# SYNOPSIS

In the realm of emotion recognition, this project centred on Knowledge Distillation (KD) for fear level detection in humans using advanced deep learning models. Knowledge distillation is a method where a "teacher" model transfers knowledge to a smaller "student" model, aiming to retain the performance of the larger model while reducing complexity and computational costs. This research builds on KD by using multiple deep neural networks and developing corresponding student models for fear level classification, specifically using the DEAP dataset. The project employs a variety of deep learning architectures, including Deep Neural Networks (DNN) with 300 and 150 hidden units, Long Short-Term Memory (LSTM) networks, a CNN-LSTM hybrid (combining convolutional and recurrent layers), and Temporal Convolutional Networks (TCN). The teacher models include DNN 300 and TCN, and corresponding student models are trained to achieve comparable accuracy but with reduced computational resources. Knowledge distillation plays a key role in the training process, allowing the student models to mimic the performance of their respective teacher models.

The DEAP dataset, which includes physiological signals such as electroencephalography (EEG) for emotion analysis, is used to classify fear levels into categories like relax, low, medium, and high. Unlike traditional methods, no machine learning models are used in this framework. Instead, deep learning-based methods are applied to process the multi-dimensional physiological data. A rigorous cross-validation process is applied to DNN 300 and TCN models, ensuring robust generalization and preventing overfitting. This methodology facilitates better accuracy in detecting fear levels while optimizing model efficiency. By comparing different deep learning models and their distilled counterparts, the research significantly advances the understanding of how to effectively utilize KD in emotion recognition and fear detection. The results provide key insights into reducing model complexity without compromising performance, with implications for real-world applications in affective computing, neuroscience, and human-computer interaction.

# LIST OF FIGURES

# LIST OF TABLES

# CONTENTS

**CHAPTER** **Page No.**

# CHAPTER 1

# INTRODUCTION

Detecting fear levels in humans is essential for a deeper understanding of emotional states, with significant applications in areas such as affective computing, human-computer interaction, and mental health diagnosis. Fear, being a fundamental emotion, can heavily influence decision-making and behavior, making its accurate detection important for various real-world applications, such as virtual reality, therapeutic interventions, and safety monitoring systems. The DEAP dataset serves as a popular benchmark for emotion recognition, providing physiological recordings, including electroencephalography (EEG), paired with annotations of valence (the degree of pleasantness), arousal (emotional intensity), and dominance (the sense of control during the emotion). These dimensions are crucial for interpreting emotional reactions, with fear generally associated with high arousal and low valence. Utilizing the DEAP dataset enables the effective classification of fear levels based on physiological data.

Deep learning models, such as Deep Neural Networks (DNN), Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and Temporal Convolutional Networks (TCN), have demonstrated significant success in capturing intricate patterns from physiological data like EEG signals. These models are particularly adept at detecting subtle signal variations, leading to improved accuracy in identifying fear levels when compared to traditional approaches. Additionally, Knowledge Distillation (KD) plays a key role in enhancing this process by transferring knowledge from larger, more complex models (teacher models) to smaller, more efficient ones (student models). This technique not only retains high accuracy in fear level detection but also lowers computational demands, making it suitable for use in environments with limited resources. By applying KD, this research advances the scalability and real-world usability of deep learning models, especially in scenarios where fear detection is crucial for understanding emotional states.

## 1.1 PROBLEM STATEMENT

This project addresses the challenge of improving the performance of emotion recognition models, particularly those leveraging EEG and ECG signals, to more accurately classify fear levels. Applications in domains including psychology, neurology, and human-computer interface require accurate categorization of fear. The focus of this project is to enhance deep learning models, aiming for greater accuracy and reliability in fear detection. By incorporating cutting-edge deep learning methods, along with knowledge distillation and multimodal data integration, the project strives to significantly improve the efficiency of emotion recognition systems. The development of these advanced models will enable a more thorough and accurate interpretation of emotional states, thereby benefiting therapeutic practices, research, and interactive technologies. In the broader context of affective computing, this project seeks to deliver tools that not only offer higher accuracy but also possess greater applicability in real-world scenarios. This progress has the potential to revolutionize the way emotional responses are analyzed and applied across different sectors.

## 1.2 SCOPE AND MOTIVATION

The goal of this project is to greatly increase the precision and generalizability of fear detection by leveraging advanced deep learning techniques, with a particular focus on knowledge distillation. The strategy involves developing a comprehensive approach that includes exploring ensemble and hybrid models, along with the integration of multimodal data sources. The project seeks to address the current limitations of emotion recognition systems by restructuring and adapting datasets for various training methods. Additionally, knowledge distillation will be employed to create lightweight models that maintain high performance while minimizing computational demands.

The primary driving force behind this endeavor is to address the limitations of current emotion recognition systems that rely on EEG and ECG signals. These existing models often face challenges in accurately classifying fear levels, limiting their practical use. By advancing deep learning methods and integrating multimodal data, this project aims to enhance classification accuracy and reliability. The ultimate objective is to make a significant contribution to the field of affective computing by offering enhanced emotion analysis tools with applications in human-computer interaction, psychology, and neuroscience.

The outcomes of this study hold significant potential for advancing affective computing technologies. By overcoming the limitations of current emotion recognition models, the project not only aims to enhance the accuracy of fear level classification but also contributes to the creation of more robust and adaptable emotion analysis tools. It is anticipated that the research would enhance the useful applications of these technologies in domains including psychology, neurology, and human-computer interaction by integrating multimodal data and utilizing sophisticated deep learning techniques, such as knowledge distillation. Its creative approach to model optimization and dataset structure may establish new benchmarks in the industry and lead to advances in our knowledge and use of emotional reactions across a range of fields.

The research findings are expected to be a useful source for industry professionals, policymakers, and researchers working in the development and application of emotion recognition technologies. By offering in-depth insights into advanced deep learning models and their effectiveness in fear level classification, the study provides practical guidance for incorporating these technologies into real-world use cases. Industry practitioners can capitalize on the improved accuracy and robustness of the models to enhance user experiences and emotional analytics across various sectors. Policymakers can utilize the findings to shape regulations and standards for emotion recognition systems. Researchers will also benefit from the project's contributions, gaining a deeper understanding of multimodal data integration and model optimization. This comprehensive approach is anticipated to fuel further innovation in affective computing, advancing the development of reliable and effective emotion analysis tools.

# CHAPTER 2

# LITERATURE SURVEY

## 2.1 RELATED WORKS

In 2024, Nagisa Masuda Eguchi and their team introduced a novel deep learning model that combines multiple inputs for classifying fear levels. This model, which combines Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks, was applied to the DEAP dataset, which includes multichannel EEG recordings and various physiological signals. The model uses LSTMs to examine the temporal dynamics of physiological signals and CNNs to extract spatial characteristics from EEG data. With this approach, the team achieved a 97% accuracy in fear level classification, showcasing the successfulness of the hybrid CNN-LSTM model in capturing both spatial and temporal patterns. This research shows the strength of such models in advancing affective computing and emotion recognition, offering a significant breakthrough in the field[1].

In 2023, Mohammed Syed Mohammed's study titled "Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset" presented a groundbreaking approach to real-time emotion recognition using EEG signals. This research expanded the capabilities of emotion recognition to a broad range of emotional states, making it applicable to various contexts. Utilizing the DEAP dataset, which includes multichannel EEG and physiological signals, the study extracted features from multiple domains to enhance emotion classification. The research employed a Support Vector Machine (SVM) as the classifier and used Leave-One-Out Cross-Validation (LOOCV) to thoroughly assess the model's performance. Despite the inherent challenges in recognizing emotions from EEG signals, the method achieved a notable 65% accuracy. This work not only underscores the potential of multi-domain feature fusion in capturing intricate emotional nuances but also paves a breakthrough in real-time emotion recognition systems. Its implications are significant for enhancing emotion-based applications in healthcare, gaming, and human-computer interaction[2].

In 2023, Soraia M. Alarcao and Manuel J. Fonseca, a Senior Member of IEEE, published the study "Emotions Recognition Using EEG Signals." This research focuses on

EEG signal based emotion recognition using the MAHNOB dataset. The authors explored novel methods to enhance emotion classification, incorporating various physiological signals such as Electromyography (EMG), Galvanic Skin Response (GSR), and Heart Rate (HR). They applied advanced classification techniques, including LSTM networks and SVM. Their proposed methods achieved an accuracy of 82%, advancing the field by refining approaches to emotion recognition based on EEG and other physiological data. This study illustrates how EEG signals can be combined with other physiological measurements to improve emotion identification systems' accuracy. A strong framework for accurately classifying emotions is offered by the combination of LSTM networks, SVM algorithms, and the integration of GSR, EMG, and HR data. The results of this study immediately aid in the creation of increasingly complex and trustworthy emotion identification technologies, which find use in a variety of industries.[4].

In 2022, the study conducted by Caglar Uyulan, Ahmet Ergun Gumus, and Zozan Guleken delved into the intricacies of fear-type emotion detection through an innovative approach that combined EEG and heart rate (HR) signals with entropy-based analysis. By leveraging a comprehensive brain wave dataset, the researchers were able to meticulously train and test a suite of advanced machine learning(ML) models, including Support Vector Machines (SVM), Random Forest (RF), and Long Short-Term Memory (LSTM) networks. These models were designed to capture the subtle yet significant patterns indicative of fear-related emotions. The evaluation process was rigorous, ensuring that the models were tested under both user-dependent and user-independent conditions, a critical step to ensure the hardiness and generalizability of the findings. The remarkable accuracy rate of 94% achieved by these models underscores the potential of entropy-based analysis in conjunction with EEG and HR signals for precise fear detection. This pioneering work not only advances our understanding of emotion recognition but also paves the way for the growth of more sophisticated and reliable emotion-based technologies in various fields, such like healthcare, virtual reality, and human-computer interaction[3].

