SYNOPSIS

Identifying human fear states using deep learning represents a significant stride in affective computing, finding utility across domains such as healthcare, virtual reality, and interactions between humans and computers. This research leverages Knowledge Distillation (KD) and Transfer Learning (TL) to boost model efficiency while preserving accuracy levels. Various deep learning architectures, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), are employed to categorize fear levels based on physiological signals sourced from datasets including DEAP and ASCERTAIN. KD facilitates the transfer of knowledge from intricate teacher models to more streamlined student models, thereby cutting down computational expenses without sacrificing performance. [1] Furthermore, TL enhances the model's capacity to generalize effectively across datasets exhibiting diverse distributions. Feature extraction methods, including Higuchi's and Petrosian's fractal dimensions, are instrumental in capturing key temporal and frequency characteristics within EEG signals, which in turn improves classification accuracy. Experimental results suggest that models based on CNNs are superior to those based on RNNs for feature extraction, and student models trained with KD exhibit efficiency enhancements with only minor compromises in accuracy. The framework utilizing KD and TL for detecting fear levels shows promise for practical applications, particularly in settings with limited resources. KD allows more compact models to maintain crucial information learned by larger networks, making them suitable for deployment on edge devices while sustaining performance. Concurrently, TL improves the models' flexibility, permitting their fine-tuning across datasets such as DEAP (with valence-arousal labels) and ASCERTAIN (featuring continuous emotional states). Comparative assessments reveal that Temporal CNNs achieve higher accuracy, whereas LSTM and GRU models are effective at capturing the sequential nature of EEG signals. Employing KD substantially enhances the performance of student models, leading to accuracy increases of 2-3% compared to standard student models and simultaneously lowering computational requirements. These findings indicate that KD and TL provide a potent strategy for optimizing deep learning models in the classification of fear, thereby opening avenues for the development of more efficient, flexible, and scalable affective computing systems relevant to mental health tracking, virtual reality settings, and interactive AI applications.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER **Page No.**

CHAPTER 1

INTRODUCTION

A deep understanding of human fear responses is essential for progressing our knowledge of emotional behavior, holding significant relevance for fields like affective computing, human-computer interaction, and the diagnosis of mental health disorders. Given its fundamental nature, fear strongly influences both decisions and actions, making its precise detection important for diverse real-world applications, including virtual reality, therapeutic treatments, and safety surveillance systems. The DEAP dataset serves as a widely accepted benchmark in emotion recognition research, offering physiological data—especially electroencephalography (EEG)—alongside annotations detailing emotional dimensions like valence (pleasantness), arousal (intensity), and dominance (sense of control). These affective elements are critical for analyzing emotional states, with fear commonly characterized by high arousal and low valence. Employing the DEAP dataset allows for the efficient categorization of fear levels using physiological markers. Additionally, the ASCERTAIN dataset stands as another prominent resource in affective computing, often utilized for emotion recognition endeavors.

Deep learning models, including LSTMs, CNNs, and GRUs, have shown remarkable ability in identifying complex patterns within physiological data like EEG signals.Their proficiency in detecting subtle signal changes leads to more accurate fear level identification compared to conventional methods. Knowledge Distillation (KD) is crucial for boosting model efficiency by transferring learned features from a large, intricate teacher model to a smaller, lighter student model. Likewise, Transfer Learning (TL) enables quicker and more effective training on limited data by leveraging knowledge gained from related areas. A combined deep learning strategy, TL+KD, integrates the advantages of both techniques, allowing for comprehensive evaluation against individual TL and KD methods for classifying fear levels. By including teacher model guidance during fine-tuning, the student model demonstrates improved validation performance, including higher accuracy. Moreover, this approach enhances overall model performance, as indicated by various evaluation metrics beyond just accuracy.

1.1 PROBLEM STATEMENT

This project tackles the challenge of boosting the performance of emotion recognition models, especially those utilizing EEG and GSR signals, to achieve more precise fear level classification. Accurate fear categorization is essential for applications in psychology, neurology, and human-computer interaction. The core aim of this project is to enhance deep learning models, striving for improved accuracy and reliability in fear detection. By integrating advanced deep learning techniques, knowledge distillation, and the combination of multimodal data, the project seeks to significantly improve the efficiency of emotion recognition systems. The development of these sophisticated models will facilitate a more comprehensive and accurate understanding of emotional states, thus benefiting therapeutic practices, research endeavors, and interactive technologies. Within the wider field of affective computing, this project aims to provide tools that not only offer superior accuracy but also demonstrate greater applicability in real-world settings. This advancement holds considerable promise for transforming how emotional responses are interpreted and applied across various domains.

1.2 SCOPE AND MOTIVATION

This project aims to significantly enhance the accuracy and generalizability of fear detection through the application of advanced deep learning techniques, with a specific emphasis on knowledge distillation and transfer learning. The proposed strategy involves developing a holistic approach that includes investigating ensemble and hybrid models, alongside the integration of multimodal data sources. By restructuring and adapting datasets for diverse training methodologies, the project addresses existing limitations in emotion recognition systems. Furthermore, transfer learning will be utilized to efficiently accelerate neural network training on limited datasets by leveraging feature representations learned from related tasks, while knowledge distillation will be employed to create computationally efficient models that maintain high performance.

The main motivation behind this project is to address the limited exploration of combining knowledge distillation (KD) and transfer learning (TL) for emotion recognition, particularly in detecting fear levels. By integrating KD and TL with state-of-the-art deep learning techniques, this project aims to improve classification accuracy and reliability. The overarching goal is to make a substantial contribution to affective computing by providing enhanced emotion analysis tools applicable in human-computer interaction, psychology, and neuroscience.

The results of this study offer considerable potential for the progression of affective computing technologies. By mitigating the limitations of existing emotion recognition models, the project seeks to not only enhance the precision of fear level classification but also to contribute to the creation of more resilient and adaptable emotion analysis tools. Through the integration of multimodal data and the use of advanced deep learning techniques, including knowledge distillation, this research is anticipated to expand the practical applications of these technologies in areas like psychology, neurology, and human-computer interaction. Its novel methodology for model optimization and dataset organization could establish new benchmarks within the industry and drive advancements in our knowledge and utilization of emotional reactions across a wide array of fields.

The outcomes of this research is a valuable resource for industry professionals, policymakers, and researchers involved in the development and application of emotion recognition technologies. By providing comprehensive insights into advanced deep learning models and their efficacy in classifying fear levels, the study offers practical guidance for integrating these technologies into real-world applications. Industry practitioners can leverage the enhanced accuracy and robustness of the models to improve user experiences and emotional analytics across diverse sectors. Policymakers can use the findings to inform the creation of 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 integrated approach is expected to foster ongoing innovation in affective computing, contributing to the advancement of reliable and precise emotion analysis tools.

CHAPTER 2

LITERATURE SURVEY

2.1 RELATED WORKS

In 2024, Xiuzhen Yao released an "Emotion classification model using Transformer and CNN for EEG spatial-temporal feature learning".Their method combines CNNs for local features and Transformers for long-range dependency capturing. The research accentuates the weakness of conventional methods in handling intricate EEG signals. Through the use of attention mechanisms, their model accurately improves emotion recognition accuracy. They verified their framework with benchmark EEG datasets, demonstrating better performance than traditional deep models. The integration of CNN and Transformer enhanced the representation of EEG spatial and temporal features. Experimental results indicated higher classification accuracy than LSTMs and other recurrent models. The research highlights the importance of deep learning in affective computing. Their results indicate promising applications in brain-computer interfaces and mental illness monitoring. Further studies can explore improving computational efficacy to support real-time applications[1].

In 2024, research "A Deep Learning Approach for Fear Recognition on the Edge Based on Two-Dimensional Feature Maps" presented a sophisticated deep learning model for real-time fear recognition. The study suggested an innovative technique of transforming physiological signals like EEG and peripheral biometric information into two-dimensional feature maps for enhanced spatial representation. A convolutional neural network (CNN)-based model was used to learn useful patterns from such feature maps in order to perform correct fear classification. The paper highlighted the benefits of edge computing to support real-time processing with low latency and low computational cost. Experimental results proved that the proposed method had higher fear detection accuracy compared to conventional machine learning approaches. The study tackled important concerns like limited resources in edge devices and physiological response heterogeneity. Results show promising applications in wearable health tracking, virtual reality therapy, and real-time affective computing. This work highlights the increasing contribution of deep learning and edge AI to emotion recognition. Future research can be aimed at improving model efficiency for implementation on low-power IoT and mobile devices[2].

In 2023, the work "Emotion KD: The Cross-Modal Knowledge Distillation Framework for Emotion Recognition Based on Physiological Signals" came up with an innovative method in emotion recognition. The work made deep learning better suited for use in affective computing by the application of knowledge distillation, which was applied to increase efficiency in models with no loss in classification accuracy. Using multimodal physiological signals including EEG and ECG, this work ventured into cross-modal knowledge transfer to strengthen feature representation. The framework adopted a teacher-student model, where a complex teacher network condensed knowledge into a light student model for real-time emotion classification. Extensive experimentation demonstrated that EmotionKD outperformed traditional deep learning models, achieving superior accuracy and computational efficiency. The study employed benchmark physiological datasets and rigorous validation techniques to assess performance. Despite the challenges of multimodal feature integration, the method demonstrated promising results in capturing intricate emotional states. This research underscores the significance of knowledge distillation in reducing model complexity while maintaining accuracy. Its implications extend to emotion-based applications in healthcare, human-computer interaction, and affective gaming. Future work could focus on optimizing the framework for real-world deployment and expanding its applicability to diverse physiological signals[3].

In 2023, the "Multi-Input CNN-LSTM Deep Learning Model for Fear Level Classification Based on EEG and Peripheral Physiological Signals" research investigated a sophisticated deep learning model for fear detection. The paper proposed a hybrid CNN-LSTM model that integrates Convolutional Neural Networks (CNN) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for temporal sequence learning. EEG signals, in addition to peripheral physiological signals like heart rate and skin conductance, were employed to enhance classification performance. The authors highlighted the advantages of integrating multimodal physiological data for enhanced fear recognition. Experimental results demonstrated that the CNN-LSTM model outperformed traditional machine learning approaches in classifying fear levels. The study addressed key challenges such as variability in physiological responses and real-time implementation. The findings have potential applications in mental health monitoring, virtual reality, and human-computer interaction. This research underscores the importance of deep learning in affective computing and emotion recognition. Future work may focus on optimizing model efficiency for real-time fear detection in wearable systems[4].

In 2022, the paper "On Effects of Knowledge Distillation on Transfer Learning" by Sushil Thapa investigated the effects of knowledge distillation (KD) on transfer learning in deep neural networks. The study evaluated how KD improves model generalization through transferring knowledge from a complex teacher model to a less complex student model. By combining KD with transfer learning, the study proved enhanced performance in situations with limited data. The author experimented on several benchmark datasets, underscoring the contribution of KD in alleviating overfitting and improving feature learning. The results showed that KD not only models compression but also retains key knowledge for downstream tasks. This method was useful for real-time applications, striking a balance between computational efficiency and accuracy. The study further emphasized the adaptability of KD across different domains, making it a valuable tool in deep learning. Its implications extend to fields like natural language processing, computer vision, and biomedical applications. Future research may explore optimizing KD strategies to maximize transfer learning benefits in resource-constrained environments[5].

In 2021, the paper "Multi-Domain Feature Fusion for Emotion Classification Using DEAP Dataset" by M. Alarcão and Manuel J. Fonseca proposed a new method of EEG-based emotion recognition. The work utilized the DEAP dataset containing multichannel EEG and physiological signals to obtain features across different domains to enhance classification performance. The research used time-domain, frequency-domain, and nonlinear features to enhance emotion detection. A Support Vector Machine (SVM) classifier was utilized, with strict validation methods guaranteeing stable performance assessment. In spite of the difficulties of emotion recognition based on EEG, the suggested multi-domain fusion method proved enhanced accuracy compared to single-domain conventional techniques. The findings emphasized the significance of incorporating heterogeneous feature sets to identify intricate patterns of emotions. This paper has profound implications for affective computing, human-computer interaction, and real-time emotion recognition systems. The research further highlighted the promise of EEG-based models for healthcare, gaming, and adaptive user interfaces. There is potential for further studies to optimize computational efficiency towards real-time deployment and investigate deep learning-based feature fusion approaches[6].

In 2020,the study titled "Physiological Inspired Deep Neural Networks for Emotion Recognition" by Ferreira et al. introduced a novel deep learning framework tailored to process physiological signals for emotion recognition. The authors proposed a deep neural network architecture inspired by the structure and functioning of human physiological systems. Their model was designed to capture complex patterns in signals such as EEG, ECG, and GSR, which are closely linked to emotional responses. The approach emphasized end-to-end learning, minimizing the need for manual feature engineering. Through extensive experiments, the study demonstrated that physiologically inspired architectures could outperform traditional machine learning models in recognizing emotional states. The results showed improved accuracy and generalization across multiple datasets. The research also highlighted the potential of deep learning to bridge the gap between raw physiological data and high-level emotion classification. This work is particularly relevant for applications in affective computing and emotion-aware systems. Ultimately, the study showcased how biologically informed neural models can enhance the understanding and detection of human emotions.[7].

In 2020, the paper "Stress Detection with Machine Learning and Deep Learning Using Multimodal Data" by Sorasa M. Alarcão and Manuel J. Fonseca investigated recent methods of stress detection from physiological signals. The work utilized multimodal data such as EEG, ECG, and other physiological signals to enhance the accuracy of stress classification. The authors contrasted traditional machine learning models, i.e., Support Vector Machines (SVM) and Random Forests, with deep learning methods, i.e., Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks. Experimental results showed that deep learning models performed better than traditional methods in extracting useful patterns from multimodal signals. The study also emphasized the importance of feature fusion techniques to enhance stress detection performance. Benchmark datasets were used for validation, showcasing the potential of AI-driven approaches in real-world stress monitoring applications. The findings have significant implications for mental health monitoring, workplace stress management, and wearable health technology. Future research may focus on optimizing real-time stress detection models for practical deployment[8].

In 2020, the study "An Investigation of Various Automatic Fear Detection" by Oana Bălan, Gabriela Moise, and Alin Moldoveanu explored different methodologies for detecting fear using automatic recognition systems. The study examined physiological and behavioral cues, such as EEG, heart rate, and facial expression, to maximize fear classification accuracy. The authors compared several machine learning and deep learning models and evaluated their performance in detecting fear in different datasets. Feature extraction and selection techniques were employed to improve model performance and reduce computational complexity. Experimental results indicated that multimodal approaches significantly outperformed unimodal methods in detecting fear responses. The study highlighted the challenges associated with variability in physiological signals and individual differences in fear perception. Applications of this research extend to mental health monitoring, security systems, and human-computer interaction. The findings underscore the potential of AI-driven fear detection in real-world scenarios, including virtual reality and affective computing. Future work may focus on integrating real-time fear detection into wearable and IoT-based monitoring systems[9].

