JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

JSCDM

Vol. 5 No. 2 (2024) 62-71
https://publisher.uthm.edu.my/ojs/index.php/jscdm

Fine-Grained Classification for Emotion Detection Using Advanced Neural Models and GoEmotions Dataset

Rohan Sharma Sitoula[1], Moumita Pramanik[1]*, Ranjit Panigrahi[1]

[1] *Department of Computer Applications*
Sikkim Manipal Institute of Technology, Sikkim Manipal University, Sikkim, 737136 INDIA

*Corresponding Author: moumita.pramanik@gmail.com
DOI: https://doi.org/10.30880/jscdm.2024.05.02.005

Article Info

Received: 1 May 2024
Accepted: 11 November 2024
Available online: 18 December 2024

Keywords

Emotion prediction, sentiment analysis, multi-labeled classification

Abstract

Emotion detection, a pivotal facet of artificial intelligence, involves deciphering and categorizing human emotions from various sources such as text, images, and audio. This process holds immense significance across industries, including mental health, customer sentiment analysis, and human-computer interaction. This abstract encompasses the essence of emotion detection, its vital role in understanding human behaviors and sentiments, and the diverse methods employed. The research explores three distinct emotion detection techniques: Bidirectional Encoder Representations from Transformers (BERT), Robustly Optimized BERT Pre-training Approach (ROBERTA), and Convolutional Neural Network (CNN). These methods are evaluated for their effectiveness in recognizing emotions, from subtle nuances to overt expressions. The results reveal ROBERTA's exceptional prowess, consistently outperforming its counterparts across various emotional categories. Particularly remarkable is its ability to predict the emotion "gratitude," with an impressive F1-score of 0.8458, underscoring its potential in capturing complex emotional states. This research emphasizes the significance of emotion detection in bridging human-computer interaction gaps and enabling a more nuanced understanding of user sentiments. The findings emphasize the prominence of ROBERTA as a powerful tool in emotion detection, offering insights into its capacity to comprehend diverse human emotions effectively.

1. Introduction

Emotions play a vital role in shaping social interactions by influencing people's reactions and nurturing the bonds of friendly relationships. With the evolution of communication, the ability to convey emotions has become simplified, relying on concise phrases and expressive emojis. The area of Natural Language Processing (NLP) has diligently collected vast volumes of data and devised sophisticated algorithms, enabling computers to classify and forecast emotions from diverse conversational inputs [1,2]. Recently, the domain of emotion prediction has witnessed remarkable strides with the help of machine learning.

Diverse machine learning algorithms have emerged to predict emotions effectively [3–6]. Among these, deep learning is a versatile tool capable of deciphering individual sentiments, organizing textual content, and interpreting images [7,8]. By scrutinizing factors like facial expressions and various datasets, machine learning algorithms have achieved impressive accuracy in anticipating an individual's emotional state. One approach involves utilizing facial recognition software and training computers to identify facial cues and infer associated emotions [9]. Alternatively, language analysis is another potent avenue for emotion prediction, achieved through

text analysis or voice modulation assessment. This understanding of emotional patterns empowers machine learning to predict emotional states accurately.

The applications of text analysis in emotion prediction extend to multiple benefits. Accurate emotional anticipation enables tailored interactions and support for individuals. Moreover, insights gained from emotional awareness can enhance marketing strategies and communication precision. In the area of machine learning, supervised and unsupervised learning emerge as fundamental categories. The supervised learning process encompasses data classification based on specific criteria, often crucial for automating business processes and deriving insights from textual data. This method holds particular significance in actions like language and emotion recognition. The classification can manifest in binary or multi-label forms, serving various purposes.

