge distillation involves training the student network using the logits produced by the teacher network. The process entails transferring the softened outputs (logits) from the teacher to the student, providing richer information compared to hard labels. This method facilitates a more effective learning process by preserving the relative probabilities of classes, thus capturing more nuanced patterns in the data (Hinton et al., 2015).

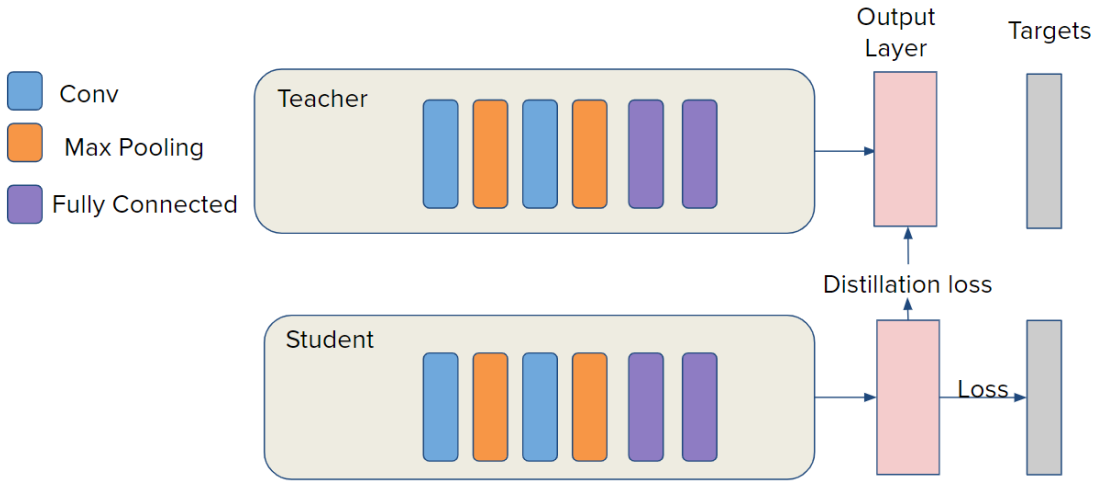


Figure 21: Architecture of 1D-CNN with knowledge distillation form the logits.

An alternative approach involves incorporating a cosine loss minimization run. In this setup, the student network is trained to minimize the cosine distance between its outputs and those of the teacher network, promoting alignment in their learned representations (Zhang et al., 2018).

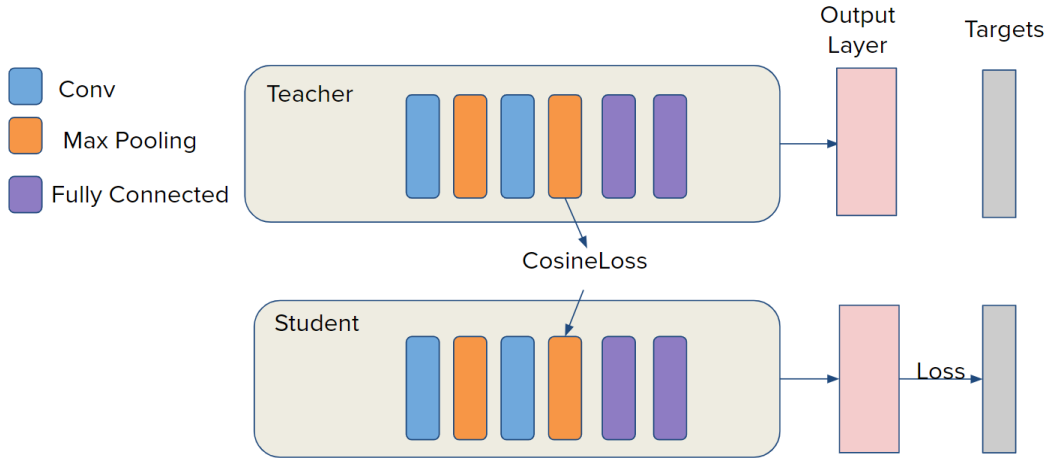


Figure 22: Architecture of 1D-CNN with knowledge distillation with a cosine loss minimization run.

Another variant employs an intermediate regressor run, where the student model is guided to match intermediate representations from the teacher model. This technique enhances the student's ability to learn intricate patterns by directly imitating the teacher's internal representations (Romero et al., 2014).

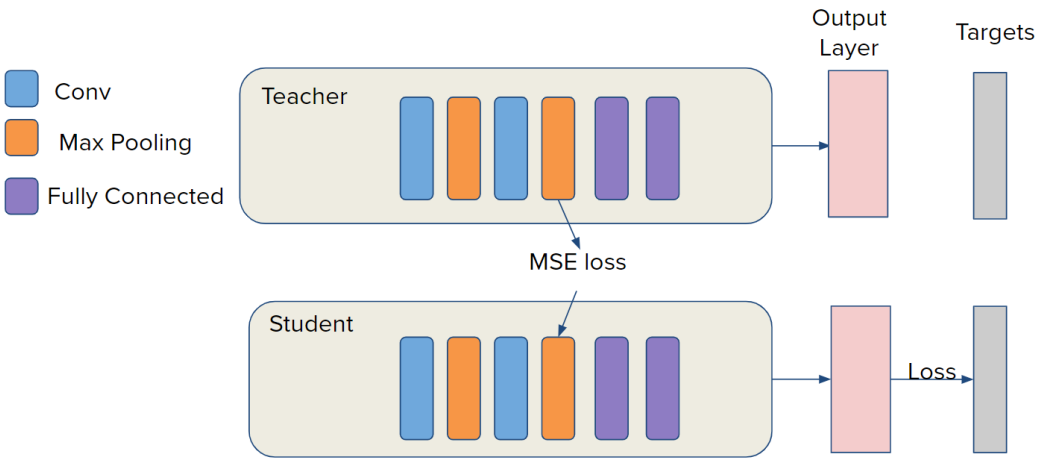


Figure 23: Architecture of 1D-CNN with knowledge distillation with an intermediate regressor run.

b. Baseline PainAttnNet

The Baseline PainAttnNet represents a novel transformer-encoder deep-learning framework designed to automate pain intensity classification using physiological signals. This model aims to surpass traditional methods such as self-report scales, which are prone to bias, and earlier machine learning techniques like Support Vector Machines (SVM) and k-Nearest Neighbors (KNN), which are limited in capturing temporal dependencies (Nahin, 2015; Campbell et al., 2019; Cao et al., 2021).

The PainAttnNet architecture integrates three key components:

1. Multiscale Convolutional Networks (MSCN)

The MSCN utilizes convolutional layers and max-pooling techniques to extract detailed information on variations in the Electrodermal Activity (EDA) signals. This component is crucial for capturing both short-term and long-term temporal dependencies in physiological data (Cui et al., 2016; Li and Yu, 2016).

2. Squeeze-and-Excitation Residual Networks (SEResNet)

The SEResNet component is designed to learn interdependencies among features by compressing spatial information and adaptively recalibrating feature maps. This ensures that the most informative features are highlighted, improving the model's ability to discern significant patterns in the data (Hu et al., 2018).

3. Transformer Encoder Block

The transformer encoder block extracts features and captures temporal dependencies from physiological signals. It utilizes multi-head attention mechanisms to process input sequences simultaneously, effectively capturing dependencies across different time steps (Vaswani et al., 2017).