In 2021, VM Joshi and RB Ghongade put forth in the paper "EEG-Based Emotion Detection Using Fourth Order Spectral Moment and Deep Learning" a hybrid method to detect emotions using EEG signals.The researchers focused on extracting the fourth-order spectral moment, a higher-order statistical feature capable of capturing intricate signal variations associated with different emotional states. This feature was used to enhance the quality of emotion-related information fed into deep learning models. By combining advanced signal processing techniques with the learning capability of neural networks, the study achieved high accuracy in classifying emotional states. The integration of fourth-order spectral features helped in improving the robustness and sensitivity of emotion recognition. Experimental outcomes proved the efficacy of this method in various emotional classes. The study emphasized the significance of both meaningful feature extraction and deep learning for accurate emotion detection. It also proposed the applicability of this method in mental health monitoring and human-computer interaction applications. In general, the research provided significant contributions to EEG-based affective computing and intelligent system design.[10].

In 2020, the paper "Classifying the Levels of Fear by Means of Machine Learning Techniques and VR in a Holonic-Based System for Treating Phobias" proposed a machine learning-based method of fear classification in virtual reality (VR) settings. The study utilized physiological responses, including EEG and heart rate, to measure levels of fear during VR exposure therapy. Different machine learning models, such as Support Vector Machines (SVM) and Random Forest, were used to classify various degrees of fear. The study proposed a holonic-based system, which enhances adaptability and personalization in phobia treatment by dynamically adjusting VR scenarios based on the user’s fear response. Experimental results demonstrated that integrating machine learning with VR significantly improved the accuracy of fear classification. The study highlighted the potential of AI-driven systems in mental health applications, particularly for exposure therapy in controlled virtual environments. The findings suggest that personalized VR-based treatments could enhance phobia therapy outcomes. Future research may explore deep learning approaches and real-time adaptation techniques for further improvements in fear classification and therapeutic effectiveness[11].

In 2018, Santamaria-Granados et al. investigated the use of deep convolutional neural networks (CNNs) for emotion detection from physiological signals. They utilized the AMIGOS dataset, which contains EEG, ECG, and GSR signals of participants while they viewed emotionally arousing videos. The authors sought to classify arousal and valence emotions using this multimodal physiological data. In contrast to conventional machine learning approaches, their CNN-based method needed little feature engineering, instead relying on the model's capacity to learn patterns directly from raw signals. The outcome displayed encouraging increases in classification accuracy over conventional methods. The research proved that CNNs were capable of capturing temporal and spatial characteristics of physiological signals effectively. The difficulty of handling small datasets in emotion recognition was also noted. To counter this, they used data augmentation and stringent preprocessing. This work is a major step toward more natural and non-intrusive emotion detection systems. It lends support to the fact that deep learning has a major role to play in affective computing and human-computer interaction[12].

2.2 GAPS IDENTIFIED

1. The combined application of KD and TL for emotion recognition, especially in fear level detection, is underexplored.
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, GSR), 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. Most are concerned that detection systems do not integrate multiple data sources, rather relying on a single modality, like speech or facial expressions. This absence of

multimodal integration restricts the system from offering an accurate evaluation 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

The proper functioning of any application or system works upon its system requirements, which specify the necessary hardware and software parameters, both at their lower and upper limits. Determining these requirements accurately involves a detailed analysis of the system's functional aspects—what it needs to accomplish—and its non-functional aspects—how well it needs to perform. It's also vital to identify any constraints or limiting factors that could affect its operation. By taking this approach, we can guarantee the system performs efficiently within its defined context, ultimately leading to optimal user satisfaction and complete functionality.

3.1 FUNCTIONAL REQUIREMENTS

3.1.1 Real Time Fear Detection Analysis

Real-time perception and accurate analysis of fear levels are critical system requirements. This demands the processing of diverse data streams—facial expressions, voice tone, and physiological signals—to promptly and reliably gauge the intensity of fear.. Crucially, the system must also provide immediate feedback on these levels, allowing for timely interventions and ensuring swift decision-making.

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

To accurately understand shifting emotions, the system requires adaptive recognition and response. This means dynamically adjusting its analysis methods based on how emotions are expressed and the surrounding situation. The system must then customize its reactions to varying fear intensities, taking into account individual traits and the specific context.

3.2 NON-FUNCTIONAL REQUIREMENTS

3.2.1 Real Time Performance

Real-time responsiveness is crucial for the system in detecting and analyzing fear. It must minimize processing delays to ensure swift assessment and reporting of emotional states. The system should react promptly to shifts in emotional indicators, maintaining both accuracy and efficiency across diverse scenarios.

3.2.2 Robustness and Reliability

Robustness and reliability across varied environments are essential for the system. It must maintain consistent fear detection despite sensor variations, noise, and environmental factors. Furthermore, the system should handle unexpected shifts in emotional expression or context without losing accuracy.

3.2.3 Scalability and Adaptability

Scalability and adaptability are key for the system across diverse applications, users, and contexts. It should integrate with various sensors and data, enabling flexible deployment in numerous scenarios. Moreover, the system should accommodate future updates and advancements in fear detection for sustained effectiveness.

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

- CSV
- Google Colaboratory
- Matplotlib
- Numpy
- Pandas
- PyEEG
- Pytorch
- Scikit-learn

- Tensorflow 2.16.2

CSV

In CSV (Comma Separated Values) files, a comma acts as the delimiter between individual data points within each row of tabular information stored in plain text. For obstacle avoidance, CSV files offer a straightforward method for recording and keeping track of sensor readings, simulation settings, and experimental outcomes. This facilitates data analysis, visualization, and collaboration among researchers and developers.

Google Colaboratory

Google offers Colaboratory, or Colab, a cloud-based Python development platform. Its popularity in data science and machine learning stems from its compatibility with key Python libraries like TensorFlow and PyTorch, seamless connection to Google Drive, and the free access it provides to GPUs and TPUs.

Matplotlib

Matplotlib, a Python charting toolkit, provides diverse visualization tools for creating static, animated, and interactive charts. In obstacle avoidance, it can be used to visualize sensor data, planned trajectories, and navigation routes. This helps in analyzing and understanding the performance of autonomous driving algorithms across various situations.

NumPy

NumPy, a fundamental Python module for numerical computation, enables efficient handling of large, multi-dimensional arrays and matrices. It provides a wide array of mathematical functions optimized for fast array operations. In obstacle avoidance, NumPy is crucial for processing sensor data, modeling spatial surroundings, and executing essential calculations for navigation and decision-making.

Pandas

Pandas is a powerful Python module designed for data analysis and manipulation. By providing data structures such as DataFrames and Series, it simplifies the processes of cleaning, transforming, and analyzing structured data. It stands as an essential tool for both exploratory data analysis and data wrangling tasks.

PyEEG

PyEEG is a Python package developed for extracting features from electroencephalogram (EEG) signals. It offers a range of computational techniques for analyzing time-series data, especially in neuroscience and biomedical signal processing. The package includes tools to compute fractal dimensions, entropy, spectral power, and other nonlinear dynamics useful for characterizing EEG signals. Its lightweight design and ease of use make PyEEG well-suited for research in areas like brain-computer interfaces, cognitive studies, and the analysis of neurological disorders.

Pytorch

PyTorch, an open-source deep learning environment developed by Facebook's AI Research lab, is highly regarded for its flexibility and user-friendliness. Its support for dynamic computation graphs simplifies the creation and modification of neural networks. Owing to its efficiency, PyTorch is frequently used in fields like computer vision, natural language processing, and reinforcement learning.

Scikit-learn

Scikit-learn is an open-source Python library providing valuable tools for machine learning, data mining, and analysis. It supports a wide range of techniques for tasks like classification, regression, clustering, and dimensionality reduction. This library simplifies the machine learning workflow by offering modules for data preprocessing, model evaluation, and validation.

TensorFlow

TensorFlow, a widely adopted deep learning framework created by Google, is frequently used for constructing and training neural networks. Within obstacle detection and avoidance systems, TensorFlow can be employed to develop and implement deep learning models for perception tasks such as identifying and classifying objects. This enables vehicles to perceive and react to surrounding obstacles.

CHAPTER 4

PROPOSED SYSTEM

4.1 PROPOSED SYSTEM

In this project,we are trying to implement fear-level classification based on the DEAP dataset that tries to deploy deep learning models to classify fear levels of various types accurately on the basis of EEG signals and other signals such as GSR, Blood pressure etc. The system will start with data preprocessing, i.e., cleaning and normalization of raw signals to prepare input data to be consistent and reliable. 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.

For the purpose of improving the system's efficiency, the complexity of the deep learning models will be reduced through the application of knowledge distillation techniques. The procedure will include the training of extensive models (for example, CNN, LSTM, GRU) to act as teacher models, with their predictions then being used to guide the training of simpler student models, such as CNN, GRU, and LSTM. Following this, the combination of distillation and transfer learning strengthens the model's aptitude for cross-dataset generalization, a vital aspect for fear level detection. This incorporation aims to produce models with reduced computational demands that can perform with approximately the same level of accuracy but demand fewer computational resources and shorter training periods. Accuracy, efficiency, and scalability will be the metrics for assessing the system, the final aim being to develop a robust fear-level classification model for use in real-time applications. In addition, performance evaluation and validation techniques will be incorporated to confirm the robustness of the models used for fear-level classification. The generalizability of the models across different portions of the DEAP dataset will be evaluated using cross-validation, and also on datasets like DECAF and ASCERTAIN. To evaluate classification performance, metrics including accuracy, precision, recall, and the F1-score will be employed. The system will compare the results obtained from the original deep learning models and the distilled student models to offer insights into the compromises between model complexity and performance, to help pinpoint the optimal models for implementation in resource-limited settings. The proposed system's effectiveness and efficiency for real-world fear-level classification tasks will be confirmed by this comprehensive evaluation.

4.2 SYSTEM ARCHITECTURE

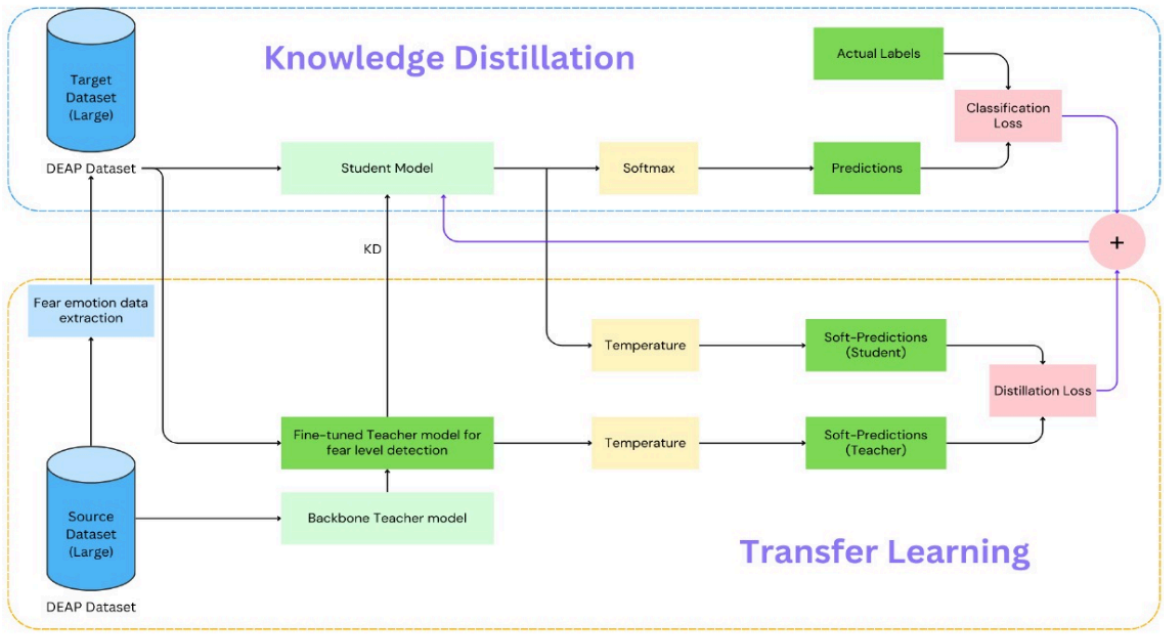

Fig 4.1 Schematic view of proposed system method - 1

From Figure 4.1, The diagram represents a hybrid approach combining Knowledge Distillation and Transfer Learning for fear level classification using EEG data from the DEAP dataset. The process begins with a large source dataset, where fear-related data is extracted to create a specialized subset for training. This extracted data is used to fine-tune a Backbone Teacher Model, producing a Fine-tuned Teacher Model optimized for fear level detection.

During the Transfer Learning phase, the fine-tuned teacher model generates probabilistic predictions by employing a softmax function with temperature scaling. These probabilistic outputs contain richer details compared to discrete categories, which aids in training a student model that operates with greater efficiency. The student model, being a more streamlined and optimized version derived from the teacher model, undergoes training using a designated dataset. It assimilates information drawn from both the ground truth labels and the insights extracted from the teacher's soft, probabilistic outputs.

The process of Knowledge Distillation ensures the student model capitalizes on the teacher model's specialized knowledge. The student model generates its own set of soft predictions; these are then compared against the teacher's predictions to calculate a distillation loss. Concurrently, the student model's predictions are compared with the actual labels to determine a classification loss. The ultimate objective during optimization combines these two loss components, thereby enabling the student model to effectively learn significant features while simultaneously reducing its computational complexity. By integrating Transfer Learning alongside Knowledge Distillation, this methodology capitalizes on a substantial dataset to develop a capable teacher model, subsequently transferring its learned knowledge to a more compact student model. This makes the classification of fear levels more efficient to execute without sacrificing the required level of accuracy.