This article focuses on the classification and prediction of 28 distinct emotions: admiration, amusement, anger, annoyance, approval, caring, confusion, curiosity, desire, disappointment, disapproval, disgust, embarrassment, excitement, fear, gratitude, grief, joy, love, nervousness, optimism, pride, realization, relief, remorse, sadness, surprise, and neutral. These emotions are categorized into three groups: positive, negative, and ambiguous. This involves a multi-labeled approach, dissecting sentences into distinct clauses for comprehensive evaluation. Multi-label text categorization entails assigning multiple categories to a single text, contrasting with single-label classification. This proves invaluable when multiple labels accurately describe a text, as observed in sentiment analysis and subject categorization. Two primary techniques govern multi-label text categorization: problem transformation and algorithm adaptation. Problem transformation divides the multi-label challenge into single-label problems, while algorithm adaptation tailors existing algorithms to address multi-label complexities. This approach circumvents binary limitations, yielding more precise conclusions by accommodating multiple groups. In this pursuit, breaking down statements into discrete categories such as negative, positive, neutral, and ambiguous holds immense significance. This facilitates comparisons like "Intel surpasses AMD" or assessment of dual-polarity statements such as "The mess's cuisine was terrible, yet accommodations were remarkably pleasant." Our research draws upon the GoEmotions dataset, a compilation of Reddit comments annotated across 28 distinct emotional categories, including neutral. These span 12 positive, 11 negative, and 4 ambiguous emotions, alongside a neutral category. The spectrum spans from admiration to surprise, encapsulating many emotional nuances. The dataset also includes 9 features: text, id, author, subreddit, link_id, parent_id, created_utc, rater_id, and example_very_unclear. These features are well-suited for predicting the 28 emotion classes.

Early algorithms in emotion categorization could merely discern binary sentiments of positive or negative, revealing a notable limitation. Subsequent progress expanded the emotional spectrum to six categories: anger, sadness, disgust, joy, fear, and surprise. Yet, the domain of emotions is far more diverse. Our research aims to predict 28 distinct emotional descriptors, encompassing awe, interest, pride, regret, compassion, approval, and beyond.

2. Literature Survey

In emotion analysis, researchers have orchestrated diverse methodologies to unravel the complex tapestry of human sentiments. One notable endeavor by Dumont et al. [10] resulted in creating an interactive online platform, giving users a window into the emotional landscapes associated with a remarkable 483 subreddits. Drawing upon a vast compilation of over 58,000 meticulously classified Reddit posts, each sorted into 28 distinct emotions utilizing the GoEmotions dataset, this innovative platform afforded users insights into the emotional currents coursing through diverse subreddit communities. This engaging interface allowed readers to navigate the amalgamation of feelings, shedding light on intersections among different subreddit groups. Yet, as they probed deeper, they confronted the challenge of categorizing each post's sentiment into a binary of positivity or negativity, acknowledging the complexities inherent in human expression.

Concurrently, Singh et al. [11] introduced a novel multi-task approach to sentiment modeling. Within this framework, emotion definition modeling was treated as an auxiliary task, while the primary focus lay in training for emotion prediction. Leveraging the capabilities of BERT, a potent language model, they sought to encapsulate the essence of varied emotion classes through their definitions. In a series of meticulous experiments, they explored different definition modeling settings, ultimately championing the BERT+CDP+MLM approach as the most robust. This approach navigated the subjective nature of emotions as some predictions veered from reality. This inquiry also extended to strategies for addressing the complex degrees of ambiguity and the objectivity entwined within emotion categorization.

Meanwhile, Huang et al. [12] charted new territory with their Seq2Emo technique, introducing emotion correlations into a bi-directional decoder. Emotion encoding was achieved through a two-layer Long Short-Term Memory (LSTM), enhanced by token-level and contextually pre-trained embeddings to represent words within phrases. This comprehensive approach, encompassing both Nested LSTM and LSTM, yielded exceptional results, culminating in a staggering accuracy score of 99.167% across seven emotion categories. These achievements shone particularly bright within the landscape of multi-labeled datasets.

J. of Soft Computing and Data Mining Vol. 5 No. 2 (2024) p. 62-71

Baruah et al. [13] embarked on a distinct exploration of the complex terrain of few-shot emotion recognition. Armed with insights from the GoEmotions Reddit dataset, they ventured to extend their knowledge to the SemEval tweets corpus, employing various emotion encoding techniques. Unveiling the nuances of supervised and unsupervised methods to encapsulate emotions at the sentence level, this analysis explored the knowledge transfer between diverse datasets, even across dissimilar label taxonomies. While insights were gleaned into knowledge application, the research did not traverse the avenues of augmenting emotion representation beyond rudimentary definitions. Further enriching the discourse, Suresh et al. [13] introduced the concept of Knowledge-Embedded Attention (KEA), harnessing emotion lexicons to enhance contextual interpretations derived from previously trained models. Their innovative model architecture established hidden regions dedicated to contextual representation and emotional encoding. External emotional lexicons were woven into the model's fabric, fostering deeper insights and understanding across diverse datasets. This approach amplified performance and addressed the intricacies of closely related emotions, showcasing the potential for enriched emotion recognition through carefully crafting model topologies to mirror the fine-grained nature of the emotions under scrutiny.