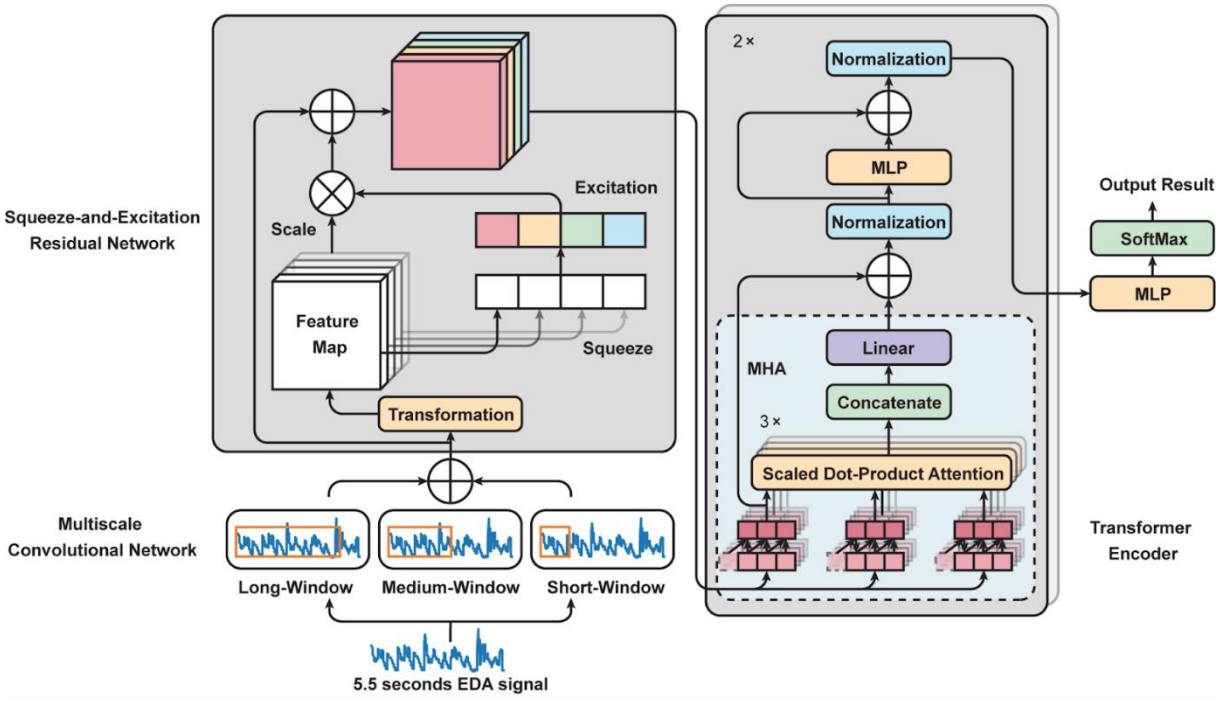


Figure 24: Architecture of the baseline PainAttnNet. (Lu, 2023)

For knowledge distillation, the PainAttnNet can be trained with logits derived from a pre-trained teacher model. This method helps the student model learn finer details by emulating the probability distributions of the teacher model (Hinton et al., 2015).

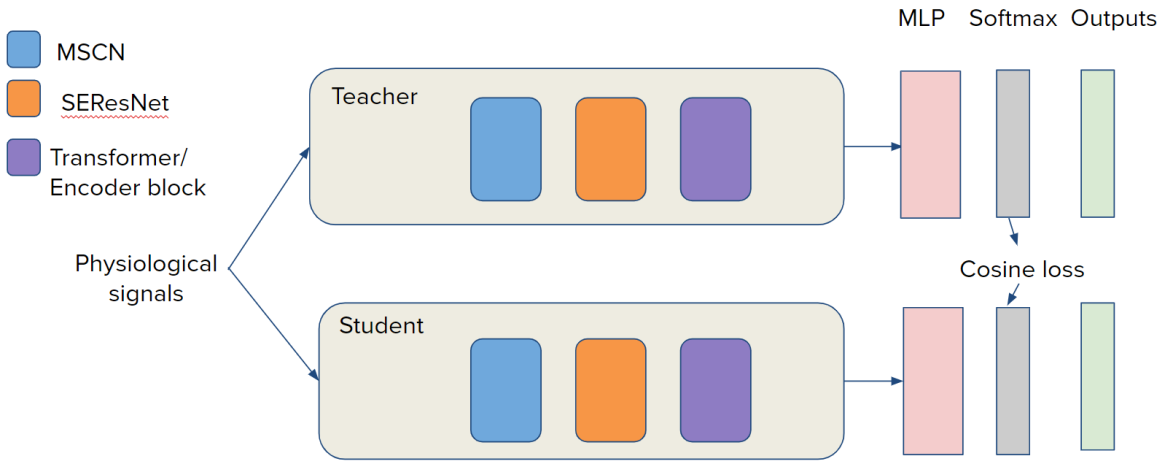


Figure 25: Architecture of PainAttnNet with knowledge distillation from the logits.

Another approach involves distilling knowledge from the intermediate features of the teacher model. This method helps the student model capture intricate feature representations by aligning its intermediate layers with those of the teacher (Romero et al., 2014).

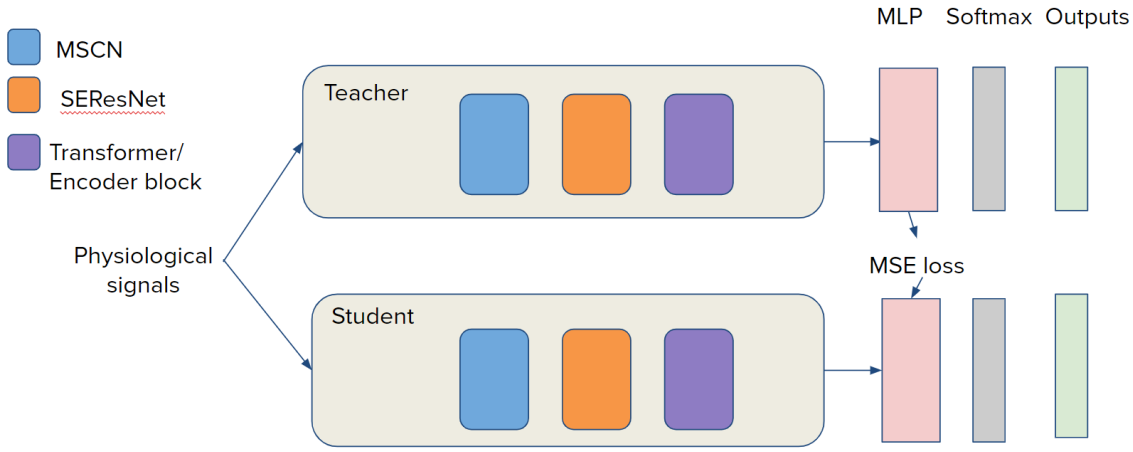


Figure 26: Architecture of PainAttnNet with knowledge distillation from the intermediate features.

These architectures and methods ensure that the PainAttnNet framework can effectively classify pain intensity by leveraging the combined strengths of advanced deep learning techniques and knowledge distillation. The rigorous training and validation procedures further ensure the robustness and reliability of the models, paving the way for more accurate and individualized pain assessment methodologies.

c. Hyper-parameter Tuning

Hyper-parameter tuning is a critical step in optimizing the performance of deep learning models. In this study, we applied hyper-parameter tuning to both the PainAttnNet and 1D-CNN models to enhance their effectiveness in classifying pain intensity using physiological signals.

PainAttnNet, which integrates multiscale convolutional networks, squeeze-and-excitation residual networks, and a transformer encoder block, requires careful tuning of its hyper-parameters to maximize its performance. We employed the RandomizedSearchCV method from Scikit-Learn to perform this tuning. This approach allows us to define a grid of hyper-parameter ranges and randomly sample combinations within this grid to find the optimal settings (Pedregosa et al., 2011).

