

Adaptive multitask emotion recognition and sentiment analysis using resource-constrained MobileBERT and DistilBERT: an efficient approach for edge devices

Muhammad Hussain¹, Caikou Chen¹, Muhammad Anwar², Sara Abdelwahab Ghorashi³, Ali Ahmed⁴, Muhammad Sheraz Arshad Malik⁵ and Iqra Yamin¹

¹ College of Information and Artificial Intelligence, Yangzhou University, Yangzhou, China

² Department of Information Sciences, Division of Science and Technology, University of Education, Lahore, Pakistan

³ Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

⁴ Department of Computer Science, Faculty of Computing and Information Technology-Rabigh, King Abdulaziz University, Rabigh, Saudi Arabia

⁵ Department of Software Engineering, Government College University Faisalabad, Faisalabad, Pakistan

ABSTRACT

Emotion recognition and sentiment analysis are crucial tasks in natural language processing, enabling machines to understand human emotions and opinions. However, the complex, nuanced relationship between emotions and sentiment in conversation poses significant challenges to accurate emotion recognition, as sentiment cues can be easily misinterpreted. Deploying emotion recognition and sentiment analysis tasks on edge devices poses substantial challenges due to computational resource constraints. We present an adaptive multitask learning approach that jointly leverages resource-constrained Mobile Bidirectional Encoder Representations from Transformers (MobileBERT) and Distilled BERT (DistilBERT) models to optimise emotion recognition and sentiment analysis. Our proposed approach utilises prototypical networks to learn effective representations of emotions and sentiment, while a focal weighted loss function effectively mitigates the class imbalance. We adaptively fine-tune the learning process to balance task importance and resource utilisation, resulting in better performance and efficiency. Our experimental results demonstrate the efficacy of our method, achieving the best results on MELD and IEMOCAP benchmark datasets while keeping a compact model size. Despite limited computational demands, our solution demonstrates that emotion and sentiment analysis can deliver performance comparable to resource-intensive large language models (LLMs). Facilitating various applications in human-computer interaction, affective computing, social media, dialogue conversion, and healthcare.

Submitted 19 March 2025

Accepted 19 September 2025

Published 3 November 2025

Corresponding author

Caikou Chen, yzcck@126.com

Academic editor

José Alberto Benítez-Andrades

Additional Information and
Declarations can be found on
page 23

DOI 10.7717/peerj-cs.3287

© Copyright

2025 Hussain et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Algorithms and Analysis of Algorithms, Artificial Intelligence, Computational Linguistics, Data Mining and Machine Learning, Sentiment Analysis

Keywords Multitask learning, Emotion recognition, Sentiment analysis, Resource-constrained, Prototypical network, Focal weighted loss

INTRODUCTION

Emotion recognition and sentiment analysis are critical tasks in natural language processing (NLP), enabling systems to interpret human emotions and opinions in conversational text. These tasks play a crucial role across multiple domains, particularly in areas such as social media monitoring, where real-time analysis of user-generated content is essential for understanding public sentiment and emotions, detecting emerging trends, and identifying potential crises. Beyond social platforms, they are also indispensable in *Amangeldi, Usmanova & Shamoi (2024)*, *Halawani et al. (2023)*, customer feedback analysis (*Abhyudhay et al., 2024*), mental health applications (*Lal & Neduncheliyan, 2024*), and human-computer interaction (*Tang, Yuan & Zhang, 2024*). The ability to accurately identify emotions and sentiments enables the creation of intelligent systems that are empathetic and context-aware (*Akhtar et al., 2019*; *Shah et al., 2025*). The rise of online communication platforms, including messenger apps, Twitter, and Facebook (*Rashid et al., 2020*; *Rehmani et al., 2024*) has generated vast amounts of conversational text data, increasing the demand for efficient and scalable solutions (*Zhang et al., 2022*).

Despite advances in NLP, traditional models for emotion recognition and sentiment analysis face challenges with imbalanced datasets and high computational requirements, which limit their deployment on resource-constrained devices, such as mobile phones and edge platforms. Lightweight transformers, such as Mobile Bidirectional Encoder Representations from Transformers (MobileBERT) and Distilled BERT (DistilBERT), offer promising solutions by balancing computational efficiency with high accuracy (*Sanh et al., 2019*; *Sun et al., 2020*). The domain of emotion recognition and sentiment analysis has gained increasing attention for several reasons. First, advanced techniques and approaches have been used to analyse data and recognise emotions and opinions. Second, the increasing adoption of technology has streamlined data collection processes. These technologies, now deeply embedded in daily life, facilitate the generation of vast datasets capturing user behaviours from diverse sources. This has fueled the rise of big data and the emergence of the Internet of behaviours (*Javed et al., 2020*). Third, there has been a marked increase in user interest and active engagement with these cutting-edge technologies, driven by their growing accessibility and tangible real-world benefits. As emotion recognition and sentiment analysis tools become more user-friendly and integrated into everyday applications (*Alslaity & Orji, 2024*), they have driven an exponential growth in the amount of available data.

While often used interchangeably in research, “sentiment” and “emotion” are tied to human subjectivity, yet they carry distinct meanings (*Chen et al., 2021*). Sentiment represents an attitude, judgment, or thought shaped by feelings, while emotion describes acute, consciously experienced affective states. The key distinction lies in the temporal duration of their experience (*Deshmukh et al., 2023*). Sentiments last longer and are more

stable than emotions (*Das et al., 2023*). Emotions are generally more complex and nuanced compared to sentiments.

Traditional approaches to emotion recognition and sentiment analysis have focused on single-task learning. These models are trained to perform a single specific task. These conventional approaches, however, overlook the demonstrated advantages of multitask learning paradigms, where a model is trained to perform multiple related tasks simultaneously. Recent research has shown that multitask learning frameworks can outperform single-task models in both accuracy and efficiency (*Huddar, Sannakki & Rajpurohit, 2020*), particularly for related tasks like emotion recognition and sentiment analysis.

Multitask learning of emotion recognition and sentiment analysis is closely correlated, often sharing overlapping features and contexts. Recognising an emotion like anger can help predict a negative sentiment and vice versa (*Chauhan et al., 2020*). Multitask learning (MTL) provides a robust paradigm to leverage these interdependencies. By training a model to perform both tasks simultaneously, MTL enables shared learning of representations, improving efficiency and generalisation. Traditional models with single tasks require distinct training pipelines and computational resources for each task, resulting in redundancy and inefficiency. MTL addresses this issue by allowing shared parameters and representations across functions, enabling the model to learn complementary information.

Despite the advantages of emotion recognition and sentiment analysis of multitask learning, existing approaches often rely on complex computational model architectures and require significant resources and amounts of labelled data. Moreover, traditional models are unsuitable for imbalanced datasets and deployment in low-resource environments, such as mobile devices and edge computing platforms. To address these challenges, our research proposes a novel multi-task learning approach that leverages lightweight transformers, such as MobileBERT and DistilBERT, to achieve state-of-the-art performance while reducing computational overhead (*Sanh et al., 2019; Ullah et al., 2023*).

To our knowledge, no prior work has combined prototypical networks with focal-weighted loss for this multitask learning problem in affective computing. Our approach addresses the key challenges and opportunities in text emotion recognition and sentiment analysis by leveraging lightweight transformers, MobileBERT and DistilBERT. Our work has the following key contributions.

- Novel Multitask Framework: we propose an adaptive MTL framework that jointly trains MobileBERT and DistilBERT for emotion recognition and sentiment analysis, leveraging task interdependencies to enhance performance.
- The proposed lightweight and efficient approach utilises MobileBERT and DistilBERT to achieve cutting-edge accuracy with substantially lower resource requirements, enabling practical implementation on mobile and edge computing systems.

- **Task-Specific Adaptations:** we incorporate optimised focal weighted loss and prototypical networks to handle skewed class distributions and improve shared learning across tasks.
- **Comprehensive Evaluation:** we evaluate our framework on MELD and IEMOCAP datasets, demonstrating superior accuracy, efficiency, and adaptability in low-resource settings.
- **Innovative Integration:** we combine prototypical networks, multi-task learning (MTL), and focal weighted loss to jointly learn emotion and sentiment, addressing key challenges in conversational natural language processing (NLP).

RELATED WORKS

Emotion recognition and sentiment analysis

Emotion recognition and sentiment analysis have gained significant attention in natural language processing (NLP) due to their wide applications, including social media content analysis, customer service analysis, human-computer interaction, and applications of healthcare analysis. The emotion recognition and sentiment analysis tasks aim to understand the emotional and sentiment-driven features of the conversational text. Still, they differ in the emotional responses they seek to recognise. Emotion recognition typically identifies emotions such as happiness, sadness, anger, or surprise, while sentiment analysis classifies text as expressing a positive, negative, or neutral sentiment. Despite their differences, these two tasks are closely related, and recent advances have explored methods to address both through multitask learning approaches ([Huddar, Sannakki & Rajpurohit, 2020](#)).

Emotion recognition and sentiment analysis are two interrelated tasks that have gained significant attention in the natural language processing (NLP) community, focusing on the dynamic nature of dialogue conversations and how emotions change over time in dialogue. Significant advancements in emotion recognition and sentiment analysis have been seen in recent years, with the development of various deep learning models and techniques. Advancements in deep learning have led to the development of various models for emotion recognition and sentiment analysis; traditional methods can be broadly categorised into three approaches: recurrent-based, graph-based, and transformer-based models.

Recurrent-based models utilise recurrent neural networks (RNNs) ([Salehinejad et al., 2017](#)) and their variants, such as long short-term memory (LSTM) ([Hochreiter & Schmidhuber, 1997](#)) and gated recurrent units (GRU) ([Dey & Salemt, 2017](#)) to capture sequential dependencies in dialogue conversion. Hierarchical Gated Recurrent Unit (HiGRU) ([Iqbal et al., 2022](#)) employs two GRU layers to model individual utterance emotions and the overall conversation context, capturing emotional dynamics at both the local word level and global utterance level. Dialogue Recurrent Neural Network (DialogRNN) ([Majumder et al., 2019](#)) uses a multi-GRU architecture to model context, speaker, and emotion states, providing a comprehensive understanding of conversation dynamics. Commonsense Knowledge for Emotion Identification in Conversations

(COSMIC) (Ghosal et al., 2020) extends DialogRNN by integrating external knowledge, enhancing the model's ability to recognise emotions by incorporating background context in dialogues. Bidirectional Long Short-Term Memory (BiLSTM)-Sentiment (Aziz Sharfuddin, Nafis Tihami & Saiful Islam, 2018) utilises BiLSTM to capture long-term dependencies in sentiment analysis and is often combined with attention mechanisms to emphasise important parts of a sentence. Sentiment Analysis Based on Attention Mechanisms (SABAM) (Zhu et al., 2019) proposes an RNN-based model with attention mechanisms for sentiment analysis, capturing the importance of different words in text sequences. Graph-based models represent conversations as networks, where utterances are nodes, and dependencies between them are captured through edges. Dialogue Graph Convolutional Network (DialogGCN) (Guesmi et al., 2023) uses graph convolutional networks (GCNs) to model conversations using directed graphs. This approach helps in capturing complex dependencies between dialogue turns. DAG-Emotion Recognition in Conversations (ERC) (Shen et al., 2021) employs a directed acyclic graph (DAG) structure to represent conversations, focusing on causal relationships between dialogue turns and information flow from previous dialogue states. Speaker and Position-Aware Graph Neural Network Model for Emotion Recognition in Conversation (S+PAGE) (Liang et al., 2021) further enhances speaker-aware graph convolutions, improving emotion recognition by emphasising speaker identity in multi-turn conversations. Transformer models have also been adapted for ERC, with the advantage of capturing long-range dependencies in dialogue. Knowledge-Enriched Transformer (KET) (Zhong, Wang & Miao, 2019) combines knowledge injection with transformer encoders to improve emotion recognition. Dialogue All-in-One XLNet for Multi-Party Conversation Emotion Recognition (DialogXL) (Shen et al., 2020) adapts the transformer architecture for ERC by applying dialogue-aware self-attention mechanisms, enhancing the model's ability to focus on appropriate parts of the conversation. BERT-ERC (Qin et al., 2023) applies a fine-tuned version of BERT to capture contextual and emotional conversation features. Pretrained language models (PLMs) such as BERT (Devlin et al., 2019), generative pre-trained transformer (GPT) (Yenduri et al., 2023), and others have wildly succeeded in emotion recognition tasks. Capturing long-range dependencies and semantic nuances in text. Supervised learning models traditional methods for sentiment classification relied on feature engineering combined with classifiers like support vector machines (SVMs), naive Bayes, and logistic regression. More recently, deep learning models like BiLSTM (Hochreiter & Schmidhuber, 1997) and convolutional neural network (CNN) (Khan & Alharbi, 2024) they have been widely used to capture long-range dependencies and local patterns in sentiment-level texts. With the advent of transformer models such as BERT, Sentence BERT (SentBERT) (Reimers & Gurevych, 2019) and Robustly Optimised Bidirectional Encoder Representations from Transformers (RoBERTa) (Liu et al., 2019) have emerged as state-of-the-art approaches for sentiment analysis. These models are fine-tuned on sentiment classification tasks to achieve high accuracy and robustness.