In a distinct research, Suresh et al. [14] undertook an innovative approach to enhance model discernment among multiple negative examples. The model could fathom the similarity and dissimilarity of class pairings by integrating inter-class interactions into a Label-aware Contrastive Loss (LCL). Their dual-model strategy, intertwining label connections within the main embedding model's contrastive objective, exhibited promising sentiment analysis and recognition results. Vanmassenhove et al.[15] introduced an ingenious technique that transmutes sentence sentiments into synthetic voices, forming an emotional text-to-speech system. Their method involved ascertaining sentence polarity through the SentiWordsTweet tool and utilizing the NRC Emotion Lexicon for sentence categorization. The outcomes showcased the system's aptitude to predict emotional intensity within texts and align it with human annotations. Akhtar et al. [16] proposed a complex framework for deep multi-task learning amalgamating sentiment and emotion analyses. Their multi-task model extracted sentiments (positive or negative) and emotions (anger, contempt, fear, joy, sadness, surprise) from video speakers, leveraging the interdependencies between these tasks to enhance predictive confidence. The conclusions of this research indicated that sentiments and emotions were mutually fortified within the multi-task framework, surpassing existing systems in various settings.

Chiorrini et al. [17] harnessed the power of bidirectional encoder representations from transformers (BERT) models to scrutinize Twitter data for sentiments and emotions. Augmenting the model with BERT-Base architecture and a final classification phase, they achieved remarkable accuracy rates of 92% for sentiment classification and 90% for emotion analysis on tweet datasets. Shenoy et al. [18] revolutionized sentiment analysis and emotion recognition by devising an RNN architecture emphasizing a multi-modal approach within conversations. In contrast to prevalent models, they addressed each modality independently while upholding contextual continuity, consistently outperforming rivals across benchmark datasets. Adoma et al. [19] embarked on research investigating transformer models' efficacy (ROBERTA, BERT, DistilBERT, XLNet) in extracting emotions from texts, yielding ROBERTA as the highest-performing model, followed by XLNet, BERT, and DistilBERT.

Tenney et al. [20] examined the depths of the BERT network's capabilities in comprehending syntactic and semantic structures within text. Employing scalar mixing weights and cumulative scoring, their exploration unveiled nuanced hierarchical information processing, revealing deep language models' prowess in complex language processing tasks. Alotaibi et al. [21] endorsed the utilization of the ISEAR emotion dataset and Logistic Regression (LR) for supervised machine learning, producing a successful classifier that achieved positive performance evaluations and showcased the proposed strategy's potential. Polignano et al. [22] proposed an effective classification method employing deep neural networks, Bi-LSTM, CNN, and self-attention, demonstrating their model's prowess on multiple datasets. Wang et al. [23] devised a neural architecture that enriches word embeddings with sentiment supervision at both document and word levels, producing a superior sentiment lexicon. Liu et al. [24] introduced Multi-Tasking Deep Neural Network (MT-DNN), a model that combined multi-task learning and language model pre-training. MT-DNN showcased state-of-the-art performances across various NLU tasks.

3. Materials & Methods

The evaluation process begins by employing the GoEmotions dataset [25], which includes a wide range of 28 distinct emotional attributes. The dataset undergoes rigorous preprocessing, which involves converting emojis into their appropriate textual representations. Furthermore, language that includes several contractions is converted into expanded forms to improve clarity. Regular expressions are utilized to correct spelling errors and address discrepancies in acronyms to achieve linguistic precision. After completing the data preparation phase, the subsequent step involves tokenization. During this stage, a systematic procedure is employed to tokenize each word within every data row. This tokenization process ensures that the neural network receives coherent inputs

that can be efficiently processed. Following this, the data is subjected to embedding, which is a crucial process that involves the representation of words. The methodology above embodies the concept that words possessing comparable meanings are represented as vectors in a common vector space positioned near one another to indicate their semantic similarity.

To assess the effectiveness of the emotion detection procedure, three specific models are utilized: Convolutional Neural Network (CNN), Bidirectional Encoder Representations from Transformers (BERT) [26], and Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach (ROBERTA) [27]. The models have been trained on data labeled with 28 different emotional classes, allowing for an analysis of how well each performs at recognizing emotions. Careful data preparation, comprehensive tokenization, embedding for complex word representation, and subsequent model deployment are all part of the scientific process for determining which emotion detection method is most effective. In the following section, the methodologies are explained in detail.