In 2020, Sorasa M. Alarcao and Manuel J. Fonseca published the study titled "Stress Detection with Machine Learning and Deep Learning Using Multimodal Data." This research utilizes the DEAP dataset, focusing on EEG signals to detect stress using cutting-edge deep learning and machine learning methods. The study employs Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) networks, and Gated Recurrent Units (GRU) to analyze the data. Key feature selection algorithms used in the study include Principal Component Analysis (PCA) and Sequential Feature Selection (SFS). The research incorporates the Model of Affect, considering dimensions such as valence, arousal, dominance, and liking. The classification models achieved an impressive accuracy rate of

95%, demonstrating the effectiveness of these methods in stress detection. This study highlights the potential of combining EEG signals with sophisticated machine learning and deep learning techniques to accurately identify stress. The use of RNN, LSTM, and GRU networks, along with PCA and SFS for feature selection, provides a robust framework for stress detection. The incorporation of the Model of Affect further enhances the understanding and classification of stress-related emotions. The high accuracy rate achieved in this research underscores the significant advancements in emotion recognition and stress detection, paving the way for more effective interventions and applications in fields such as mental health, workplace wellness, and personalized healthcare[6].

In 2020, Vikrant Doma and Matin Pirouz conducted a study titled “Comparative Analysis of Machine Learning Using EEG”, focusing on emotion recognition through EEG and peripheral physiological signals from the DEAP dataset. The study employed various machine learning models, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression, to classify emotional states. Preprocessing techniques were used to enrich the accuracy of emotion recognition. With these approaches, the study achieved an accuracy of 75%, outlines the potential of machine learning in making better emotion detection systems, particularly for applications in mental health, human-computer interaction, and neurofeedback[8].

In 2020, Oana Bălan, Gabriela Moise, and Alin Moldoveanu conducted a study on automatic fear detection using machine learning techniques. Utilizing the MAHNOB dataset, which includes physiological signals such as Electroencephalogram (EEG), Heart Rate (HR), and Galvanic Skin Response (GSR), they explored the effectiveness of classifiers like Support Vector Machines (SVM), Decision Trees, and Random Forests. Fear was modeled as a complex emotion, defined by a combination of three key dimensions: valence (emotional positivity or negativity), arousal (intensity), and dominance (level of control). By analyzing these emotional states, their models achieved an accuracy of 89% in fear classification. This study highlights the potential of machine learning models in detecting fear through physiological signals, offering valuable insights for real-time emotion recognition. The research represents a significant step in enhancing automatic emotion detection systems, with implications for areas such as security, mental health, and human-computer interaction[9].

In 2019, Leordeanu and Florica Moldoveanu published a study titled "Fear Level Classification: Dimensions, Emotional and Machine Learning", which focused on classifying fear levels using Deep Neural Networks (DNN) and Support Vector Machines (SVM). The study utilized data from the DEAP database, specifically analyzing EEG signals and biophysical recordings to accurately detect and classify emotional states related to fear. Their innovative approach aimed to achieve a system capable of serving in the treatment of phobias, with potential applications in both medical and behavioral therapies. The classification model developed in the study achieved a significant accuracy of 85%, highlighting its effectiveness in recognizing fear-related emotions. This research paves the way for advanced emotion recognition systems that could be used in therapeutic settings, helping patients manage fear and anxiety through personalized treatments[5].

In 2019, M. Alarcao and Manuel J. Fonseca conducted a study titled “Multi-Domain Feature Fusion for Emotion Classification Using the DEAP Dataset”, focusing on improving emotion recognition accuracy through advanced signal processing techniques. The research utilized EEG signals from the widely used DEAP dataset and employed Support Vector Machines (SVM) for classification. A key innovation of the study was the use of Hjorth parameters, which capture signal complexity and mobility, integrated into a multi-domain feature fusion approach. This method allowed for the combination of time-domain, frequency-domain, and spatial features to enhance the system’s understanding of emotional states. The model achieved an impressive 94% accuracy, which, according to the authors, represents the highest performance for emotion classification on the DEAP dataset to date. This breakthrough demonstrates the potential of feature fusion in advancing emotion recognition technology, with applications in mental health monitoring, affective computing, and personalized emotion-based interventions[10].

In 2019, Nakısaa and Bahareh Chandranb published a study titled “Automatic Emotion Recognition Using Temporal Learning”, which focused on emotion recognition using EEG signals from the MAHNOB dataset. The study employed a combination of Convolutional Neural Networks (ConvNet) and Long Short-Term Memory (LSTM) networks to address the challenge of classifying emotions along the valence dimension (positive vs. negative emotions). By leveraging multimodal physiological signals, the model was able to capture both spatial and temporal patterns, providing a complete understanding of emotional states. This approach achieved a classification accuracy of 71%, showcasing the potential of deep learning techniques like ConvNet and LSTM in improving automatic emotion recognition. The study contributes to advancing emotion detection systems for applications in human-computer interaction and emotion-aware systems[7].

## 2.2 GAPS IDENTIFIED

1. Current methods for emotion recognition often exhibit limited accuracy. This is primarily due to the inherent complexity of emotional states and the limitations of the models used to capture these states effectively.
2. Many models face difficulties in generalizing their results across different datasets due to being trained and tested on a single dataset that often lacks the diversity found in real-world scenarios.
3. Although some studies integrate multiple data modalities (e.g., EEG, ECG), there is still a lack of effective techniques to fully utilize this diverse information. Models often fail to effectively combine these different sources for improved emotion recognition.
4. The quality of the features that are retrieved has a significant impact on how well emotions are detected. Poor model performance can be caused by subpar feature extraction and preprocessing procedures, highlighting the necessity of optimal approaches.
5. Fear detection systems may have difficulty operating effectively in low-light environments where visual cues are diminished, thereby reducing the system's accuracy and reliability.
6. Many fear detection systems don't combine numerous data sources and instead rely on a single approach, such speech or facial movements. This lack of multimodal integration limits the system’s ability to provide a comprehensive assessment of fear.
7. Existing technologies might struggle to accurately measure the intensity of fear, which is crucial for understanding the severity of the emotion and tailoring appropriate responses.
8. In emotion recognition datasets, certain emotional states are often underrepresented, leading to class imbalance. This imbalance can cause models to become biased, performing well on the more frequent classes while struggling with less represented emotions
9. The deployment of fear detection technology raises significant privacy and ethical issues, particularly regarding the monitoring and recording of individuals' emotional states without adequate safeguards.

10. Current fear detection systems may not adapt well to new or emerging indicators of fear, making them less effective as the nature of fear cues evolves over time or varies in different contexts.

# CHAPTER 3

# SOFTWARE REQUIREMENTS SPECIFICATIONS

System requirements refer to the minimum and maximum hardware and software specifications necessary for a system or application to function correctly. To accurately determine these requirements, it is essential to thoroughly analyze the functional and non-functional needs of the system, as well as to identify any constraints or factors that could impact its performance. This process ensures that the system operates efficiently within its intended environment, providing optimal user experience and functionality.

## 3.1 FUNCTIONAL REQUIREMENTS

### 3.1.1 Real Time Fear Detection Analysis

The system must be able to perceive and accurately analyze fear levels in real-time. This requires processing data from multiple input sources, including facial expressions, voice tone, and physiological signals, to detect and assess the intensity of fear swiftly and reliably. Additionally, the system should deliver timely feedback on fear levels to enable immediate interventions or responses, ensuring that critical decisions can be made without delay.

### 3.1.2 Adaptive Emotion Recognition and Response

The system should possess adaptive emotion recognition and response capabilities to accurately interpret changing emotional states. This involves adjusting analysis algorithms dynamically based on variations in emotional expression and contextual factors. The system must be able to tailor its responses to different levels of fear while considering individual differences and situational context..

### 3.1.3 Multimodal Integration and Interpretation

The system must integrate data from multiple modalities, such as visual, auditory, and physiological sensors, to provide a comprehensive assessment of fear levels. This includes implementing algorithms that can combine and interpret diverse data sources effectively, ensuring a holistic understanding of the individual's emotional state. The system should also be capable of updating its analysis in real-time as new data becomes available.

## 3.2 NON-FUNCTIONAL REQUIREMENTS

### 3.2.1 Real Time Performance

The system must exhibit real-time responsiveness in detecting and analyzing fear levels. This involves minimizing processing delays to ensure that emotional states are assessed and reported swiftly. The system should be capable of reacting promptly to changes In Emotional Indicators, maintaining accuracy and efficiency in various situations.

### 3.2.2 Robustness and Reliability

The system should demonstrate robustness and reliability across diverse environments and conditions. It should be resilient to variations in sensor input quality, noise, and environmental disturbances, ensuring consistent performance in fear detection. The system should also be capable of handling unexpected changes in emotional expression or context without compromising accuracy.

### 3.2.3 Scalability and Adaptability

The system should be scalable and adaptable to different applications, user demographics, and operational contexts. It should be able to integrate with various types of sensors and data sources, allowing for flexible deployment in a range of scenarios. Additionally, the system should be designed to accommodate future updates, enhancements, and advances in fear detection technology to ensure long-term effectiveness and relevance.

## 3.3 HARDWARE REQUIREMENTS

- GPU: NVIDIA GeForce RTX or AMD Radeon RX series.
- Graphics Card: DirectX 11 compatible.
- Storage: Minimum 110 GB free disk space.
- RAM: 16 GB
- OS: Windows 10 or above
- Processor: 64-bit processor

## 3.4 SOFTWARE REQUIREMENTS

- Google Colaboratory
- Numpy
- Pytorch
- Tensorflow 2.16.2

- Matplotlib
- Pandas
- Scikit-learn
- CSV

### Google Colaboratory

Google offers a Python development environment in the cloud called Google Colaboratory, or Colab. Because it supports well-known Python libraries like TensorFlow and PyTorch, integrates seamlessly with Google Drive, and offers free access to GPUs and TPUs, it is a popular choice for data science and machine learning applications.

### NumPy

A key Python module for numerical computation, NumPy makes it possible to handle big, multi-dimensional arrays and matrices effectively. Numerous mathematical functions designed for quick array operations are available. In obstacle avoidance systems, NumPy processes sensor data, models spatial environments, and performs critical calculations for navigation and decision-making.