Multitask learning emotion recognition and sentiment analysis is an actively growing research trend. Multitask learning emotion recognition and sentiment analysis are closely correlated tasks, often sharing the features and context. Recognising an emotion like happy

can help predict a positive sentiment and vice versa. Multitask learning (MTL) trains both tasks jointly; models can leverage shared representations and improve overall performance. MTL allows emotion and sentiment sharing of knowledge across tasks, improving generalisation and performance on emotion recognition and sentiment analysis. MTL model Enhanced Representation through kNowledge IntEgration (ERNIE) (Sun et al., 2019), a pre-trained language model that incorporates knowledge from external sources to improve both emotion recognition and sentiment analysis simultaneously. Dual-task transformers (Aziz et al., 2023) fine-tune PLMs for both tasks simultaneously, leveraging shared representations for better performance in social media and other text data analysis. Joint-encoding Emotion Recognition (JER)-Sentiment (Delbrouck et al., 2020) uses a joint emotion recognition and sentiment analysis approach, leveraging the strengths of both tasks to improve performance. Multi-task Model for Sentiment and Emotion Analysis (MMSEA) (Kumar et al., 2019) introduces a multitask emotion recognition and sentiment analysis model, using a shared encoder to capture commonalities between the tasks.

However, large language models are computationally complex, requiring substantial resources and memory, which limits their use in resource-constrained settings. To address this challenge in multitask emotion recognition and sentiment analysis, we leverage lightweight, resource-efficient pre-trained language models, MobileBERT and DistilBERT. By integrating these models, we achieve a robust synergy that enhances the system's capacity to detect nuanced emotions and sentiments while improving generalisation to unseen data, thereby elevating the reliability and efficacy of emotion recognition and sentiment analysis systems.

METHODOLOGY

Problem statement

Given the MELD and IEMOCAP datasets, comprising conversational utterances labelled with seven emotion classes for MELD and six for IEMOCAP, along with sentiment labels for three classes: positive, neutral, and negative, our goal is to develop a multitask learning model that jointly predicts the emotion and sentiment states of speakers in a conversation. Each conversation is a sequence of utterances, $[(s_1, u_1), (s_2, u_2), \dots, (s_n, u_n)]$, where s_i is a speaker and u_i is their utterance. The model leverages lightweight transformers MobileBERT and DistilBERT, and prototypical networks with focal weighted loss to achieve high accuracy and efficiency on resource-constrained edge devices, addressing class imbalance and conversational context. Predictions for a given utterance are based on the Euclidean distance (Ji et al., 2020) between its embedding and the prototypes. Focal weighted loss, combined with the prototypical network, addresses class imbalance and enhances performance in minority classes.

Model architecture multitask

Multitask learning Models like MobileBERT and DistilBERT serve as feature extractors. Both are lightweight, low-resource transformer models optimised for efficiency and performance. The pre-trained MobileBERT (Sun et al., 2020) model is designed for

Algorithm 1 Training multitask MobileBERT/DistilBERT prototypical network.

Require: Training data $\mathcal{D} = \{(x_i, y_{e,i}, y_{s,i})\}_{i=1}^N$, hyperparameters: $\eta = 2 \times 10^{-5}$, $\epsilon = 1 \times 10^{-8}$, $T = 16$, $B = 64$, $\alpha_e, \alpha_s, \gamma, \lambda_1 = \lambda_2 = 1.0$

- 1: Initialise model θ with MobileBERT/DistilBERT weights, Adam optimiser (η, ϵ) , scheduler with 10% warmup
- 2: Compute class weights $w_{e,c}, w_{s,c}$ for $c \in \mathcal{E}, \mathcal{S}$: $w_c = \frac{N}{\sum_{i=1}^N I\{y_i = c\} \cdot |\mathcal{E}| \text{ or } |\mathcal{S}|}$
- 3: **for** $t = 1$ to T **do**
- 4: Shuffle $\mathcal{D}$ and create batches of size B
- 5: **for** each batch (x, y_e, y_s) **do**
- 6: $h \leftarrow \text{Model}(x; \theta)[:, 0, :], h \in \mathbb{R}^{B \times d}$
- 7: $h \leftarrow \text{Normalise}(h, p = 2)$
- 8: $p_{e,c} \leftarrow \text{ComputePrototypes}(h, y_e, \mathcal{E}), p_{s,c} \leftarrow \text{ComputePrototypes}(h, y_s, \mathcal{S})$
- 9: $\text{logits}_e \leftarrow -\text{CDist}(h, p_{e,c})/0.1, \text{logits}_s \leftarrow -\text{CDist}(h, p_{s,c})/0.1$
- 10: $L_e \leftarrow \text{WeightedFocalLoss}(\text{logits}_e, y_e, w_e, \alpha_e, \gamma), L_s \leftarrow \text{WeightedFocalLoss}(\text{logits}_s, y_s, w_s, \alpha_s, \gamma)$
- 11: $L \leftarrow \lambda_1 L_e + \lambda_2 L_s$
- 12: Clip gradients: $\nabla_\theta \leftarrow \text{Clip}(\nabla_\theta, 1.0)$
- 13: Update θ using Adam and scheduler
- 14: **end for**
- 15: Update $p_{e,c}, p_{s,c}$ with mean embeddings per class
- 16: Adjust α_e, α_s based on misclassification rates
- 17: Increase γ for harder examples
- 18: Evaluate on validation set: $L_{\text{val}} \leftarrow \text{Validate}(\theta, \mathcal{D}_{\text{val}})$
- 19: **end for**
- 20: **Output:** Trained model θ , prototypes $p_{e,c}, p_{s,c}$

resource-constrained environments; it provides a compact architecture with minimal performance degradation compared to BERT. MobileBERT incorporates bottleneck layers, inverted residual structures, and knowledge distillation for computational efficiency. It extracts robust contextual representations for utterances while minimising latency and memory usage. Whereas pre-trained DistilBERT (Sanh et al., 2019) is a smaller version of BERT, it balances speed and accuracy by retaining ~97% of BERT's performance with a 40% smaller model size. It captures semantic nuances and context in utterances, which are crucial for emotion recognition and sentiment analysis. Both models encode the input utterance into dense vector representations, further processed by downstream multitask layers.

Prototypical networks for emotion recognition and sentiment analysis

Prototypical networks are a type of metric-based learning designed for few-shot tasks (Ji et al., 2020). They operate by computing a prototype class centre for each class in the embedding space. Prototypes are generated for each of the seven emotion classes. Each utterance representation is compared with these prototypes using a distance metric euclidean distance. The class with the closest prototype is selected as the prediction. Sentiment task prototypes are computed for the three sentiment classes positive, neutral, negative. The model predicts sentiment by determining the nearest prototype in the sentiment space. This approach is robust for class-imbalanced datasets as prototypes adapt dynamically to the class distribution and ensure discriminative embeddings.

Shared layers and task-specific layers

The architecture employs shared layers to leverage common patterns across emotion recognition and sentiment analysis tasks while maintaining task-specific layers to preserve task distinctions. Shared layers consist of transformer layers from MobileBERT and DistilBERT and a shared self-attention mechanism that captures shared contextual and semantic information. Task-specific layers separate fully connected layers are added for emotion recognition and sentiment analysis, with independent weights. The prototypical layer calculates class prototypes and performs classification for each task. Attention mechanisms are integrated to enhance task performance. Multi-head attention improves the model's ability to focus on critical parts of the utterance, such as emotionally charged words or phrases. Cross-attention facilitates interaction between emotion and sentiment tasks, allowing the model to use sentiment signals to refine emotion recognition and vice versa.

Robust loss function for multitask learning

We use the focal weighted loss to mitigate the class imbalance and attention on difficult-to-predict classes, which adapts the conventional cross-entropy loss (Zhang & Sabuncu, 2018). Unlike standard cross-entropy loss, focal weighted loss includes a modulating factor that dynamically adjusts each class's contribution to the total loss based on its classification difficulty.

examples.

$$FL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t) \quad (1)$$

where:

p_t : model's predicted probability for the correct class.

α_t : a weight assigned to a class t , dynamically adjusted based on its frequency or misclassification rate.

γ : modulates the loss for well-classified examples to focus more on hard examples.

Enhances learning for minority classes while reducing the bias toward majority classes

Prototypical loss integrates the distance-based loss used in prototypical networks.

$$L_{proto} = \frac{1}{N} \sum_{i=1}^N d(f(x_i), \mu_{y_i}) \quad (2)$$

where:

$f(x_i)$: model's embedding of sample x_i .

μ_{y_i} : prototype of the true class y_i .

d : distance metric Euclidean distance.

The prototypical loss ensures that embeddings are tightly clustered around their respective class prototypes, improving inter-class separability and task performance. Total loss is the final loss for multitask learning, a weighted combination of focal weighted loss and prototypical loss for both tasks

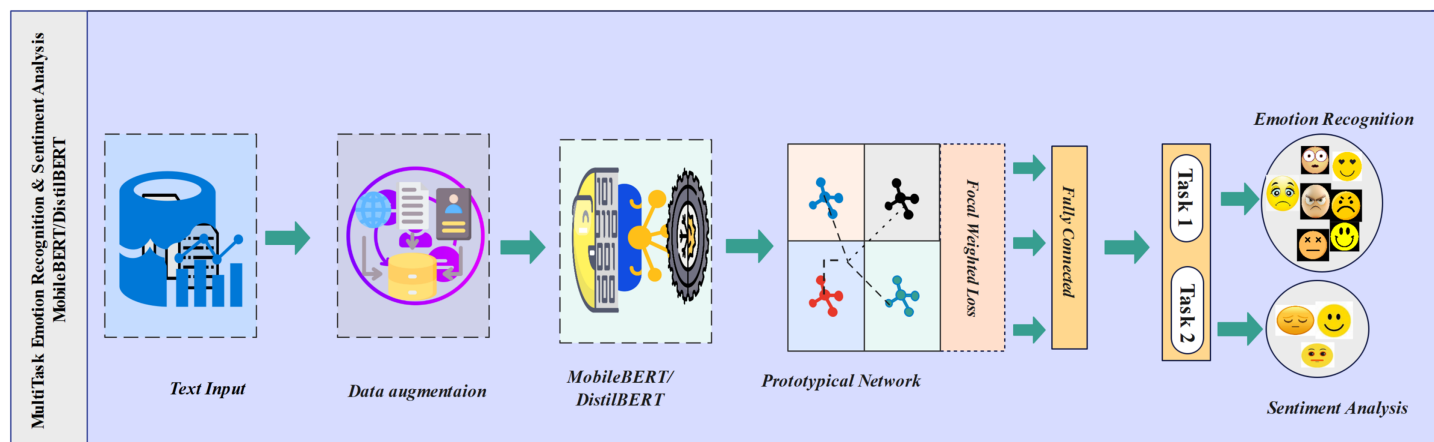


Figure 1 Multitask emotion recognition and sentiment analysis.

Full-size DOI: 10.7717/peerj-cs.3287/fig-1

$$L_{total} = \lambda_1 L_{emotion} + \lambda_2 L_{sentiment} \quad (3)$$

where λ_1 and λ_2 are weights balancing the two tasks.