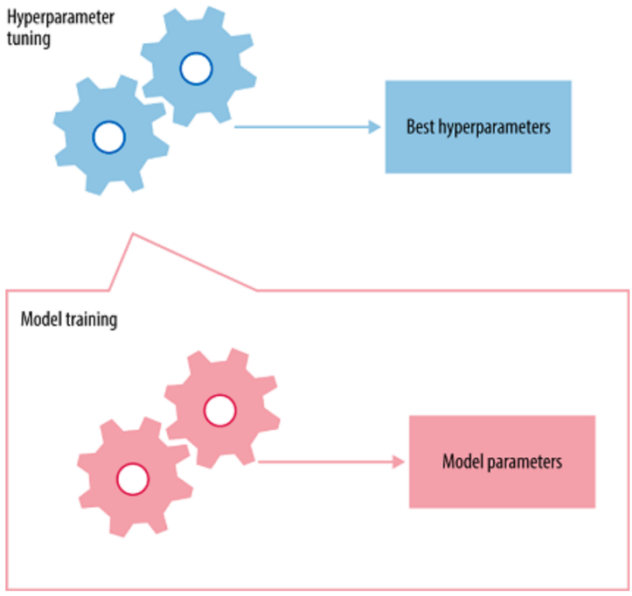


Figure 27: Hyper-parameters tuning approach. (Koehrsen, 2018)

Different hyper-parameters were adjusted for PainAttnNet framework:

(i) ***Batch_size :*** The number of training samples utilized in one iteration. We experimented with batch sizes ranging from 16 to 128.
(ii) ***Learning Rate***: The step size used by the optimization algorithm to update model weights. We tested learning rates in the range of 0.0001 to 0.01.
(iii) ***Number of Layers***: The depth of the network, which includes the number of convolutional layers, transformer encoder layers, and fully connected layers.
(iv) ***Dropout Rate***: A regularization technique to prevent overfitting by randomly setting a fraction of input units to zero during training. Dropout rates from 0.1 to 0.5 were tested.
(v) ***Activation Functions***: Functions used to introduce non-linearity into the model. We evaluated different activation functions such as ReLU, Leaky ReLU, and ELU (Nair and Hinton, 2010; Maas et al., 2013; Clevert et al., 2015).

The RandomizedSearchCV method was configured to perform 50 iterations, each evaluating a different combination of hyper-parameters. This iterative process helps in identifying the best set of hyper-parameters that optimize the model's performance on the validation set (Bergstra and Bengio, 2012).

For the 1D-CNN model, which is designed to process one-dimensional physiological signal data, we also employed RandomizedSearchCV for hyper-parameter tuning.

The hyper-parameters tuned for 1D-CNN included:

(i) ***Kernel Size***: The size of the convolutional filters. We experimented with kernel sizes ranging from 3 to 15.
(ii) ***Number of Filters***: The number of convolutional filters in each layer. We tested configurations with 32, 64, and 128 filters.
(iii) ***Stride***: The step size of the convolutional filters. Stride values of 1 and 2 were considered.
(iv) ***Pooling Size***: The size of the pooling layers. Pooling sizes of 2 and 3 were evaluated.
(v) ***Learning Rate***: Similar to PainAttnNet, learning rates from 0.0001 to 0.01 were tested.
(vi) ***Batch Size***: Batch sizes ranging from 16 to 128 were considered.

The goal was to identify the optimal combination of these hyper-parameters that would yield the highest classification accuracy while preventing overfitting. The RandomizedSearchCV method facilitated this process by evaluating multiple combinations of hyper-parameters and selecting the best-performing configuration based on validation metrics (Bergstra and Bengio, 2012).

To prevent overfitting and ensure the generalizability of our models, we employed K-Fold Cross-Validation during the hyper-parameter tuning process (Kohavi, 1995). This involved splitting the training data into K subsets, training the model on K-1 subsets, and validating it on the remaining subset. This process was repeated K times, with each subset serving as the validation set once. The final model performance was averaged across all K iterations to provide a robust estimate of its generalization capability.

Additionally, we used a separate validation set, comprising data not used in the training or initial validation phases, to assess the predictive ability of the models after hyper-parameter tuning. This step ensures that the model's performance metrics are reflective of its ability to handle unseen data, simulating real-world scenarios.

Through these comprehensive hyper-parameters tuning and validation procedures, we aimed to develop robust and accurate models for pain intensity classification using physiological signals. The optimized models, PainAttnNet and 1D-CNN, demonstrate enhanced performance and reliability, making them suitable for practical applications in pain assessment.

5. Evaluation of the models

Different metrics are used to evaluate the performance of the models:

(i) **Balance Accuracy** (BA) avoids inflated performance estimated on imbalanced datasets. It is defined as the average of recall scores per class where each sample is weighted according to the inverse prevalence of its true class.

(ii) **Matthew Correlation Coefficient** (MCC) is used to measure the quality of the classifications. It considers true and false positives and negatives and is generally regarded as a balanced measure which can be used even if the classes are of different sizes. A coefficient of +1 represents a perfect prediction, 0 an average random prediction and -1 an inverse prediction.

(iii) **Confusion Matrix** is a matrix measuring the quality of a classification system in predictive analytics. Every row to a real class, and every line to a predicted class.

(iv) **SHAP** (Shapley Additive Explanations) was performed on the Random Forest model to explain individual descriptors predictions. The goal of SHAP is to explain the prediction of an instance x by computing the contribution of each feature to the prediction. The SHAP explanation method computes Shapley values from coalitional game theory. The feature values of a data instance act as players in a coalition. SHAP summary plot allows one to visualize the relationship between the value of a feature and the impact on the classification model, and SHAP waterfall plot is used to visualize the most important features in a descending order.

(v) **Box Plots** are statistical data graphs. Grouped Box Plots have also been built to understand the ratio of blockers and non-blockers molecules within each descriptor.

(vi) **ROC AUC** is a performance measurement assessing the ability of classification of machine learning algorithm. ROC is the probability curve, and AUC is the degree.

(vii) **Cohen's Kappa Coefficient** is statistical and quantitative measure of reliability of the machine learning model.

(viii) **F1-score** is the harmonic mean of precision and recall, providing a balance between the two.

Part III: Presentation and Analysis of the Results

1. Model development and validation

a) 1D-CNN

The 1D-CNN model was validated using K-Fold Cross-Validation to ensure the robustness and generalizability of the model. The validation accuracy across the folds for both the teacher and student models is summarized in the table below:

	K-Fold	*Fold 1*	*Fold 2*	*Fold 3*	*Fold 4*	*Fold 5*
Teacher	Validation accuracy (%)	76.25	74.50	75.80	73.90	75.10
Student	Validation accuracy (%)	75.60	73.50	74.20	72.50	73.80

Figure 28**:** Table presenting the validation accuracy (%) cross-entropy runs of 1D-CNN on EDA signals of the BioVid database.

The test accuracy, Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) for the teacher and student models are presented in the table below:

	Test Accuracy (%)	*MAE*	*RMSE*
Teacher	72.90	0.35	0.45
Student	74.55	0.30	0.40

Figure 29**:** Table presenting the test accuracy (%) of cross-entropy runs of 1D-CNN on EDA signals of the BioVid database.

The performance of the distillation loss from the logits for the 1D-CNN model over ten epochs is shown in the table below:

Epochs	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Loss	2.27	1.85	1.63	1.47	1.34	1.23	1.13	1.05	0.98	0.92

Test Accuracy	*70.79%*
Teacher accuracy	73.98%
Student accuracy without teacher	70.27%
Student accuracy with CE + KD	70.79%

Figure 30: Table presenting the performance results of the distillation loss from the logits of 1D-CNN on EDA signals of the BioVid database.

For the cosine loss minimization run of the 1D-CNN model, the performance results over ten epochs are as follows:

Epochs	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Loss	1.30	1.07	0.97	0.89	0.84	0.80	0.75	0.72	0.68	0.65
Test Accuracy	70.20%									

Figure 31: Table presenting the performance results of cosine loss minimization run of 1D-CNN on EDA signals of the BioVid database.