3.1 Dataset Used

GoEmotions [25] is a corpus of 58k Reddit comments manually analyzed by humans and classified into 28 different emotion categories. The following list of emotions is organized into different categories: amusement, caring, love relief, nervousness, excitement, curiosity, surprise approval, gratitude, fear, disapproval, grief, joy, pride, disappointment, anger, remorse, sadness, realization, desire, optimism, embarrassment, disgust, confusion, admiration, and annoyance. The attributes found in the GoEmotions dataset have been presented in Table 1.

Table 1 *Attributes of GoEmotions dataset*

Attributes	Meaning
text	The reddit comments.
id	The comment's special identifier.
author	The commentator's Reddit username.
subreddit	The subreddit to which the comment is associated to.
link_id	The comment's link_id.
parent_id	The comment's parent id.
created_utc	The comment's timestamp.
rater_id	The annotator's distinctive ID.
example_very_unclear	Whether the annotator indicated that the comment was difficult to label or was very unclear.
Class labels	With binary labels for the emotion categories (0 or 1) in one hot encoded form.

All annotations and comment metadata are included in three different CSV files that make up the dataset. We were able to correlate 211225 Reddit comments with their corresponding emotions (anger, realization, relief, disappointment, pride, nervousness, caring, grief, gratitude, desire, excitement, surprise joy, annoyance, optimism, admiration, remorse, curiosity, embarrassment, disgust, disapproval, approval, amusement, fear, confusion, sadness, love) in one hot encoded form.

3.2 Data Preprocessing

The following procedures were part of the data processing that we performed: -

3.2.1 Converting the Emojis to the Corresponding Textual Form

Many comments in the dataset contained emojis, which were neglected in the past research. One emoji is enough to explain a line of text. These emojis are crucial in emotion prediction and need to be handled carefully. To translate emojis into their related textual data, the *demojize*() function from the Python "emoji" library was used. For eg: 👍 was converted to *thumbs_up* & ❤️ was converted to *red_heart*.

Function used: Demojize()
Library: python emoji library.
Function Description: The demojize function determines the short name of an emoji. The short name of the emoji will be returned once the emoji is sent as an argument to the demojize () function.

J. of Soft Computing and Data Mining Vol. 5 No. 2 (2024) p. 62-71

3.2.2 Expanding the Contractions

Words or word combinations may be abbreviated by omitting letters and adding an apostrophe, is known as a contraction. (eg. I will becomes I'll). These days, we rely on abbreviations and shortened form of words for social interactions.

The dataset contained multiple contractions like "$We've\ been\ waiting\ for\ this\ day\ for\ so\ long$". Which was converted into natural form like "$We\ have\ been\ waiting\ for\ this\ day\ for\ so\ long$" using the $contraction.fix()$ method from contractions library.

Function used: $contraction.fix()$
Library: contractions
Function Description: The $contraction.fix()$ library takes shortened texts as argument and returns the normal form of the text.

3.2.3 Fixing Various Acronyms Errors and Misspellings

The acronyms errors and misspellings were fixed using the regular expressions. Some of the transformations are shown in Table 2.

Table 2 *Before and after fixing acronyms and spelling errors*

Before Fix	After Fix
Cuz,coz	because
Ikr	I know right
Faux pas	mistake

Along with these steps we also lowercase the texts and replaced some words with multiple occurrences of a letter, example "$coooool$" turned into $-> cool$. After the pre-processing phase, clean text for the corresponding text column was obtained. The emojis were converted to textual form, contractions were expanded, acronym errors and misspellings were fixed, the texts were lowercase, and some words with multiple occurrences of letters were replaced.

3.2.4 Tokenization

We tokenized each word in every row of the data to give the neural network a legitimate input. As a result, we obtained mapping to a huge lexicon of Token: Word. e.g., a sentence "$I\ like\ riding$ " was tokenized to 1:"i", 2: "like", 3: "riding".
Function used: $Tokenizer()$
Library: Keras.
Function Description: By converting each text into either a sequence of integers (each integer representing the index of a token in a dictionary) or a vector with a binary coefficient for each token based on word count or term frequency-inverse document frequency, this function enables the vectorization of a corpus of text.