### Pytorch

Facebook's AI Research lab created the open-source deep learning framework PyTorch, which is renowned for its adaptability and user-friendliness. Because of its support for dynamic computation graphs, creating and adjusting neural networks is simple. Because of its effectiveness, PyTorch is frequently used for applications like computer vision, natural language processing, and reinforcement learning.

### TensorFlow

TensorFlow is a well-known deep learning framework that was created by Google and is frequently used to create and train neural networks. In obstacle avoidance, TensorFlow can be employed to develop and deploy deep learning models for perception tasks, such as object detection and classification, enabling vehicles to detect and react to obstacles in their surroundings.

### Matplotlib

A Python charting toolkit called Matplotlib offers a variety of visualization features for making static, animated, and interactive charts. In obstacle avoidance, Matplotlib can be utilized to visualize sensor data, trajectory planning, and navigation paths, aiding in the analysis and visualization of autonomous driving algorithms and their performance in different scenarios.

### Pandas

A robust Python module for data analysis and manipulation. Pandas makes it simple to clean, modify, and analyze structured data by offering data structures like DataFrames and Series. It is a vital tool for exploratory data analysis and data wrangling.

### Scikit-learn

Scikit-learn is a popular Python machine learning framework that provides effective data mining and analysis tools. For problems including classification, regression, clustering, and dimensionality reduction, it offers a variety of techniques. The library streamlines the entire machine learning process by including modules for data preprocessing, model evaluation, and validation.

### CSV

A comma separates each value in the simple CSV (Comma Separated Values) file format, which is used to store tabular data in plain text format. Each line of the file represents a row of data. In obstacle avoidance, CSV files can be used for logging and storing sensor data, simulation parameters, and experimental results, facilitating data analysis, visualization, and sharing among researchers and developers.

# CHAPTER 4

# PROPOSED SYSTEM

## 4.1 PROPOSED SYSTEM

In this project,we aim to do fear-level classification using the DEAP dataset that aims to leverage deep learning models to accurately classify different fear levels based on EEG signals and other signals like GSR, Blood pressure etc. The system will begin with data preprocessing, which involves cleaning and normalizing the raw signals to ensure consistency and reliability in the input data. The core of the system will involve implementing and training several deep learning models to classify the fear levels. Each model will be evaluated based on its accuracy and computational performance.

To enhance the system's efficiency, knowledge distillation techniques will be employed to reduce the complexity of the deep learning models. This process will involve training larger models (such as Temporal CNN, DNN 150, DNN 300) as teacher models and using their predictions to guide the training of smaller student models, such as Simple ANN, DNN 150, DNN 300, and LSTM. The goal of this distillation process is to create lightweight models that can perform with similar accuracy but require less computational resources and training time. The system will be evaluated in terms of accuracy, efficiency, and scalability, with the ultimate goal of developing a reliable fear-level classification model that can be deployed in real-time applications. Additionally, incorporation of performance evaluation and validation techniques to ensure the robustness of the fear-level classification models. Cross-validation will be employed to test the generalizability of the models across different subsets of the DEAP dataset, and metrics such as accuracy, precision, recall, and F1-score will be used to assess classification performance.By comparing the results of the original deep learning models with the distilled student models, the system will provide insights into the trade-offs between model complexity and performance, helping identify the most suitable models for deployment in resource-constrained environments. This comprehensive evaluation will ensure that the proposed system is both effective and efficient for real-world fear-level classification tasks.

## 4.2 SYSTEM ARCHITECTURE

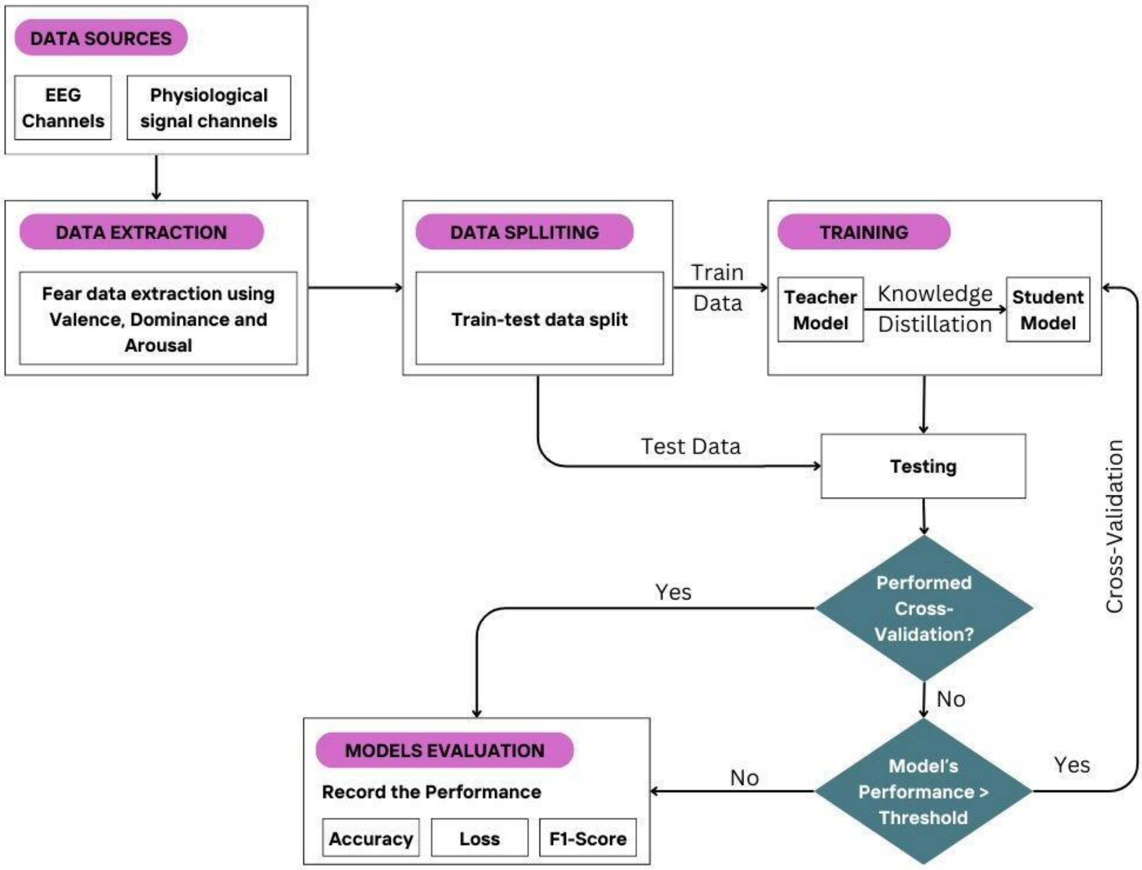


**Fig 4.1 Schematic view of proposed system**

From Figure 4.1, the process begins with the collection of the DEAP dataset, which comprises EEG signals and other physiological signals such as EMG, GSR, and heart rate, which are used for classifying emotional states like fear. This dataset serves as the foundation for the fear-level classification system. Once the dataset is collected, it undergoes a preprocessing phase to ensure it is ready for analysis. Preprocessing involves steps like filtering to remove noise, normalization to standardize the data, and splitting the dataset into training and testing sets. Additionally, feature extraction is performed to identify key features from the raw physiological data, which are most relevant to fear-level classification. This step might involve extracting time-domain, frequency-domain, or statistical features, depending on the model's requirements. These features can enhance the accuracy of the models by providing more structured input data. In cases where deep learning models are applied without feature engineering, raw data will be directly fed into the models for training, allowing the networks to learn relevant patterns on their own.

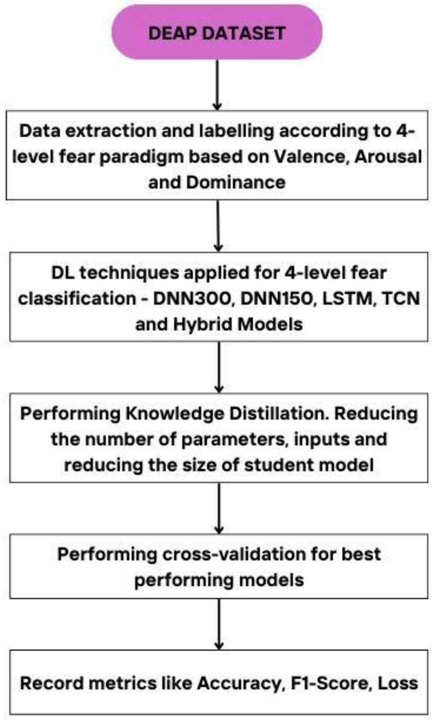


**Fig 4.2 Steps involved to obtain classifier**

After preprocessing, the training phase begins, where the deep learning models such as Temporal CNN, RNN, GRU, or even lighter models after knowledge distillation are trained on the dataset. This training is conducted with and without feature engineering to compare the impact of pre-extracted features versus end-to-end learning from raw signals. Knowledge distillation is also incorporated to reduce model complexity and improve efficiency, where larger models (like RNNs or GRUs) act as teacher models to train smaller, more efficient student models, such as ANNs or DNNs. Once the training is complete, the models are tested and validated using the reserved testing dataset to ensure they perform well on unseen data. This validation phase helps to fine-tune the models and ensures they generalize well to new inputs. The Figure 4.2 outlines the process of fear level detection using the DEAP dataset. It includes data extraction based on valence, arousal, and dominance, followed by deep learning classification with models like DNN, LSTM, and TCN. Knowledge distillation reduces the student model’s complexity, cross-validation optimizes performance, and metrics like accuracy are recorded.

## 4.3 DATASET

The dataset used for this project is the **DEAP (Database for Emotion Analysis using Physiological Signals) Dataset**, a standard benchmark in emotion recognition research using EEG (Electroencephalogram) and other physiological signals. This dataset is particularly designed for the analysis of emotional states, including fear, based on various stimuli. Below is a detailed breakdown of the dataset and its components.