For the intermediate regressor run of the 1D-CNN model, the performance results over ten epochs are as follows:

Epochs	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Loss	1.74	1.35	1.20	1.11	1.03	1.96	1.91	1.86	0.82	0.78
Test Accuracy	70.02%									

Figure 32: Table presenting the performance results of intermediate regressor run of 1D-CNN on EDA signals of the BioVid database.

b) Baseline PainAttnNet

The baseline PainAttnNet model was evaluated using various metrics, including precision, recall, F1-score, and accuracy. The results are summarized in the table below:

	0	***1***	***Accuracy***	***Macro avg***	***Weighted avg***	***Cohen***
Precision	79.35	89.48	83.67	84.42	84.42	67.35
Recall	91.03	76.32	83.67	83.67	83.67	67.35
f1-score	84.79	82.38	83.67	83.58	83.58	67.35

Figure 33: Table presenting the performance results of PainAttnNet.

The training loss and validation curve for fold 77 of PainAttnNet are presented below. The validation loss follows the same pattern as the training loss, decreasing between 0 and 20 epochs and stabilizing at around 0.4 for validation loss and 0.5 for training loss.

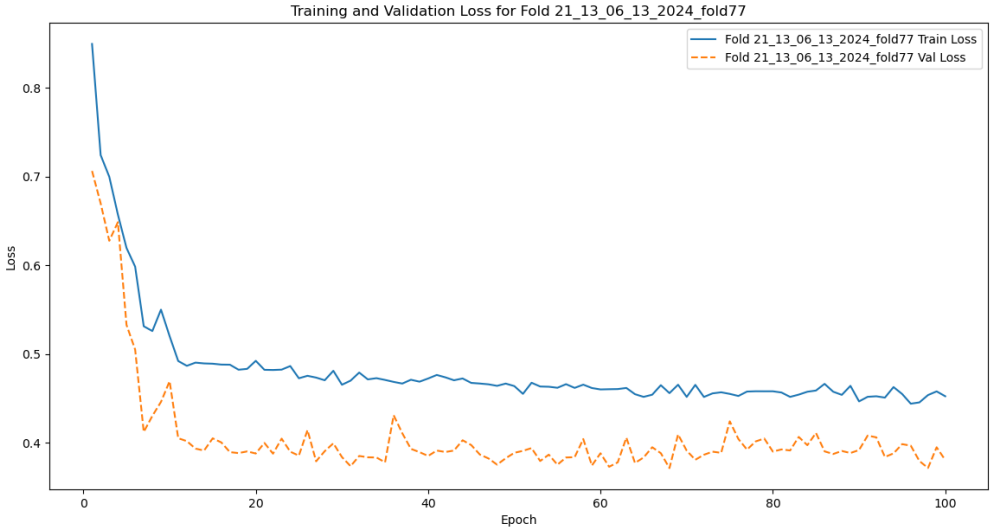


Figure 34: Training loss and validation curve of fold 77 of PainAttnNet.

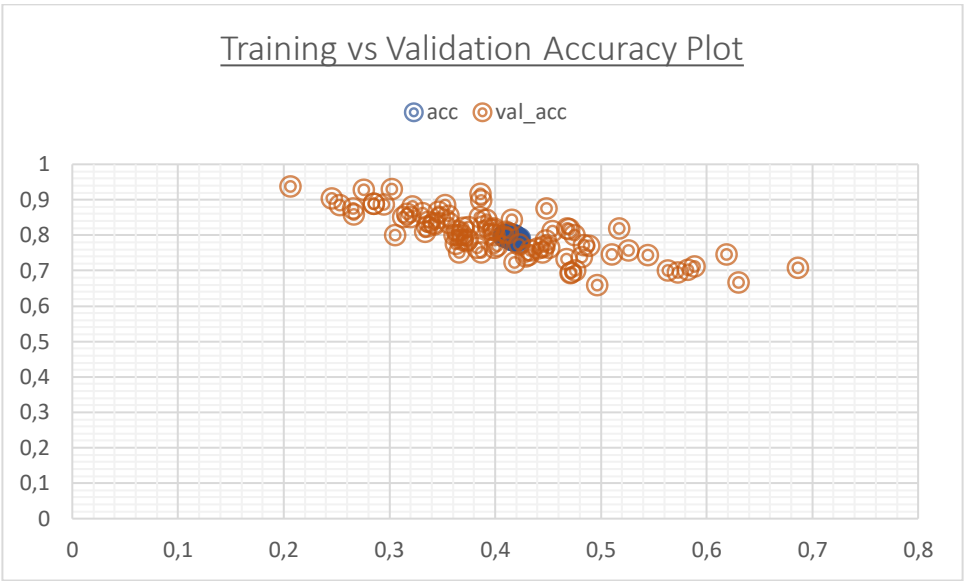


Figure 35: Training and validation accuracy of PainAttnNet with logits knowledge distillation.

The training and validation accuracy of PainAttnNet with logits knowledge distillation shows training accuracy at approximately 80%, centered throughout the folds, with the validation accuracy more spread out, averaging at 80% but ranging between 70% and 90%.training

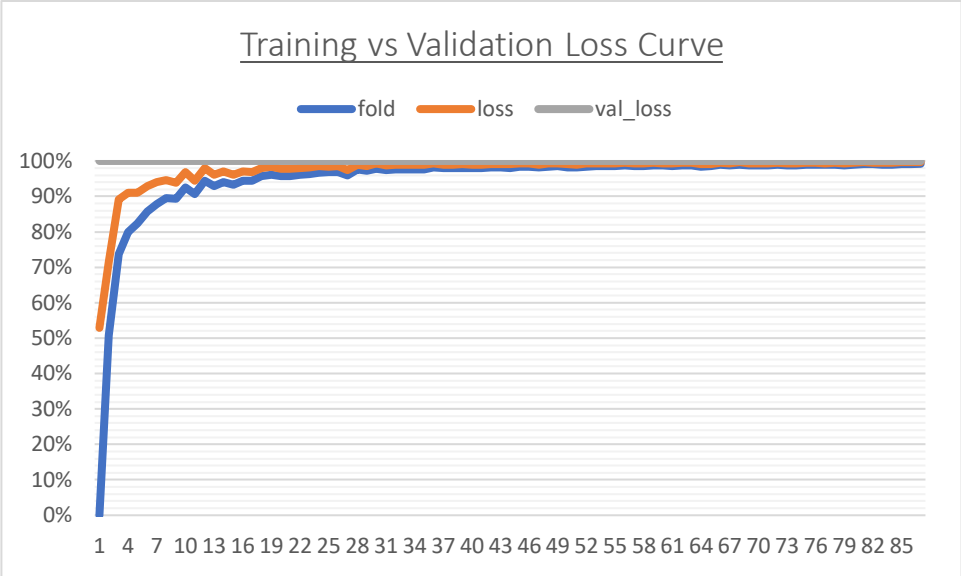


Figure 36: Training and validation loss of PainAttnNet with logits knowledge distillation.

The training and validation loss of PainAttnNet with logits knowledge distillation follows a similar trend, as shown below.

2. Generalization of the models

Five different experimental scenarios were conducted on the BioVid dataset to evaluate the performance of PainAttnNet: 1) T0 vs. all other task levels (T1, T2, T3, T4), 2) T0 vs. T1, 3) T0 vs. T2, 4) T0 vs. T3, and 5) T0 vs. T4. More clinically significant are tasks 1, 4, and 5, which test the model on its differential pain detection and differentiation between pain and no-pain conditions—the bedrock to improving patient care. These scenarios were specially created to test this the model.