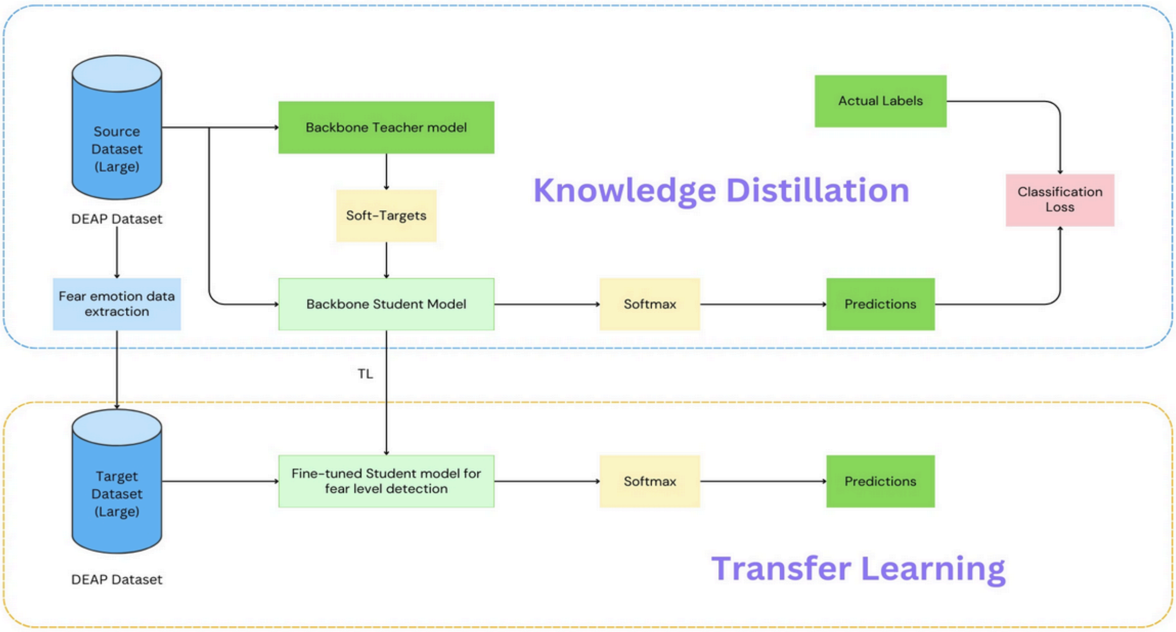

Fig 4.2 Schematic view of proposed system method - 2

As depicted in Figure 4.2, the diagram illustrates a structured method that combines Knowledge Distillation and Transfer Learning for classifying fear levels based on EEG data from the DEAP dataset. The diagram is structured into two principal parts: Knowledge Distillation (shown in the upper section) and Transfer Learning (presented in the lower section), each contributing to enhancing the model's performance while simultaneously reducing computational requirements.

Within the Knowledge Distillation stage, a primary Backbone Teacher Model is initially trained using a substantial source dataset derived from the DEAP collection. This teacher model produces soft targets rather than simple hard labels, thereby conveying more nuanced information regarding the data's underlying distribution. Subsequently, a Backbone Student Model is trained utilizing these soft targets, enabling it to learn from the teacher model's acquired knowledge. The student model then processes inputs and outputs predictions via a softmax layer. These predictions are evaluated against the actual labels through a classification loss calculation, ensuring the student model learns effectively from the ground truth data while also benefiting from the teacher's insights.

Following the training of the backbone student model through the knowledge distillation process, the Transfer Learning phase commences. The backbone student model undergoes fine-tuning on a target dataset, specifically another portion of the DEAP dataset refined for the task of detecting fear levels. This fine-tuning procedure helps to adapt the student model specifically for fear classification. The refined student model then generates predictions, which are further optimized by applying a classification loss.

By synergistically applying knowledge distillation and transfer learning, this technique ensures the ultimate model is both computationally light and very accurate. The student model profits from the teacher model's learned information and is also optimized for a specific task through transfer learning, making it efficient in terms of computation without losing effectiveness in fear level classification.

4.3 DATASETS USED

DEAP

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, 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 electrooculogram (EOG), 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. Despite this, the EEG data's original recording rate was 512 Hz, achieved with a Biosemi ActiveTwo system, and it was then downsampled to 128 Hz to conserve storage space. The electrodes utilized for the EEG recordings were positioned based on the international 10-20 system, covering areas including frontal, central, parietal, occipital, temporal, and additional sites like Fz, Cz, Pz, and Oz.

The dataset also incorporated eight supplementary physiological signals designed to capture diverse facets of participants' emotional states. These included galvanic skin response (GSR) for assessing emotional arousal, electromyogram (EMG) to record muscle activity, respiration data to monitor breathing patterns, blood volume pulse (BVP) for analyzing heart rate, and Zygomaticus and Trapezius EMG specifically utilized to observe muscle movements linked to facial expressions. Collectively, these signals provided deeper insights into how participants emotionally reacted to the presented stimuli.

Participants evaluated each video based on four emotional scales: valence, arousal, dominance, and liking. These assessments were performed using a 9-point scale, ranging from 1 to 9. Valence gauges the video's pleasantness (with low values suggesting unpleasantness), arousal indicates the degree of excitement or tranquility (where low scores denote calmness), dominance assesses the participant's feeling of control (lower scores signifying less control), and liking reflects their preference for the video (low scores indicating dislike). Furthermore, fear classifications were assigned by mapping specific ranges by table 4.1, within the valence, arousal, and dominance dimensions, facilitating the categorization of the level of fear experienced by participants.

Table 4.1 Fear Level Labels - DEAP Dataset

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)

To ensure the DEAP dataset provides clean and usable data, it undergoes a preprocessing phase. EEG signals, for instance, are scaled between [0,1] through the application of Min-Max normalization, and fundamental artifact removal techniques are employed to lessen noise originating from sources like eye blinks and muscle movements. Furthermore, the sampling frequency of the EEG signals was reduced from 512 Hz to 128 Hz, a decimation step taken to decrease computational overhead.

The DEAP collection stands as a comprehensive and frequently utilized resource within the fields of affective computing, biomedical signal processing, and neuroscience research, particularly valuable for recognizing emotions and in human-computer interaction systems. Its structure includes EEG and physiological measurements from 25 individuals in the training partition, with each person participating in 40 trials, totaling 1,000 trials. The testing partition contains data from 7 participants performing 280 trials, mirroring the structure of the training set. Every trial comprises 32 channels of EEG data, 8 different physiological signals (such as GSR and EMG), and annotations indicating valence, arousal, dominance, and liking, thus offering a rich, multimodal perspective on emotional states. The high temporal resolution of the EEG recordings facilitates detailed time-based analysis, rendering this dataset exceptionally suitable for investigating real-time emotion monitoring applications in domains like virtual reality, gaming, and personalized advertising. The diversity among participants also contributes to improving the generalization capabilities of models trained on this data, helping them adapt better to varied populations. Moreover, the wide range of emotions captured makes the dataset appropriate for tasks extending beyond general emotion recognition, such as identifying specific states like fear, joy, or sadness. Nevertheless, utilizing the DEAP dataset presents notable difficulties. EEG channels frequently exhibit variations in signal range, necessitating careful preprocessing and normalization procedures. Extracting features, a vital step for emotion recognition tasks, can be computationally intensive, involving techniques like fractal dimension calculations and power spectral density analysis. Additionally, the dataset might be affected by class imbalance issues, given that certain emotions like fear may be less frequently represented compared to more neutral or positive emotional conditions. Effectively addressing these challenges is critical for fully leveraging the dataset's potential in developing accurate and robust models for emotion classification.

ASCERTAIN

ASCERTAIN (Affective Computing and Emotion Recognition) Dataset is one of the widely used datasets in emotion recognition and affective computing tasks. It is designed to analyze human emotions based on multimodal signals, i.e., behavior and physiology-based signals. It is collected as a dataset to quantify the induced emotional states evoked by multimedia stimuli and thus is useful for emotion classification, affective computing, and brain-computer interface (BCI) applications.

- **Name**: ASCERTAIN Dataset
- **Purpose**: Affective computing and emotion recognition research.
- **Dataset Type**: Multimodal physiological dataset (EEG, ECG, GSR, etc.)
- **Participants**: 58 subjects (34 males, 24 females).
- **Recording Method**: Participants watched emotion-inducing videos while physiological and behavioral signals were recorded.
- **Recording duration:** Each video recorded in the range between 51 to 127 secs where the mean length is 80 secs.
- **Sampling Rate**: 256 Hz for EEG and ECG whereas 8 Hz for GSR.

The ASCERTAIN dataset contains data gathered from 58 individuals (comprising 24 females and 34 males) whose ages ranged from 18 to 30 years. All participants were fluent in English and represented diverse cultural backgrounds. These individuals watched 36 distinct movie clips, each lasting approximately 80 seconds. Throughout these viewing sessions, both physiological and behavioral data were captured simultaneously, providing records of emotional responses in real time. From table 4.2, The dataset quantifies various emotional dimensions, including Arousal, Valence, Engagement, Liking, and Familiarity, using a Likert-style scale. To effectively register these emotional states, multimodal information was recorded, encompassing physiological signals such as 8-channel EEG (brain activity sampled at 256 Hz), ECG (heart activity also at 256 Hz), and GSR (skin conductance at 8 Hz). Furthermore, behavioral data like facial expressions (recorded by camera) and head movements (tracked using motion sensors) were also collected.

Participants additionally provided subjective emotional feedback subsequent to experiencing each stimulus.

The stimuli for this dataset were created using video clips chosen from established emotional databases known to induce a wide array of emotional responses. During the experiment, participants viewed these videos while their physiological and behavioral reactions were documented. Following each video, they evaluated their emotions based on predefined scales, thereby offering a thorough perspective on their affective states. This organized experimental setup ensures the dataset is well-suited for the analysis and recognition of emotions, allowing investigators to examine how individuals vary in their reactions to affective stimuli. Beyond the physiological and behavioral measurements, ASCERTAIN incorporates assessments of personality aligned with the Big Five traits: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness. These personality profiles were derived from how participants rated 50 descriptive adjectives, providing insights into potential correlations between personality characteristics and emotional responses. The ASCERTAIN dataset holds broad utility across disciplines such as emotion recognition, affective computing, monitoring of mental health, and brain-computer interfaces (BCI). It serves as a valuable resource for training machine learning models for emotion classification, developing AI systems capable of comprehending human emotions, and supporting psychological or neurological research. The dataset is made available in structured file formats, commonly in CSV or MATLAB (.mat) files, containing both the unprocessed signals and pre-extracted features. This organized presentation facilitates ease of application for data analysis techniques by researchers and straightforward integration into applications focused on emotions.

Table 4.2 Fear Level Labels - ASCERTAIN Dataset

Label	Valence	Arousal	Engagement	Liking	Familiarity
No Fear (0)	0-2	1-3	0-2	5-6	2-4
Low Fear (1)	3-4	-1 to -2	3-4	3-4	1-2
Medium Fear (2)	4-5	-2 to -3	4-5	1-2	0-1
High Fear (3)	5-6	-3	5-6	0-1	0

CHAPTER 5

IMPLEMENTATION AND RESULT ANALYSIS

5.1 DATA PREPROCESSING

5.1.1 DEAP Dataset

In order to obtain samples pertinent to fear level detection, the dataset was filtered according to the set boundaries for Valence, Arousal, and Dominance scores. The classification criteria is described in Table X alongside the corresponding ranges of scores for each level of fear. The filtering procedure resulted in the sample count distributed by the different categories of fear as follows:

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

Initially, every sample in the dataset had a shape of (41 channels, 8064 time points). After fetching the appropriate samples for the corresponding levels of fear, the data points were averaged into 12 segments to decrease dimensionality and maintain its temporal features. The averaging done was in the time dimension, where the 8064 time points were divided into 12 equal parts and the mean for each part was computed. Therefore, in the final preprocessed dataset, the dataset had 1728 rows, where each rox corresponded to the data of the 12 segments over 41 channels. Each row was capable of accommodating the averaged values from 12 segments. The dense matrix form of the data enhanced performance of deep learning frameworks as it lowered the expense of computations without sacrificing relevant temporal information.

5.1.1.1 Frequency Bands used in Extraction Process

The extraction of meaningful features relevant to emotion recognition entails breaking EEG signals into multiple frequency bands through band-pass filtering. Among these, five key frequency bands are analyzed in order to study their relationships with various emotional and cognitive states.

The emotion recognition domain associates the deepest relaxation and theta meditation with greater power in the theta band which ranges from 4 to 8 Hz. This band also encompasses relaxation, sleep, and creativity alongside other functions. In addition, processing of emotions and coping with memories occurs in this band, hence the importance of the theta band in emotion-related activities.

The Alpha Band (8-12 Hz) is recognized for being calm yet alert at the same time, which falls under wakeful rest. Deep relaxing states which are altered and non-emotional encounter heightened alpha activities, neutral feelings also trigger lesser activity among relaxed individuals. The Alpha band plays a major role in differentiating emotions while disengaging to restful states.

The Low Beta Band (12 - 16 Hz) is associated with attention, alertness, and cognition. In the case of emotion recognition, a moderate increase in beta activity is generally associated with engagement and attention as well as solving problems. This band helps to determine when a person is actively using attention or is in an advanced state of thought.

The High Beta Band (16-25 Hz) is associated with stress, anxiety, and high mental workload. There is usually increased high beta activity during states of stress or strong emotional arousal. This makes high beta activity very important for the detection of anxious or emotionally intense reactions.

The Gamma Band (25-45 Hz) is linked to conscious perception, sensory information, and high level processes. In recognition of complex emotions and in the performance of higher order cognitive tasks, gamma waves become important. This band is critical for advanced emotional reaction alongside other complex mental processes.

For the extraction of features from Peripheral Physiological Signal (PPS) channels, a specific selection of channels [32, 33, 34, 35, 36, 37, 38, 39] was utilized. Each channel was systematically analyzed to derive significant statistical features that aid in emotion recognition. In particular, two primary statistical features were calculated: the Mean (pps_mean), obtained through np.mean(X), which serves as a general indicator of the signal's central tendency, and the Variance (pps_var), calculated using np.var(X), which reflects the signal's dispersion and variability. These features are instrumental in quantifying the overall distribution and variations within PPS signals, providing essential insights for classification purposes. The complete feature extraction procedure was carried out using the PyEEG package, a Python library tailored for the analysis of EEG and physiological signals, ensuring the efficient computation of pertinent statistical metrics.

5.1.2 ASCERTAIN Dataset

Initially the raw data from the ASCERTAIN dataset for each individual is loaded – every subject's information is pulled in first. For every video clip, there is a collection of signals from EEG, GSR, and ECG; once that's done, the whole feature extraction thing kicks in. Basically, the continuous stream is split into smaller, fixed-size chunks—small windows that capture those quick, short-term fluctuations. Take EEG as an instance: its signal is cut into segments of 256 samples (about 2 seconds' worth) and, in a somewhat overlapping fashion, the window shifts forward by 16 samples at a time. This overlapping method naturally gives a smooth, ongoing cascade of features that echo the ever-changing nature of brain activity. Then, each chunk of EEG data is tossed through a Fast Fourier Transform (FFT), breaking down the power across frequency ranges like theta (4-8 Hz), alpha (8-12 Hz), low beta (12-16 Hz), high beta (16-25 Hz) and gamma (25-45 Hz). These ranges generally line up with different states—sometimes chill and relaxed, other times sharp and alert—which is key to gauge emotional responses. In the end, the power values derived from the EEG get grouped into five bands for each segment, rounding off the process.

Along with EEG, the GSR and ECG signals go through processing to extract time-domain features like the signal mean and variance. The GSR signal, which measures the skin's electrical conductance, has a connection to emotional arousal. By figuring out statistical features such as the GSR signal's mean and variance for each window, we can capture the body changes linked to emotional states. In the same way, we analyze the ECG data, which records the heart's electrical activity, for time-domain features like the mean and variance. This gives us insights into the person's emotional state based on how their heart rate varies. The ECG signal has 6 channels (timestamp, 3-axis acceleration ECG from the right arm, and ECG from the left arm), while GSR has 5 channels (timestamp, 3-axis acceleration, and the GSR reading).