- **Name**: DEAP Dataset
- **Purpose**: Emotion analysis and classification using physiological signals.
- **Dataset Type**: Multimodal physiological dataset (EEG, EMG, GSR, ECG, etc.)
- **Size**: 1.9 GB
- **Participants**: 32 subjects (16 males, 16 females).
- **Recording Duration**: Each participant's EEG signals were recorded for 63 seconds per video clip (60 seconds of stimuli and 3 seconds pre-trial baseline).
- **Sampling Rate**: 512 Hz (downsampled to 128 Hz for preprocessed data).
- **Total Data Points**: 32 participants × 40 video trials per participant = 1,280 trials.

The dataset consists of recordings from 32 participants (16 males and 16 females) who watched 40 music videos chosen to elicit different emotional responses. Each participant's physiological signals were recorded while they watched the videos. The data is stored in .dat files, with each file containing data for all 40 trials of a single participant. This structure includes 32-channel EEG data and additional peripheral signals such as galvanic skin response (GSR) and electrocardiogram (ECG), among others, recorded for each trial. The dataset also includes annotations for valence, arousal, dominance, and liking scores for each trial.

In terms of data format, the EEG data is represented in the shape (40, 40, 8064), where 40 refers to the number of trials, 40 represents the number of channels, and 8064 corresponds to the number of time points per trial. The labels for each trial, including valence, arousal, dominance, and liking scores, are structured in a (40, 4) format. EEG signals and peripheral signals are both sampled at 128 Hz. However, the EEG data was originally recorded at 512 Hz using a Biosemi ActiveTwo system and downsampled to 128 Hz for storage. The electrodes used for the EEG recordings were placed according to the international 10-20 system, covering frontal, central, parietal, occipital, temporal, and additional areas such as Fz, Cz, Pz, and Oz.

The dataset also includes 8 additional physiological signals that capture various aspects of the participant's emotional state. These include galvanic skin response (GSR) to measure emotional arousal, electromyogram (EMG) to capture muscle activity, electrocardiogram (ECG) to monitor heart activity, respiration to track breathing patterns, blood volume pulse (BVP) for heart rate analysis, and zygomaticus and trapezius EMG to monitor muscle movements related to facial expressions. These signals provide further insights into participants' emotional reactions to the stimuli.

Participants rated each video on four emotional dimensions: valence, arousal, dominance, and liking. These ratings were done on a scale from 1 to 9. Valence measures the pleasantness of the video (with low scores indicating unpleasantness), arousal reflects the level of excitement or calmness (low scores indicate calmness), dominance measures the participant's sense of control (with lower scores meaning less control), and liking reflects how much they liked the video (low scores meaning they disliked it). Additionally, fear labels are mapped based on specific ranges of valence, arousal, and dominance dimensions, helping classify the level of fear participants experienced.

**Table 4.1 Fear Level Labels**

| Label | Valence | Arousal | Dominance |
|---|---|---|---|
| **No Fear (0)** | [7:9] | [1:3) | [7:9] |
| **Low Fear (1)** | [5:7) | [3:5) | [5:7) |
| **Medium Fear (2)** | [3:5) | [5:7) | [3:5) |
| **High Fear (3)** | [1:3) | [7:9] | [1:3) |

The DEAP dataset undergoes preprocessing to ensure clean and usable data. EEG signals are normalized using Min-Max normalization to scale them between [0, 1], and basic artifact removal methods are applied to reduce noise from sources like eye blinks and muscle movements. To reduce computational overhead, EEG signals are downsampled from 512 Hz to 128 Hz. The DEAP dataset is a comprehensive and widely used resource in affective computing, biomedical signal processing, and neuroscience research, particularly for emotion recognition and human-computer interaction systems. It consists of EEG and physiological data from 25 participants in the training set, with each participant undergoing 40 trials, resulting in a total of 1,000 trials.

The testing set includes 7 participants with a total of 280 trials, maintaining the same structure. Each trial contains 32-channel EEG data, 8 physiological signals (such as GSR, ECG, EMG), and annotations for valence, arousal, dominance, and liking, offering a rich multimodal perspective on emotional states. The high-resolution EEG recordings allow for detailed temporal analysis, making the dataset ideal for studying real-time emotion monitoring applications in fields like virtual reality, gaming, and personalized marketing. Its diversity across participants improves the generalization of models trained on this dataset, allowing them to better adapt to various populations. The dataset's broad emotional spectrum also makes it suitable for tasks beyond emotion recognition, such as detecting specific emotional states like fear, joy, or sadness. However, working with the DEAP dataset comes with significant challenges. EEG channels often exhibit variability in signal range, requiring careful preprocessing and normalization. Feature extraction, essential for emotion recognition tasks, can be computationally demanding as it involves techniques like fractal dimension calculation and power spectral density analysis. Additionally, the dataset may suffer from class imbalance, as emotions like fear may be less frequently represented compared to more neutral or positive emotional states. Addressing these challenges is crucial for leveraging the full potential of the dataset in developing accurate and robust emotion classification models.

# CHAPTER 5

# IMPLEMENTATION AND RESULT ANALYSIS

## 5.1 DATA PREPROCESSING

To extract samples relevant to fear level detection, the dataset was filtered based on predefined ranges for Valence, Arousal, and Dominance scores. The classification criteria are outlined in Table X, where each fear level corresponds to specific score ranges. The filtering process yielded the following sample counts for each fear category:

- **Relax (No Fear):** 7 samples
- **Low Fear**: 60 samples
- **Medium Fear**: 42 samples
- **High Fear**: 35 samples

Each sample in the dataset originally had a shape of (41 channels, 8064 time points). After extracting the relevant samples for each fear level, the data points were averaged into 12 segments to reduce dimensionality while preserving temporal characteristics. This averaging was done across the time dimension, splitting the 8064 time points into 12 equal segments, and then the mean of each segment was computed. As a result, the final preprocessed dataset consisted of 1728 rows, where each row represented the averaged data from the 12 segments, and each row contained 41 channels. This transformation was essential for feeding the data into deep learning models efficiently, as it reduced computational complexity while maintaining key temporal patterns.

## 5.2 ARTIFICIAL NEURAL NETWORK RESULTS

### 5.2.1 ANN model using 40 Channels for Both Teacher and Student Model

From figure 5.1, the teacher model was implemented as a simple Artificial Neural Network (ANN) with 40 input channels, which correspond to the features extracted from the dataset. The model architecture includes two hidden layers, each with 64 neurons and ReLU activation. A softmax output layer with 4 neurons predicts the fear level classes. The model was trained for 100 epochs, using the training set and tested on the validation set. After training, the teacher model achieved a 51% accuracy, reflecting moderate performance in classifying fear levels based on input features.

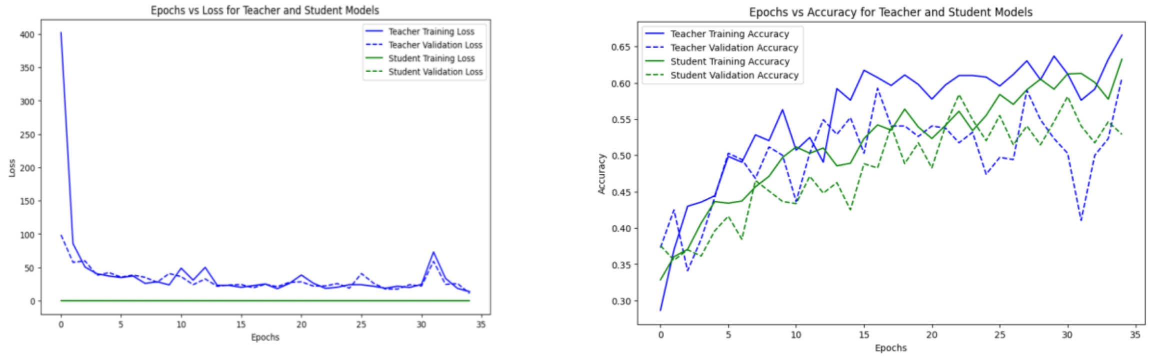


**Fig 5.1 ANN 40 Channel for both teacher and student models**

For the knowledge distillation process, a student model with a simpler architecture was designed with 40 input channels, consisting of two hidden layers with 32 neurons each, also using ReLU activation functions and a softmax output layer. The knowledge distillation was performed using a custom training loop, where the student model's predictions were compared to the teacher model's soft outputs using a distillation loss function. This process allows the student model to mimic the teacher's behavior, while also considering the true labels. A temperature of 3 and an alpha value of 0.1 were used to balance between the distillation loss and the true label loss. After training, the student model achieved an accuracy of 53%, which slightly outperformed the teacher model, highlighting the effectiveness of knowledge distillation in transferring knowledge from the larger model to the smaller one.

## 5.2.2 ANN model using 40 Channels for Teacher and 8 PPS for Student Model

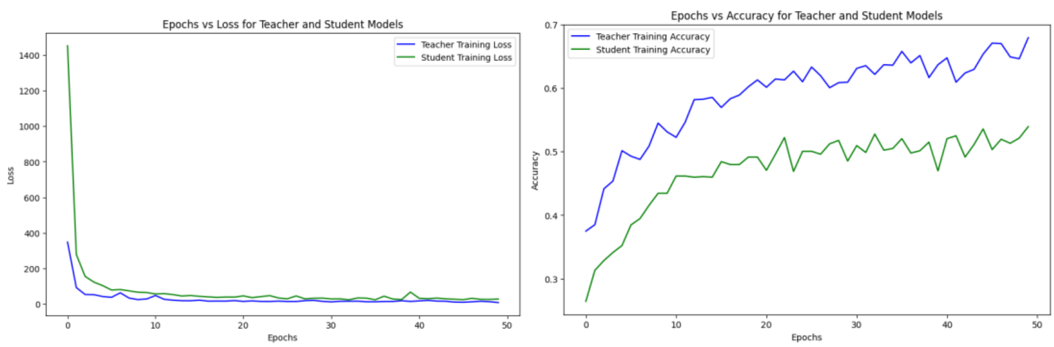


**Fig 5.2 ANN- 40 channels for teacher and 8 PPS for student**

From figure 5.2 , knowledge distillation was applied using a teacher-student architecture, where both models were implemented using Artificial Neural Networks (ANNs). Initially, the loss value was high at the beginning of training, but it progressively decreased as the number of epochs increased, indicating effective learning over time. The teacher model achieved an accuracy of 68% using 40 channels which includes eeg signals and physiological signals, while the student model reached 62% using only 8 physiological signals. While the performance was satisfactory, the student model's lower accuracy suggests that further tuning or enhancements in the distillation process may be necessary to improve the student's performance.The metrics used for evaluating the models included accuracy, loss, and other performance measures. The loss function was employed to capture the error during training, where a high initial loss was observed, gradually reducing as the epochs progressed. Despite the promising results, there is room for improvement in the knowledge distillation process. The significant gap between the teacher model's 68% accuracy and the student model's 62% suggests that the student model did not fully capture the teacher's knowledge. This could be addressed by fine-tuning hyperparameters, increasing the training epochs, or exploring more effective loss functions during distillation.