3.2.5 Embedding

Word embeddings, in their simplest form, are a word representation that links a computer's understanding of language to that of a person. This suggests that two equivalent words are represented in a vector space by nearly identical vectors close to one another. These are critical for most natural language processing issues. In this case, vectors near one another indicate identical words, such as king-queen and man-woman. Stanford's Glove library was used for embedding. The co-occurrence of each word with other words in the corpus is determined using this procedure, which is repeated until a co-occurrence matrix is created. When two words are adjacent, they are assigned a value of 1 if separated by one word, a value of 1/2, two words, a value of 1/3, and so on. For instance, for the corpus "It was the best evening., Good Evening! Was it the best evening?", the co-occurrence matrix generated is presented in Table 3.

Table 3 *The co-occurrence matrix for the corpus "It was the best evening., Good Evening! Was it the best evening?"*

	it	was	the	best	evening	good
it	0					
was	1+1	0				
the	1/2+1	1+1/2	0			
best	1/3+1/2	1/2+1/3	1+1	0		
evening	1/4+1/3	1/3+1/4	1/2+1/2	1+1	0	
good	0	0	0	0	1	0

As a result, instead of an input array mapping certain sentences to text, a massive matrix which mapped each word to dimensional values is obtained. So, rather than just a list of words, a matrix of meanings for each word in the corpus is obtained.

3.2.6 Model Preparation

Model 1: Convolutional Neural Networks

Convolution Neural Networks (CNNs) are interartificial networks that can recognize complex properties in data, such as identifying features in image and text data. In computer vision software, CNNs are mostly used in picture segmentation, object recognition, and classification tasks. However, CNNs have recently been used to address text-related difficulties. The CNN implementation was achieved using the following steps:

Step 1 To receive the embedding matrix, add an embedding layer.

Step 2 Add a convolutional layer, which will train the model to iteratively search through 256 possible "meanings" of the text.

Step 3 To prevent premature convergence in the model, include a dropout layer.

Step 4 Set the learning rate of 0.0002 and epoch of 12.

Step 5 Combine the model with an Adam optimizer and binary cross-entropy loss function to train the neural network more quickly and effectively.

Model 2: BERT Classifier

A machine learning method called BERT (Bidirectional Encoder Representations from Transformers) is built on transformers and attentional components that can identify the connections between words in context. We employed the PRE-Trained Bert Model in our research. The Bert implementation was achieved using the following steps:

Step 1 Use the Bert tokenizer to turn natural language sentences into a vector representation.

Step 2 As suggested in GoEmotions' official documentation, using BERT, tune the model with a learning rate of 0.0006 and an epoch of 10.

Step 3 Feed the vector representations into the model for a classification task.

Model 3: ROBERTA

ROBERTA is an advancement of BERT that modifies the pretraining process. The changes include giving the model a longer training period, larger batch sizes, and more data. eliminating the goal of the following phrase prediction. exercise with longer sequences. To implement ROBERTA, we utilized the hugging face library transformers. The ROBERTA implementation was achieved using the following steps.

Step 1 Use the ROBERTA-base tokenizer to tokenize the text into vectors.

Step 2 Design a trainer helper class to facilitate the finetuning of models using the Transformers library.

Step 3 Define all the hyperparameters like epoch=10, learning rate= 5e-5.

Step 4 Combine the model with the Adam optimizer.

Step 5 Combine the model with the Adam optimizer.

J. of Soft Computing and Data Mining Vol. 5 No. 2 (2024) p. 62-71

4. Results & Discussion

The results of the text classifiers CNN, BERT, and ROBERTA were obtained in terms of F1-score. The F1 score is one of the most important assessment metrics in machine learning. It succinctly distills a model's predictive power by integrating precision and recall, two measurements that often compete. The F1-score may be quite significant when data is out of balance, such as when there are many more items in one class than the other, as in our dataset. The F1-Score was determined from the precision and recall of the models, where precision was used to determine how many predictions for the favorable class fell into the positive class. On the other hand, Recall quantifies the number of accurate class predictions generated from the whole dataset's successful examples.

The harmonized means of recall and precision is known as the F1 Score. It considers FPs and FNs and uses the following algorithm to combine precision and recall into a single number: 2*(Recall * Precision) / (Recall + Precision).

After classifying the dataset using the CNN model and BERT and ROBERTA, we obtained the results outlined in Table 4.

