A Fine-Tuned Distilled Zero-Shot Student Model for Emotion Detection in Academic-Related Responses

Mary Joy P. Canon[1,*], Lany L. Maceda[1], Thelma D. Palaoag[2], and Mideth B. Abisado[3]

[1]Computer Science and Information Department, Bicol University, Legazpi City, Philippines
[2]College of Information Technology and Computer Science, University of the Cordilleras, Baguio City, Philippines
[3]College of Computing and Information Technologies, National University, Manila, Philippines
Email: mjpcanon@bicol-u.edu.ph (M.J.P.C.); llmaceda@bicol-u.edu.ph (L.L.M.); tdpalaoag@uc-bcf.edu.ph (T.D.P.); mbabisado@national-u.edu.ph (M.B.A.)
[*]Corresponding author

Manuscript received January 2, 2024; revised February 14, 2024; accepted March 6, 2024; published July 23, 2024

***Abstract*—Exploring emotions in academic settings, particularly in feedback on educational programs, is critical for understanding student experiences and improving educational policies. While the fine-tuning of pre-trained models has consistently delivered state-of-the-art results in emotion detection tasks, the potential of zero-shot learned model in this area remains largely unexplored. This paper presents a novel approach to emotion detection by fine-tuning a distilled zero-shot student model for classifying emotions in text, specifically focusing on feedback from beneficiaries of a free education program in the Philippines. Basic data cleaning and tokenization were performed, while retaining the stopwords in the corpus. Stopwords, in this work, contributes in understanding emotional expressions within academic-related texts. Our experiments highlight the superior performance of the distilled zero-shot student model achieving 84.21% accuracy and 84.23% F1-Score, notably outperforming the EmoRoberta model. The model exhibits excellent predictive ability in identifying emotions like desire, gratitude, and neutral, but encounters confusion in classifying optimism and relief. We deployed the model for automatic emotion labeling of feedback texts. Analysis revealed a predominantly positive reception on the program among its beneficiaries, with feelings of relief, approval, and gratitude being the most prominent. However, the presence of neutral and disappointment also highlights areas where the program needs improvement. These insights can be valuable for policymakers to understand the program impact of and to make data-driven decisions for its improvement and targeted interventions.**

***Keywords*—text-based emotion detection, transfer learning, natural language processing, zero-shot model, free higher education feedback**

I. Introduction

The ability to understand and interpret emotions from text has become one of the developing subjects in the field of Natural Language Processing (NLP). This discipline, known as emotion detection, examines a writer's input to infer the underlying emotions that are affecting them. It operates on the premise that a person's emotional state influences their choice of words. Happiness, for instance, is often manifested using positive words, while feelings of sadness, frustration, or anger, tend to be reflected in more negatively connotated expressions. This connection helps us better understand the emotions hidden in what people express in text.

Emotion detection can be considered as a logical development of sentiment analysis as it predicts the unique emotion present in the text rather than simply identifying positive, negative, or neutral sentiments. Through time, emotion detection models have been developed to effectively discern the emotions conveyed in a human text or speech. These models significantly improved the Chatbot's communication with a human [1], providing more empathetic and effective responses. Furthermore, these models have proven to be valuable tools in digital marketing [2] by assessing customer reviews [3] and maximizing customer satisfaction [4]. In the education sector, these models demonstrated to be transformative, improving the quality of decision-making process [5] and enriching teaching-learning experience by determining the emotional content of texts.

Various computational methods to detect and analyze emotions of text have been proposed and have evolved over time–from lexicon-based approaches [6], machine learning techniques [7], deep learning [8] and neural networks [9] to hybrid models [10]. Recent prominent success in emotion classification and NLP in general has been brought about by transfer learning or through the development of pre-trained models for language understanding. Transfer learning allows the transfer of knowledge acquired from one task or domain to another related task or domain.

Bidirectional Encoder Representations from Transformers (BERT) [11] and its successors are some of the transfer learning models used to improve several domain-specific tasks, effectively capturing knowledge from massive labeled and unlabeled data. In the context of emotion detection, a model initially trained on a vast corpus of text data, such as a language modeling task, can learn valuable linguistic features and contextual understanding to be adopted to the target task of identifying emotions from data. Emotion detection models based on BERT architecture have demonstrated state-of-the-art results in improving various target tasks. For instance, these transformer-based models have demonstrated effectiveness in identifying emotions in conversational utterances [12, 13] and Twitter discussions [14], classifying emotions as joyful, sad, furious, or others. By fine-tuning these models to domain-specific dataset, they were able to generate contextual emotion classifiers.