## 5.3 DENSE NEURAL NETWORK - 300 RESULTS

### 5.3.1 DNN Model using 40 Channels for Teacher and 8 PPS for Student Model

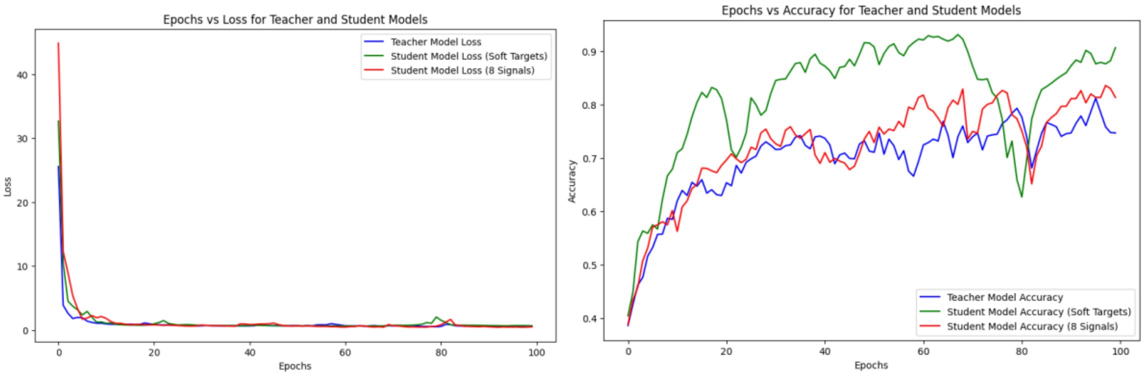


**Fig 5.3 DNN 300 - 40 channels for teacher 8 PPS for student**

In Figure 5.3, two models were developed: a teacher model and a student model. The teacher model was designed to handle 40 input features, including EEG and other physiological signals, using a deep neural network (DNN) with 5 hidden layers, each consisting of 300 neurons. The layers were initialized with a normal distribution, and the ReLU activation function was applied to all hidden layers, while the softmax function was used in the output layer to predict 4 fear levels (Relax, Low Fear, Medium Fear, High Fear). The model was trained with categorical cross-entropy loss and optimized using Adam over 100 epochs with a batch size of 32. The teacher model achieved an accuracy of 69.36% and an F1 score of 68.51%.

The student model was trained through knowledge distillation, using only 8 physiological signals as input. Its architecture matched the teacher model, with 5 hidden layers of 300 neurons and ReLU activations. The output layer used softmax for 4-class predictions. Trained with soft targets from the teacher model, the student model underwent 100 epochs with a batch size of 32. Despite fewer input features, it reached 61.27% accuracy and an F1 score of 60.37%. A baseline student model, trained directly on the hard labels without knowledge distillation, was also developed. Following the same architecture and input features, it slightly outperformed the distillation-based model, achieving 63.29% accuracy and an F1 score of 62.68%. This suggests that in this case, knowledge distillation did not significantly enhance the student model's performance.

### 5.3.2 DNN model using 40 Channels for Teacher and 8 PPS for Student Model with Cross Validation

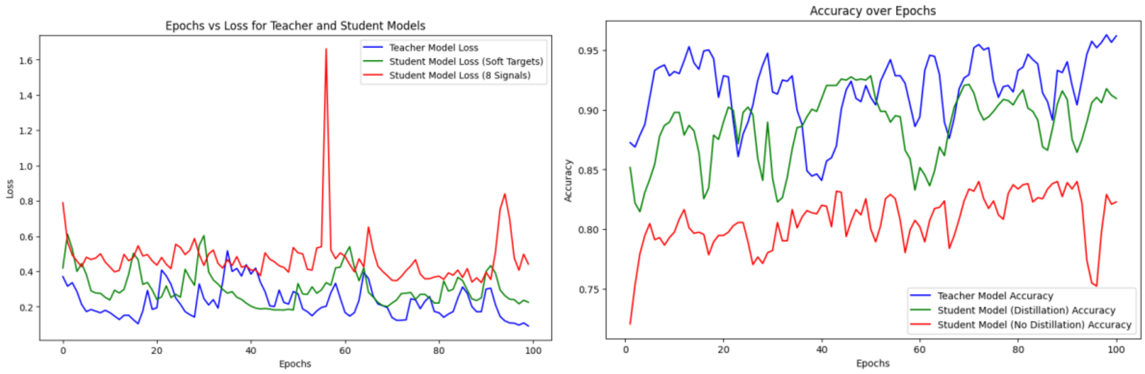


**Fig 5.4 DNN 300 result with cross validation**

In figure 5.4 , a knowledge distillation framework was applied to train a student model using a teacher model for the task of fear level detection. The teacher model was designed to take 40 input features, while the student model utilized only 8 physiological signals for prediction. Both models were based on fully connected deep neural networks with similar architectures. The teacher model consisted of six layers: five dense layers with 300 neurons each, using the ReLU activation function, and a final output layer with four neurons corresponding to the four fear levels (relax, low fear, medium fear, high fear). The output layer used a softmax activation function to classify the input into one of the four fear levels. The model was compiled with the categorical cross-entropy loss function and the Adam optimizer, and it was trained for 100 epochs with a batch size of 32. After training, the teacher model achieved an accuracy of 75.43% and an F1-score of 75.28%.

The student model, which used only 8 physiological input features, followed a similar structure, with five dense layers of 300 neurons each and ReLU activation functions. The final output layer had four neurons with a softmax activation function. The key difference was the knowledge distillation (KD) technique applied: instead of using hard labels, the student model was trained on the soft targets generated by the teacher model. Soft targets were obtained by passing the training data through the teacher model to capture its probabilistic output distribution. The student model was then trained for 100 epochs using these soft targets, also with categorical cross-entropy loss and the Adam optimizer.The results of the knowledge distillation process showed that the student model achieved an accuracy of 73.12% and an F1-score of 72.95%, which is close to the teacher model's performance. Interestingly, training the student model directly on the hard labels without KD resulted in a slightly better performance, with an accuracy of 75.43% and an F1-score of 75.24%. This suggests that while the KD approach preserved the student model’s accuracy, it did not significantly improve performance, possibly due to the simplicity of the architecture or dataset size.

## 5.4 DNN-150 RESULTS

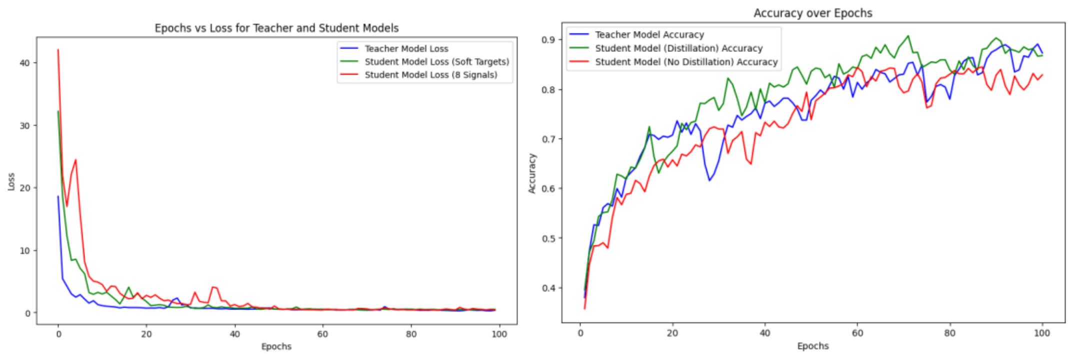


**Fig 5.5 DNN 150 result analysis**

In figure 5.5, both a teacher and student model were created for knowledge distillation in the context of fear level detection. The teacher model is a fully connected neural network with an input dimension of 40, corresponding to 32 EEG signals and 8 additional physiological signals. The network consists of three hidden layers, each with 150 neurons, using the ReLU activation function. The output layer consists of 4 neurons, using the softmax activation function to predict the four fear levels (relax, low fear, medium fear, high fear). The model is compiled using the Adam optimizer and the categorical cross-entropy loss function. The teacher model was trained over 100 epochs with a batch size of 32, achieving an accuracy of 69.94% and an F1-score of 69.16%. The student model uses a similar architecture but with a reduced input dimension of 8, representing only the physiological signals. The student model consists of two hidden layers, each with 150 neurons and ReLU activation, and an output layer with 4 neurons and a softmax activation function. The student model was trained in two ways:

first using soft targets from the teacher model (knowledge distillation) and second without distillation, directly using the 8 physiological signals.

For the knowledge distillation approach, soft targets were generated from the teacher model's predictions. The student model was then trained using these soft targets to mimic the teacher's output, encouraging the student to generalize better. Despite this approach, the student model trained with distillation reached an accuracy of 58.09% and an F1-score of 57.36%, which was lower than the teacher model's performance. Interestingly, training the student model without knowledge distillation (i.e., directly using the labels) yielded better results, with an accuracy of 67.63% and an F1-score of 67.50%. So, while knowledge distillation provided a useful training paradigm.