EXPERIMENTAL CONFIGURATION

Model architecture and hyperparameter tuning

We have implemented the prototypical network and proposed loss function on MobileBERT and DistilBERT MTL for emotion recognition and sentiment analysis. The models are trained on the MELD and IEMOCAP datasets. We leverage pre-trained MobileBERT and DistilBERT models to extract contextualised word embeddings from text transcriptions. We fine-tune the pre-trained models, MobileBERT and DistilBERT, emotion recognition, and sentiment analysis. The models are trained using the Adam optimizer (Kingma & Ba, 2015) with a learning rate of $2e-5$. The training process spans different epochs 4, 8, 10, 12, and 16 with batch sizes 64 for multitasking. We apply data augmentation (Wei & Zou, 2019) through random word deletion to introduce variability into the training data, enhancing model generalisation and robustness. To address the class imbalance issues in the datasets, we employ focal weighted loss, dynamically adjusting the weights based on the difficulty of predicting specific classes. Prototypical networks are used for emotion recognition and sentiment analysis to learn task-specific embeddings. Where prototypes for each class are computed, the model classifies new data based on similarity to these prototypes. Our multitask learning model achieves robust and optimal results on MELD and IEMOCAP datasets. Our proposed approach requires fewer computational resources than large-scale models, offering a balance between performance and efficiency. Figure 1 is the representation of our multitask model architecture, illustrating how it processes both emotion recognition and sentiment analysis tasks simultaneously.

Model training configuration

We fine-tuned MobileBERT and DistilBERT models for multitask learning on the MELD and IEMOCAP datasets, enabling the simultaneous recognition of emotions and

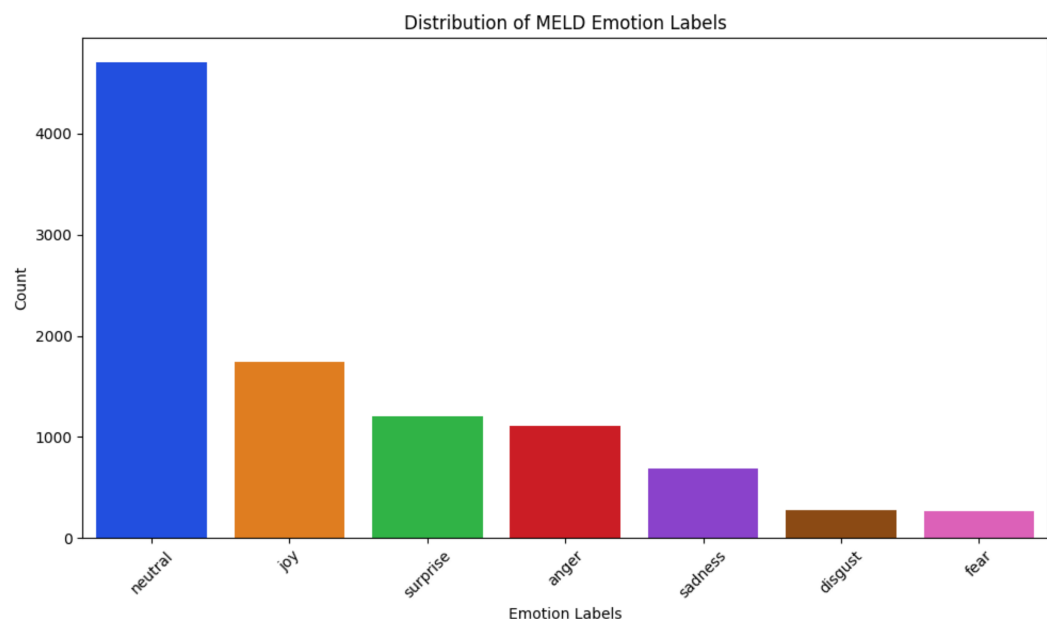


Figure 2 MELD emotion classes.

Full-size DOI: [10.7717/peerj-cs.3287/fig-2](https://doi.org/10.7717/peerj-cs.3287/fig-2)

sentiment analysis. Training was performed on a Windows 10 operating system with an RTX 3060 6 GB GPU, 32 GB of RAM, and a Core i7-2.60 GHz processor, using Python 3.8.2 as the environment. The Adam optimiser was employed with an optimised learning rate of $2e-5$, and models were trained over different epochs for better configuration. Specifically, 16 epochs were optimised with a batch size of 64. The models shared a standard backbone for feature extraction, followed by task-specific heads for emotion and sentiment classification. To address class imbalance in both tasks, we utilised the focal weighted loss function, which dynamically adjusts the loss based on the difficulty of class prediction. The focal weighted loss, combined with the prototypical network function, helps the model focus more on difficult-to-predict, minority classes. Additionally, we used prototypical networks to learn task-specific embeddings for emotion and sentiment categories, where each task has its own set of prototypes for classification. DistilBERT captures better nuance than MobileBERT. DistilBERTs have a larger parameter count and more compact architecture, providing them with better representational capabilities and enabling them to capture more intricate patterns and relationships in the data. DistilBERT's larger hidden size and additional layers allow it to capture contextual cues and relationships more effectively, resulting in improved performance on emotion recognition and sentiment analysis tasks. Whereas, MobileBERT's smaller parameter count and more efficient architecture make it more suitable for deployment on mobile devices or in resource-constrained environments.

Dataset descriptions

MELD (Poria et al., 2020)—The dataset, a multi-party *corpus* conversation collected from the Friends TV show, consists of approximately 1,400 conversations and over 13,000

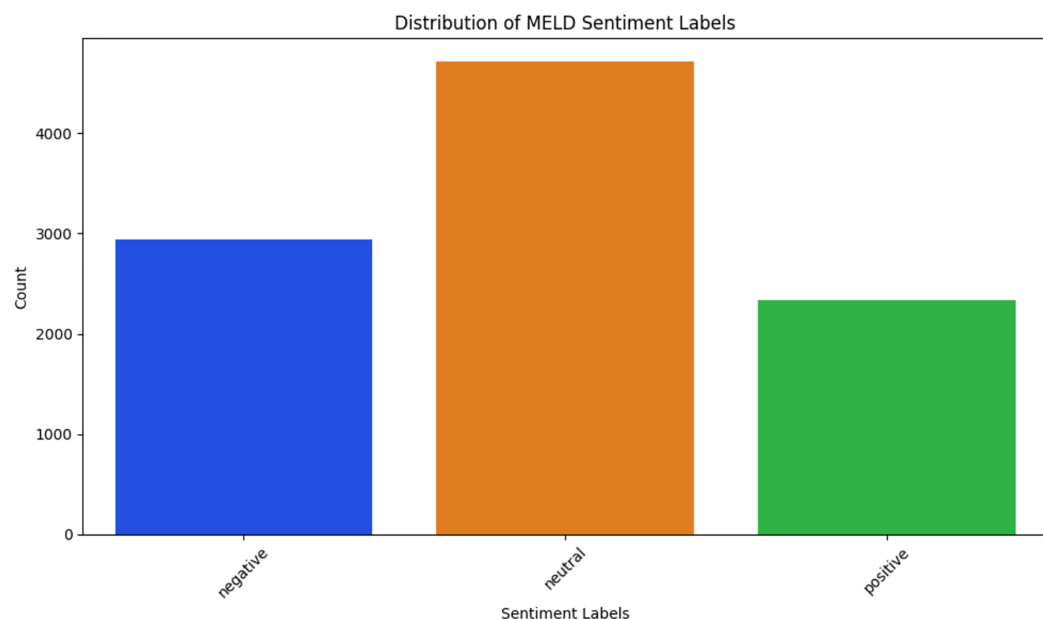


Figure 3 MELD sentiment classes.

Full-size DOI: 10.7717/peerj-cs.3287/fig-3

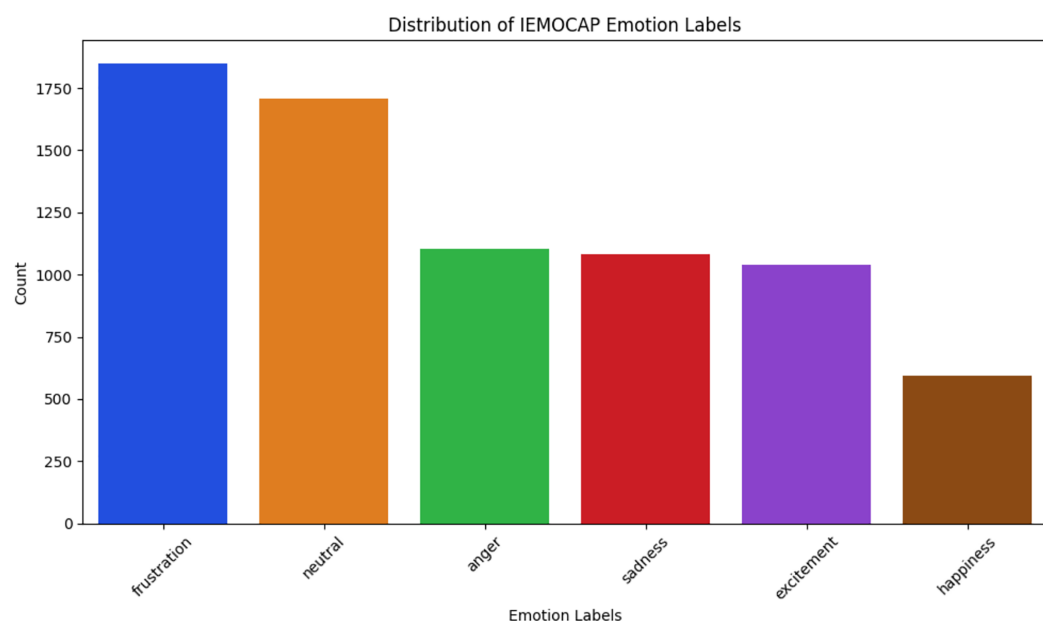


Figure 4 IEMOCAP emotion classes.

Full-size DOI: 10.7717/peerj-cs.3287/fig-4

utterances. The dialogues feature multiple speakers, with each utterance annotated with one of seven emotion classes, as depicted in Fig. 2: neutral, joy, surprise, anger, sadness, disgust, and fear. For multitask learning, each utterance is classified into one of these emotion categories for emotion recognition and assigned one of three sentiment labels, as illustrated in Fig. 3: positive, neutral, and negative for sentiment analysis.

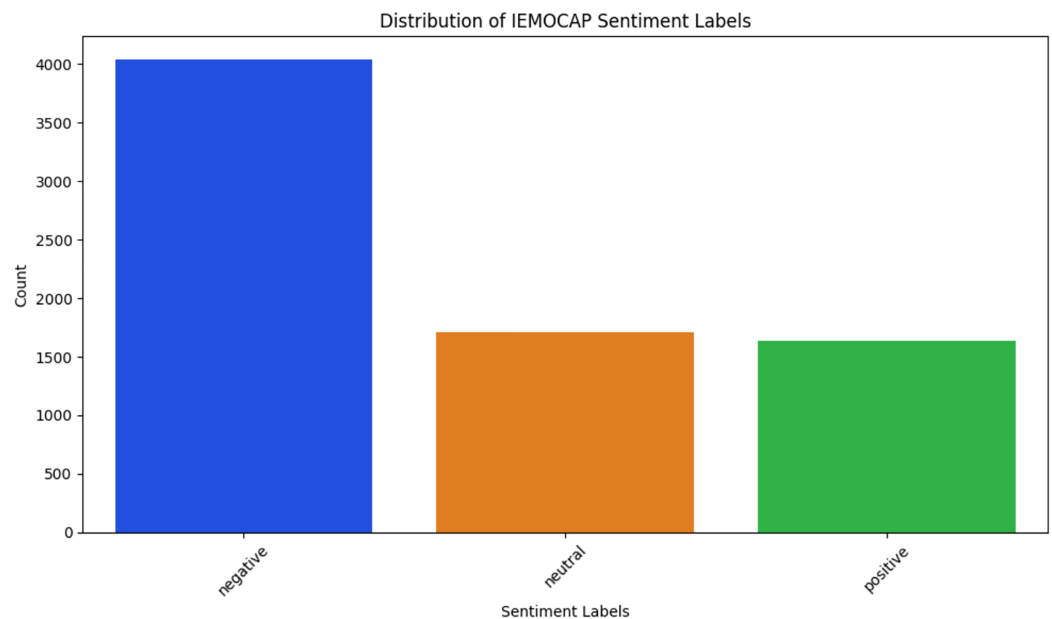


Figure 5 IEMOCAP sentiment classes.