Table 4 *Classification results were obtained by CNN, BERT, and ROBERTA for 28 emotions*

Emotions	F1 Score			Emotions	F1 Score		
	BERT	CNN	ROBERTA		BERT	CNN	ROBERTA
admiration	0.5529	0.5318	0.6274	fear	0.4562	0.4194	0.5965
amusement	0.6787	0.5759	0.7815	gratitude	0.7191	0.7655	0.8458
anger	0.3727	0.3471	0.4428	grief	0.2845	0.2222	0.3123
annoyance	0.2578	0.2834	0.2286	joy	0.4673	0.3331	0.5283
approval	0.4039	0.2363	0.3159	love	0.5416	0.6168	0.4537
caring	0.0216	0.3562	0.2496	nervousness	0.1388	0.1119	0.1561
confusion	0.2781	0.3261	0.2913	optimism	0.3864	0.3699	0.2537
curiosity	0.3555	0.4171	0.4817	pride	0.4896	0.1348	0.5834
desire	0.3289	0.3554	0.4156	realization	0.3751	0.1608	0.4233
disappointment	0.1311	0.1525	0.1833	relief	0.2512	0.1692	0.3281
disapproval	0.2437	0.2773	0.2673	remorse	0.5026	0.4599	0.6124
disgust	0.3965	0.2963	0.4571	sadness	0.4882	0.3568	0.5198
embarrassment	0.2587	0.3057	0.2089	surprise	0.4207	0.4425	0.3655
excitement	0.3275	0.2876	0.3511	neutral	0.5067	0.5281	0.5141

The tabulated F1 scores across a myriad of distinct emotions, each associated with varying methodologies—BERT, CNN, and Roberta—illuminate the nuanced performance nuances of each approach. For every emotion, like "admiration," a specific F1-Score is assigned to each method. For instance, considering "admiration," we observe F1 scores of 0.5529, 0.5318, and 0.6274 for BERT, CNN, and Roberta, respectively. These scores encapsulate the models' dexterity in achieving precise positive predictions (precision) while comprehensively encompassing the universe of actual positives (recall) for the given emotion category. This evaluation pattern extends to all other emotions, wherein each method exhibits a unique competency profile. To examine specifics, emotions such as "amusement" yield F1-Scores of 0.6787, 0.5759, and 0.7815 for BERT, CNN, and Roberta, respectively. Conversely, sentiments like "anger" correspond to scores of 0.3727, 0.3471, and 0.4428, respectively. Analogous scrutiny across all emotions manifests in distinct arrays of F1 scores for each method, unraveling the complexity of emotional classification.

Overall, all three text classifiers—BERT, CNN, and ROBERTA—reveal exceptional results for the emotion "gratitude," boasting F1 scores of 0.7191, 0.7655, and 0.8458, respectively. Among these, ROBERTA stands out with an impressive 84.58% accuracy in classifying this emotion. Conversely, CNN exhibits a notable deficiency in detecting the emotion "nervousness," garnering low F1-Scores of 0.1119 and 0.1561, while BERT struggles to accurately detect the emotion "caring" with an F1-Score as low as 0.0216. This complex interplay of F1 scores across various emotions is a comprehensive evaluation framework, ultimately underscoring Roberta's proficiency as the exemplar classifier.

5. Comparative Analysis

In the domain of fine-grained emotion classification, recent advances have showcased varying levels of effectiveness across different methodologies, as reflected in their F1-Scores. Table 5 details the F1-Score of the recent methods and the F1-Score received in this work. Singh et al. [11] explored multiple configurations of BERT, incorporating techniques such as Class Definition Prediction (CDP) and Masked Language Modeling (MLM). Their approach yielded F1-Scores of 51.96% for BERT+CDP+MLM, 51.25% for BERT+MLM, and 52.34% for BERT+CDP. Among these, BERT+CDP demonstrated the highest performance, suggesting that CDP may provide a slight edge in emotion classification tasks by better contextualizing data. Huang et al. [12] presented Seq2Emo, a method that achieved an F1-Score of 59.57%, surpassing the standalone BERT model, which had an F1-Score of 58.49%. This improvement highlights the potential of Seq2Emo in capturing emotional nuances more effectively than BERT alone. In contrast, Baruah et al. [13] utilized a combination of WordNet and Sentence-BERT (SBERT FT) to reach an F1-Score of 49.00%, and Suresh et al. [13] applied KEA-ELECTRA to achieve 49.60%. These results suggest that while these methods are robust, they fall short compared to newer approaches. Notably, Suresh et al. [14] achieved a significant improvement with ELECTRA+LCL, reaching an F1-Score of 64.80%, underscoring the efficacy of integrating Label-aware Contrastive Loss (LCL) with ELECTRA in capturing fine-grained emotions. However, the most remarkable performance is demonstrated by the current work using ROBERTA, which attained an impressive F1-Score of 84.58%. This substantial leap in performance highlights ROBERTA's advanced capabilities in handling the complexities of emotion classification, making it a superior choice compared to the other methods reviewed.