## 5.5 LSTM RESULTS

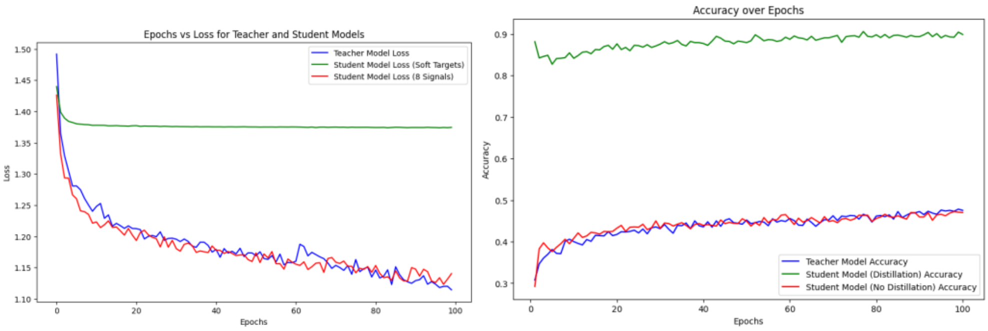


**Fig 5.6 LSTM result analysis**

The teacher model was built using an LSTM network with 64 units, followed by a Dropout layer with a 0.5 rate to minimize overfitting. It includes two Dense layers with 32 and 16 neurons, each using the ReLU activation function, with Dropout layers applied after both. The final layer, consisting of 4 neurons with a softmax activation, predicts the four fear levels. The model uses the categorical cross-entropy loss function and Adam optimizer, with accuracy as the evaluation metric. The input data was reshaped into a 3D format [samples, time steps, features] to fit LSTM requirements, from the figure 5.6, the model achieved an accuracy of 51.16% and an F1-score of 42.67%.

The student model was similarly designed but with a reduced number of neurons to match its smaller input size. It consists of a single LSTM layer with 32 units, followed by a Dense layer with 16 neurons, using ReLU activations, and a softmax output layer. Dropout was applied with the same 0.5 rate. The knowledge distillation process involved using the soft labels predicted by the teacher model, which were generated by applying the softmax function to the teacher's outputs. The student model was trained on these soft labels for 100 epochs, achieving an accuracy of 47.40% and an F1-score of 35.85%. Then the student model was trained without knowledge distillation, it showed a slightly higher performance as shown in figure 5.6, achieving 47.69% accuracy and an F1-score of 39.04%, suggesting that in this instance, the distillation did not considerably enhance the performance of the student model.

## 5.6 TEMPORAL - CNN RESULTS

### 5.6.1 TCN model using 40 Channel for Teacher and 8 PPS for Student Model

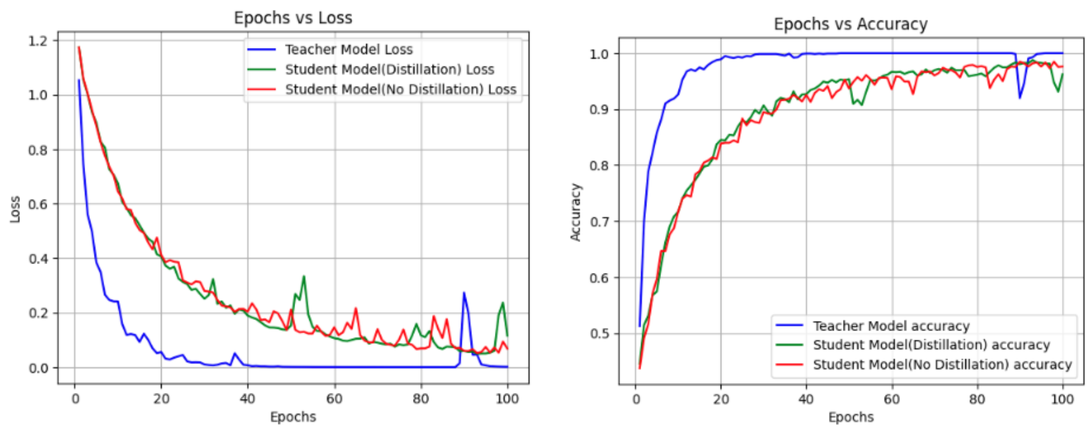


**Fig. 5.7 TCN result analysis for teacher 40 signals and student 8 PPS signals**

The teacher model was implemented using a Temporal Convolutional Network (TCN) architecture designed to process 40 physiological signals. The model consists of two Conv1D layers with 64 filters each. The first layer uses a kernel size of 3 and a dilation rate of 1, while the second layer has the same kernel size with a dilation rate of 2, enabling the model to capture long-range dependencies in the data. Following the convolutional layers, the data is flattened using a Flatten layer, and a Dense layer with 128 neurons and ReLU activation is added. The output layer is a Dense layer with 4 neurons and a softmax activation function to classify the samples into four fear levels. The model was compiled with categorical cross-entropy loss, the Adam optimizer, and accuracy as a performance metric. This teacher model achieved an accuracy of 92.49% and an F1 score of 92.45 after training. The student model was designed similarly but was trained on only 8 physiological signals. It consists of two Conv1D layers with 64 filters each, using a kernel size of 2 and dilation rates of 1 and 2, respectively. After the convolutional layers, the data is flattened, followed by a Dense layer with 128 neurons and ReLU activation, and an output layer with 4 neurons for classification using softmax activation. The model was trained both with and without Knowledge Distillation (KD) from the teacher model. The student model with KD achieved an accuracy of 76.01% and an F1 score of 76.02, while the model trained without KD reached an accuracy of 79.19%

and an F1 score of 79.08. The comparison from figure 5.7 suggests that knowledge distillation did not enhance performance, possibly due to the reduced number of input signals affecting the model's ability to learn the complex representations provided by the teacher.

### 5.6.2 TCN model using 40 Channels for Both Teacher and Student Model

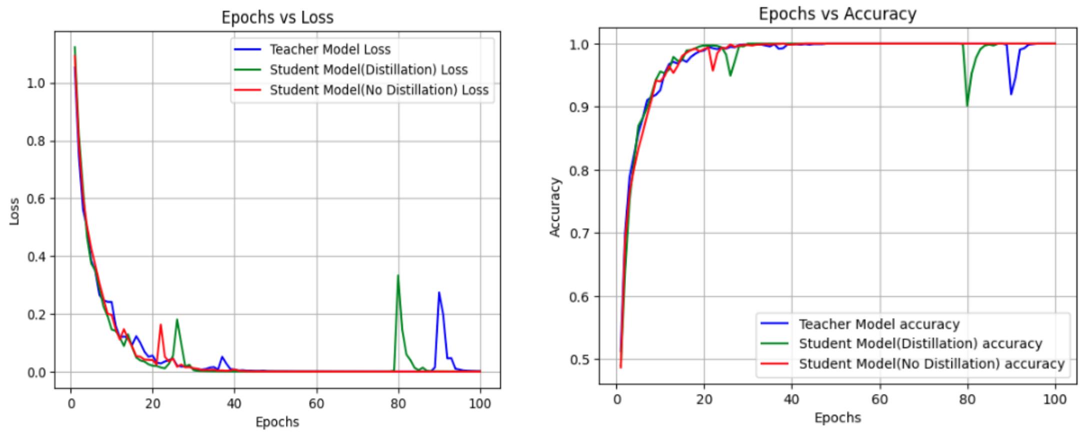


**Fig. 5.8 TCN result analysis for teacher 40 signals and student 40 signals**

From figure 5.7, the Temporal Convolutional Network (TCN)-based teacher model for fear level detection is built with two Conv1D layers, each having 64 filters and kernel sizes of 3. The dilation rates of 1 and 2 allow the model to capture both fine-grained and larger temporal patterns. After these layers, the output is flattened and passed through a fully connected layer with 128 neurons and ReLU activation. The final layer, consisting of 4 neurons with softmax activation, classifies the fear levels into relax, low fear, medium fear, and high fear. Categorical cross-entropy was used as the loss function in the model's compilation, and accuracy and the Adam optimizer were used as performance metrics. The model demonstrated an amazing 92.49% accuracy and 92.45% F1-score after 100 epochs of training.

The student model, which also uses a TCN architecture, incorporates an additional Conv1D layer with 64 filters, further refining temporal feature extraction while maintaining computational efficiency by reducing the number of parameters. Knowledge distillation (KD) is employed to train the student model by generating soft targets from the teacher model's output. The student model trained with KD achieved a slightly higher accuracy of 93.35% and F1-score of 93.35%, surpassing both the teacher model and the student model trained without KD (which achieved an accuracy of 92.77% and F1-score of 92.76%). This demonstrates the effectiveness of the additional convolutional layer and the KD technique in improving model performance.