Full-size DOI: [10.7717/peerj-cs.3287/fig-5](https://doi.org/10.7717/peerj-cs.3287/fig-5)

Table 1 Statistics of the MELD and IEMOCAP datasets.

Dataset	No. dials	Train	Dev	Test	No. Uttrs	Train	Dev	Test	No. CLS	Evolution metrics
MELD	1,432	1,038	114	280	13,708	9,989	1,109	2,610	7	Weighted avg F1
IEMOCAP	151	100	20	31	7,333	4,810	1,000	1,523	6	Weighted avg F1

Table 2 Deployment efficiency of baseline and proposed models on IEMOCAP and MELD datasets.

Model	Dataset	Emotion F1	Sentiment F1	Avg inference latency (s/sample)	Throughput (samples/sec)	CPU RAM usage (MB)	GPU VRAM usage (MB)
Bert-base-uncased	IEMOCAP	79	83	0.004551	219.733	1,421.253	1,678.595
Roberta-base	IEMOCAP	78	82	0.004546	219.9691	1,459.195	1,901.089
Distilbert-base-uncased	IEMOCAP	77	79	0.002487	402.0103	810.7658	1,145.097
Mobilebert-uncased	IEMOCAP	75	74	0.00197	507.4894	591.7427	640.5503
Bert-base-uncased	MELD	82	78	0.002402	416.2334801	1,467.072	1,599.713
Roberta-base	MELD	81	78	0.002268	440.8412999	1,991.656	1,768.6
Distilbert-base-uncased	MELD	81	77	0.000583	1,713.963	1,779.5	980.9394
Mobilebert-uncased	MELD	68	68	0.00103	970.6744	1,549.402	487.477

IEMOCAP (Busso et al., 2008)—The Interactive Emotional Dyadic Motion Capture dataset comprises 151 dialogue conversations, each recorded as video and featuring two speakers, resulting in a total of 7,333 conversation utterances. The dataset features distinct speakers participating in dyadic interactions. Each annotated utterance is labelled with one of six emotion categories, as shown in Fig. 4: frustrated, neutral, anger, sadness, excitement, and happiness, for multi-task learning. Additionally, each utterance is

Table 3 Models performance multitask emotion and sentiment.

Configuration	Model	MELD emotion F1	MELD sentiment F1	IEMOCAP emotion F1	IEMOCAP sentiment F1
Full configuration	DistilBERT	81%	77%	77%	79%
Full configuration	MobileBERT	68%	68%	75%	74%
Without augmentation	DistilBERT	78%	74%	74%	76%
Without augmentation	MobileBERT	65%	65%	72%	71%
Without prototypical network	DistilBERT	63%	60%	60%	62%
Without prototypical network	MobileBERT	53%	52%	60%	59%
Without focal weighted loss	DistilBERT	76%	72%	72%	74%
Without focal weighted loss	MobileBERT	64%	63%	71%	70%
With oversampling	DistilBERT	72%	71%	73%	73%
With oversampling	MobileBERT	63%	62%	63%	63%
With class weighting	DistilBERT	75%	75%	76%	76%
With class weighting	MobileBERT	66%	64%	66%	66%

classified into one of these emotion classes for emotion recognition and assigned one of three sentiment labels, as depicted in Fig. 5: positive, neutral, and negative, for sentiment analysis.

Performance metrics

We identify significant imbalances in the classes of MELD and IEMOCAP, popular public benchmark datasets, as detailed in Table 1. To ensure a rigorous and fair evaluation of our multitask learning framework for text emotion recognition and sentiment analysis, we adopt tailored evaluation metrics for both datasets, as outlined in Table 3. The subsequent table reports the weighted F1-scores for emotion recognition and sentiment analysis tasks across the MELD and IEMOCAP datasets for DistilBERT and MobileBERT under various configurations. The full configuration includes prototypical networks, focal weighted loss and random word deletion augmentation. Other configurations modify one component while retaining the others, with oversampling and class weighting replacing focal loss.

Model efficiency analysis

To validate suitability for resource-constrained environments, we evaluated the model size and parameter counts of MobileBERT and DistilBERT, comparing them to larger models BERT, RoBERTa, MobileBERT (~451 MB, 25M parameters, 68% emotion Weighted F1), and DistilBERT (~1.08 GB, 66M parameters, 81% emotion Weighted F1). These models have significantly lower resource demands than BERT (~2.7 GB, 110M parameters) and RoBERTa (~3.4 GB, 125M parameters), making them ideal for mobile devices. To further validate deployment suitability, we extend our analysis beyond model size and parameter counts by reporting inference latency, throughput, CPU RAM usage, and GPU VRAM usage for all baseline and proposed models as detailed in Table 2. The results highlight a trade-off between accuracy and efficiency. BERT-base and RoBERTa-base achieve the highest F1-scores up to 83% sentiment and 79% emotion on IEMOCAP but exhibit the

highest inference latency, 0.004551 for BERT-base and memory usage 1678.595 MB VRAM for BERT-base, rendering them less viable for edge deployment. DistilBERT offers a balanced approach, maintaining competitive F1-scores 81% emotion on MELD with reduced latency 0.000583 and memory usage 980.9394 MB VRAM, demonstrating its versatility. MobileBERT stands out with the lowest inference latency 0.00103 on MELD and memory footprint 487.477 MB VRAM, achieving over 2× throughput improvement 970.6744 samples/sec on MELD compared to larger baselines. Despite a modest F1-score drop 68% on MELD, MobileBERT's efficiency gains make it optimal for real-time inference in resource-constrained settings. This detailed breakdown underscores the practical advantages of distilled models, particularly MobileBERT, for deployment efficiency.

RESULTS AND IMPLICATIONS

We evaluate the proposed multitask learning approach for emotion recognition and sentiment analysis. Our results are presented in [Table 3](#). By leveraging MobileBERT and DistilBERT models with prototypical networks and focal weighted loss, our approach achieves superior performance on MELD and IEMOCAP. These results highlight the effectiveness of our multitask learning setup in capturing both emotional and sentiment dynamics in conversations. MobileBERT and DistilBERT offer computational efficiency while maintaining high accuracy, and a resource-friendly approach. Our model outperforms existing methods across key metrics, including weighted F1-score, [Table 3](#).

Ablation analysis

We conducted ablation experiments to evaluate the contributions of prototypical networks, focal weighted loss, and random word deletion augmentation. We conducted ablation with the full configuration, all components enabled, serving as the baseline. DistilBERT achieved 81% emotion-weighted F1 and 77% sentiment-weighted F1 on MELD, and 77% emotion-weighted F1 and 79% sentiment-weighted F1 on IEMOCAP. In comparison, MobileBERT achieves 68% emotion-weighted F1 and 68% sentiment-weighted F1 on MELD, and 75% emotion-weighted F1 and 74% sentiment-weighted F1 on IEMOCAP. Below, we describe the performance drops when each component is removed and analyse their relative impacts. Without random word deletion augmentation, disabling augmentation results in a consistent drop of approximately 3% in weighted F1-scores across datasets. Without prototypical networks, removing them causes the most significant performance drop, ranging from 15% to 18% in weighted F1-scores, highlighting their critical role in learning robust class prototypes for imbalanced datasets. Without focal weighted loss, disabling focal weighted loss leads to moderate drops of 4% to 5% in weighted F1-scores, confirming its importance (with $\gamma = 2$, $\alpha = 0.25$ for minority classes) in prioritising hard-to-classify examples.

Relative impact of components. The ablation results reveal that prototypical networks have the most considerable effect on performance, with F1-score drops of 15–18% when disabled, underscoring their essential role in enabling robust representation learning for

emotion and sentiment classification on imbalanced datasets. Focal weighted loss has a moderate impact, with drops of 4–5%, indicating its critical contribution to addressing class imbalance by focusing on complex examples. Random word deletion augmentation has the most negligible impact, with drops of approximately 3%, providing incremental improvements in generalisation, particularly for minority classes, when combined with focal loss. The synergy of all components in the full configuration maximises performance, with DistilBERT outperforming MobileBERT due to its larger parameter count, achieving up to 81% emotion-weighted F1 and 77% sentiment-weighted F1 on MELD.

Combining prototypical network with focal weighted loss

Our results demonstrate that the proposed multitask learning framework, combining lightweight models with prototypical networks and focal weighted loss, achieves high accuracy, 81% emotion-weighted F1-score on MELD with DistilBERT and efficiency of 66M parameters on conversational datasets. Unlike prior multitask emotion sentiment models often focus solely on accuracy, our approach explicitly addresses two critical gaps: class imbalance and deployment efficiency. By integrating prototypical networks, we improve representation learning for low-resource classes, while focal loss further enhances robustness against skewed label distributions. This combination enables consistent gains across both tasks, where recent multitask approaches typically underperform on minority emotions or require significantly larger computational resources.

Low resource requirements make it ideal for mobile and edge environments, where state-of-the-art multitask models remain impractical due to high memory and latency demands. Our ablation experiments confirm that both the prototypical module and focal loss are essential to these improvements, validating the originality of our methodology. Together, these contributions advance the field by demonstrating that efficient multitask learning with robust imbalance handling is possible without sacrificing accuracy, positioning our work as a practical and novel multitask transformer-based approaches.

Comparison of class imbalance techniques

To justify the use of focal weighted loss for addressing class imbalance in our multitask framework, we compared it with class weighting and oversampling techniques. We applied random word deletion as a preprocessing step for all experiments to mitigate initial class imbalance. Focal weighted loss outperforms class weighting by approximately 5% in emotion-weighted F1 and 1.5% in sentiment-weighted F1 on MELD, as it dynamically adjusts loss contributions based on prediction difficulty, prioritising hard-to-classify minority classes in conjunction with prototypical networks that learn robust class centroids. Class weighting, which applies static weights based on class frequencies, is less effective for conversational datasets with nuanced imbalances. Oversampling underperforms by 8.2% in emotion-weighted F1 and 4.3% in sentiment-weighted F1 on MELD, likely due to overfitting from synthetic data, which introduces noise despite the random word deletion preprocessing. Focal loss also exhibits lower test loss, implicitly lower than 0.672 for class weighting and 0.776 for oversampling. It avoids the

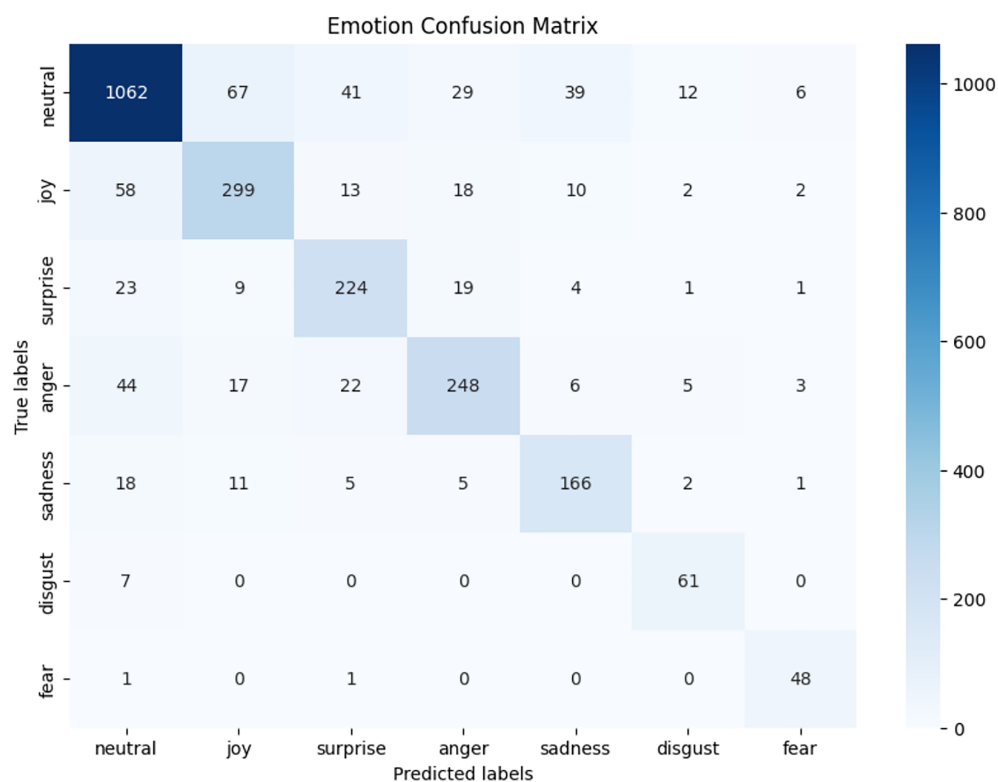


Figure 6 MELD emotion confusion matrix.