Table 5 *Comparative analysis of this work with other related works*

Author et al.	Method	F1-Score
Singh et al. [11]	BERT+CDP+MLM	51.96
Singh et al. [11]	BERT+MLM	51.25
Singh et al. [11]	BERT+CDP	52.34
Huang et al. [12]	Seq2Emo	59.57
Huang et al. [12]	BERT	58.49
Baruah et al. [13]	WordNet + SBERT FT	49.00
Suresh et al. [13]	KEA-ELECTRA	49.60
Suresh et al. [14]	ELECTRA+LCL	64.80
This work	ROBERTA (Max)	84.58

6. Conclusion

Emotion prediction, a complex endeavor involving the training of computers to anticipate emotions from varied inputs like texts, facial expressions, and audio cues, marks a significant stride in artificial intelligence. Traditionally, the scope of such studies was confined to binary emotions: positive and negative. Our undertaking sought to transcend these limitations by venturing into predicting various emotions from textual inputs. In pursuit of this, we harnessed the GoEmotions dataset, which boasts an extensive collection of 28 distinct emotional categories, thus offering a comprehensive platform for the challenge of multi-label classification. Within this endeavor, three distinct methodologies emerged as our tools of choice: ROBERTA, BERT, and CNN. A compelling conclusion emerged through rigorous training and analysis of these three models where the ROBERTA exhibited a slight edge over its counterparts. While emotions like amusement, gratitude, and love presented a relatively higher degree of predictability, the subtler emotional nuances encompassing feelings like disappointment, realization, and relief posed more complex challenges. Nevertheless, it's essential to note that the course of analysis brought to light a significant aspect, i.e., approximately 30% of mislabeled data within the dataset. This mislabeling phenomenon undeniably contributed to lower F1 scores across numerous emotion labels, thus introducing complexity and a potential avenue for refining future model performances. The journey through this comprehensive research highlighted ROBERTA's prowess in emotion prediction. It illuminated the imperative need for meticulous data quality assurance as a pivotal facet of such computational undertakings.

Several methods can improve emotion detection model further. First, improving annotations and active learning to reduce 30% mislabeled data can increase dataset quality. Synonym replacement and paraphrasing can help improve model robustness. On top of these, fine-tuning hyperparameters and trying ensemble approaches with CNN, BERT, and ROBERTA predictions may improve performance. Better preprocessing for emojis and idiomatic expressions and model validation on external datasets would improve emotion recognition accuracy and generalizability.

Acknowledgement

The authors declare that no funds, grants, or other support was received during the preparation of this manuscript.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Author Contribution

All the authors have designed the research, developed the methodology, performed the analysis, and written the manuscript. All authors have contributed equally.

References

[1] Koleck, T.A., Dreisbach, C., Bourne, P.E. and Bakken, S. (2019) Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review. Journal of the American Medical Informatics Association, 26, 364–79. https://doi.org/10.1093/jamia/ocy173

[2] Alslaity, A. and Orji, R. (2024) Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. Behaviour & Information Technology, 43, 139–64. https://doi.org/10.1080/0144929X.2022.2156387

[3] Ameer, I., Bölücü, N., Siddiqui, M.H.F., Can, B., Sidorov, G. and Gelbukh, A. (2023) Multi-label emotion classification in texts using transfer learning. Expert Systems with Applications, 213, 118534. https://doi.org/10.1016/j.eswa.2022.118534

[4] Kumari, N., Anwar, S. and Bhattacharjee, V. (2023) A Comparative Analysis of Machine and Deep Learning Techniques for EEG Evoked Emotion Classification. Wireless Personal Communications, 128, 2869–90. https://doi.org/10.1007/s11277-022-10076-7