### 5.6.3 TCN model using 40 Channels for Both Teacher and Student Model with Cross Validation

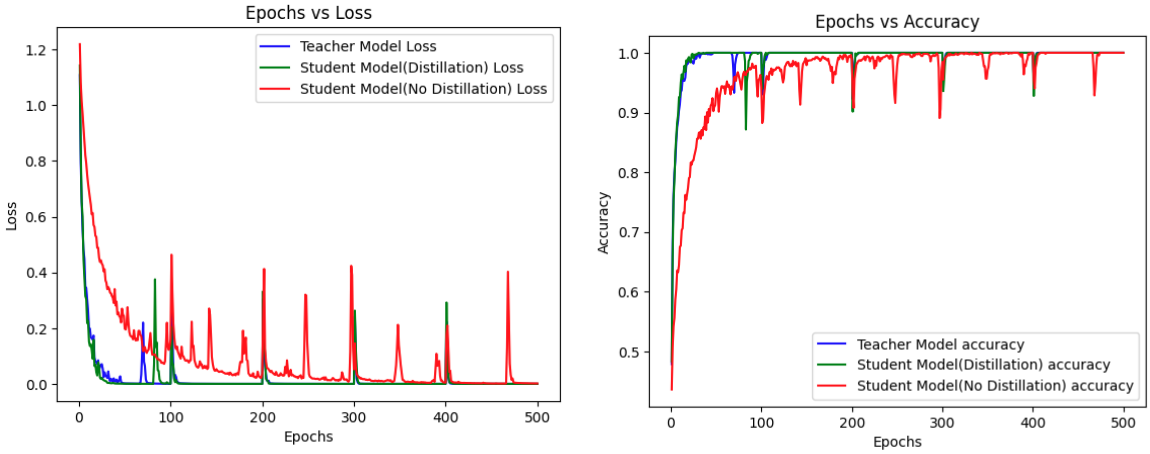


**Fig. 5.9 TCN result analysis for teacher 40 signals and student 40 signals with cross validation**

From figure 5.8, the fear level detection system was developed using a Temporal Convolutional Network (TCN)-based architecture. Two models were implemented, a teacher model and a student model. Both models utilized a sequential structure with 1D convolutional layers. The teacher model consisted of an input layer with a shape of (40, 1), followed by two Conv1D layers with 64 filters, kernel sizes of 3, and dilation rates of 1 and 2, respectively. The convolutional layers in the model were activated using ReLU functions. Following these, the data was flattened and passed through a dense layer with 128 neurons. The final softmax layer, with 4 neurons, classified the four fear levels: relax, low, medium, and high. The model was compiled using the categorical cross-entropy loss function and optimized with Adam, with accuracy serving as the evaluation metric.

The student model, designed to be smaller while maintaining performance, included a similar architecture but added an additional Conv1D layer with a dilation rate of 2. Like the teacher model, the student model concluded with a dense layer of 128 neurons and a softmax output layer. The model was trained using knowledge distillation (KD), where the student model learned from the soft targets generated by the teacher model. A custom pipeline was built using a transformer to flatten the input data, followed by standardization and model training. The model was trained for 100 epochs with a batch size of 20.

In the results, the teacher model achieved a 5-fold cross-validated accuracy and F1 score of 0.9306. The student model with 40 channels closely matched the teacher model, achieving an accuracy of 0.9219 and an F1 score of 0.9220, demonstrating the efficacy of knowledge distillation. However, the student model using only 8 physiological signals (pps) showed a lower performance, with an accuracy of 0.7687 and an F1 score of 0.7681. When trained without knowledge distillation, the student model with 8 signals achieved a slightly higher accuracy of 0.7832 and F1 score of 0.7832. These results highlight the impact of KD in maintaining model performance even with reduced input features.

## 5.7 RESULTS OF HYBRID CNN-LSTM

A hybrid CNN-LSTM teacher model was designed to classify fear levels based on EEG and physiological signals. The teacher model takes two input streams: one for 32 EEG channels and another for 8 physiological signals. The EEG branch utilizes a 1D Convolutional Neural Network (CNN) with two Conv1D layers having 64 and 128 filters, kernel size 3, and ReLU activation. Each convolutional layer is followed by max-pooling layers with a pool size of 2 to down-sample the feature maps. After flattening, the EEG branch is merged with the physiological signal branch, which uses an LSTM layer with 64 units to process the time-series data. The two data streams are concatenated and fed into a Dense layer with 128 neurons, utilizing ReLU activation. The output layer employs softmax for multi-class classification across four classes. The teacher model is trained with the Adam optimizer and uses sparse categorical cross-entropy as the loss function. In terms of performance, the model achieved an impressive accuracy of 91.33% and an F1 score of 91.47%, demonstrating the effectiveness of the hybrid CNN-LSTM architecture in capturing both spatial and temporal features from EEG and physiological signals.

### 5.7.1 Hybrid CNN - LSTM model with LSTM as Student Model

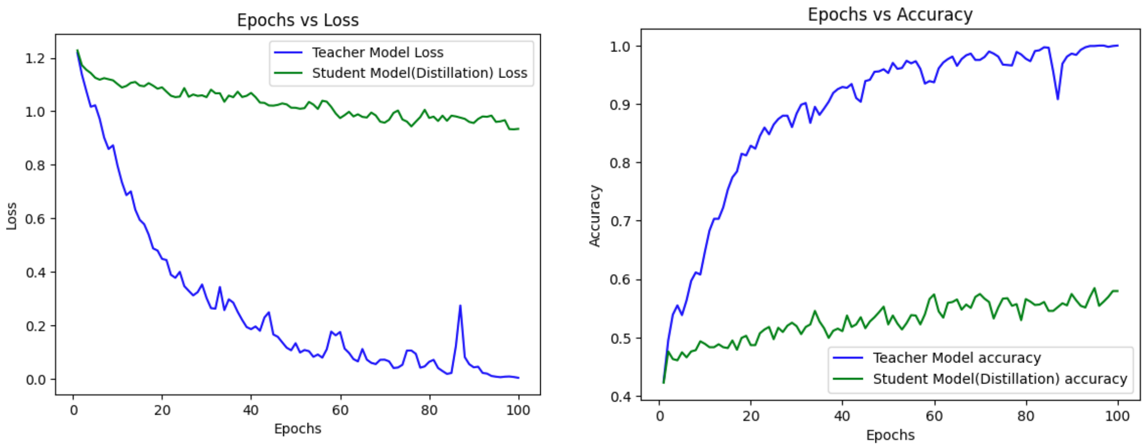


**Fig. 5.10 Result analysis of CNN - LSTM as teacher model and LSTM as student model**

From figure 5.9 for the student model only the physiological signals (8 channels) were used. The student model employs a single LSTM layer with 64 units to extract temporal features. It is followed by a fully connected network comprising Dense layers of 128, 64, and 32 neurons, each using ReLU activation. Dropout layers with a 0.3 dropout rate were included for regularization to prevent overfitting. The output layer, similar to the teacher model, uses a softmax activation function for predicting the 4 fear levels. The student model also used the Adam optimizer and sparse categorical cross entropy as the loss function. The student model, relying solely on the physiological signals and LSTM architecture, resulted in a significantly lower accuracy of 51.73% and an F1 score of 44.26%. This gap highlights the importance of EEG data in fear level detection and suggests that additional knowledge distillation techniques could improve the performance of the student model by transferring more knowledge from the teacher model to compensate for the missing EEG input.

### 5.7.2 Hybrid CNN - LSTM model with DNN-300 as Student Model

A deep neural network (DNN) architecture was implemented using only the 8 physiological signals as input. The student model had three fully connected layers, each with 300 neurons, initialized using a normal distribution and employing the ReLU activation function. The final output layer, identical to the teacher model, consisted of 4 neurons with softmax activation. This model was also trained using the Adam optimizer, with sparse categorical cross entropy loss. The student model was trained as part of the knowledge distillation process, where the teacher model's predictions were used to guide the training of the student model, ensuring that the student could learn patterns despite using fewer input signals. After training, the student model, which only used physiological signals, achieved an accuracy of 64.45% and an F1 score of 64.34%. This gap in performance highlights the importance of including EEG signals for improved accuracy. However, the knowledge distillation technique allowed the student model to learn with reduced input, which could be beneficial in situations where collecting EEG data is not feasible.

## 5.8 RESULT ANALYSIS

The table 5.1 compares the performance of different models without using Knowledge Distillation (KD). KD is a technique in which a smaller, simpler "student" model is trained to replicate the behavior of a larger, more complex "teacher" model. In this table, models like Artificial Neural Networks (ANN), Deep Neural Networks (DNN), and LSTMs were trained with all 40 input channels from the EEG data, but without any knowledge distillation.

The ANN model struggled with generalizing to the test set, reflecting a significant gap between training and testing performance. The DNN 150 model performed better, demonstrating improved accuracy both during training and testing compared to ANN. However, there is still some degree of overfitting. The LSTM model had a surprisingly low performance, indicating difficulties in learning from the available features in this configuration. This suggests that LSTM may require further tuning or feature engineering to perform well on this task. The table 5.1 highlights the challenges models face in generalizing when trained with large numbers of channels, leading to the exploration of Knowledge Distillation techniques to improve performance with fewer inputs. Table 5.2 presents the results after applying Knowledge Distillation (KD), where smaller "student" models with fewer channels (8 instead of 40) were trained to mimic the behavior of the larger "teacher" models. This approach reduces computational complexity while aiming to preserve model performance. The ANN model showed a modest improvement in accuracy after applying KD, demonstrating the effectiveness of the distillation process in retaining key information with fewer input channels.

**Table 5.1 Result Analysis for Deep Learning Models without Knowledge Distillation**

| Model Name | Without Knowledge Distillation - Deep Learning Models | | | | |
|---|---|---|---|---|---|
| | **Channels** | **Epoch** | **Layers** | **Training Accuracy** | **Testing Accuracy** |
| **ANN** | 40 | 100 | 3 | 72.14 | 57.51 |
| **DNN 150** | 40 | 100 | 6 | 82.19 | 69.94 |
| **DNN 300** | 40 | 100 | 4 | 78.24 | 69.36 |
| **LSTM** | 40 | 100 | 4 | 47.78 | 48.84 |

In the case of DNN 150, the distilled model performed slightly worse after applying KD. This could be due to the reduced capacity of the student model, which may not have been sufficient to capture the complexity of the EEG data. The LSTM model's performance remained unchanged after applying KD, suggesting that the model may need further optimization in terms of architecture or hyperparameters to benefit from knowledge distillation.

Overall, the table 5.2 emphasizes that KD can help in creating smaller, efficient models with reduced computational requirements, though the impact on accuracy varies across different models. It also presents a performance comparison of several deep learning models, including DNN 300, CNN-LSTM Hybrid, and Temporal Convolutional Network (TCN), with and without Knowledge Distillation (KD). The focus here is on the application of KD to reduce the number of input channels from 40 to 8 and the subsequent impact on model performance. e DNN 300 student model, after applying KD, maintained a high level of performance, with only a small drop in accuracy compared to the teacher model. This result shows that KD was highly effective for this model, allowing the smaller student model to mimic the larger teacher with minimal performance loss. The CNN-LSTM Hybrid model demonstrated excellent performance as a teacher model, but the student model experienced a significant drop in accuracy after distillation. This drop could be attributed to the complexity of the hybrid model, where the integration of convolutional layers (to capture spatial patterns) and LSTM layers (for temporal dependencies) may have been difficult for the smaller student model to fully replicate.