Full-size DOI: 10.7717/peerj-cs.3287/fig-6

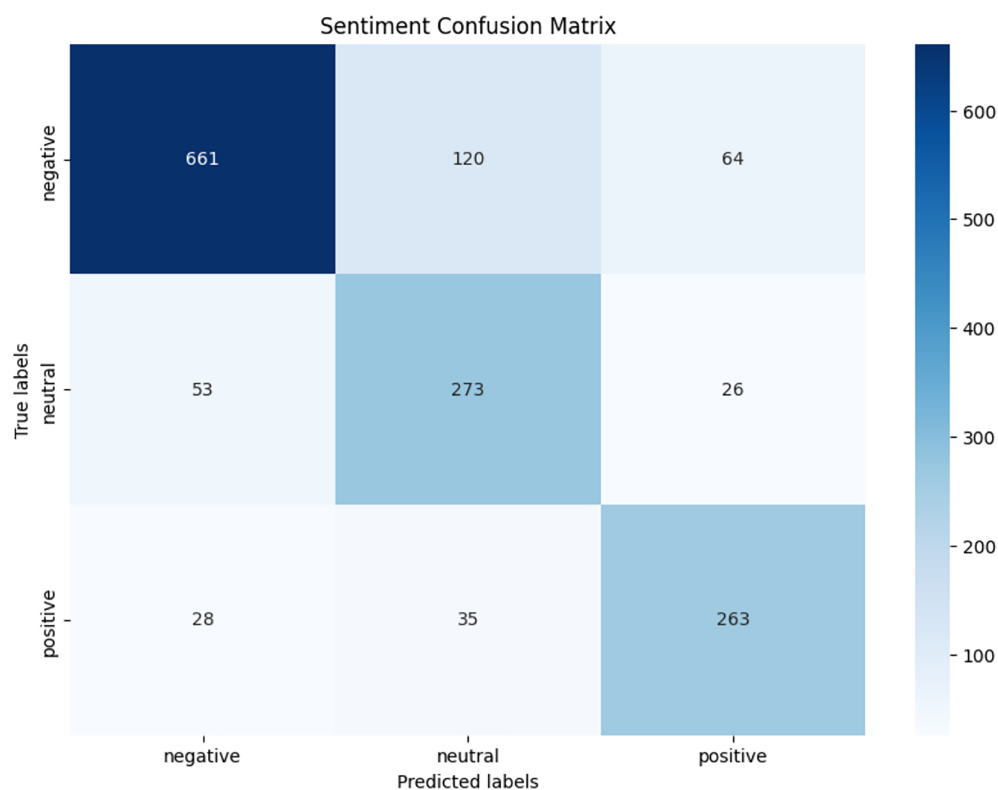


Figure 7 MELD sentiment confusion matrix.

Full-size DOI: 10.7717/peerj-cs.3287/fig-7

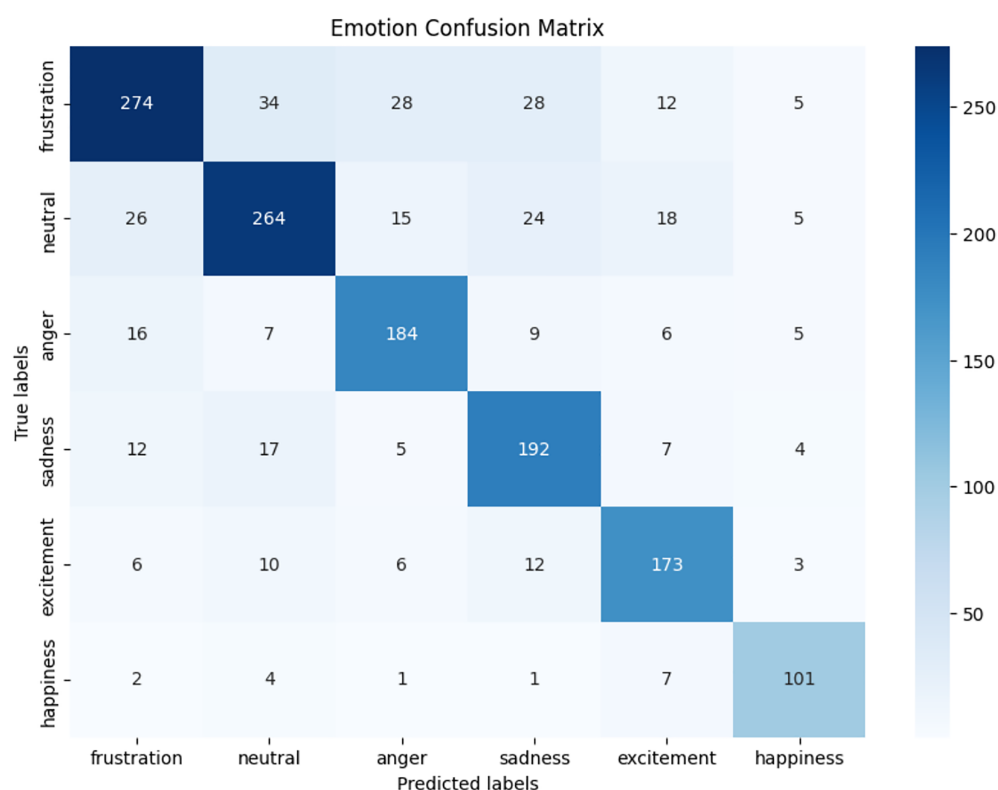


Figure 8 IEMOCAP emotion confusion matrix.

[Full-size](#) DOI: 10.7717/peerj-cs.3287/fig-8

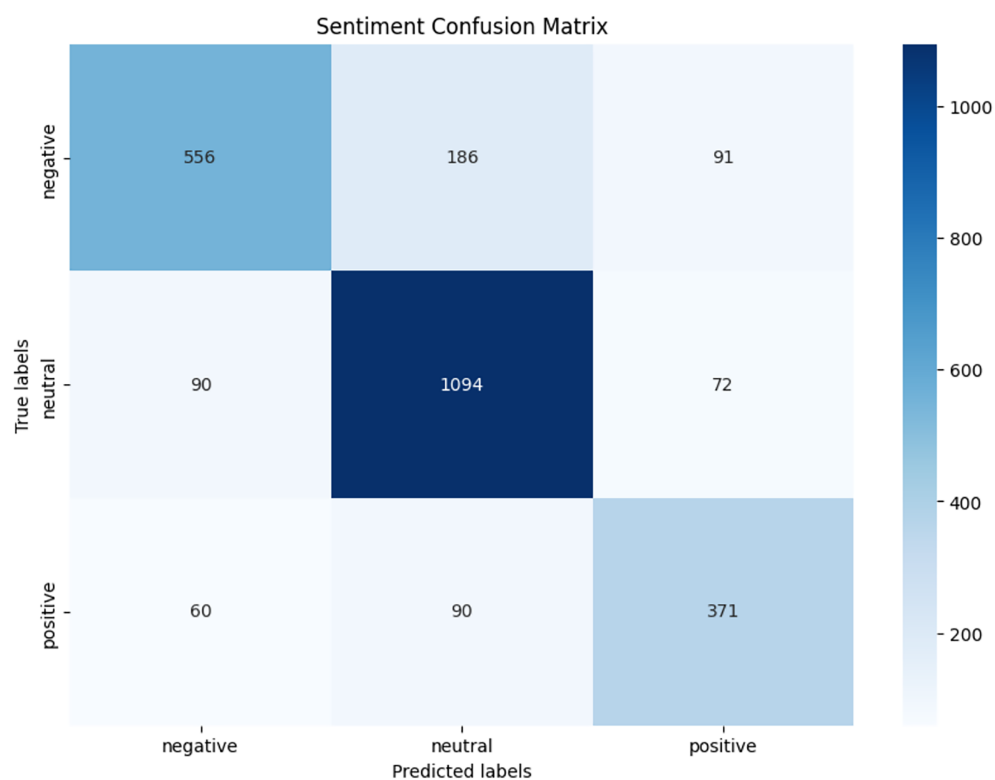


Figure 9 IEMOCAP sentiment confusion matrix.

[Full-size](#) DOI: 10.7717/peerj-cs.3287/fig-9

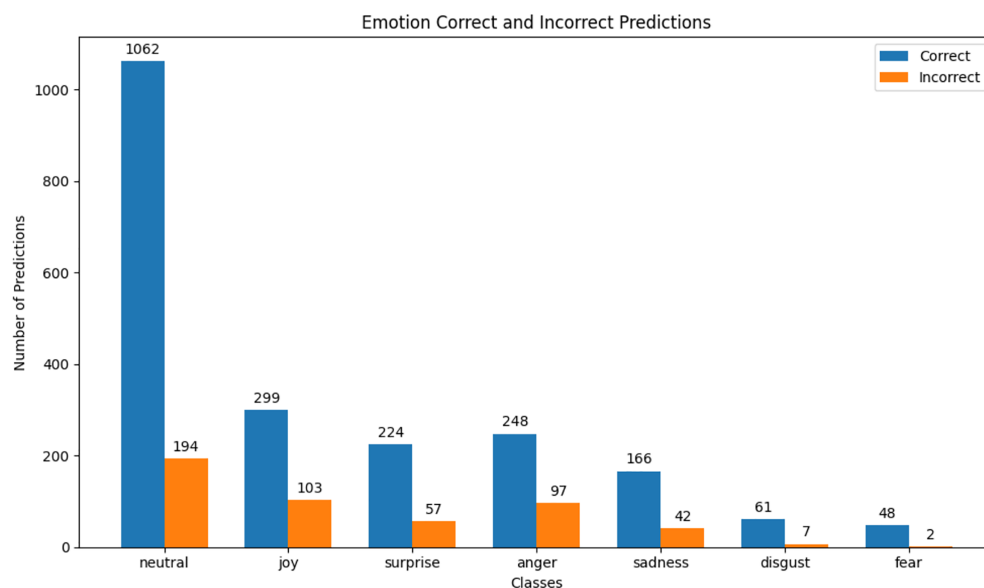


Figure 10 MELD emotion correct & incorrect prediction.

Full-size DOI: 10.7717/peerj-cs.3287/fig-10

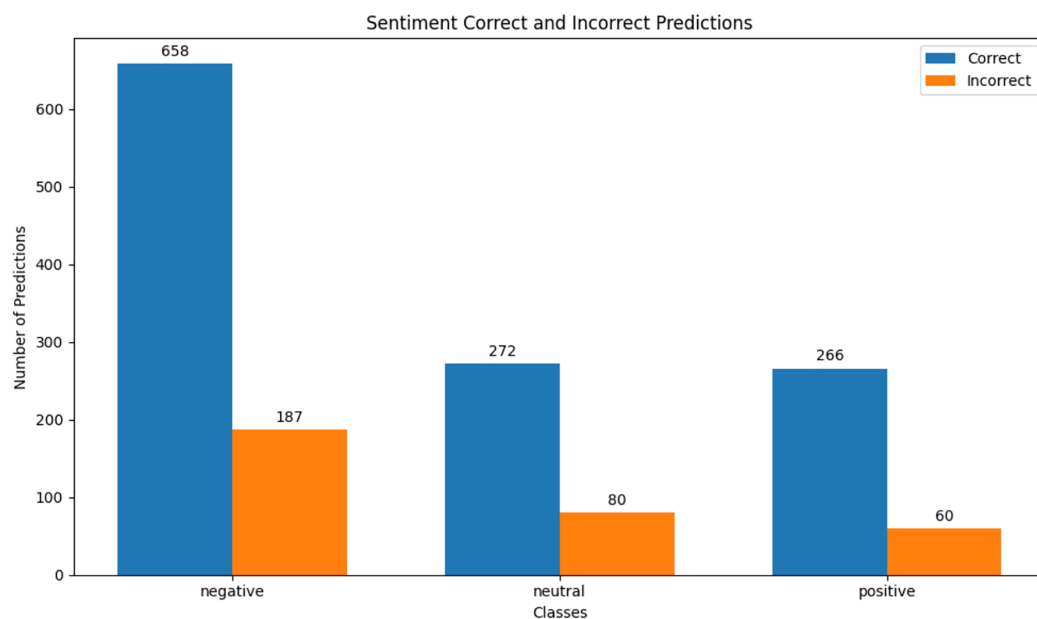


Figure 11 MELD sentiment correct & incorrect prediction.