[5] Ayetiran, E.F. (2022) Attention-based aspect sentiment classification using enhanced learning through cnn-Bilstm networks. Knowledge-Based Systems, 252, 109409. https://doi.org/10.1016/j.knosys.2022.109409

[6] Zhang, J. and Guo, Y. (2024) Multilevel depression status detection based on fine-grained prompt learning. Pattern Recognition Letters, 178, 167–73. https://doi.org/10.1016/j.patrec.2024.01.005

[7] Zhang, L., Wang, S. and Liu, B. (2018) Deep learning for sentiment analysis: A survey. WIREs Data Mining and Knowledge Discovery, 8. https://doi.org/10.1002/widm.1253

[8] Dang, N.C., Moreno-García, M.N. and De la Prieta, F. (2020) Sentiment Analysis Based on Deep Learning: A Comparative Study. Electronics, 9, 483. https://doi.org/10.3390/electronics9030483

[9] Bahreini, K., van der Vegt, W. and Westera, W. (2019) A fuzzy logic approach to reliable real-time recognition of facial emotions. Multimedia Tools and Applications, 78, 18943–66. https://doi.org/10.1007/s11042-019-7250-z

[10] Dumont, F. and Facen, T. Visualing fine-grained emotions in Reddit posts through the GoEmotions dataset.

[11] Singh, G., Brahma, D., Rai, P. and Modi, A. (2021) Fine-Grained Emotion Prediction by Modeling Emotion Definitions. 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), p. 1–8.

[12] Huang, C., Trabelsi, A., Qin, X., Farruque, N., Mou, L. and Zaiane, O.R. (2021) Seq2Emo: A sequence to multi-label emotion classification model. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, p. 4717–24.

[13] Suresh, V. and Ong, D.C. (2021) Using knowledge-embedded attention to augment pre-trained language models for fine-grained emotion recognition. 2021 9th International Conference on Affective Computing and Intelligent Interaction (ACII), p. 1–8.

[14] Suresh, V. and Ong, D.C. (2021) Not all negatives are equal: Label-aware contrastive loss for fine-grained text classification. ArXiv Preprint ArXiv:210905427,.

[15] Vanmassenhove, E., Cabral, J.P. and Haider, F. (2016) Prediction of Emotions from Text using Sentiment Analysis for Expressive Speech Synthesis. SSW, p. 21–6.

[16] Akhtar, M.S., Chauhan, D.S., Ghosal, D., Poria, S., Ekbal, A. and Bhattacharyya, P. (2019) Multi-task learning for multi-modal emotion recognition and sentiment analysis. ArXiv Preprint ArXiv:190505812,.

[17] Chiorrini, A., Diamantini, C., Mircoli, A. and Potena, D. (2021) Emotion and sentiment analysis of tweets using BERT. EDBT/ICDT Workshops,.

[18] Shenoy, A. and Sardana, A. (2020) Multilogue-net: A context aware rnn for multi-modal emotion detection and sentiment analysis in conversation. ArXiv Preprint ArXiv:200208267,.

[19] Adoma, A.F., Henry, N.-M. and Chen, W. (2020) Comparative analyses of bert, roberta, distilbert, and xlnet for text-based emotion recognition. 2020 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), p. 117–21.
[20] Tenney, I., Das, D. and Pavlick, E. (2019) BERT rediscovers the classical NLP pipeline. ArXiv Preprint ArXiv:190505950.
[21] Alotaibi, F.M. (2019) Classifying text-based emotions using logistic regression.
[22] Polignano, M., Basile, P., de Gemmis, M. and Semeraro, G. (2019) A comparison of word-embeddings in emotion detection from text using bilstm, cnn and self-attention. Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, p. 63–8.
[23] Wang, L. and Xia, R. (2017) Sentiment lexicon construction with representation learning based on hierarchical sentiment supervision. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, p. 502–10.
[24] Liu, X., He, P., Chen, W. and Gao, J. (2019) Multi-task deep neural networks for natural language understanding. ArXiv Preprint ArXiv:190111504,.
[25] Alon, D. and Ko, E. (2021) GoEmotions: A Dataset for Fine-Grained Emotion Classification. Google Research.
[26] Alaparthi, S. and Mishra, M. (2020) Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. ArXiv Preprint ArXiv:200701127,.
[27] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D. et al. (2019) Roberta: A robustly optimized bert pretraining approach. ArXiv Preprint ArXiv:190711692.