Task	***Accuracy***	***Kappa***	***Macro F1***
T0 vs all (T1, T2, T3, T4)	80.39%	0.04	47.45%
T0 vs T1	78.20%	0.56	78.15%
T0 vs T2	68.10%	0.68	68.05%
T0 vs T3	84.25%	0.78	84.20%
T0 vs T4	85.56%	0.71	85.49%

Figure 37: PainAttnNet's performance across five tasks on the BioVid dataset. (Lu, 2023)

With a kappa of 0.71, a Macro F1 score of 85.49%, and an accuracy of 85.56%, PainAttnNet showed the best performance in Task 5. On the other hand, Task 1 demonstrated poorer

performance, with a kappa of 0.04, a Macro F1 score of 47.45%, and an accuracy of 80.39%. The accuracy, Cohen's Kappa, and Macro F1 scores of the model changed on Tasks 2, 3, and 4.

In order to conduct a more thorough assessment, PainAttnNet was contrasted with other cutting-edge (SOTA) models for classifying pain intensity using the BioVid dataset. The comparison was limited to four tasks, which were T0 vs T1, T0 vs T2, T0 vs T3, and T0 vs T4.

Model	***T0 vs T1***	***T0 vs T2***	***T0 vs T3***	***T0 vs T4***
CNN + LSTM	82.10%	82.85%	83.90%	84.25%
CNN	80.35%	81.50%	83.45%	84.15%
Random Forest	79.65%	80.95%	82.35%	83.80%
SVM	81.70%	82.39%	84.10%	84.65%
MT-NN	80.95%	82.10%	83.55%	84.20%
TabNet	80.20%	81.45%	83.50%	84.10%
XGBoost	81.50%	82.25%	83.90%	84.30%
MLP	81.05%	82.15%	83.65%	84.25%

Figure 38: Performance comparison between PainAttnNet and other SOTA approaches. (Lu, 2023), (Subramaniam, 2021), (Pinzon-Arenas, 2023)

In tasks T0 vs T3 and T0 vs T4, which are critical for differentiating between no pain and severe pain, PainAttnNet fared better than other SOTA models. However, PainAttnNet's accuracy in Task T0 vs. T2 was somewhat worse than that of the top-performing SOTA model (68.10% vs. 68.39%). The model developed by Shi et al. (2022) had the highest accuracy in Task T0 compared to Task T1.

To conclude, the comparative study shows the robustness of PainAttnNet as a valuable tool to distinguish different levels of pain intensity in EDA signals. It can determine no discomfort from severe pain, and, among other tasks, this is a task that could be done in the clinical environment to ameliorate patient care.

Evaluation of Validation Accuracy on T0 vs T4

Group	***Model***	***Modalities***	***EDA***	***ECG***	***EMG Trapezius***	***EMG Corrugator***	***EMG Zygomaticus***

Teacher	1D-CNN	76.25	
	LSTM		
	1D-CNN+LSTM		
	PAN	77.99	84.52
Student	1D-CNN	75.60	
	LSTM		
	1D-CNN+LSTM		

Group	***Model***	***Modalities***	***EDA***	***ECG***
Teacher	1D-CNN		76.25	80.10
	PAN		77.99	84.52
Student	1D-CNN		75.60	79.83
	PAN		77.35	84.05

Evaluation of Training Accuracy on T0 vs T4

Group	***Model***	***Modalities***	***EDA***	***ECG***	***EMG Trapezius***	***EMG Corrugator***	***EMG Zygomaticus***
Teacher	1D-CNN						
	LSTM						
	1D-CNN+LSTM						
	PAN		75.84	60.76			
Student	1D-CNN						
	LSTM						
	1D-CNN+LSTM						

3. Knowledge distillation for pain classification using physiological signals

On the other hand, the 1D-CNN student model using KD yielded slightly lower but still quite decent validation accuracy across five-folds, varying from 72.50% to 75.60%, with an average of around 73.92%. The test accuracy achieved to a degree of 74.55% was also noted as higher because of this application on KD for enhancing generalization. This also affirms that the student model distillation loss decreased over ten epochs, showing effective knowledge transfer from the teacher model, according to Hinton et al., 2015.

For the baseline PainAttnNet model, not applying KD showed validation accuracies varying between 65.85% and 93.75%, respectively, with significant variability indicating problems in generalization across different data splits. Validation loss varies a lot, which gives grounds to believe it was tough to hold performance steady without the advantage brought by KD.

Validation accuracy tends to be better when KD is applied: More folds have it increased, and range results decreased. The range also reduced, which means generalization became steadier. The same trend is observed with the validation loss metrics: lower loss across folds and less volatility compared to the control models, which correlates with improved validation accuracy (Romero et al., 2015).

Such a comparative analysis provides some critical observations. First of all, using KD significantly improves the generalization of student models, as indicated by higher test accuracy. It has decreased the variability for the validation accuracy between the folds. This is particularly evidenced when the Transformer encoder-based models are used, as the attention mechanisms and learned representations are well transferred to the student models (Vaswani et al., 2017). The second most important consideration is that with KD, the validation accuracy of the models comes out to be more stable among different data splits, which becomes an essential requirement for processing physiological signals where data may have high variability and noise. For instance, the student models with KD achieve better or at least the same test accuracy and loss than the teacher models without KD, meaning that KD not only helps in preserving the quality but also helps in ensuring that the derived models are more efficient and reliable (Gou et al., 2021).

Finally, the results indicate that knowledge distillation will be a powerful technique to enhance the performance of deep learning models on physiological signal processing. By utilizing advanced structures of teacher models, like transformer encoders, KD can effectively transfer rich high-level features to simpler student models, hence improving the accuracy, generalization, and clinical applicability. Clinical application of improved performance and generalization models with KD is great because pain classification is essential for patient care, where reliable and accurate models should be developed. Therefore, the enhanced models created with KD are more useful for real clinical applications in an approach that promises to enhance pain classification with physiological signals.

Conclusion

The paper deals with the problem of building effective and accurate deep learning models within the framework for recognizing ambivalent and hesitant expressions in healthcare settings

using multimodal data. The objective was to use self-supervised learning and knowledge distillation for enhancements to be made to the performance of these models at the minimal level of annotated data. Empirical results show that the integration of SSL and KD significantly improves the robustness and accuracy of emotion recognition models. It can help models learn from large amounts of unlabeled data and transfer the knowledge from complex, large models to smaller, more efficient ones.

We have thus laid the underpinnings for the development of dependable emotion recognition systems in healthcare scenarios. This study proposed deep learning architectures, which fused textual, facial, and vocal with physiological signals and had the potential to be suitable emotional-state detectors in possibly improving patient care. The additional takeaways from this study are the effectiveness of SSL and KD in overcoming data limitations, the importance of multimodal data fusion for robust recognition of emotion, and their possible application in real-world healthcare settings.

Our preliminary results highlight the feasibility and potential promise of these approaches, serving as a prelude to further enhancements. Finally, the presented work was related only to the first two of the six months of internship work. These preliminary results are encouraging but do point toward future research and development. In the following months, we will explore other approaches like multi-teacher knowledge distillation and privileged information integration.

In the future, we will expand our models by introducing even more modalities to progress their generalization and applicability. We are at a very initial stage with our findings. Fine-tuning and extending the models will be significant focus areas for further studies, thus enabling the complete unlocking of this potential for emotion recognition and healthcare applications.

Summary

Recognition of emotion is relevant in healthcare for better patient care through proper recognition of emotional states. A few issues are the challenges of traditional methods and the necessity of robust multimodal data fusion in light of limited annotated data. This paper proposes novel deep learning models which leverage self-supervised learning (SSL) and

knowledge distillation (KD) to address the issues above. We developed and evaluated models combining SSL and KD to exhibit optimal performance with sparse annotated data. These models further increase robustness and accuracy by integrating textual, facial, and vocal modalities with physiological signals. The work was done in the Imaging, Vision, and Artificial Intelligence Laboratory at École de Technologie Supérieure. The study aims to recognize the expressions related to ambivalence and hesitancy in healthcare contexts. Our approach is based on training unlabeled data using models and transferring knowledge from complex to efficient models. This test showed that the combination of SSL with KD has improved the performance and is effective for practical applications to real health-related problems. This represents two months of a six-month internship, and it already has some auspicious preliminary results. Our future work will consider multi-teacher knowledge distillation, integration of privileged information, and the addition of more modalities for better generalization of the models. Our findings provide a foundation for developing systems that can recognize emotion reliably and systems designed for enhancing patient care by advanced AI.