Full-size DOI: 10.7717/peerj-cs.3287/fig-11

computational overhead of data duplication in oversampling, making it more suitable for lightweight models, such as DistilBERT 66M parameters. The combination of focal loss, prototypical networks, and random word deletion preprocessing enhances the performance of the minority class.

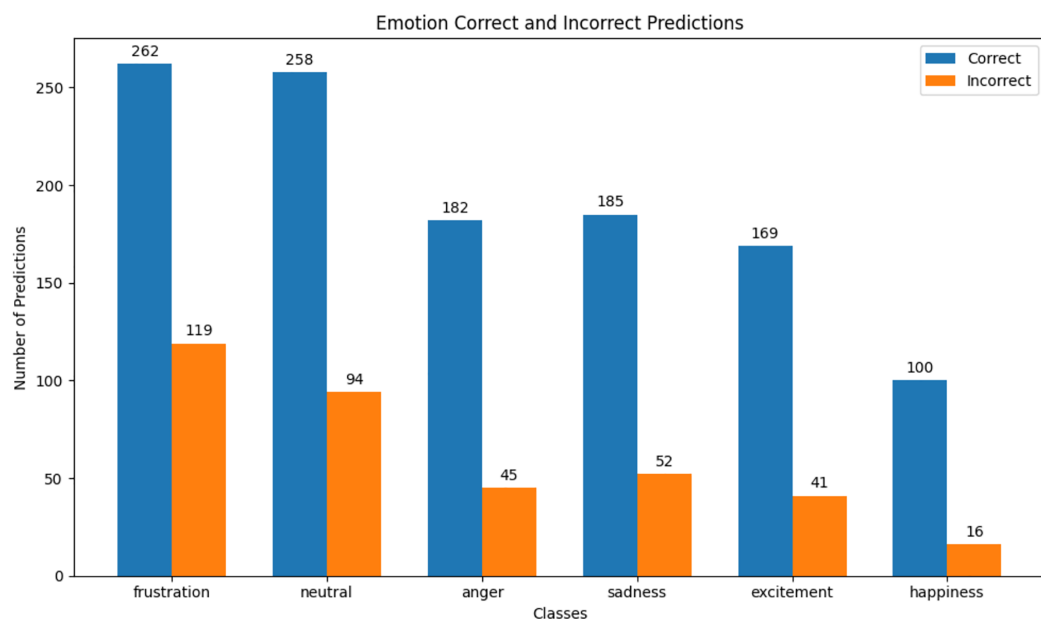


Figure 12 IEMOCAP emotion correct & incorrect prediction.

Full-size DOI: 10.7717/peerj-cs.3287/fig-12

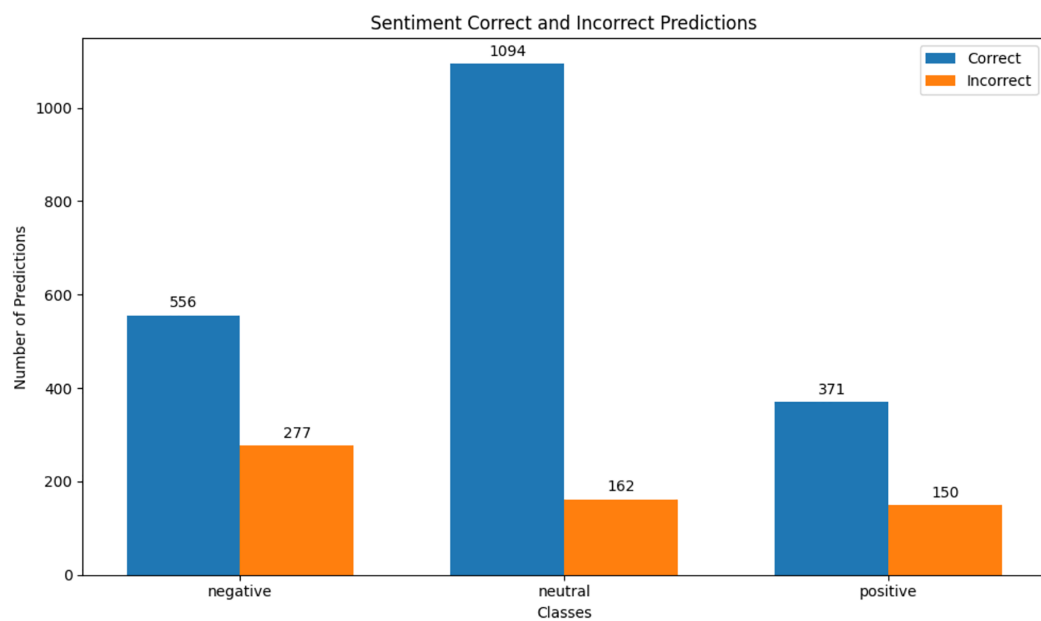


Figure 13 IEMOCAP sentiment correct & incorrect prediction.

Full-size DOI: 10.7717/peerj-cs.3287/fig-13

Hyperparameter tuning

We performed an extensive hyperparameter optimisation study to evaluate the influence of various hyperparameters on the efficacy of our multitask learning model for emotion recognition and sentiment analysis. Our findings indicate that an optimal configuration of

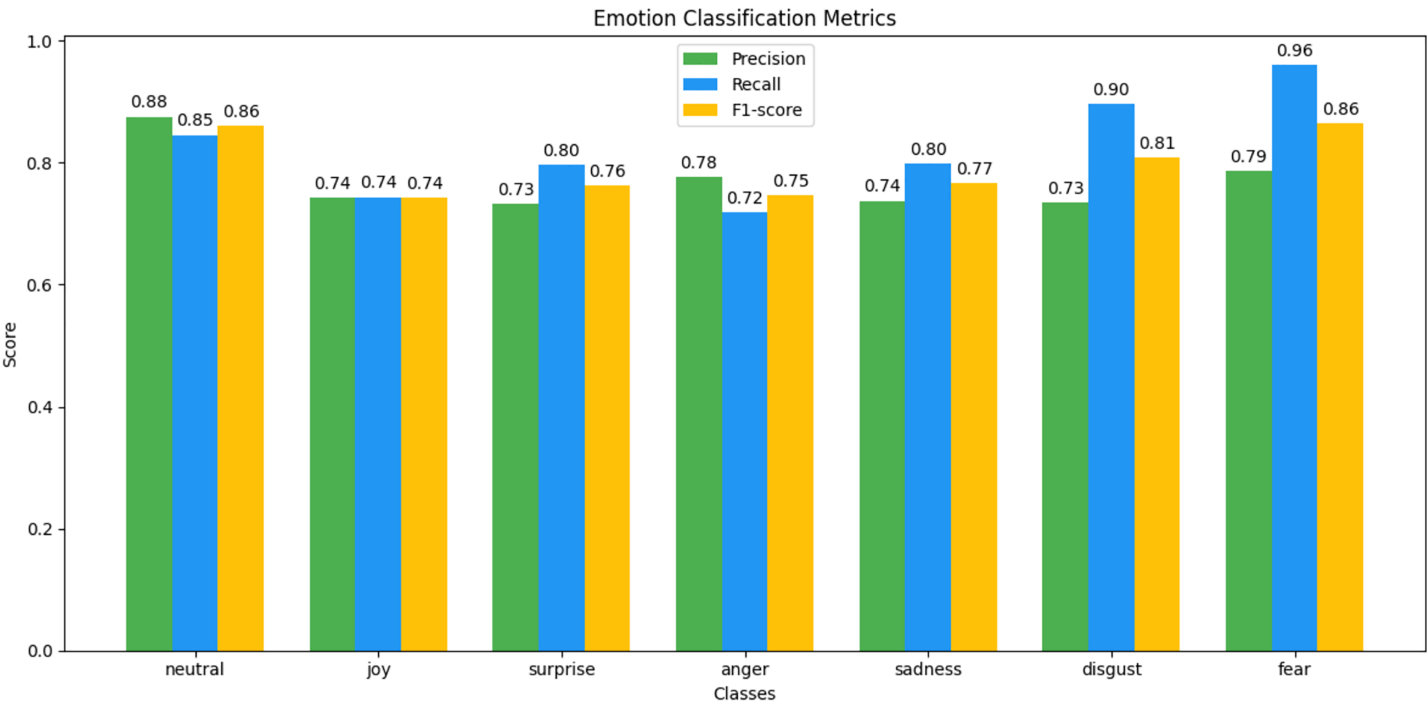


Figure 14 MELD emotion classification matrix.

Full-size DOI: 10.7717/peerj-cs.3287/fig-14

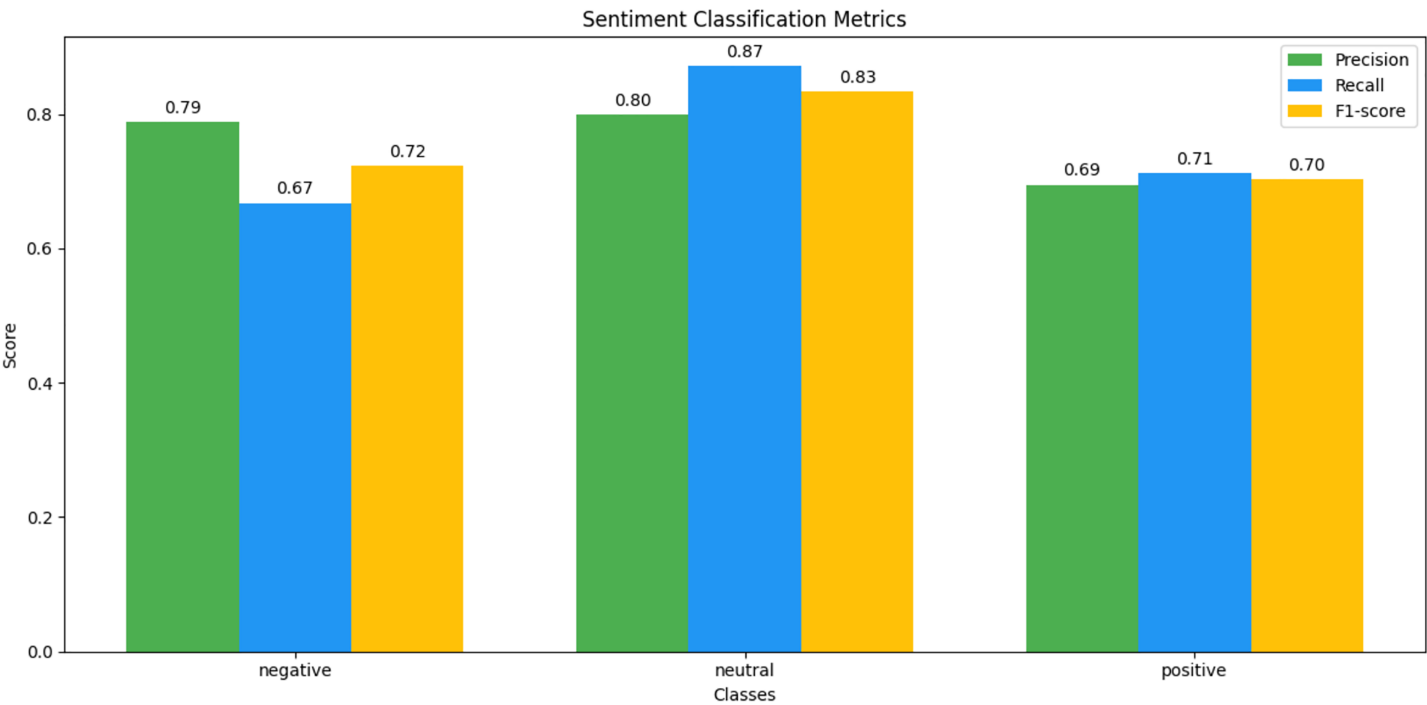


Figure 15 MELD sentiment classification matrix.