Another significant advancement in the capabilities of NLP is the zero-shot learning which can be viewed as an extension [15] of the concept of transfer learning. This approach showcases a more advanced level of knowledge transfer and generalization. Particularly, zero-shot classification involves predicting a class that the model has not encountered during training. In this paradigm, the model is provided with a prompt and a sequence of text that

describes what task it is intended to do [16], typically by associating attributes with classes. Zero-shot models use no supervised examples during the transfer protocol, which more to mean, a model is trained to perform tasks for which it has not been explicitly trained. Learning herein relies on the model's ability to generalize from its training on one set of tasks or categories to entirely new tasks or categories without any direct examples. In the context of emotion detection, this approach is relevant for it allows the identification of emotions that may not have been explicitly covered during training of the model. The use of zero-shot models offers a richer semantic meaning of words–an important factor in identifying emotional expressions within text.

While the fine-tuning of pre-trained models, a standard architecture in transfer learning, has consistently delivered state-of-the-art results in emotion detection tasks, the potential of zero-shot learned model in this area remains largely unexplored. This paper is an attempt to build a novel approach to emotion detection by fine-tuning a zero-shot learned student model to a domain-specific task of classifying emotions present in the responses related to the implementation of the Universal Access to Quality Education (UAQTE) program in the Philippines. The UAQTE program, aimed at providing quality tertiary education across the country, implements programs which provide: (a). free tuition and school fees to state universities and colleges, and (b). education subsidy to private higher education institutions. The program's enactment is part of the government's effort to promote inclusivity and remove financial barriers to college education, thereby expanding career opportunities and advancing national development. Despite the program's years of implementation, there remains lack of comprehensive understanding on the recipient's feedback towards the program. Aiding the government's current efforts in evaluating the program's implementation, this research was undertaken to understand the experiences and sentiments of the stakeholders, and to provide insights for a more student-centric approach in policy enhancement.

Through fine-tuning experiments, we compared the performance of EmoRoberta model and distilled zero-shot student model in the target task of determining emotions present in academic responses. Ultimately, the objective is to deploy the best emotion classifier to automatically label scholars' feedback responses and systematically analyze their emotional experiences in the implementation of the UAQTE program. Result of the research does not only contribute to the growing body of educational research, but more importantly, it offers practical value by empowering program administrators to make informed decisions and provide targeted interventions for the stakeholders. Actionable insights generated can resonate with policymakers at regional and national level, prompting broader efforts to make education more accessible in the Philippines.

II. Literature Review

A. Emotion Taxonomy and Text Corpora

Different emotion taxonomies and approaches represent the foundational framework for understanding and categorizing human emotions in NLP and related fields. Initially, emotions were often categorized into 6 universal and basic classes: joy, sadness, anger, fear, disgust, and surprise [17]. Independently, Plutchik and Conte [18] introduced a similar taxonomy that added the emotions of anticipation and trust, mapped in the devised emotion wheel. Ortony *et al.* [19] presented an Ortony, Clore, and Collins (OCC) model with 22 categories where some emotions are created by combining two or three types of emotions, while Cambria *et al.* [20] provided 20 categories claiming that this model explains the entire emotional experience that happens to everyone.

To build text-based emotion detection systems, text corpus annotated with emotion labels is necessary. Text data capturing various context and domains enable the development and benchmarking of state-of-the-art models, which is critical for advancing research in emotion analysis. To facilitate research in affective text analysis, a dataset was introduced [21], annotated using Ekman's emotion anatomy. Based on self-reported emotional experience, International Survey on Emotion Antecedents and Reactions (ISEAR) dataset was also used in psychology studies and sentiment analysis tasks.

Collected from various online platforms, EmoReact [22] is a corpus designed for emotion detection in conversations, which focuses on dyadic(two-person) conversations and includes interactions from job interviews and therapy sessions. Friends Corpus [23] provides a finer-grained understanding of the intensity of emotions expressed in the dialogues. Capturing the language in informal setting, Affect in Tweets (AIT-2018) [24] corpus consists of tweets annotated with emotion labels.