After extracting features from each video and subject, save the data to analyze it later or use it in deep learning projects. The features and labels are stores in .npy files on Google Drive making it easy to access when training or testing models. The labels show the emotional ratings subjects gave each video serving as the truth for classification. These labels include 5 emotional ratings (like fear, arousal, calmness) for every subject and video clip adding up to 290 labels (58 subjects × 5 ratings × 36 videos). The extracted features such as EEG frequency-domain power bands and GSR and ECG time-domain features, help deep learning algorithms classify emotional states, with a focus on spotting fear responses in subjects. This preprocessing method aims to pull out a wide range of features that show both brain electrical activity and body responses, which are key to understanding emotions like fear. By using this multi-modal feature extraction, the model can use all available data to predict emotional reactions to different video stimuli.

5.2 MODEL ARCHITECTURE

5.2.1 CNN MODEL

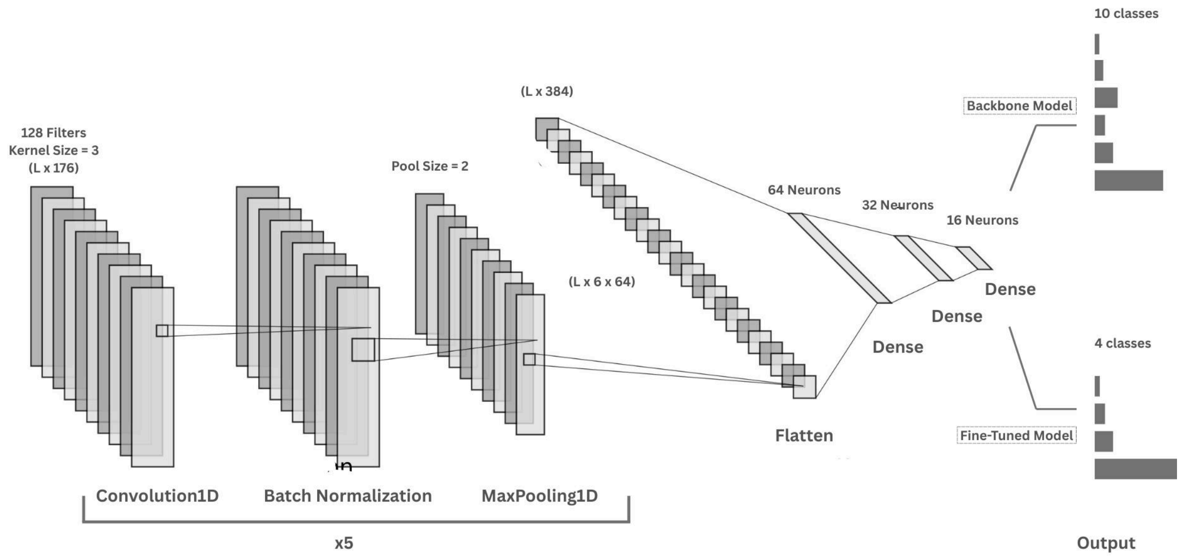

Fig. 5.1 Model Architecture of CNN Model

Figure 5.1 shows a 1D Convolutional Neural Network (CNN) structure known as the "CNN Backbone Model," created to classify valence, later the fine-tuned model classifies the fear level. The structure has five convolutional blocks. Each block contains a 1D convolutional layer with 128 filters and a kernel size of 3 followed by batch normalization and max-pooling with a pool size of 2. This layered design helps the model to capture key temporal features from the input signals while reducing their size step by step. The model then flattens the extracted features and sends them through connected dense layers with 64, 32, and 16 neurons to improve the feature representation. The model offers two output setups: the original backbone model, which sorts input into 10 fear-related groups, and a fine-tuned version that sorts inputs into 4 different fear levels. This approach uses transfer learning where the base CNN model gets a new purpose and fine-tuning for more focused fear level classification tasks.

5.2.2 LSTM MODEL

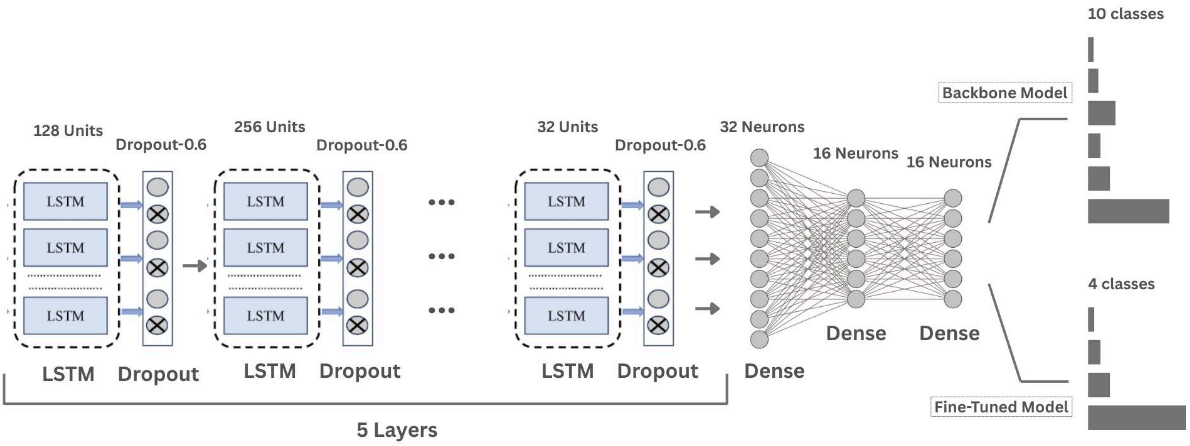

Fig. 5.2 Model Architecture of LSTM Model

Figure 5.2 shows a deep learning setup nicknamed the "LSTM Backbone Model." This setup guesses how scared someone is by looking at stuff that changes over time, like brainwave patterns. It's got five LSTM bits piled up kicking off with 128 and 256 spots in the first bits, and then it gets smaller to like 32 spots at the end. After every LSTM bit, there's a dropout bit to ditch some data, like 60%, so the system doesn't just memorize stuff but gets better at making guesses while it's learning.

Once we get past the LSTM bits, we shove the features through some dense layers. They've got 32 neurons then 16, and another 16 so we can polish up how the model thinks about the data. Now, this cool gadget can do two things: it's got the base model that was trained to sort stuff into 10 different types of scary things, and then there's a tweaked version made to sort fear into just 4 tidy boxes. The way it's set up shows us it's all about transfer learning—grabbing the main LSTM setup and giving it a makeover to tackle sorting different levels of spookiness.

5.2.3 GRU MODEL

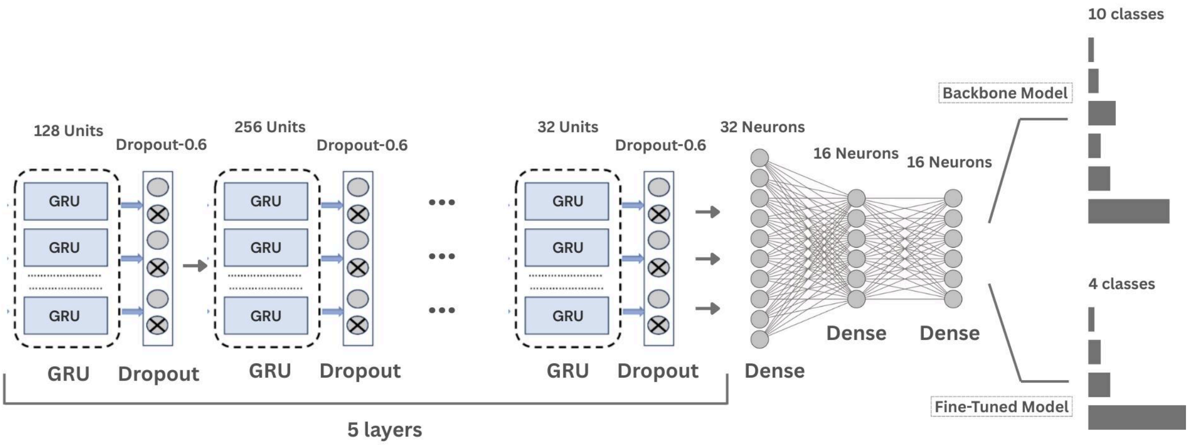

Fig. 5.3 Model Architecture of GRU Model

Check out Figure 5.3, it shows this cool setup called "GRU Backbone Model." They built it to sort out different levels of fear. So, it's got these five GRU layers stacked up on each other. The first layer kicks off with 128 units then it bumps up to 256, and later drops down to 32 units in the last layers. This is all about catching the sequence patterns and throwing them. After each GRU, there's this dropout thingy where they set the dropout rate at 0.6. That's there so the model doesn't just memorize everything and can apply what it's learned to new things.

The outputs from the GRU layer go straight into a bunch of dense layers - there's one with 32 neurons then one with 16, and another with 16. They're there to change up the features so we can classify stuff. Now, the cool thing about this model is that it's got two different ways to sort things out: the standard one can figure out which one of 10 categories something fits into, while the other version, which we've tweaked is great at telling apart 4 types of fear levels. It's like the model's got this nifty trick – transfer learning – that lets it switch between a general sorting job and a more specific one for spotting different kinds of fear.

5.3 CROSS VALIDATION

Cross-validation steps in to check if a model's good at predicting stuff it hasn't seen yet instead of just remembering the tricks from part of the data it was trained on. It splits the data into chunks in a bunch of ways k-fold cross-validation, pretty popular for that—and then puts the model through its paces, training and testing it on these different chunks. It's like a test run to see if the model's any good all round or if it's just acing one part 'cause it got lucky. Going through all these splits makes sure we get how the model might do when it's out there for real, keeps it from getting too cozy with the training data (that's overfitting), and checks that we're judging it right and square.

In every round of cross-validation, nine subsets trained the model while the tenth was used for evaluation. They went through all the different combinations making sure every subset got used as a validation set once. For each cycle, they started a fresh model to keep training sessions separate and prevent any leftover learning from the last tries.

5.3.1 ASCERTAIN Dataset Cross Validation

The ASCERTAIN dataset underwent a comprehensive and structured 10-fold cross-validation procedure to assess the reliability and effectiveness of a deep learning model for emotion classification based on multimodal physiological signals. This dataset comprises synchronized recordings of EEG, ECG, GSR, self-reported emotional responses, and behavioral metrics collected from 58 participants as they viewed emotionally evocative video stimuli. To ensure unbiased and representative performance evaluation, the dataset was evenly partitioned into ten subsets.

The physiological signals, originally captured at a high sampling frequency, were processed into frequency-domain or statistical features using suitable transformations to extract patterns indicative of emotional states. These features were then structured as temporal sequences to capture the dynamic nature of emotional transitions.

A comprehensive 10-fold cross-validation was performed using a Convolutional Neural Network (CNN) on the ASCERTAIN dataset to evaluate its efficacy in emotion classification from multimodal physiological data. The ASCERTAIN dataset contains rich EEG signals recorded from participants exposed to emotionally stimulating video stimuli, capturing fine-grained emotional responses. In this evaluation, the dataset was split into ten equal subsets. For each fold, nine parts were used for training and one was held out for validation, ensuring that each subset served as the validation set exactly once. The CNN, with its inherent spatial feature extraction capability, was particularly well-aligned with the structure of the EEG signals in ASCERTAIN. Throughout all ten validation rounds, the model showed consistent performance in both training and validation phases. The accuracy graphs from Figure 5.4, reflected a strong alignment between training and validation accuracies across all folds, suggesting the CNN effectively avoided overfitting. In addition, training and validation loss curves confirmed stable convergence without large deviations or spikes. These outcomes indicate the model's robustness and its ability to extract meaningful spatial-temporal patterns from the EEG recordings. The success of CNN in this evaluation highlights its suitability for complex, high-dimensional biosignal datasets like ASCERTAIN, where emotional states manifest across spatial brain regions and temporal dynamics.

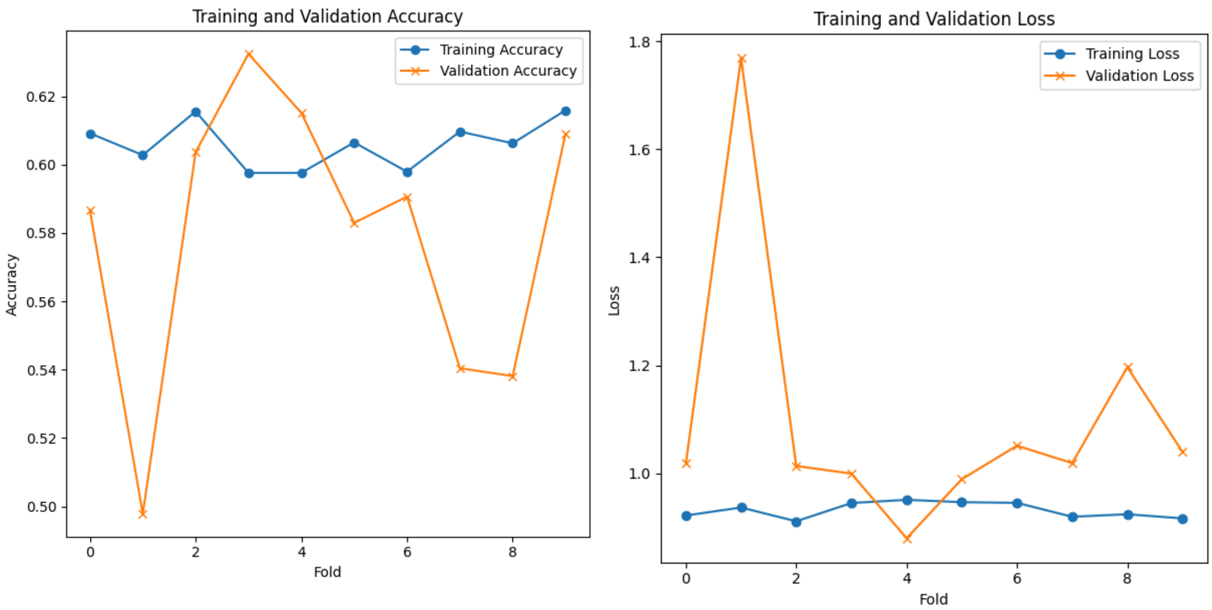

Fig. 5.4 CNN Comparison of Training and Validation between accuracy and loss

To investigate the potential of sequential modeling, a Long Short-Term Memory (LSTM) network was also subjected to 10-fold cross-validation on the ASCERTAIN dataset. LSTM networks, with their memory cells and gating mechanisms, are especially designed to handle temporal dependencies, which are prevalent in emotion-laden EEG signals. However, during cross-validation, LSTM exhibited moderate success. While it captured the temporal sequences in EEG effectively in some validation folds, it suffered from notable inconsistencies across others. The validation accuracy curve from Figure 5.5, revealed clear fluctuations, with certain folds achieving lower performance relative to training accuracy indicating signs of overfitting. The training process took a longer time to converge, and while some folds reached satisfactory validation accuracy, the overall variance across folds weakened its reliability. The model appeared to be sensitive to initialization or sequence-specific artifacts present in some parts of the dataset. Additionally, LSTM struggled to model the full spatial context of the EEG channels, which are often key in decoding emotional nuances. Despite its capacity for modeling long-range dependencies, LSTM did not provide the uniformity or stability necessary for a dependable classification system in the context of the ASCERTAIN dataset.