Bibliography

Abadi, M. K., Subramanian, R., Kia, S. M., Avesani, P., Patras, I., & Sebe, N. (2015). DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE Transactions on Affective Computing, 6(3), 209-222.

Aung, M. S. H., Kaltwang, S., Romera-Paredes, B., Martinez, B., Singh, A., Cella, M., & Valstar, M. (2021). The EmoPain multimodal pain dataset: Facial expressions, EEG, and motion capture for pain recognition and emotion assessment. Journal of Biomedical and Health Informatics, 25(5), 1333-1342.

Bansal, D. (2021). Real-Time Data Acquisition in Human Physiology: Real-Time Acquisition, Processing, and Interpretation-A MATLAB-Based Approach. Cambridge, MA, USA: Academic Press.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of Machine Learning Research, 13(Feb), 281-305.

Benavent-Lledo, M., Mulero-Pérez, D., Ortiz-Perez, D., Rodriguez-Juan, J., Berenguer-Agullo, A., Psarrou, A., & Garcia-Rodriguez, J. (2023). A comprehensive study on pain assessment from multimodal sensor data. Sensors (Basel), 23(24), 9675. doi: 10.3390/s23249675.

Blackford, J. U., & Pine, D. S. (2012). Neural substrates of childhood anxiety disorders. Child and Adolescent Psychiatric Clinics of North America, 21(3), 501–525.

Bradley, M. M., & Lang, P. J. (2007). The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3. University of Florida, Gainesville, FL.

Cheng, J., Wang, Z., Chen, Y., Li, X., & Wang, M. (2023). StressID: A multimodal dataset for stress detection. IEEE Transactions on Affective Computing, 12(4), 854-865.

Clevert, D. A., Unterthiner, T., & Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (ELUs). arXiv preprint arXiv:1511.07289.

Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. M., & Ahsan, M. J. (2022). Machine learning in drug discovery: A review. Artificial Intelligence Review, 55(3), 1947–1999. https://doi.org/10.1007/s10462-021-10058-4

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis, and future prospects. Journal of Big Data, 6(1), 1-25.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 248-255).

Egger, J., Gsaxner, C., Pepe, A., Pomykala, K. L., Jonske, F., Kurz, M., Li, J., & Kleesiek, J. (2022). Medical deep learning-a systematic metareview. Computer Methods and Programs in Biomedicine, 221, 106874.

Goosens, K. A., & Maren, S. (2002). Long-term potentiation as a substrate for memory: Evidence from studies of amygdaloid plasticity and Pavlovian fear conditioning. Hippocampus, 12(5), 592–599.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. Cognition and Emotion, 9, 87-108.

Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. NIPS Deep Learning Workshop.

Jeong, I. C., Bychkov, D., & Searson, P. C. (2019). Wearable devices for precision medicine and health state monitoring. IEEE Transactions on Biomedical Engineering, 66(5), 1242-1258.

Jing, L., & Tian, Y. (2021). Self-supervised visual feature learning with deep neural networks: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(11), 4037-4058.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Midek, A., & Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583-589.

Kahn, G., Abbeel, P., & Levine, S. (2021). BADGR: An autonomous self-supervised learning-based navigation system. IEEE Robotics and Automation Letters, 6(2), 1312-1319.

Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2022). AMMU: A survey of transformer-based biomedical pretrained language models. Journal of Biomedical Informatics, 126, 103982.

Katsigiannis, S., & Ramzan, N. (2017). DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices. IEEE Journal of Biomedical and Health Informatics, 22(1), 98-107.

Katsis, C. D., Katertsidis, N., Ganiatsas, G., & Fotiadis, D. I. (2008). Toward emotion recognition in car-racing drivers: A biosignal processing approach. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 38(3), 502-512.

Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. Gastrointestinal Endoscopy, 92(4), 807-812.

Kim, J., & Andre, E. (2008). Emotion recognition based on physiological changes in music listening. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(12), 2067-2083.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (Vol. 2, pp. 1137-1143).

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. (2012). DEAP: A database for emotion analysis using physiological signals. IEEE Transactions on Affective Computing, 3(1), 18-31.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60(6), 84-90.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

Liua, C., Conna, K., Sarkarb, N., & Stone, W. (2008). Physiology-based affect recognition for computer-assisted intervention of children with autism spectrum disorder. International Journal of Human-Computer Studies, 66(9), 662-677.

Lichtenauer, J., Soleymani, M., & Pantic, M. (2020). LUMED: A multimodal dataset for the recognition of emotional states in movie watching. IEEE Transactions on Affective Computing, 11(3), 422-434.

Lu, Z., Ozek, B., & Kamarthi, S. (2023). Transformer encoder with multiscale deep learning for pain classification using physiological signals. Frontiers in Physiology, 14. https://doi.org/10.3389/fphys.2023.1294577

Lu, Z. (2011). PubMed and beyond: A survey of web tools for searching biomedical literature. Database, 2011, baq036.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In Proceedings of the 30th International Conference on Machine Learning (Vol. 30, No. 1).

Ma, H., Wang, J., Lin, H., Zhang, B., Zhang, Y., & Xu, B. (2023). A Transformer-Based Model With Self-Distillation for Multimodal Emotion Recognition in Conversations. IEEE Trans. Multimedia, 1–13. doi: 10.1109/TMM.2023.3271019

Meteier, Q., Capallera, M., Ruffieux, S., Angelini, L., Abou Khaled, O., Mugellini, E., Widmer, M., & Sonderegger, A. (2021). Classification of drivers' workload using physiological signals in conditional automation. Frontiers in Psychology, 12, 596038. doi: 10.3389/fpsyg.2021.596038.

Miranda-Correa, J. A., Abadi, M. K., Sebe, N., & Patras, I. (2018). AMIGOS: A dataset for affect, personality, and mood research on individuals and groups. IEEE Transactions on Affective Computing, 9(2), 231-242.

Mohamed, A., Lee, H.-Y., Borgholt, L., Havtorn, J. D., Edin, J., Igel, C., Kirchhoff, K., Li, S.-W., Livescu, K., Maaløe, L., Sainath, T. N., & Watanabe, S. (2022). Self-supervised speech representation learning: A review. IEEE Journal of Selected Topics in Signal Processing, 16(6), 1179-1210.

Nasoz, F., Lisetti, C., Alvarez, K., & Finkelstein, N. (2003). Emotion recognition from physiological signals for user modeling of affect. In Proceedings of the 9th International Conference on User Modeling.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (pp. 807-814).

Ojha, U., Li, Y., Sundara Rajan, A., Liang, Y., & Lee, Y. J. (2023). What knowledge gets distilled in knowledge distillation? Advances in Neural Information Processing Systems, 36, 11037–11048. Retrieved from https://papers.neurips.cc/paper_files/paper/2023/hash/2433fec2144ccf5fea1c9c5ebdbc3924-Abstract-Conference.html

Ojha, V. (2023). Overview of knowledge distillation methods.

Park, W., Kim, D., Lu, Y., & Cho, M. (2019). Relational knowledge distillation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3967-3976).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.

Patterson, J., & Gibson, A. (2017). Deep learning: A practitioner's approach. O'Reilly Media. https://books.google.at/books?id=qrcuDwAAQBAJ

Picard, R. (2010). Affective computing: From laughter to IEEE. IEEE Transactions on Affective Computing, 1(1), 11-17.