Full-size DOI: 10.7717/peerj-cs.3287/fig-15

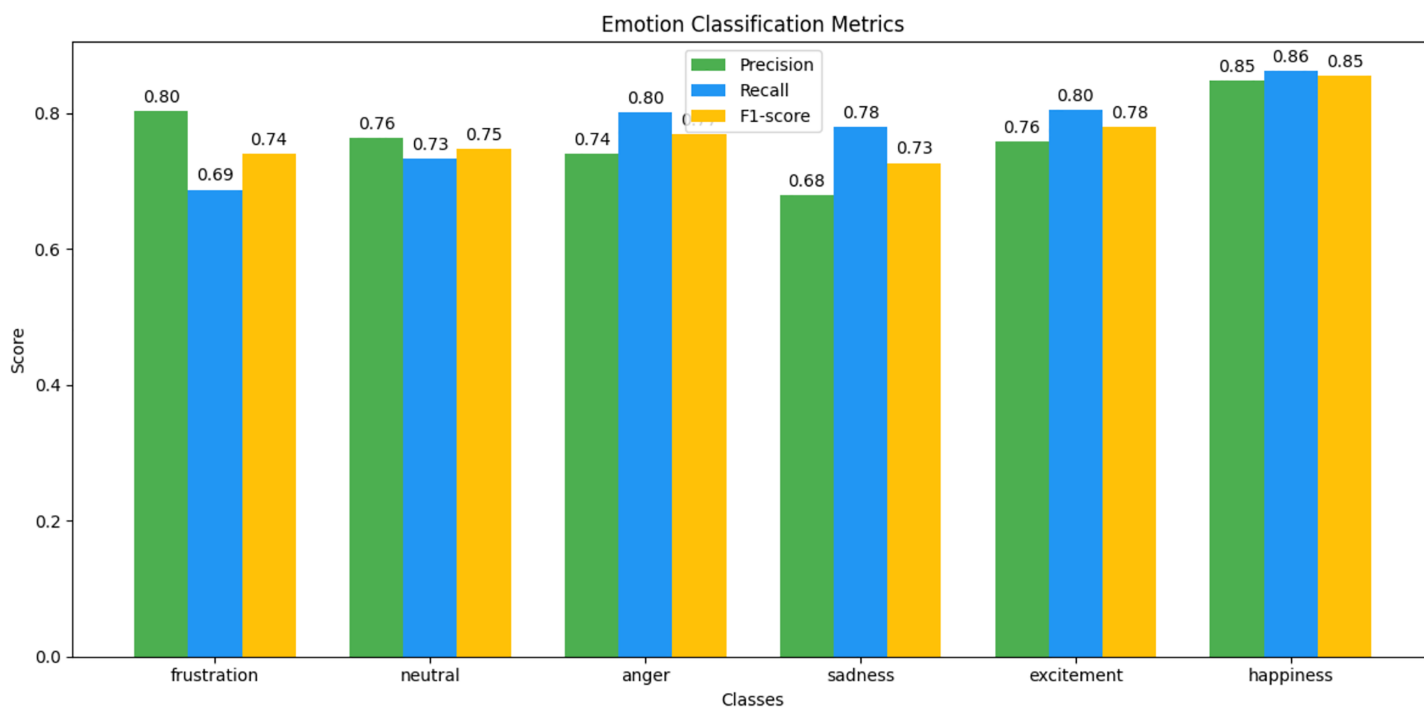


Figure 16 IEMOCAP emotion classification matrix.

Full-size DOI: 10.7717/peerj-cs.3287/fig-16

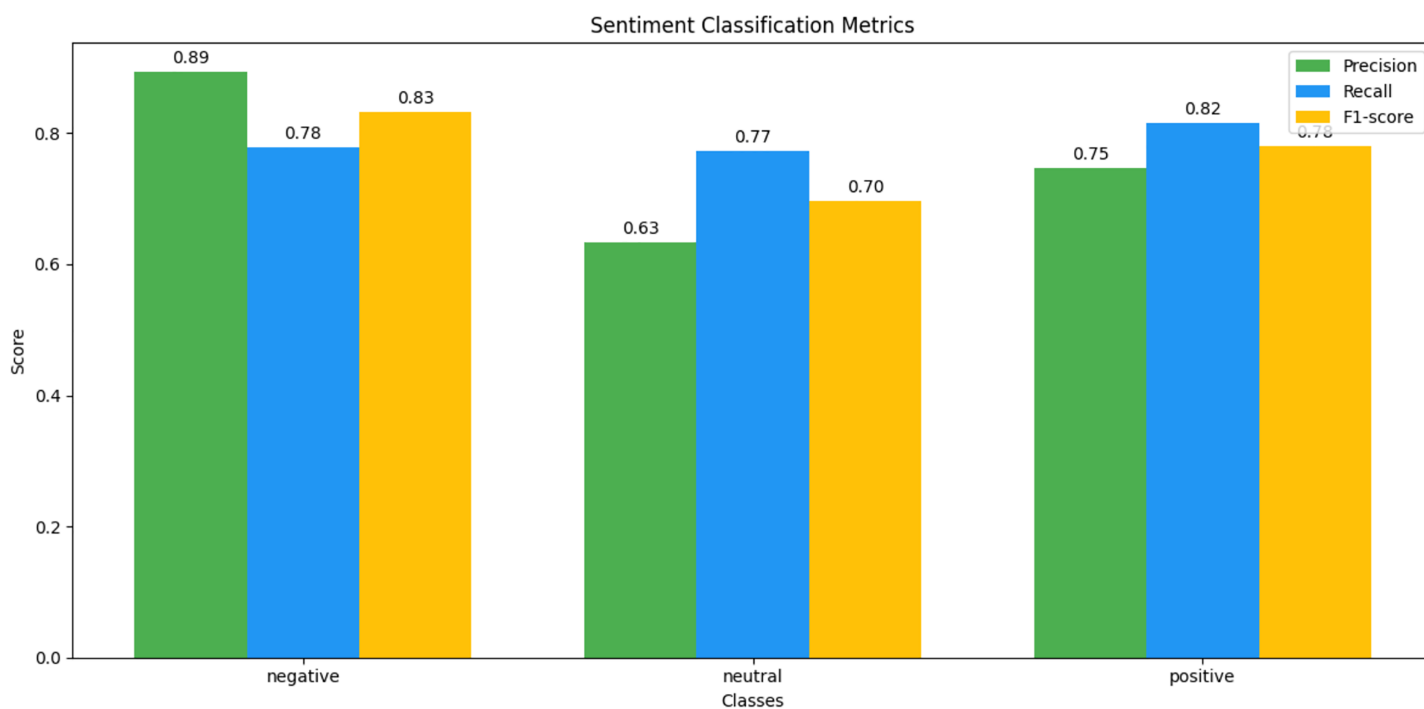


Figure 17 IEMOCAP sentiment classification matrix.

Full-size DOI: 10.7717/peerj-cs.3287/fig-17

hyperparameters encompassing the optimiser, learning rate, and batch size is critical for maximising model performance across both tasks. The Adam optimiser provided the best results due to its efficient handling of sparse gradients and adaptive learning rates. A learning rate of $2e-5$ and a batch size of 64 achieved optimal performance, striking a balance between training.

To tackle the issue of imbalanced datasets, we applied random deletion as a data augmentation strategy prior to experimentation. This preprocessing step reduced class imbalance, enhancing the model's generalisation across all classes in the MELD and IEMOCAP datasets. During experimentation, we further mitigated class imbalance by incorporating a focal weighted loss to emphasise challenging minority classes and utilising prototypical networks to establish robust class centroids in the feature space, as detailed in 'Combining prototypical Network with focal weighted loss'.

We observed that integrating prototypical networks with focal weighted loss, MobileBERT, and DistilBERT achieved superior performance across benchmark datasets MELD and IEMOCAP. The results, depicted in Figs. 6–17, are analysed as follows: Fig. 6, MELD emotion confusion matrix, shows the model's performance across seven emotion classes neutral, joy, surprise, anger, sadness, disgust, and fear. High diagonal values indicate strong classification performance. Misclassifications often occur between similar emotions, highlighting the challenge of detecting nuanced emotions in imbalanced datasets. Figure 7 MELD sentiment confusion matrix for sentiment classification positive, neutral, negative. The model excels at identifying neutral and positive sentiments more effectively than negative sentiments. The focal weighted loss mitigates class imbalance by prioritising negative samples, which are less frequent. Figure 8 IEMOCAP emotion confusion matrix across six emotion classes frustration, neutral, anger, sadness, excitement, happiness, the model performs well on minority classes as shown in confusion matrix. Figure 9 IEMOCAP sentiment confusion matrix the model accurately classifies positive and neutral sentiments as compare to the negative sentiment prediction, improved by focal weighted loss. Figures 10–13 correct & incorrect predictions Fig. 10 MELD emotion and Fig. 12 IEMOCAP emotion show correct predictions for emotions using DistilBERT, with MobileBERT slightly underperforming due to its smaller parameter count. Figure 11 MELD sentiment and Fig. 13 IEMOCAP sentiment indicate correct & incorrect sentiment predictions, with DistilBERT outperforming MobileBERT due to its ability to capture nuanced sentiment cues. Figures 14–17 classification metrics. Figure 14 MELD emotion and Fig. 16 IEMOCAP emotion report precision, recall, and F1 scores per class, for DistilBERT, demonstrating robust performance despite class imbalance. Figure 15 MELD sentiment and Fig. 17 IEMOCAP, show sentiment precision, recall, and F1 scores per class, respectively, with DistilBERT outperforming MobileBERT due to its larger architecture. These results validate the effectiveness of combining prototypical networks with focal weighted loss to address class imbalance and achieve competitive performance on resource-constrained models.

CONCLUSION AND FUTURE WORK

In this article, we present an efficient and practical approach for multitasking text emotion recognition and sentiment analysis, optimised for low-resource mobile and edge devices. By integrating adaptive multitask learning, prototypical networks, and a focal-weighted loss function, our method achieves best performance on benchmark datasets while maintaining a compact model size. Key findings demonstrate that adaptive multitask learning enables joint optimisation of emotion recognition and sentiment analysis, enhancing both performance and efficiency. Prototypical networks provide robust representations for these tasks, and the focal-weighted loss effectively mitigates class imbalance, further improving model performance.

Future work will focus on refining our approach for emotion recognition and sentiment analysis by exploring advanced, efficient transformer architectures and compact models optimised for low-resource environments. We aim to enhance performance by incorporating multimodal inputs, such as text, images, and audio, to enable more comprehensive and accurate analysis. Additionally, we plan to investigate novel loss functions and curriculum learning strategies to address class imbalance and extreme samples further, improving robustness on benchmark datasets.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work received funding from Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R755), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia: PNURSP2025R755.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Muhammad Hussain conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Caikou Chen conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Muhammad Anwar analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Sara Abdelwahab Ghorashi analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

- Ali Ahmed analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Muhammad Sheraz Arshad Malik performed the computation work, prepared figures and/or tables, and approved the final draft.
- Iqra Yamin performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The data used in this study are publicly available benchmark datasets commonly used for emotion recognition and sentiment analysis.

MELD (Multimodal Multi-Party Dataset) is available at: <https://affective-meld.github.io>.

The IEMOCAP (Interactive Emotional Dyadic Motion Capture) dataset is available at: <https://sail.usc.edu/iemocap>.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.3287#supplemental-information>.