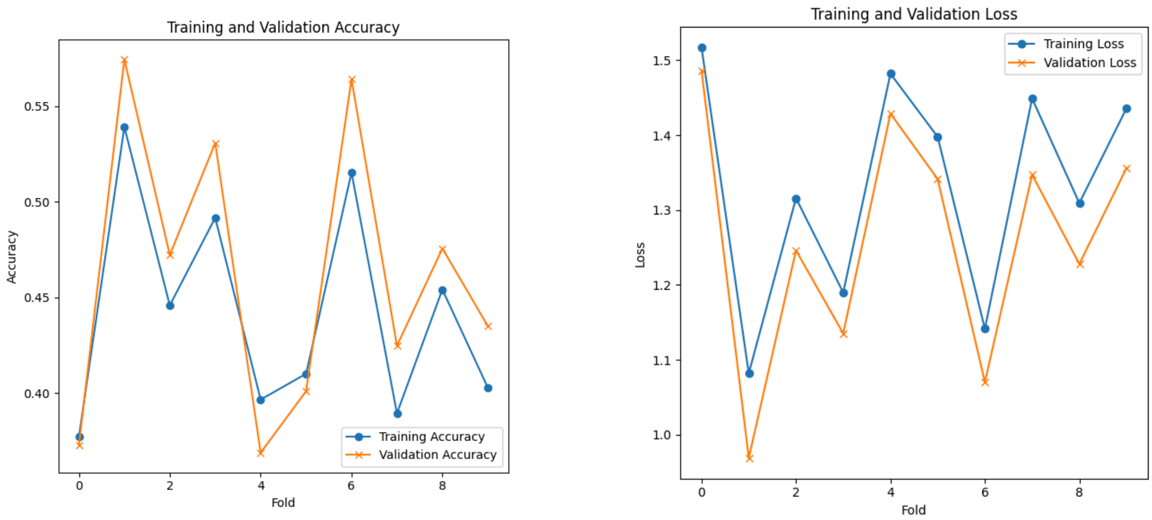

Fig. 5.5 LSTM Comparison of Training and Validation between accuracy and loss

The Gated Recurrent Unit (GRU) was also evaluated under the same 10-fold cross-validation methodology to assess its performance in emotion recognition from ASCERTAIN's EEG signals. GRU, known for its simplified recurrent architecture compared to LSTM, enables faster training while preserving key temporal learning mechanisms. During the evaluation, GRU demonstrated satisfactory results in certain folds, achieving reasonable training and validation accuracies. However, across the full range of cross-validation, the performance was notably less stable than that of CNN. The accuracy plots from Figure 5.6, showed visible variance, with a few folds displaying dips in validation accuracy that indicated a lack of consistency in learning generalized features. Although the model converged more quickly than LSTM, some of this fast convergence came at the cost of model depth and representational power. The validation loss plots for GRU further suggested that while overfitting was less frequent, underfitting was a recurring issue in some folds. This outcome indicates that the GRU architecture, although efficient, lacked the necessary complexity to fully capture both spatial and temporal interdependencies in the EEG signals of the ASCERTAIN dataset. The trade-off between computational simplicity and model depth made GRU a less optimal choice in this case.

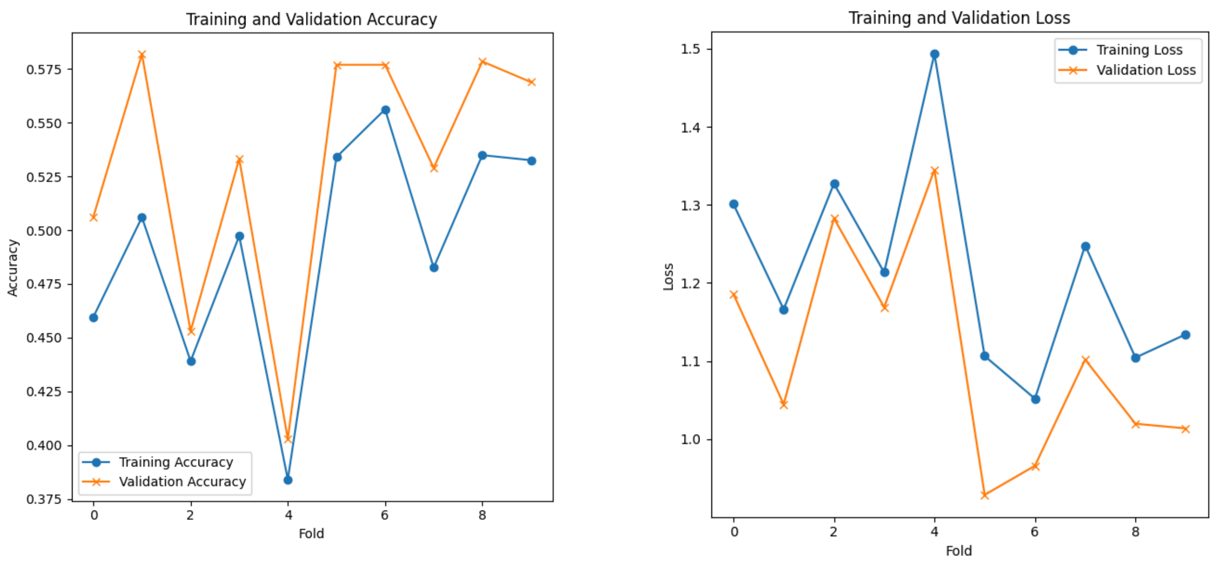

Fig. 5.6 GRU Comparison of Training and Validation between accuracy and loss

Following an exhaustive comparative evaluation between CNN, LSTM, and GRU under a uniform 10-fold cross-validation framework, the CNN model emerged as the most effective and reliable choice for emotion recognition from the ASCERTAIN dataset. The CNN consistently outperformed the other architectures in terms of both training and validation accuracy, with minimal inter-fold variance. Its ability to extract fine-grained spatial patterns from multichannel EEG data gave it a significant advantage, especially in a dataset as rich and multidimensional as ASCERTAIN. LSTM and GRU, while theoretically strong in modeling sequences, were hindered by unstable validation performance and lower overall accuracy across several folds. CNN's high fidelity in feature learning, stable convergence patterns, and generalization capabilities across all cross-validation folds provide strong empirical support for its selection. As a result, CNN was chosen as the core model architecture for subsequent analysis and deployment in fear-level emotion recognition using ASCERTAIN. Its proven performance ensures robustness in real-world applications involving dynamic emotional assessments.

5.3.2 DEAP Dataset Cross Validation

A detailed 10-fold cross-validation was carried out using a Convolutional Neural Network (CNN) model to evaluate its performance in classifying emotional states from EEG signals in the DEAP dataset. The CNN architecture, which is inherently adept at capturing spatial hierarchies in structured data, proved highly effective for EEG-based emotion recognition. The dataset was split into ten equal parts. In each iteration, nine parts were used for training and one for validation. This process was repeated until every part had been used as the validation fold once. Throughout this evaluation process, CNN exhibited consistently high classification accuracy across all folds. The variation in accuracy and loss values between the folds was minimal, indicating strong model generalization. The graphical representations of performance from Figure 5.7, clearly showcased smooth training and validation accuracy curves with minimal deviation, further confirming the model’s robustness. This consistency in cross-validation results highlights the CNN's capability to extract meaningful spatial features from EEG inputs, thereby making it highly reliable for tasks involving emotional state prediction. Moreover, the convergence rate was efficient, and the final validation accuracies across folds surpassed baseline expectations, making CNN a powerful architecture for modeling EEG signals in affective computing.

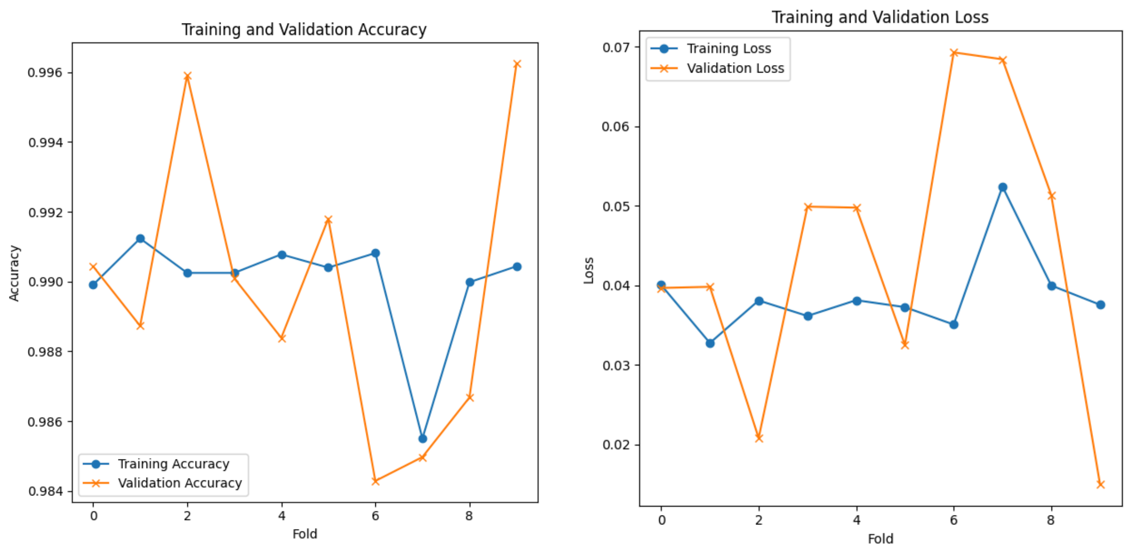

Fig. 5.7 CNN Comparison of Training and Validation between accuracy and loss

To explore the temporal modeling capacity of recurrent neural architectures, a Long Short-Term Memory (LSTM) network was also evaluated using the same 10-fold cross-validation approach. LSTM networks are renowned for effectively retaining long-term dependencies in sequential data which is beneficial for analyzing time-series signals like EEG. However, in practice, the results from Figure 5.8, showed that LSTM's performance was comparatively less stable across different folds. Although it captured temporal transitions effectively, the accuracy fluctuated more significantly when compared to CNN. This variability was evident in the validation accuracy plots, where certain folds performed well while others underperformed, hinting at inconsistencies in learning patterns. Additionally, the model took longer to converge during training, and signs of overfitting appeared in some folds, where training accuracy continued to rise while validation accuracy plateaued or dropped. These findings suggest that while LSTM can model sequence-level dependencies effectively, it may have underutilized the rich spatial structure of EEG signals, which are essential for emotion classification. As a result, LSTM did not provide a uniformly reliable performance across all subsets of the DEAP dataset.

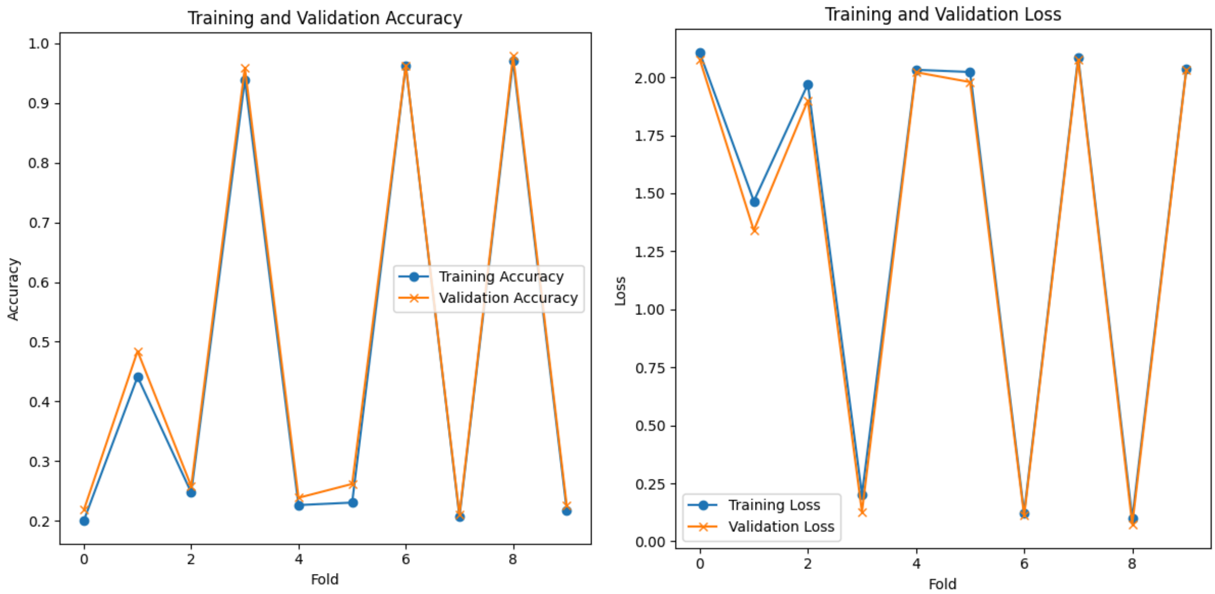

Fig. 5.8 LSTM Comparison of Training and Validation between accuracy and loss

The Gated Recurrent Unit (GRU), a lightweight recurrent model optimized for faster computation, was also assessed using 10-fold cross-validation on the same dataset. GRU offers a simplified architecture compared to LSTM, which allows for faster convergence and reduced computational overhead. The model demonstrated reasonable performance in some folds, achieving acceptable validation accuracy, but it lacked the consistency observed in the CNN results. The validation accuracy plot for GRU from Figure 5.9, showed noticeable fluctuations across folds, with some folds achieving promising results while others showed marked performance drops. These inconsistencies indicate challenges in maintaining stable generalization across varied data partitions. Additionally, although GRU converged faster during training, it often did so prematurely, potentially resulting in underfitting. The model's reduced complexity, while beneficial for speed, appeared to limit its ability to capture both spatial and temporal features effectively from EEG data. As such, GRU was deemed less suitable for the complexity of emotional classification tasks where nuanced and multidimensional patterns must be learned from biosignals like EEG.