The set of emotion labels that were used in this research is based on GoEmotions taxonomy. GoEmotions [25] dataset is the largest fully annotated fine-grained emotion dataset currently available, consists of 58,000 English Reddit comments categorized into 27 distinct emotion categories. Designed with both psychology and data, this taxonomy has been employed in various deep learning [26] and pre-training experiments [27, 28]. The selection of GoEmotions taxonomy for this research is motivated by its ability to provide detailed and contextually appropriate framework for classifying emotions in textual feedback related to the UAQTE program. Specifically, this emotion taxonomy captures wide range of emotions that is suited for analyzing diverse and complex emotional responses. Moreover, given that GoEmotions dataset is derived from Reddit comments, it is inherently suited for analyzing text generated in similar contexts, such as online feedback forms or any platform where stakeholders might express their opinions and experiences related to the UAQTE program.

B. NLP Advancements in Text-Based Emotion Detection

The journey of emotion detection in NLP began with simple lexicon-based methods, where emotions were inferred based on the presence of predefined words associated with specific feelings [29] or based on commonsense knowledge and interaction [30]. Such approaches, although foundational, were limited by their inability to grasp context or the nuances of language use [31]. Despite these limitations, these early

methods laid the groundwork for understanding the association between language and emotion.

The advent of machine learning and, subsequently, deep learning models marked a significant evolution in emotion detection. Machine learning techniques, utilizing algorithms like Support Vector Machines (SVM) and Naïve Bayes, represented an advance over lexicon-based methods by incorporating features such as n-grams and syntactic structures to predict emotions [32]. Deep learning further propelled the field forward through neural networks to process text data in layers, thereby capturing more complex language patterns and contextual cues [33]. These models demonstrated a greater ability to understand the subtleties of language and significantly improved the accuracy of emotion detection.

The introduction of pre-trained language models such as BERT and Generative Pre-Trained Transformer (GPT) marked a major breakthrough for emotion detection in NLP. These models, trained on vast corpora of text, have shown unprecedented understanding of language context and semantics [11, 34]. By capturing the intricacies of human language, they enable more nuanced detection and classification of emotions, even in complex or ambiguous texts.

C. The Role of Pre-Trained Language Models in Emotion Detection

By fine-tuning a pre-trained language model on a specific emotion detection task, researchers can harness the knowledge acquired during pre-training to achieve impressive performance on tasks related to emotion analysis. Yang *et al.* [12] effectively used transfer language modeling to analyze utterances and their contextual information. Similarly, a model [13] utilizing dialogue utterances from Friends dataset [23] incorporated a three-phase framework for detecting emotions: casual utterances, pre-training, and fine-tuning phases. This model involved preprocessing the dataset for Masked Language Modeling (MLM) and Next Sentence Prediction (NSP) tasks, followed by fine-tuning using the BERT architecture, which yielded better performance results compared to Yang *et al.*'s [12] approach. Additionally, novel approaches proposed by Ameer *et al.* [35] proved to outperform the previous deep learning models in multi-label emotion classification by adapting the RoBERTa-based transformer network.

Beyond formal text, transfer learning also delivered robustness and effectiveness in informal texts. Huang *et al.* [36] demonstrated transfer learning by successfully classifying emotions in Twitter conversations where the model captured the nuances of informal and slang text. Furthermore, using the same architecture, the DeepEmotex-BERT model [37], specifically designed to detect emotions in text, outperformed Bi-LSTM model by 23% on large, labeled Twitter datasets. The work of Kodati and Tene [38] focused on developing models using BERT mechanism for suicide detection, identifying negative emotions in suicidal posts. Their results revealed a remarkable performance against conventional methods used to identify emotions in text sequences. Moreover, the BERT-based model excelled in sarcasm detection [39], outperforming variants of LSTM and XLNet. This work revealed that the number of sentences in a conversation contribute to the degree of sarcasm in a text.

Several studies have employed the GoEmotions taxonomy and dataset in generating their emotion classification models. The proposed EmoRoBERTa [27] approach involves two key tasks: integrating a pre-trained, fine-tuned RoBERTa model, and adapting it to the GoEmotions dataset [25]. This model performed well when tested across three different emotion taxonomies. To address the issue of imbalanced samples, the GoEmotions dataset was also used in applying data augmentation techniques [40]. The processed data samples were trained on BERT and ELECTRA language models specifically for emotion classification which led to more balanced and effective model training.