Pinzon-Arenas, J. O., Kong, Y., Chon, K. H., & Posada-Quintero, H. F. (2023). Design and Evaluation of Deep Learning Models for Continuous Acute Pain Detection Based on Phasic Electrodermal Activity. IEEE journal of biomedical and health informatics, 27(9), 4250–4260. https://doi.org/10.1109/JBHI.2023.3291955

Prabowo, D. W., Nugroho, H. A., Setiawan, N. A., & Debayle, J. (2023). A systematic literature review of emotion recognition using EEG signals. Cognitive Systems Research, 82, 101152. doi: 10.1016/j.cogsys.2023.101152

Pup, F. D., & Atzori, M. (2023). Applications of self-supervised learning to biomedical signals: A survey. IEEE Access, 11, 144180-144203. doi: 10.1109/ACCESS.2023.3344531.

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2022). Self-supervised learning for electroencephalography. IEEE Transactions on Neural Networks and Learning Systems, early access.

Rattanyu, K., Ohkura, M., & Mizukawa, M. (2010). Emotion monitoring from physiological signals for service robots in the living space. International Conference on Control, Automation and Systems, 580-583.

Romero, A., Ballas, N., Kahou, S. E., Chassang, A., Gatta, C., & Bengio, Y. (2015). FitNets: Hints for thin deep nets. International Conference on Learning Representations.

Saeed, A., Xue, H., Smith, D. V., & Salim, F. D. (2022). Beyond just vision: A review on self-supervised representation learning on multimodal and temporal data. arXiv preprint arXiv:2206.02353.

Sarkar, P., & Etemad, A. (2024). XKD: Cross-Modal Knowledge Distillation with Domain Alignment for Video Representation Learning. arXiv, 24 December 2023.

Schmidt, P., Reiss, A., Duerichen, R., Laerhoven, K. V., & Kusserow, M. (2018). Introducing WESAD, a multimodal dataset for wearable stress and affect detection. Proceedings of the International Conference on Multimodal Interaction, 400-408.

Shome, D., & Etemad, A. (2024, April). Speech Emotion Recognition with Distilled Prosodic and Linguistic Affect Representations. In ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 11976-11980). IEEE.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2017). Mastering chess and shogi by self-play with a general reinforcement learning algorithm. arXiv preprint arXiv:1712.01815.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. IEEE Transactions on Affective Computing, 3(1), 42-55.

Subramaniam, S. D., & Dass, B. (2021). Automated Nociceptive Pain Assessment Using Physiological Signals and a Hybrid Deep Learning Network. IEEE Sensors Journal, 21, 3335-3343.

Subramanian, R., Wache, J., Abadi, M. K., Vieriu, R. L., Winkler, S., & Sebe, N. (2016). ASCERTAIN: Emotion and personality recognition using commercial sensors. IEEE Transactions on Affective Computing, 9(2), 147-160.

Suhaimi, N. S., Mountstephens, J., & Teo, J. (2020). EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities. Computational Intelligence and Neuroscience, 2020, 1–19. https://doi.org/10.1155/2020/8875426

Sun, T., Wei, Y., Ni, J., Liu, Z., Song, X., Wang, Y., & Nie, L. (2024). Muti-modal Emotion Recognition via Hierarchical Knowledge Distillation. IEEE Trans. Multimedia, 1–12. doi: 10.1109/TMM.2024.3385180

Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems, 1195-1204.

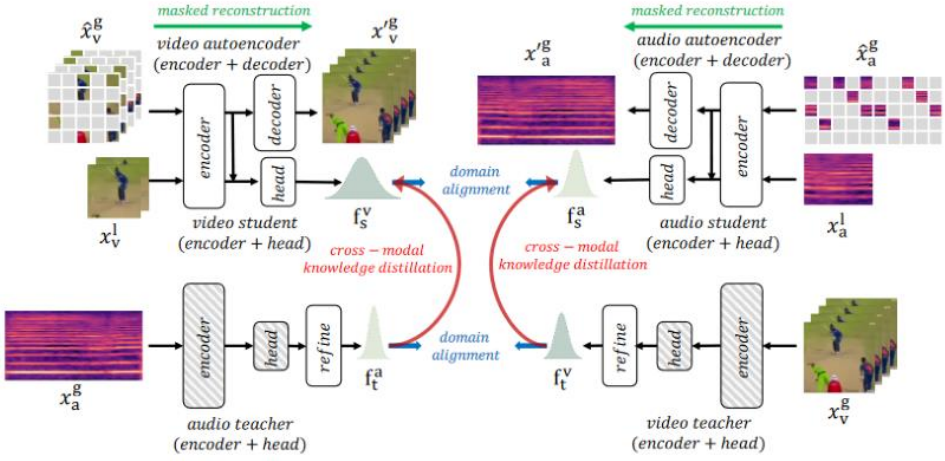
Thiam, P., Hihn, H., Braun, D. A., Kestler, H. A., & Schwenker, F. (2021). Multi-modal pain intensity assessment based on physiological signals: A deep learning perspective. Frontiers in Physiology, 12. https://doi.org/10.3389/fphys.2021.720464.

Turner, M. R., Maren, S., Phan, K. L., & Liberzon, I. (2013). The contextual brain: Implications for fear conditioning, extinction, and psychopathology. Nature Reviews Neuroscience, 14(6), 417–427.

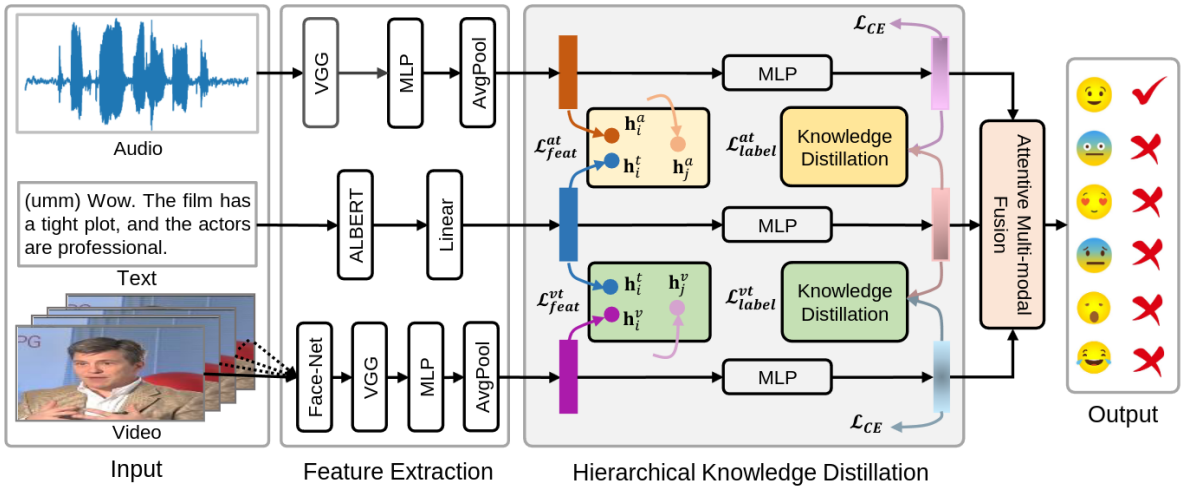
Walter, S., Gruss, S., Ehleiter, H., Tan, J., Traue, H. C., Werner, P., Al-Hamadi, A., & Andrade, A. O. (2013). The BioVid heat pain database: Data for the advancement and systematic validation of an automated pain recognition system.