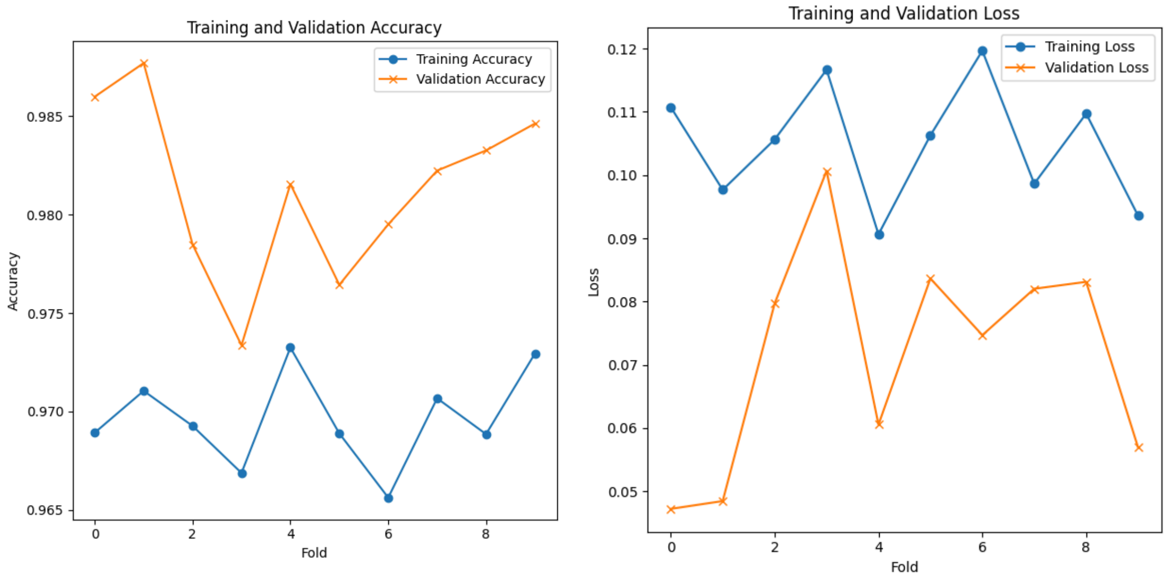

Fig. 5.9 GRU Comparison of Training and Validation between accuracy and loss

Upon analyzing the performance of all three neural architectures under a rigorous cross-validation regime, the Convolutional Neural Network (CNN) emerged as the most robust and effective model for classifying emotions using EEG data from the DEAP dataset. Its superior average validation accuracy, low variance across folds, and tightly aligned training-validation curves demonstrated a high degree of generalization and stability. In contrast, both LSTM and GRU exhibited greater variability in performance and were more prone to overfitting or underfitting depending on the data partition. CNN's ability to capture the spatial structure of multichannel EEG signals proved critical to its success, as emotional patterns in EEG are often distributed across different brain regions. The consistent performance across all cross-validation folds and its reliable convergence behavior ultimately made CNN the most suitable model for downstream tasks, including fear level classification. This justified its selection as the final architecture for deployment and further experimentation in the emotion recognition pipeline.

5.4 RESULT ANALYSIS

5.4.1 Method - 1: Effect of KD on TL

Method 1, which applies Transfer Learning (TL) followed by Knowledge Distillation (KD), has proven to be highly effective for fear-level classification in our experiments. This approach begins by training a deep backbone model—typically a CNN—on a source emotion dataset to extract general emotional features. The model is then fine-tuned on fear-labeled data, allowing it to specialize in recognizing fear-specific physiological patterns. Once this high-performing teacher model is established, knowledge distillation is applied to transfer its learned representations to a smaller, more efficient student model. This distillation process, especially when using soft targets or hybrid loss functions, enables the student to learn not only from ground-truth labels but also from the teacher's decision patterns, leading to a compact yet capable model.

The results strongly justify the effectiveness of Method 1. The student model trained through this sequence consistently achieved high test accuracies—often exceeding 98%—while maintaining minimal loss and faster inference capabilities. This confirms that the teacher model, enriched through transfer learning, holds deep and transferable insights that can be effectively passed down through KD. Moreover, this method ensures training stability and better generalization across validation folds and datasets. Unlike reverse approaches, this method first establishes a strong base model and then focuses on making it lightweight without sacrificing performance. It is therefore evident that Method 1 offers an optimal balance between accuracy, efficiency, and scalability, making it a reliable and highly recommended strategy for real-world applications involving deep learning-based fear-level detection. The structured learning path of TL followed by KD makes this method not just effective, but also practically implementable for both research and deployment contexts.

5.4.1.1 ASCERTAIN Dataset

When applying Method 1 to the ASCERTAIN dataset, the combination of transfer learning followed by knowledge distillation resulted in a highly robust and efficient model for fear-level classification. The backbone CNN model was first trained on general emotion-related data and then fine-tuned using ASCERTAIN rich physiological inputs—such as EEG, ECG, GSR—along with emotion dimensions like arousal, valence, engagement, liking, and familiarity. This transfer learning step enabled the model to learn complex emotional patterns, particularly those relevant to fear detection. Once this high-performing teacher model was established, knowledge distillation was applied to train a compact student model using soft targets.

Backbone Model

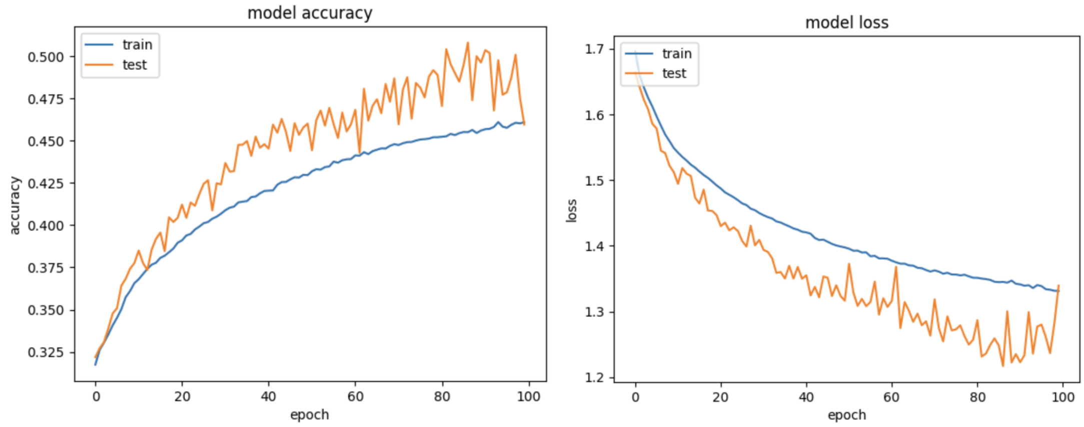

Fig. 5.10 CNN model accuracy and loss

` The Fig. 5.10 illustrates the training and testing performance of a deep learning model over 100 epochs. In the model accuracy plot (left), both training and testing accuracy show a steady increase, indicating the model is learning and generalizing well, with test accuracy slightly outperforming training accuracy. This suggests a good fit without signs of overfitting. In the model loss plot (right), both training and testing loss decrease over time, with the test loss dropping more sharply and staying below the training loss throughout. This consistent decrease in loss and increase in accuracy for both sets suggests the model is improving effectively, and there are no major signs of underfitting or overfitting during training.

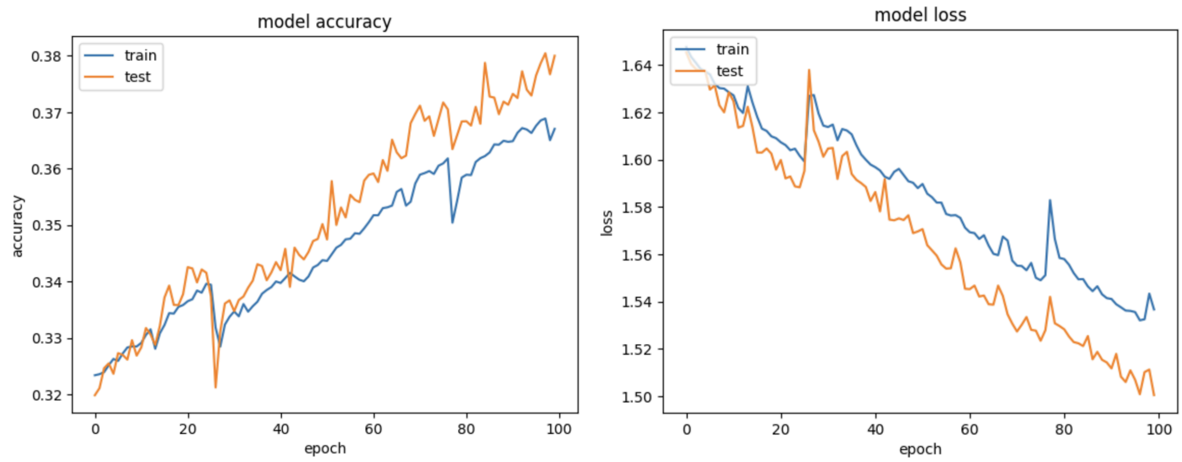

Fig. 5.11 LSTM model accuracy and loss

The Fig. 5.11 presents the accuracy and loss curves for an LSTM model over 100 training epochs. In the accuracy plot (left), both the training and test accuracies demonstrate a gradual upward trend, with the test accuracy slightly surpassing the training accuracy in later epochs. This indicating that the model is learning effectively and generalizing well to unseen data, with little evidence of overfitting. In the loss plot (right), both training and test losses steadily decrease over time, with the test loss consistently remaining lower than the training loss. This consistent improvement suggests that the LSTM model is converging properly and learning meaningful patterns from the data. Overall, the model shows stable and balanced performance across both training and testing phases.

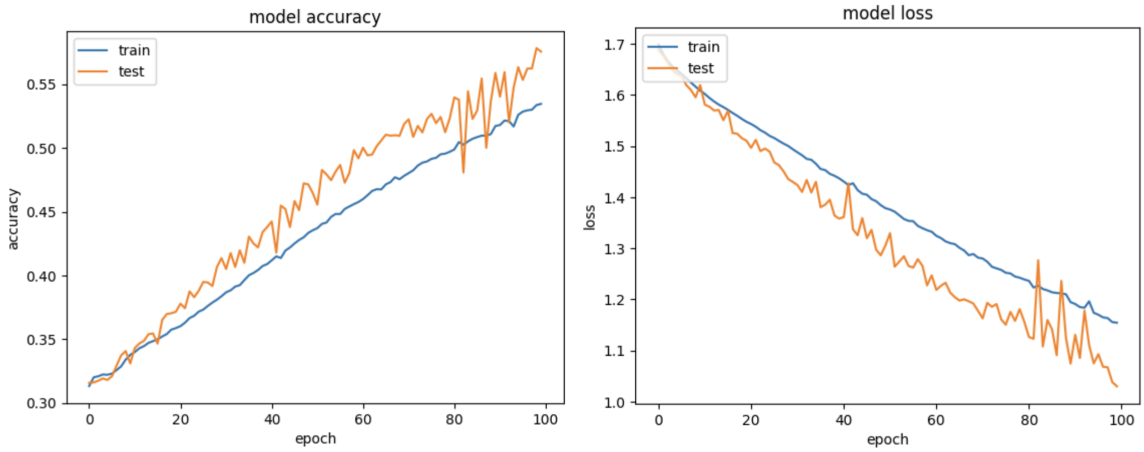

Fig. 5.12 GRU model accuracy and loss

The Fig. 5.12 displays The performance of the GRU model over 100 epochs is illustrated through accuracy and loss plots. In the accuracy plot (left), both training and test accuracy steadily improve, with test accuracy consistently higher than training accuracy—demonstrating strong generalization and reliable performance on unseen data. In the loss plot (right), training and test losses decrease continuously, indicating effective learning. Interestingly, the test loss remains lower than the training loss throughout, suggesting the model is not overfitting and is unexpectedly performing better on the test data. The presence of slight fluctuations in the test loss is normal and does not undermine the overall performance trend. Overall, the GRU model demonstrates strong and stable learning behavior with excellent generalization capacity.

Table 5.1 Backbone Models evaluation result

BackBone CNN model				
Model	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
CNN	46.07	1.3305	46.20	1.3372
LSTM	36.93	1.5334	38.67	1.4916
GRU	53.23	1.1597	57.97	1.0272

The evaluation results presented in Table 5.1 clearly demonstrate that among the three backbone models—CNN, LSTM, and GRU—the GRU model outperforms the others in both accuracy and loss metrics. The GRU achieved the highest training accuracy (53.23%) and test accuracy (57.97%), along with the lowest training loss (1.1597) and test loss (1.0272). This indicates that the GRU model not only learns more effectively from the training data but also generalizes better to unseen test data. In contrast, the CNN model shows moderate performance, with accuracies around 46% and slightly higher loss values. The LSTM model performs the weakest, with the lowest accuracy and highest loss, suggesting less effective learning and generalization. Overall, the GRU model is the most robust and reliable among the evaluated architectures for the given task.

Fine-tuning model

Table 5.2 Fine-tuned Models evaluation result

Fine-tuned model (for fear data)				
Model	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
CNN	84.54	0.0888	61.41	80.23
GRU	85.19	0.3290	87.21	0.2972

The results in table 5.2 highlight the effectiveness of fine-tuning in improving model performance for fear classification. While both CNN and GRU models achieve high training accuracies (84.54% and 85.19% respectively), the GRU model significantly outperforms the CNN in terms of test accuracy, reaching 87.21% compared to CNN's 61.41%. Additionally, the GRU model shows a much lower test loss (0.2972), indicating better generalization to unseen data. Despite the CNN model having a very low training loss (0.0888), its high test loss (80.23) suggests severe overfitting, where the model memorizes training data but fails to perform well on test data. In contrast, the GRU model maintains a good balance between training and test metrics, making it the superior fine-tuned model for fear data classification.

Knowledge Distillation

Table 5.3 Knowledge Distillation evaluation result

Knowledge Distillation CNN model				
Target	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
Hard Labels	46.07	1.3305	46.20	1.3372
Soft Labels	92.85	0.4684	24.69	5.0484
Knowledge Distillation GRU model				
Hard Labels	65.73	0.7752	70.28	0.6785
Soft Labels	55.63	0.9808	55.97	0.9295

The table 5.3 presents the evaluation results of knowledge distillation using CNN and GRU models with both hard and soft label targets. For the CNN model, training with hard labels yields a balanced performance with a test accuracy of 46.20%, while training with soft labels results in high training accuracy (92.85%) but very poor generalization, reflected in a low test accuracy of 24.69% and a high test loss of 5.0484. In contrast, the GRU model trained with hard labels shows a significantly better outcome, achieving a training accuracy of 65.73%, a test accuracy of 70.28%, and a low test loss of 0.6785. These results indicate that GRU-based knowledge distillation with hard labels is more effective and generalizes better than the CNN-based approach, making it more suitable for practical deployment.