D. Emotion Detection Models for Education

Recent studies highlight the effectiveness of emotion detection models in understanding student sentiments and enhancing educational methodologies. Maceda *et al.* [5] successfully applied a fine-tuned multilingual Bidirectional Encoder Representations from Transformers (mBERT) model for sentiment analysis in the Philippine UAQTE free education program, achieving notable accuracy and f1-score. This study stands out for its effective handling of code-mixed data and its insights into the positive perceptions of the free education program among social media users. In another undertaking [41], a Multifaceted Sentiment Detection System (MSDS) was introduced to analyze emotions in student feedback during online classes. This system, employing multi-class classifiers, was aimed at predicting student dropouts through sentiment analysis. It demonstrated better accuracy than previous methods, underscoring its potential utility in dropout prevention strategies.

Additionally, Tzacheva and Easwaran [42] linked positive student feedback to the implementation of innovative teaching methods like Flipped Classroom, indicating a positive impact on student emotions. Baron-Estrada *et al.* [43] developed a sentiment analyzer which focused on emotion recognition in an educational setting. This study highlighted the efficacy of the CNN architecture in recognizing complex, non-basic emotions and emphasized the importance of further work in classifying secondary emotions.

E. Zero-Shot Learning in Text Classification

Zero-shot learning enables models to perform tasks for which they have received no explicit training. By providing models with descriptions or attributes associated with classes, they can generalize their understanding and make predictions for entirely new tasks. Ye *et al.* [44] addressed the challenge of knowledge transfer between seen and unseen classes by transforming zero-shot learning into a semi-supervised problem, leveraging unlabeled data for improved performance. Chen *et al.* [45] enhanced zero-shot learning by incorporating knowledge graphs, significantly outperforming traditional models in classifying large social text datasets, particularly in COVID-19 contexts. Complementing these approaches, Gera *et al.* [46] employed large pre-trained language models for zero-shot classification which utilized a self-training approach that adapts the model to specific tasks with significant performance gains.

Our study builds upon the existing body of research in

emotion detection within NLP by specifically focusing on the application of zero-shot learning models to analyze emotional feedback in the educational context, particularly concerning the UAQTE program. While previous studies have extensively explored emotion detection using various NLP techniques, including supervised learning models that require large labeled datasets for specific emotions, our research identifies and addresses a significant gap–the lack of studies leveraging zero-shot learned models for emotion detection in educational feedback. This gap is crucial as it pertains to the inherent capabilities of zero-shot learned models to classify emotions present in text.

Furthermore, our research distinguishes itself by focusing on the fine-tuning and application of zero-shot learned model without the necessity for extensive labeled data specific to the educational domain. This approach not only showcases the versatility and potential of fine-tuning these models in NLP but also offers practical value in educational settings underscoring its merits to previous researches.

III. Text Corpus

A. Academic-Related Feedback Responses

One major contributing factor to the success of transfer learning lies on task-specific dataset that provides direct relevance to the task at hand. In this research on developing an emotion classifier to help assess the implementation of the Universal Access to Quality Education (UAQTE) program in the Philippines, we used feedback responses from the program beneficiaries. This data encapsulates the context and nuances anchored on our research goals to understand and analyze the emotional expressions and sentiments of the target population towards the free education program.

In gathering the dataset, we utilized the BosesKo application, a citizen's participation toolkit developed as part of eParticipation 2.1: Harnessing Natural Language Processing (NLP) for Community Participation project. This is a government-funded project directly aimed at digitizing data collection of community-level feedback related to UAQTE for analysis and informed decision making. Through the BosesKo application, the beneficiaries of the free tuition and education subsidy participated in a survey designed to capture the necessary information for assessing and evaluating the UAQTE program's impact. The survey questionnaire was formulated through a workshop participated by program implementers, technical experts, and various stakeholders. Subsequently, to ensure the reliability and validity of the data collection process, pilot testing of the questionnaire was done. This step allowed the researchers to identify and fix potential issues on specific questions, fitting them better to the goal of evaluating the UAQTE program.