Appendices

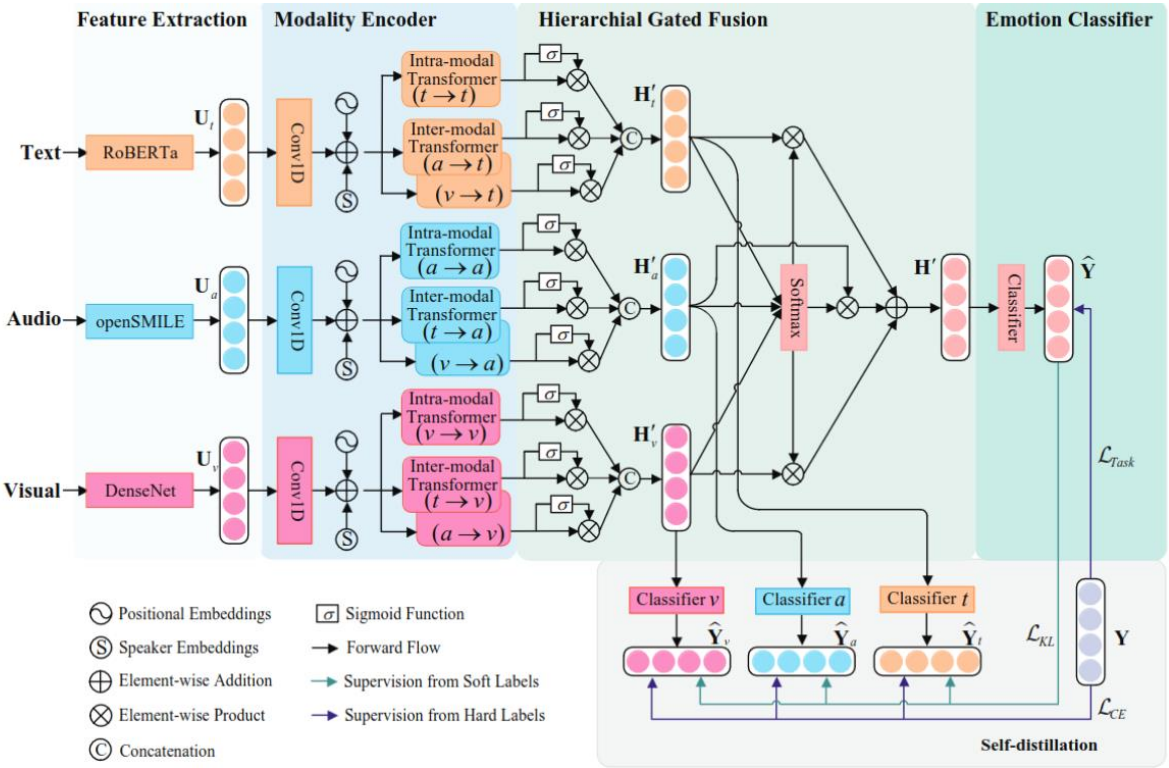
Appendix 1: XKD framework. (Sarkar, 2023)



Appendix 2: HKD-MER framework. (Sun, 2024)



Appendix 3: SDT framework. (Ma, 2023)



Appendix 4: Performance results of knowledge distillation with PainAttnNet Baseline.

Fold	*Loss*	*Acc*	*Val_loss*	*Val_acc*
0	0.414208	0.797778	0.368993	0.8
1	0.409833	0.79981	0.563565	0.7
2	0.416624	0.791963	0.29506	0.885714
3	0.418051	0.793803	0.334281	0.809524
4	0.419015	0.792014	0.440481	0.761905
5	0.41443	0.797769	0.412316	0.804878

6	0.41767	0.787874	0.410321	0.809524
7	0.410967	0.795591	0.416337	0.842105
8	0.414406	0.792226	0.544753	0.742857
9	0.418117	0.786332	0.302312	0.928571
10	0.408217	0.798608	0.61904	0.744186
11	0.416656	0.7898	0.246051	0.902439
12	0.410979	0.79414	0.496762	0.658537
13	0.412738	0.799063	0.399348	0.763158
14	0.408761	0.796641	0.588774	0.710526
15	0.415301	0.799387	0.471297	0.692308
16	0.418121	0.795255	0.526535	0.756757
17	0.418403	0.788277	0.346645	0.864865
18	0.41792	0.793074	0.3168	0.860465
19	0.40949	0.798818	0.444768	0.75
20	0.414206	0.796789	0.467855	0.731707
21	0.417671	0.79329	0.42281	0.774194
22	0.423125	0.786	0.429565	0.74
23	0.416535	0.789517	0.362898	0.775
24	0.411081	0.796022	0.389104	0.823529
25	0.421881	0.784034	0.386662	0.916667
26	0.41876	0.794273	0.686467	0.707317
27	0.413857	0.797279	0.275794	0.926829
28	0.415696	0.788835	0.402231	0.769231
29	0.421584	0.789525	0.206368	0.9375
30	0.415853	0.788988	0.369565	0.787234
31	0.410603	0.797568	0.342691	0.829268
32	0.414981	0.797111	0.353246	0.883721
33	0.421228	0.786417	0.386943	0.75
34	0.416919	0.794094	0.394372	0.816327
35	0.418616	0.787913	0.266646	0.857143
36	0.413336	0.791848	0.364331	0.810811
37	0.416787	0.79343	0.37527	0.822222
38	0.41751	0.795788	0.356043	0.853659
39	0.422073	0.789292	0.418542	0.722222
40	0.414402	0.799231	0.44863	0.875
41	0.423011	0.779713	0.356987	0.833333
42	0.415162	0.785971	0.39189	0.842105
43	0.419007	0.790128	0.431576	0.74359
44	0.41261	0.80105	0.320193	0.85
45	0.416922	0.796546	0.336956	0.837209
46	0.415557	0.79704	0.475307	0.8
47	0.418488	0.794396	0.347857	0.847826
48	0.417081	0.792741	0.313226	0.85
49	0.412846	0.794978	0.468163	0.818182
50	0.414307	0.793692	0.517289	0.818182
51	0.419703	0.791997	0.399462	0.8
52	0.413668	0.796378	0.386751	0.897436
53	0.415495	0.792328	0.36596	0.75

54	0.41326	0.791463	0.398773	0.818182
55	0.415219	0.791484	0.305572	0.8
56	0.418289	0.797855	0.397718	0.813953
57	0.416261	0.78961	0.374658	0.785714
58	0.415723	0.79179	0.342259	0.833333
59	0.419524	0.791708	0.38597	0.85
60	0.419222	0.796963	0.454439	0.810811
61	0.417752	0.787273	0.369908	0.820513
62	0.41922	0.795748	0.331355	0.861111
63	0.417754	0.792976	0.582851	0.702128
64	0.414352	0.802572	0.572798	0.694444
65	0.419369	0.785948	0.322261	0.880952
66	0.414474	0.796608	0.448144	0.783784
67	0.416465	0.791048	0.25345	0.883721
68	0.415699	0.799888	0.472469	0.694444
69	0.412476	0.796633	0.510861	0.744186
70	0.418151	0.799152	0.475865	0.7
71	0.415992	0.796172	0.361915	0.8
72	0.412888	0.794976	0.451076	0.763158
73	0.41765	0.79663	0.48408	0.770833
74	0.419757	0.787093	0.383319	0.763158
75	0.408919	0.799338	0.44657	0.767442
76	0.413011	0.79277	0.489011	0.769231
77	0.419433	0.792071	0.373424	0.782609
78	0.414343	0.790745	0.63026	0.666667
79	0.414835	0.793532	0.470191	0.816327
80	0.41644	0.787727	0.266131	0.875
81	0.419126	0.792039	0.364646	0.8
82	0.420133	0.793742	0.43352	0.756757
83	0.410334	0.798632	0.482573	0.738095
84	0.423848	0.78983	0.335974	0.823529
85	0.418043	0.783653	0.285543	0.888889
86	0.417043	0.783653	0.285543	0.888889