REFERENCES

- Abhyudhay SR, Aditya GM, Upadya AK, Naik A, Ushashree P. 2024. Customer feedback and sentiment analysis for hotel services. In: *International Conference on Distributed Computing and Optimization Techniques, ICDOT 2024* DOI 10.1109/ICDOT.2024.10516070.
- Akhtar MS, Chauhan DS, Ghosal D, Poria S, Ekbal A, Bhattacharyya P. 2019. Multi-task learning for multi-modal emotion recognition and sentiment analysis. In: *NAACL HLT, 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference*, 1 DOI 10.18653/v1/n19-1034.
- Alsaity A, Orji R. 2024. Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour and Information Technology* 43(1):139–164 DOI 10.1080/0144929X.2022.2156387.
- Amangeldi D, Usmanova A, Shamoi P. 2024. Understanding environmental posts: sentiment and emotion analysis of social media data. *IEEE Access* 12(5):33504–33523 DOI 10.1109/ACCESS.2024.3371585.
- Aziz A, Chowdhury NK, Kabir MA, Chy AN, Siddique MJ. 2023. MMTF-DES: a fusion of multimodal transformer models for desire, emotion, and sentiment analysis of social media data. ArXiv DOI 10.48550/arXiv.2310.14143.
- Aziz Sharfuddin A, Nafis Tihami M, Saiful Islam M. 2018. A deep recurrent neural network with BiLSTM model for sentiment classification. In: *2018 International Conference on Bangla Speech and Language Processing, ICBSLP 2018* DOI 10.1109/ICBSLP.2018.8554396.
- Busso C, Bulut M, Lee CC, Kazemzadeh A, Mower E, Kim S, Chang JN, Lee S, Narayanan SS. 2008. IEMOCAP: interactive emotional dyadic motion capture database. *Language Resources and Evaluation* 42(4):335–359 DOI 10.1007/s10579-008-9076-6.
- Chauhan DS, Dhanush SR, Ekbal A, Bhattacharyya P. 2020. Sentiment and emotion help sarcasm? A multi-task learning framework for multi-modal sarcasm, sentiment and emotion analysis. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics* DOI 10.18653/v1/2020.acl-main.401.

- Chen Z, Cao Y, Yao H, Lu X, Peng X, Mei H, Liu X. 2021.** Emoji-powered sentiment and emotion detection from software developers' communication data. *ACM Transactions on Software Engineering and Methodology* 30(2):1–48 DOI 10.1145/3424308.
- Das RK, Islam M, Hasan MM, Razia S, Hassan M, Khushbu SA. 2023.** Sentiment analysis in multilingual context: comparative analysis of machine learning and hybrid deep learning models. *Heliyon* 9(9):e20281 DOI 10.1016/j.heliyon.2023.e20281.
- Delbrouck JB, Tits N, Brousmiche M, Dupont S. 2020.** A transformer-based joint-encoding for emotion recognition and sentiment analysis. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics* DOI 10.18653/v1/2020.challengehtml-1.1.
- Deshmukh Rushali A, Amati V, Bhamare A, Jadhav A. 2023.** Visual sentiment analysis: an analysis of emotions in video and audio. In: *International Conference on IoT Based Control Networks and Intelligent Systems*. Singapore: Springer Nature Singapore, 313–326.
- Devlin J, Chang MW, Lee K, Toutanova K. 2019.** BERT: pre-training of deep bidirectional transformers for language understanding. In: *NAACL HLT, 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies—Proceedings of the Conference*, 1.
- Dey R, Salemt FM. 2017.** Gate-variants of gated recurrent unit (GRU) neural networks. In: *Midwest Symposium on Circuits and Systems, 2017-August*, 1597–1600 DOI 10.1109/MWSCAS.2017.8053243.
- Ghosal D, Majumder N, Gelbukh A, Mihalcea R, Poria S. 2020.** COSMIC: COMmonSense knowledge for eMotion Identification in Conversations. In: *Findings of the Association for Computational Linguistics Findings of ACL: EMNLP 2020*, 2470–2481 DOI 10.18653/V1/2020.findings-emnlp.224.
- Guesmi T, Al-Janfawi F, Guesmi R, Alturki M. 2023.** Efficient social media sentiment analysis using confidence interval-based classification of online product brands. *International Journal of Advanced and Applied Sciences* 10(10):94–102 DOI 10.21833/ijaas.2023.10.011.
- Halawani HT, Mashraqi AM, Badr SK, Alkhalaf S. 2023.** Automated sentiment analysis in social media using Harris Hawks optimisation and deep learning techniques. *Alexandria Engineering Journal* 80(23):433–443 DOI 10.1016/j.aej.2023.08.062.
- Hochreiter S, Schmidhuber J. 1997.** Long short-term memory. *Neural Computation* 9(8):1735–1780 DOI 10.1162/neco.1997.9.8.1735.
- Huddar MG, Sannakki SS, Rajpurohit VS. 2020.** Multi-level context extraction and attention-based contextual inter-modal fusion for multimodal sentiment analysis and emotion classification. *International Journal of Multimedia Information Retrieval* 9(2):103–112 DOI 10.1007/s13735-019-00185-8.
- Iqbal MJ, Iqbal MW, Anwar M, Khan MM, Nazimi AJ, Ahmad MN. 2022.** Brain tumor segmentation in multimodal MRI using U-Net layered structure. *Computers, Materials & Continua* 74(3):5267–5281 DOI 10.32604/cmc.2023.033024.
- Javed M, Ziauddin, Kamal K, Nasir JA, Raza AA, Habib A. 2020.** Socio monitoring framework (SMF): efficient sentiment analysis through informal and native terms. *International Journal of Advanced and Applied Sciences* 7(12):113–126 DOI 10.21833/ijaas.2020.12.013.
- Ji Z, Chai X, Yu Y, Pang Y, Zhang Z. 2020.** Improved prototypical networks for few-Shot learning. *Pattern Recognition Letters* 140(4):81–87 DOI 10.1016/j.patrec.2020.07.015.
- Khan SS, Alharbi Y. 2024.** Sentiment analysis of movie review classifications using deep learning approaches. *International Journal of Advanced and Applied Sciences* 11(8):146–157 DOI 10.21833/ijaas.2024.08.016.

- Kingma DP, Ba JL. 2015. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR, 2015—Conference Track Proceedings*.
- Kumar A, Ekbal A, Kawahra D, Kurohashi S. 2019. Emotion helps sentiment: a multi-task model for sentiment and emotion analysis. In: *Proceedings of the International Joint Conference on Neural Networks, 2019-July* DOI 10.1109/IJCNN.2019.8852352.
- Lal M, Neduncheliam S. 2024. Enhanced V-Net approach for the emotion recognition and sentiment analysis in the healthcare data. *Multimedia Tools and Applications* 83(29):72765–72787 DOI 10.1007/s11042-024-18364-z.
- Liang C, Yang C, Xu J, Huang J, Wang Y, Dong Y. 2021. S+PAGE: a speaker and position-aware graph neural network model for emotion recognition in conversation. ArXiv DOI 10.48550/arXiv.2112.12389.
- Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, Levy O, Lewis M, Zettlemoyer L, Stoyanov V. 2019. RoBERTa: a robustly optimized BERT pretraining approach. ArXiv DOI 10.48550/arXiv.1907.11692.
- Majumder N, Poria S, Hazarika D, Mihalcea R, Gelbukh A. 2019. DialogueRNN: an attentive RNN for emotion detection in conversations. Available at www.aaii.org.
- Poria S, Hazarika D, Majumder N, Naik G, Cambria E, Mihalcea R. 2020. MELD: a multimodal multi-party dataset for emotion recognition in conversations. In: *ACL 2019—57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference* DOI 10.18653/v1/p19-1050.
- Qin X, Wu Z, Zhang T, Li Y, Luan J, Wang B, Wang L, Cui J. 2023. BERT-ERC: fine-tuning BERT is enough for emotion recognition in conversation. *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023* 37(11):13492–13500 DOI 10.1609/AAAI.V37I11.26582.
- Rashid U, Iqbal MW, Skiandar MA, Raiz MQ, Naqvi MR, Shahzad SK. 2020. Emotion detection of contextual text using deep learning. In: *4th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT, 2020—Proceedings* DOI 10.1109/ISMSIT50672.2020.9255279.
- Rehmani F, Shaheen Q, Anwar M, Faheem M, Bhatti SS. 2024. Depression detection with machine learning of structural and non-structural dual languages. *Healthcare Technology Letters* 11(4):218–226 DOI 10.1049/htl2.12088.
- Reimers N, Gurevych I. 2019. Sentence-BERT: sentence embeddings using siamese BERT-networks. In: *EMNLP-IJCNLP 2019—2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference* DOI 10.18653/v1/d19-1410.
- Salehinejad H, Sankar S, Barfett J, Colak E, Valaee S. 2017. Recent advances in recurrent neural networks. ArXiv DOI 10.48550/arXiv.1801.01078.
- Sanh V, Debut L, Chaumond J, Wolf T. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv DOI 10.48550/arXiv.1910.01108.
- Shah SZA, Abdulkader OA, Jan S, Shah MA, Anwar M. 2025. A holistic evaluation of machine learning algorithms for text-based emotion detection. *International Journal of Advanced and Applied Sciences* 12(7):55–75 DOI 10.21833/ijaas.2025.07.006.
- Shen W, Chen J, Quan X, Xie Z. 2020. DialogXL: all-in-one XLNet for multi-party conversation emotion recognition. *35th AAAI Conference on Artificial Intelligence, AAAI 2021* 15(15):13789–13797 DOI 10.1609/aaai.v35i15.17625.
- Shen W, Wu S, Yang Y, Quan X. 2021. Directed acyclic graph network for conversational emotion recognition. In: *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational*

- Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 1551–1560 DOI [10.18653/V1/2021.ACL-LONG.123](https://doi.org/10.18653/V1/2021.ACL-LONG.123).
- Sun Y, Wang S, Li Y, Feng S, Chen X, Zhang H, Tian X, Zhu D, Tian H, Wu H. 2019.** ERNIE: enhanced representation through knowledge integration. *ArXiv* DOI [10.48550/arXiv.1904.09223](https://doi.org/10.48550/arXiv.1904.09223).
- Sun Z, Yu H, Song X, Liu R, Yang Y, Zhou D. 2020.** MobileBERT: a compact task-agnostic BERT for resource-limited devices. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 2158–2170 DOI [10.18653/v1/2020.acl-main.195](https://doi.org/10.18653/v1/2020.acl-main.195).
- Tang L, Yuan P, Zhang D. 2024.** Emotional experience during human-computer interaction: a survey. *International Journal of Human-Computer Interaction* **40**(8):1845–1855 DOI [10.1080/10447318.2023.2259710](https://doi.org/10.1080/10447318.2023.2259710).
- Ullah KA, Rehman F, Anwar M, Faheem M, Riaz N. 2023.** Machine learning-based prediction of osteoporosis in postmenopausal women with clinical examined features: a quantitative clinical study. *Health Science Reports* **6**(10):e1656 DOI [10.1002/hsr2.1656](https://doi.org/10.1002/hsr2.1656).
- Wei J, Zou K. 2019.** EDA: easy data augmentation techniques for boosting performance on text classification tasks. In: *EMNLP-IJCNLP 2019—2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 6382–6388 DOI [10.18653/v1/d19-1670](https://doi.org/10.18653/v1/d19-1670).
- Yenduri G, Ramalingam M, Selvi GC, Supriya Y, Srivastava G, Maddikunta PKR, Raj GD, Jhaveri RH, Prabadevi B, Wang W, Vasilakos AV, Gadekallu TR. 2023.** Generative pre-trained transformer: a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access* **12**(1):54608–54649 DOI [10.1109/ACCESS.2024.3389497](https://doi.org/10.1109/ACCESS.2024.3389497).
- Zhang Y, Rong L, Li X, Chen R. 2022.** Multi-modal sentiment and emotion joint analysis with a deep attentive multi-task learning model. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 13185 LNCS, 518–532 DOI [10.1007/978-3-030-99736-6_35](https://doi.org/10.1007/978-3-030-99736-6_35).
- Zhang Z, Sabuncu M. 2018.** Generalized cross entropy loss for training deep neural networks with noisy labels. In: *Advances in Neural Information Processing Systems* **31**.
- Zhong P, Wang D, Miao C. 2019.** Knowledge-enriched transformer for emotion detection in textual conversations. In: *EMNLP-IJCNLP 2019—2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 165–176 DOI [10.18653/V1/D19-1016](https://doi.org/10.18653/V1/D19-1016).
- Zhu Y, Wang M, Liu S, Song C, Wang Z, Wang P, Qin X. 2019.** Sentiment analysis based on attention mechanisms and bi-directional LSTM fusion model. In: *Proceedings—2019 IEEE SmartWorld, Ubiquitous Intelligence and Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Internet of People and Smart City Innovation, SmartWorld/UIC/ATC/SCALCOM/IOP/SCI 2019* DOI [10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00177](https://doi.org/10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00177).