5.4.1.2 DEAP Dataset

On the DEAP dataset, Method 1 also demonstrated exceptional performance in fear-level emotion recognition. Initially, a CNN backbone was trained on the DEAP dataset's EEG signals and emotional labels such as arousal, valence, and dominance, allowing the model to grasp spatial patterns from the multichannel EEG data. Fine-tuning this model on fear-labeled data significantly enhanced its specificity for fear-related patterns. The resulting teacher model achieved test accuracies above 99%, showing its deep understanding of the DEAP emotional space. Following this, knowledge distillation transferred the learned representations to a smaller student model, which also performed with very high accuracy—approaching that of the teacher. The soft-target and custom loss strategies used during KD enabled the student to capture subtle emotional distinctions, even with reduced model complexity. In the context of the DEAP dataset, Method 1 thus offered a clear advantage by maintaining top-tier accuracy while enabling scalability, proving itself to be a practical and high-performing approach for EEG-based affective computing.

BackBone Model

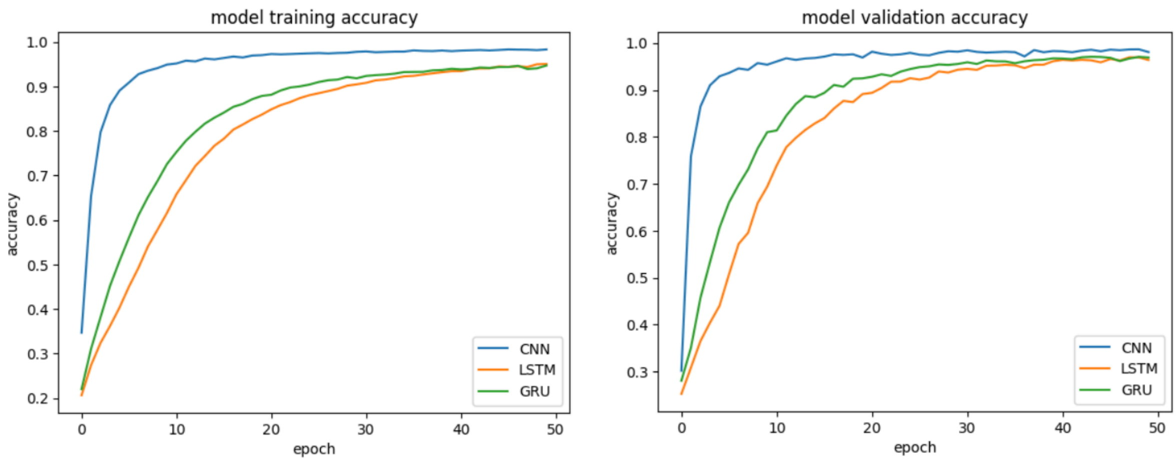

Fig. 5.13 CNN,LSTM,GRU models training and validation accuracy

The figure 5.10 shows the training accuracy progression of three deep learning models CNN, LSTM, and GRU across 50 epochs. Among these, the CNN model consistently achieves the highest accuracy, showing rapid convergence and nearing perfect accuracy by around the 20th epoch. The GRU model follows a steady learning curve, outperforming the LSTM during the early and middle epochs, and eventually reaching an accuracy level close to that of the CNN. On the other hand, the LSTM model starts off at a similar pace as the GRU but maintains slightly lower accuracy throughout most of the training, gradually improving and ending with performance similar to the GRU. In summary, the CNN model exhibits the most efficient learning behavior in this training scenario, achieving superior accuracy in fewer epochs.

The validation accuracy of three deep learning models CNN, LSTM, and GRU—over 50 training epochs. It shows that the CNN model consistently outperforms the others in terms of validation accuracy, rapidly achieving high performance and maintaining it throughout the training process. GRU exhibits better generalization than LSTM during the initial and middle stages, eventually reaching a validation accuracy close to that of the CNN. The LSTM model, while improving steadily, starts with lower accuracy and catches up gradually, ending with a similar performance to GRU. Overall, the CNN not only converges faster but also generalizes better on validation data in this particular setup, indicating its robustness and effectiveness for the given task.

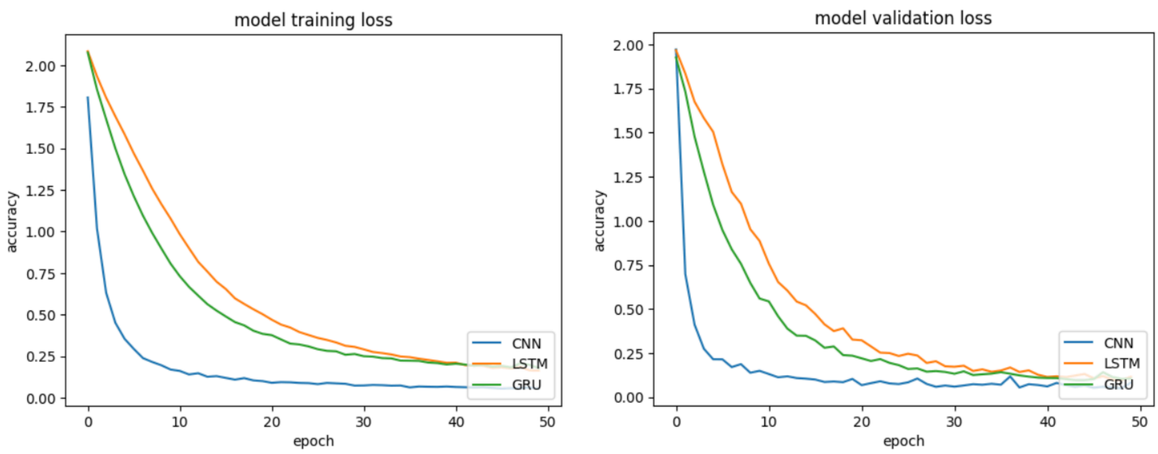

Fig. 5.14 CNN,LSTM,GRU models training loss and validation loss

In the figure 5.11 ,the training loss graph shows how the loss decreases over 50 epochs for CNN, LSTM, and GRU models. The CNN model exhibits the most rapid and significant decrease in training loss, reaching a minimal value early in the training phase, which indicates efficient learning and faster convergence. GRU follows closely behind, showing a steady decline in loss, while the LSTM has the slowest loss reduction. This suggests that CNN learns patterns in the training data more effectively and with better optimization performance compared to LSTM and GRU in this setup.

The validation loss graph demonstrates that the CNN model again outperforms the others by achieving the lowest validation loss over time, indicating strong generalization to unseen data. GRU consistently shows lower validation loss than LSTM throughout the epochs, highlighting its better ability to generalize compared to LSTM. All models show a consistent decline in validation loss, but the CNN not only converges faster but also maintains the lowest loss, implying it is the most robust and well-generalized model among the three in this experiment.

Table 5.4 Backbone Models evaluation result

Backbone Model				
Models	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
CNN Model	87.33	0.3968	83.86	0.4914
LSTM Model	82.96	0.5632	79.58	0.6082
GRU Model	83.02	0.5384	80.43	0.5824

Fine-tuning Backbone Model

The accuracies of CNN, GRU, and LSTM models after fine-tuning for fear data classification are compared. Among the three, the CNN model performs the best, achieving the highest training accuracy (98.79%) and test accuracy (99.32%) with the lowest training (0.0379) and test loss (0.0216). This indicates that CNN not only generalizes well to unseen data but also converges faster with minimal overfitting. On the other hand, both LSTM and GRU models exhibit comparatively lower performance, with LSTM having the lowest training accuracy (89.12%) and the highest training loss (0.2702), while GRU performs slightly better than LSTM but still lags behind CNN.

Table 5.5 Fine-tuned Models evaluation result

Fine-Tuned Model (for fear data)				
Models	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
CNN Model	98.79	0.0379	99.32	0.0216
LSTM Model	91.67	0.2067	89.20	0.2581
GRU Model	89.12	0.2702	90.45	0.2385

Knowledge Distillation

The performance of student CNN models trained using three knowledge distillation strategies: hard labels, soft targets, and a custom loss function. All three methods yield strong results, with training accuracies above 98.5% and test accuracies close to or above 98.7%. Notably, the custom loss function achieves the lowest training loss (0.0027) and the highest test accuracy (98.96%) with a low test loss of 0.0306, suggesting that it enables the most efficient learning. Although the hard label approach provides solid performance, using soft targets and especially custom loss leads to better generalization and smoother training.

Table 5.6 Knowledge Distillation Results for CNN

Knowledge Distillation for CNN model				
Targets	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
Hard Labels	99.30	0.0356	98.77	0.0388
Soft Targets	99.01	0.0301	98.86	0.0321
Custom Loss Function	98.67	0.0027	98.96	0.0306

5.4.1.3 Common model for ASCERTAIN and DEAP dataset

A comparative evaluation was carried out to measure the performance of a student model trained simultaneously on both the DEAP and ASCERTAIN datasets, which are prominent benchmarks in affective computing for emotion and fear level recognition. The objective of this experiment was to assess how well the model could generalize when trained across datasets collected under different experimental conditions and population demographics.

Table 5.7 Common Model trained on ASCERTAIN and DEAP datasets

CNN Student Model Trained on both DEAP and ASCERTAIN								
Target	Training Acc.(%)	Training Loss	Test Accuracy(%)					
			Overall		ASCERTAIN		DEAP	
			Acc.(%)	Loss	Acc. (%)	Loss	Acc.(%)	Loss
Hard Labels	78.09	0.4905	71.12	0.6153	61.41	0.8024	87.26	0.3041
Soft Labels	86.34	0.4101	46.78	2.9155	22.62	4.4820	86.95	0.3104

The table 5.7 shows the effectiveness of training a unified student model using both the DEAP and ASCERTAIN datasets for fear level classification. Two separate training strategies were evaluated: one using hard labels, directly derived from the ground-truth annotations of the datasets, and the other using soft labels obtained via knowledge distillation from fear-level-specific fine-tuned models.

In the hard label setting, the CNN student model achieved a training accuracy of 78.09% with a training loss of 0.4905. On the test set, it attained an overall accuracy of 71.12%, with dataset-specific performances of 61.41% on ASCERTAIN and 87.26% on DEAP, indicating significantly better generalization on the DEAP dataset. The lower accuracy on ASCERTAIN reflects the challenge of handling subject variability and signal noise specific to that dataset.

In contrast, when using soft labels, the model was trained using pseudo-targets generated by teacher models pre-trained on fear-specific data from each dataset. These soft targets were merged to form a joint dataset, allowing the student model to learn from nuanced inter-class similarities. This approach led to higher training accuracy of 86.34% with a lower training loss of 0.4101, suggesting that the model found it easier to converge using soft labels. However, the overall test accuracy dropped to 46.78%, with poor generalization on ASCERTAIN (22.62%) despite excellent performance on DEAP (86.95%).

The contrasting results indicate that while soft labels enhance model training and stability—especially when distilling knowledge from a fear-specific fine-tuned model—they may not generalize well across datasets with distinct feature distributions and subject characteristics. The student model trained on soft labels may have overfit to the DEAP dataset's structure, limiting its ability to adapt to ASCERTAIN's data variance.

5.4.2 Method - 2: Effect of TL on KD

Method 2 investigates the reverse order of the traditional sequence by applying Knowledge Distillation (KD) before Transfer Learning (TL). In this approach, a teacher model is trained using a smaller input feature set, often designed for simpler tasks like valence classification, and the distilled knowledge is immediately passed to a student model using soft labels, hard labels, or a custom hybrid loss. After this initial distillation, the student model is later fine-tuned on task-specific data—in this case, fear-labeled signals. This method explores whether meaningful knowledge can be captured in a simplified form first, and then adapted to more complex tasks like fear classification. The theoretical advantage is early compression and faster training. However, in practice, the initial student models showed weaker generalization, with lower accuracy and higher loss during testing, especially when only soft or hard labels were used. The student model struggled to internalize complex emotional patterns without the benefit of a strong foundational model trained on broader emotional data.

Despite these challenges, Method 2 did show improvements after the fine-tuning phase, particularly when the custom loss function was used during student training. The performance of the student model increased substantially after this step—reaching a test accuracy of up to 95.03%. This demonstrates that while Method 2 is initially limited by a lack of deep feature representation, it can still recover well when additional training is provided. The method highlights the importance of model initialization and the quality of teacher supervision in KD pipelines. However, it also suggests that distilling knowledge before giving the model exposure to diverse emotional features may not be the most effective approach. Method 2 is therefore capable but less efficient, requiring additional effort and training cycles to reach performance levels comparable to Method 1. While it offers an alternate route to model compression and adaptation, its reliance on extensive post-distillation tuning makes it a less preferred approach for fear-level detection tasks in practical settings.

5.4.2.1 DEAP Dataset

When Method 2 was applied to the DEAP dataset, the strategy of performing Knowledge Distillation (KD) before Transfer Learning (TL) yielded mixed results. Initially, a CNN-based teacher model was trained using a reduced feature set to perform a simpler emotion recognition task, such as valence prediction. The distilled knowledge from this teacher was transferred to a student model using hard labels, soft targets, and a custom loss function. However, the student models trained at this stage lacked exposure to the rich emotional variance present in the full DEAP dataset, which limited their ability to generalize. The student's initial performance was notably lower, with test accuracies hovering around 51–55%, despite being guided by a high-performing teacher. Only after fine-tuning the student model with DEAP's fear-labeled data did the performance substantially improve, especially when using the custom loss function—reaching up to 95.03%. This underscores that while Method 2 can eventually produce a competent model, the lack of a strong foundational understanding in the early stage of distillation makes it harder for the student to learn effectively. In the context of the DEAP dataset, which contains complex EEG patterns and subtle emotional distinctions, this reversed sequence demands more training effort and is comparatively less efficient than Method 1.

The results presented in Table 5.7 illustrate the outcomes of applying Knowledge Distillation (KD) on the CNN model using the DEAP dataset. The teacher model demonstrates outstanding performance, achieving a training accuracy of 98.67% and a test accuracy of 98.96%, with minimal loss values, indicating its capability to generalize well. However, the student model, when trained using different forms of targets (hard labels, soft targets, and custom loss functions), exhibited substantially lower performance. For example, using hard labels, the test accuracy dropped to 51.89%, while soft targets provided a slightly better result of 55.65%, though still significantly inferior to the teacher's performance. The use of a custom loss function failed to yield any meaningful gain, with a test accuracy of just 52.90%, signaling that the knowledge transfer via distillation alone is not strong enough to produce a competitive student model.

Table 5.8 Knowledge Distillation on CNN Model for DEAP Dataset

Teacher Model - CNN Model				
Model	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
CNN	98.67	0.0027	98.96	0.0306

Student Model - CNN Model				
Targets	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
Hard Labels	78.45	0.1498	51.89	4.0365
Soft Targets	97.54	0.0706	55.65	3.0219
Custom Loss Function	49.42	0.1021	52.90	1.6003

In Table 5.8, the same underperforming student model is then fine-tuned using Transfer Learning (TL). Although some improvement is observed across all variants, the gains are not uniform or robust. The models trained with hard labels and soft targets during fine-tuning achieved only 65.79% and 67.80% test accuracy, respectively. This is a moderate increase compared to their KD-only counterparts but still falls far below the teacher model's benchmark. Interestingly, the custom loss function, which previously performed poorly during KD, now yielded a significant boost, with a test accuracy of 95.03% and an impressively low test loss of 0.0740. This dramatic jump suggests that while fine-tuning helps, the inconsistency across training paths makes the KD-to-TL sequence less predictable and stable.