**Table 5.2 Result Analysis for Deep Learning Models with Knowledge Distillation**

<table>
<tr><th rowspan="3">Model Name</th><th colspan="9">With Knowledge Distillation</th></tr>
<tr><th colspan="4">Teacher Model</th><th colspan="5">Student Model</th></tr>
<tr><th>Channels</th><th>Epoch</th><th>Layers</th><th>Testing Accuracy</th><th colspan="2">Channels</th><th>Epoch</th><th>Layers</th><th>Testing Accuracy</th></tr>
<tr><td>ANN</td><td>40</td><td>100</td><td>3</td><td>57.51</td><td colspan="2">8</td><td>100</td><td>2</td><td>49.71</td></tr>
<tr><td>DNN 150</td><td>40</td><td>100</td><td>6</td><td>69.94</td><td colspan="2">8</td><td>100</td><td>5</td><td>58.09</td></tr>
<tr><td>DNN 300</td><td>40</td><td>100</td><td>3</td><td>75.43</td><td colspan="2">40</td><td>100</td><td>2</td><td>73.12</td></tr>
<tr><td>LSTM</td><td>40</td><td>100</td><td>4</td><td>51.16</td><td colspan="2">8</td><td>100</td><td>2</td><td>48</td></tr>
<tr><td>TCN</td><td>40</td><td>100</td><td>4</td><td>92.03</td><td colspan="2">40</td><td>100</td><td>5</td><td>93.03</td></tr>
<tr><td rowspan="2">CNN-LSTM Hybrid</td><td rowspan="2">40</td><td rowspan="2">100</td><td rowspan="2">5</td><td rowspan="2">91.33</td><td>DNN 300</td><td>8</td><td>100</td><td>4</td><td>64.45</td></tr>
<tr><td>LSTM</td><td>8</td><td>100</td><td>5</td><td>51.73</td></tr>
</table>

The TCN model performed exceptionally well, with the teacher model achieving the highest test accuracy among all models at 92.03%. Interestingly, after applying KD, the student model surpassed the teacher model's accuracy, achieving 93.03%. This improvement suggests that the TCN architecture is highly compatible with KD and can effectively transfer knowledge to a smaller model without losing performance. TCN's ability to capture long-term dependencies through its hierarchical convolutional structure likely contributed to its superior performance in this context. The table 5.2 also emphasizes the role of cross-validation in ensuring the robustness of these results. Cross-validation helped in mitigating the risk of overfitting, providing a more reliable assessment of model generalization.

# CHAPTER 6

# CONCLUSION AND FUTURE WORK

## 6.1 CONCLUSION

In our project on fear level classification using the DEAP dataset, with knowledge distillation using teacher-student architectures. Our experiments revealed that different combinations of teacher and student models yielded varying results. Notably, when both the teacher and student models employed the Temporal CNN architecture, trained with 40 channels, the teacher model achieved an accuracy of 92%, while the student model slightly outperformed it, achieving 93%. In another setup, both the teacher and student models were based on a DNN-300 architecture, with the teacher model reaching 75% accuracy and the student model 73%. These results demonstrate that the knowledge distilled from the teacher model effectively transfers to the student model, maintaining or even improving performance while enabling smaller, more efficient implementations. The project highlights the potential of knowledge distillation for enhancing model performance while improving computational efficiency, making it suitable for real-time or near-real-time processing in fear classification tasks.

Furthermore, the ability to transfer knowledge from complex teacher models to more lightweight student models without sacrificing accuracy suggests broader applicability in other domains involving non-linear and high-dimensional data, such as healthcare, neuroscience, and biometrics. Additionally, by reducing the computational burden, knowledge distillation opens up the possibility of deploying these models on edge devices, such as smartphones or wearable sensors, allowing for continuous monitoring and emotion tracking in everyday environments. This makes it feasible to develop more personalized and context-aware systems that can adapt to users' emotional states in real time. Overall, this project not only validates the efficacy of deep learning and knowledge distillation for emotion classification but also underscores the growing role of affective computing in creating smarter, more responsive systems that can enhance user experiences across various industries. It paves the way for future research into optimizing both model architecture and feature engineering to handle complex emotional states with greater precision and speed.

## 6.2 FUTURE ENHANCEMENT

In advancing the study of fear level classification using deep learning models with knowledge distillation, several promising research directions can be pursued to improve the models' accuracy and applicability by adapting feature engineering. One significant area of exploration is the incorporation of additional features, such as multiscale and wavelet transform features. By employing multiscale entropy and fractal dimensions across various scales, it may be possible to gain a more detailed understanding of the complexity in EEG signals, thereby enhancing classification precision. Wavelet transform features, which encompass both time and frequency information, could also play a crucial role in differentiating between varying levels of fear more effectively. Additionally, data augmentation, particularly through advanced techniques like Generative Adversarial Networks (GANs), could address the challenges associated with limited datasets. These methods can generate synthetic EEG data, increasing the robustness of the models. Moreover, developing cross-subject data augmentation techniques could lead to the creation of generalized models that perform well on previously unseen subjects, expanding the models' applicability.

The interpretability and transparency of these models are also critical areas for future research. Employing methods like SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations) could enhance the models' transparency, enabling researchers and clinicians to discern which features or EEG signal components significantly influence the models' predictions. Furthermore, creating visualization tools to illustrate the relationship between fractal features and different fear levels could offer valuable insights into the physiological mechanisms underlying fear. Additionally, developing personalized fear detection models that can adapt to individual differences in EEG patterns may result in more accurate and tailored interventions. Expanding the dataset to include larger, more diverse populations and validating the models across various domains will be crucial for enhancing their robustness and generalizability. These research directions have the potential to significantly advance the field of fear level classification, making the models more accurate, interpretable, and applicable in real-world contexts.

# BIBLIOGRAPHY

[1] N. Masuda and I. E. Yairi, "Multi-Input CNN-LSTM deep learning model for fear level classification based on EEG and peripheral physiological signals," *Frontiers in Psychology*, vol. 14, art. no. 1141801, 2023. doi: 10.3389/fpsyg.2023.1141801.

[2] M. Z. I. Ahmed, N. Sinha, E. Ghaderpour, S. Phadikar, and R. Ghosh, "A novel baseline removal paradigm for subject-independent features in emotion classification using EEG," *Bioengineering*, vol. 10, art. no. 54, 2023. doi: 10.3390/bioengineering10010054.

[3] O. Bălan, G. Moise, A. Moldoveanu, M. Leordeanu, and F. Moldoveanu, "An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy," *Sensors*, vol. 20, art. no. 496, 2020. doi: 10.3390/s20020496.

[4] A. Dziedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: review of sensors and methods," *Sensors*, vol. 20, art. no. 592, 2020. doi: 10.3390/s20030592.

[5] R. Ghosh, S. Phadikar, N. Deb, N. Sinha, R. Das, and E. Ghaderpour, "Automatic eye-blink and muscular artifact detection and removal from EEG signals using k-nearest neighbour classifier and long short-term memory networks," *IEEE Sensors Journal*, vol. 23, pp. 5422–5436, 2023. doi: 10.1109/JSEN.2023.3237383.

[6] X. Li, S. Chen, X. Hu, and J. Yang, “Understanding the disharmony between dropout and batch normalization by variance shift,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Silver Spring, MD, USA, 2019, pp. 2682–2690.

[7] J. A. Miranda, F. C. Manuel, L. Gutiérrez-Martín, J. M. Lanza-Gutierrez, M. Portela-García, and C. López-Ongil, “Fear recognition for women using a reduced set of physiological signals,” *Sensors*, vol. 21, no. 5, Art. no. 5, 2021. doi: 10.3390/s21051587.

[8] L. Yang et al., “Edgetb: A hybrid testbed for distributed machine learning at the edge with high fidelity,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 10, 2022.

[9] C. E. Orozco-Mora, D. Oceguera-Cuevas, R. Q. Fuentes-Aguilar, and G. Hernández-Melgarejo, “Stress level estimation based on physiological signals for virtual reality applications,” *IEEE Access*, vol. 10, pp. 68755–68767, 2022. doi: 10.1109/ACCESS.2022.3186318.

[10] L. Petrescu, C. Petrescu, A. Oprea, O. Mitruț, G. Moise, A. Moldoveanu, et al., “Machine learning methods for fear classification based on physiological features,” *Sensors*, vol. 21, Art. no. 4519, 2021. doi: 10.3390/s21134519.

[11] M. Sajjad, M. Nasir, K. Muhammad, S. Khan, Z. Jan, A. K. Sangaiah, et al., "Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities," *Future Gener. Comput. Syst.*, vol. 108, pp. 995–1007, 2020, doi: 10.1016/j.future.2017.11.013.

[12] J. Šalkevicius, R. Damaševičius, R. Maskeliunas, and I. Laukiene, "Anxiety level recognition for virtual reality therapy systems using physiological signals," *Electronics*, vol. 8, no. 9, p. 1039, 2019, doi: 10.3390/electronics8091039.

[13] P. Schmidt, A. Reiss, R. Duerichen, C. Marberger, and V. K. Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in *Proc. 20th ACM Int. Conf. Multimodal Interaction* (ICMI), New York, NY, USA, 2018, pp. 400–408.

[14] L. Shu, J. Xie, M. Yang, Z. Li, D. Liao, X. Xu, et al., "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, 2018, doi: 10.3390/s18072074.

[15] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe, “ASCERTAIN: emotion and personality recognition using commercial sensors,” *IEEE Trans. Affect. Comput.*, vol. 9, no. 2, pp. 147–160, 2018, doi: 10.1109/TAFFC.2016.2625250.