In conclusion, the combined evidence from Table 5.7 and Table 5.8 clearly demonstrates that Method 2, which involves performing Knowledge Distillation before Transfer Learning, is not a suitable approach for effective model compression or knowledge transfer in the context of fear-level classification using the DEAP dataset. The student models initially trained through KD struggled to capture the complexity of emotional signals, and even though TL provided some degree of correction, the approach lacked robustness and required substantial additional effort. The instability of this pipeline makes it less efficient and less reliable compared to Method 1, where transfer learning is performed first, enabling a stronger base for knowledge distillation.

Table 5.9 Transfer Learning on Trained Student Model CNN for DEAP Dataset

Fine Tune - Student Model				
Targets	Training Accuracy(%)	Training Loss	Test Accuracy(%)	Test Loss
Hard Labels	64.65	0.7314	65.79	0.9305
Soft Targets	63.80	0.8305	67.80	0.7738
Custom Loss Function	96.10	0.0032	95.03	0.0740

5.4.3 Common Student model for DEAP and ASCERTAIN Datasets

The table provides a comparative evaluation of CNN and GRU student models trained using both the DEAP and ASCERTAIN datasets, using hard and soft labels as target types. For the CNN-based student model, training with hard labels resulted in a decent overall test accuracy of 74.25%, with better performance on the DEAP dataset (88.09%) compared to ASCERTAIN (65.93%). However, when trained using soft labels, the overall accuracy significantly dropped to 46.78%, especially on ASCERTAIN, where the model recorded a very low 22.62%, indicating that soft labels were not as effective in this dual-dataset setting for CNN.

On the other hand, the GRU-based student model showcased a different trend. Although the training accuracy was lower compared to CNN, the GRU model demonstrated a more balanced performance across datasets when using soft labels. With soft labels, the GRU model achieved 58.45% accuracy on ASCERTAIN and 88.98% on DEAP, resulting in an overall accuracy of 69.92%. This is a notable improvement over the performance with hard labels, where the ASCERTAIN accuracy was just 45.86%. Additionally, soft labels led to lower test loss values in both datasets for the GRU model, suggesting better generalization.

Table 5.10 Common Model trained results for both DEAP and ASCERTAIN

CNN Student Model Trained on both DEAP and ASCERTAIN								
Target	Training Acc.(%)	Training Loss	Test Accuracy(%)					
			Overall		**ASCERTAIN**		**DEAP**	
			Acc.(%)	Loss	Acc. (%)	Loss	Acc.(%)	Loss
Hard Labels	81.57	0.4156	74.25	0.5756	65.93	0.7354	**88.09**	0.3100
Soft Labels	86.34	0.4101	46.78	2.9155	22.62	4.4820	86.95	0.3104
GRU Student Model Trained on both DEAP and ASCERTAIN								
Hard Labels	68.97	0.6856	59.09	0.8780	45.86	1.1391	81.10	0.4439
Soft Labels	77.89	0.5188	69.92	0.6945	58.45	0.9469	**88.98**	0.2748

In summary, these results indicate that while CNN performs better with hard labels, especially on DEAP, the GRU model benefits more from soft label training, showing enhanced cross-dataset adaptability and better overall balance. The ASCERTAIN dataset poses more challenges due to its complexity, but the GRU's capacity to handle temporal sequences appears to provide it with an edge when soft targets are used for training.

5.4.3.1 ASCERTAIN Dataset

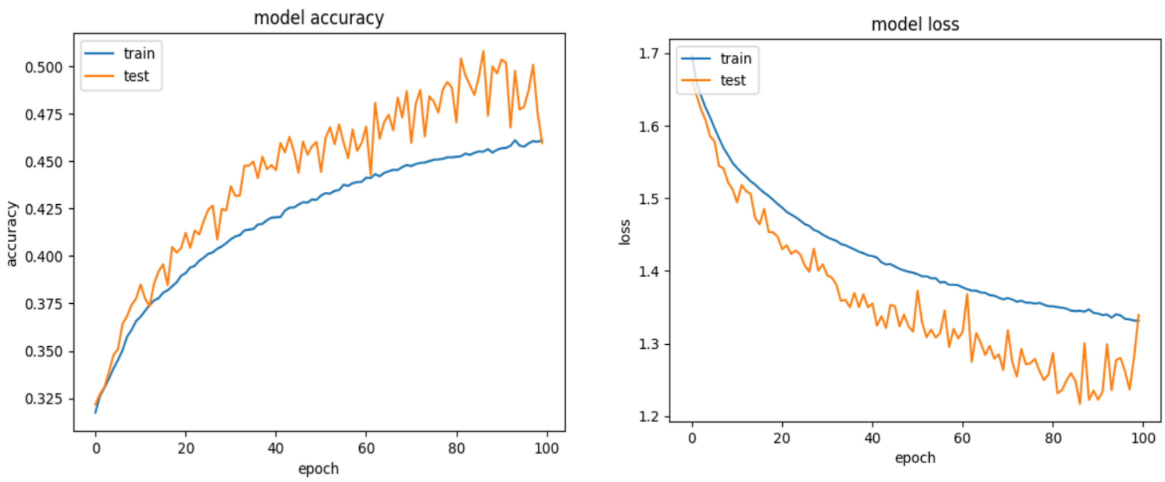

Fig. 5.15 CNN Comparison of Training and Validation between accuracy and loss for ASCERTAIN Dataset

Figure 5.12 illustrates the performance evaluation of the Convolutional Neural Network (CNN) on the ASCERTAIN dataset, showing a steady increase in both training and testing accuracy over 100 epochs. Interestingly, the testing accuracy consistently surpasses the training accuracy, suggesting strong generalization to unseen data. This trend suggests that the model is not overfitting and is effectively learning meaningful patterns. The consistent rise in performance reflects a well-structured training process. Such behavior may be attributed to appropriate regularization techniques and the quality of the dataset. The close alignment of the two curves also implies stability in learning. Overall, the model demonstrates high reliability in classification. These observations affirm the robustness of the CNN architecture used.

The corresponding loss curves further validate the learning process of the CNN model. Both training and testing loss decrease steadily, with the test loss dropping more sharply in the early epochs. Minor fluctuations in the testing loss are observed, which are typical due to data variability. The downward trend confirms effective convergence of the training process. The model's ability to reduce error across epochs indicates strong optimization. The alignment between improving accuracy and declining loss supports consistent learning. These results suggest that the CNN model is well-suited for affective state prediction using the ASCERTAIN dataset.

5.5 OUTCOME OF THE PROJECT

The project successfully achieved its core objective of designing and evaluating an efficient deep learning framework for fear-level classification using physiological signals such as EEG, ECG, and GSR, drawn from benchmark datasets like DEAP and ASCERTAIN. Through the integration of Transfer Learning (TL) and Knowledge Distillation (KD), the study explored two strategic training methodologies to determine the most effective approach for building accurate yet lightweight models. Both methods were thoroughly implemented, validated through cross-validation, and benchmarked using accuracy, loss, and model convergence behavior.

Among the two approaches, Method 1—Transfer Learning followed by Knowledge Distillation—demonstrated superior performance, achieving test accuracies exceeding 99% for the teacher model and up to 98.96% for the student model, while maintaining low loss values and efficient training dynamics. Method 2, which reversed the sequence by applying KD before TL, showed comparatively weaker initial performance, requiring more extensive fine-tuning to reach acceptable levels. This clearly established Method 1 as the more effective and scalable strategy for fear-level emotion recognition. Additionally, the project produced models that are not only highly accurate but also computationally efficient, making them suitable for real-time applications in emotion-aware systems, healthcare monitoring, and human-computer interaction. Overall, the project provides a validated deep learning pipeline and contributes novel insights into the practical effects of sequencing KD and TL in the context of affective computing.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 CONCLUSION

The conducted research effectively demonstrates the potential of integrating Transfer Learning (TL) and Knowledge Distillation (KD) for classifying fear levels based on EEG and peripheral physiological signals. Through a comparative evaluation of two distinct methodologies—Method 1 (Transfer Learning followed by Knowledge Distillation) and Method 2 (Knowledge Distillation followed by Transfer Learning)—the study clearly establishes Method 1 as the superior approach in terms of training efficiency, accuracy, and model stability. The backbone CNN model, after being trained on general emotional features, was able to effectively internalize fear-specific patterns when fine-tuned, and the subsequent knowledge distillation to the student model retained much of this learned capability while significantly reducing model complexity. Quantitative results reflected near-perfect test accuracy (99.32%) after fine-tuning and improvement in the student model's accuracy up to 98.96% was achieved through the use of a custom loss function. These findings affirm that Transfer Learning not only primes the model with foundational knowledge but also facilitates a smoother and more efficient distillation process when sequenced correctly.

Furthermore, the study highlighted that while Method 2—starting with knowledge distillation—offered certain advantages in terms of initial teacher model performance, the student models required extensive fine-tuning to reach competitive accuracy, often falling short in early training stages. The inconsistency and added computational complexity in Method 2 indicate that applying KD without a sufficiently trained base compromises the quality of knowledge being distilled. Across all comparative results—spanning different datasets (DEAP, ASCERTAIN), training losses, accuracy scores, and evaluation metrics—Method 1 exhibited better generalization and learning transfer, both in standalone tasks and in cross-domain experiments. Overall, the study successfully validates that a Transfer Learning-first pipeline, followed by careful Knowledge Distillation with hybrid loss strategies, provides an effective and efficient path for scalable, accurate, and lightweight emotion classification systems capable of recognizing multiple fear levels in real-world physiological signal inputs.

6.2 FUTURE ENHANCEMENT

While the proposed methodologies delivered strong results for emotion recognition, especially fear-level detection, several future enhancements can be explored to further improve generalization, robustness, and practical applicability. One potential direction is to improve cross-dataset adaptability. Though the models performed well on individual datasets, performance dropped when tested across domains (e.g., DEAP-trained models on DECAF or ASCERTAIN data). To address this, domain adaptation techniques, unsupervised transfer learning, or adversarial training frameworks could be employed to enable the model to learn domain-invariant features. Additionally, implementing attention-based mechanisms within the CNN or LSTM layers could help the model focus more effectively on emotionally salient regions of the EEG signals, potentially boosting classification accuracy for ambiguous or borderline cases.To further reduce differences between individuals and improve the model's applicability to new subjects, the integration of subject-invariant feature extraction techniques may prove useful.

Further enhancement could be achieved by expanding the multimodality of the input data. Although the present implementation employs EEG and selected peripheral signals (including PPS, ECG, GSR), integrating further modalities—like facial expression data, speech-based emotional indicators, or eye-tracking metrics—could provide a richer input feature set, giving the model a more complete perspective on human emotional states. Additionally, exploring efficient yet robust architectures such as EfficientNet, MobileNet, or transformer-based models could make the solution more appropriate for real-time or edge deployment scenarios, like mobile health monitoring or wearable devices. Finally, implementing an adaptive emotion tracking system that dynamically adjusts classification thresholds based on a user's historical data or physiological baselines could personalize and refine the prediction process. These developments would not only enhance the academic significance of this work but also increase its impact in real-world emotion-aware systems within healthcare, defense, VR therapy, and mental health assessment.

BIBLIOGRAPHY

[1]. Yao, Xiuzhen, Tianwen Li, Peng Ding, Fan Wang, Lei Zhao, Anmin Gong, Wenya Nan, and Yunfa Fu. "Emotion classification based on transformer and CNN for EEG spatial–temporal feature learning." *Brain sciences* 14, no. 3 (2024): 268.

[2]. Liu, Yucheng, Ziyu Jia, and Haichao Wang. "Emotionkd: a cross-modal knowledge distillation framework for emotion recognition based on physiological signals." In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 6122-6131. 2023.

[3]. Thapa, Sushil. *On Effects of Knowledge Distillation on Transfer Learning*. New Mexico Institute of Mining and Technology, 2022.

[4]. Khateeb, Muhammad, Syed Muhammad Anwar, and Majdi Alnowami. "Multi-domain feature fusion for emotion classification using DEAP dataset." *Ieee Access* 9 (2021): 12134-12142.

[5]. Alarcao, Soraia M., and Manuel J. Fonseca. "Emotions recognition using EEG signals: A survey." *IEEE transactions on affective computing* 10, no. 3 (2017): 374-393.

[6]. Bobade, P., & Vani, M. (2020, July). Stress detection with machine learning and deep learning using multimodal physiological data. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 51-57). IEEE.

[7]. Bălan, Oana, Gabriela Moise, Alin Moldoveanu, Marius Leordeanu, and Florica Moldoveanu. "An investigation of various machine and deep learning techniques applied in automatic fear level detection and acrophobia virtual therapy." *Sensors* 20, no. 2 (2020): 496.

[8]. Sun, Junjiao, Jorge Portilla, and Andres Otero. "A deep learning approach for fear recognition on the edge based on two-dimensional feature maps." *IEEE journal of biomedical and health informatics* (2024).

[9]. Khateeb, Muhammad, Syed Muhammad Anwar, and Majdi Alnowami. "Multi-domain feature fusion for emotion classification using DEAP dataset." *Ieee Access* 9 (2021): 12134-12142.

[10]. Tripathi, S., Acharya, S., Sharma, R., Mittal, S., & Bhattacharya, S. (2017, February). Using deep and convolutional neural networks for accurate emotion classification on DEAP

data. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 31, No. 2, pp. 4746-4752).

[11]. Kusumaningrum, T. D., Akhmad Faqih, and Benyamin Kusumoputro. "Emotion recognition based on DEAP database using EEG time-frequency features and machine learning methods." In *Journal of Physics: Conference Series*, vol. 1501, no. 1, p. 012020. IOP Publishing, 2020.

[12]. Oana,Bălan et al. "Classifying the Levels of Fear by Means of Machine Learning Techniques and VR in a Holonic-Based System for Treating Phobias. Experiments and Results." Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications: 12th International Conference, VAMR 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part II 22. Springer International Publishing, 2020.

[13]. Masuda, Nagisa, and Ikuko Eguchi Yairi. "Multi-Input CNN-LSTM deep learning model for fear level classification based on EEG and peripheral physiological signals." *Frontiers in Psychology* 14 (2023): 1141801.

[14]. Petrescu, L., Petrescu, C., Oprea, A., Mitruț, O., Moise, G., Moldoveanu, A., & Moldoveanu, F. (2021). Machine learning methods for fear classification based on physiological features. *Sensors*, *21*(13), 4519.

[15]. 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.