In this study, we utilized responses to the question, "*What is your overall feedback on the implementation of the UAQTE program?*". These responses reflect the beneficiaries' experiences, perceptions, and emotional reactions to the initiative. This corpus is particularly relevant to our research objective, as it encapsulates the emotions and sentiments of the scholars. It helps us gain a deeper understanding of the program's reception and areas for improvement. Using a substantial number of 2,812 text samples, this study used the feedback responses in the fine-tuning experiments. The alignment of the data collection approach and the research objective provides targeted and relevant basis for assessing the practical effects of the program on its stakeholders.

IV. The Proposed Approach in Transfer Learning for Emotion Detection in Text

This section presents our proposed approach in building an emotion classifier from a fine-tuned zero-shot learned model. The proposed approach in transfer learning is depicted in Fig. 1. In the discussion, we first describe the essential cleaning and encoding activities done in data preparation. Then, we highlight the fine-tuning experiments conducted wherein we adopted existing text-based emotion detection models. This is followed by the discussion on the evaluation and testing of the models. Lastly, we tackle how we built an emotion classifier designed to generate the emotions and sentiments present on the feedback responses of the scholars of the UAQTE program.

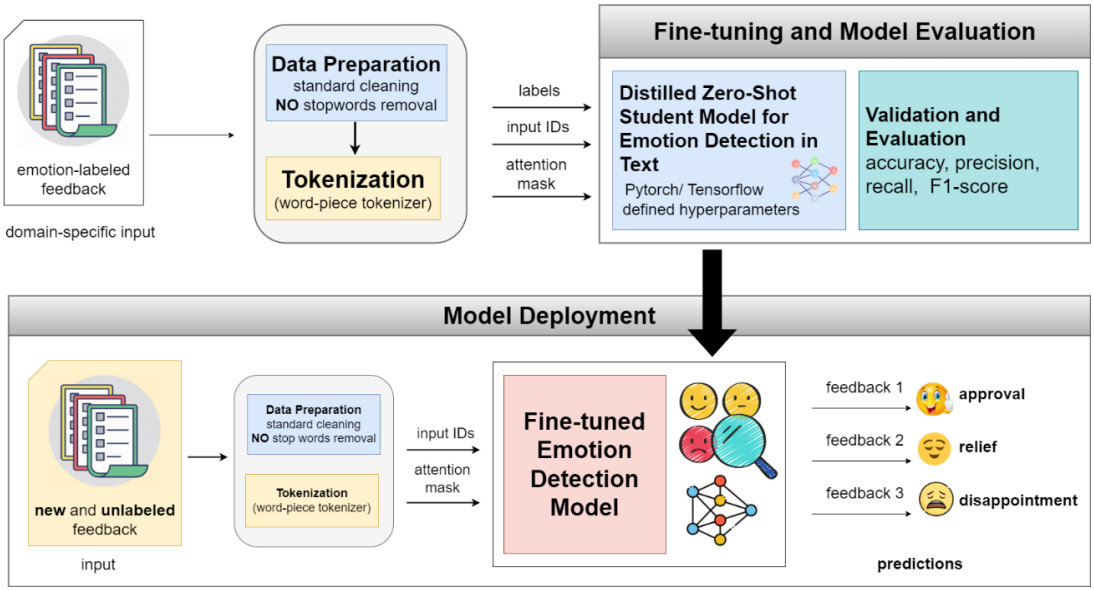

Fig. 1. Proposed approach in transfer learning for text-based emotion detection using a fine-tuned zero-shot student model.

A. Data Preparation and Tokenization

Majority of the collected samples consist of single-sentence instances. To standardize the sample length and augment the sample size, the 2,812 responses underwent sentence tokenization. This step was deemed essential to fit the data into a single-label emotion detection task, as it was observed that feedback with more than one sentence tends to express multiple emotions. Sentence tokenization resulted in an increase of 525 samples in the training dataset.

Each data point was manually annotated using GoEmotions taxonomy, focusing exclusively on the nine domain-relevant emotions: admiration, approval, desire, disappointment, gratitude, optimism, realization, relief and neutral. Through the help of social scientist, these labels were selected after thorough examination of the emotional context present in the feedback dataset. The mentioned taxonomy was instrumental in the research process as it offered comprehensive selection for categorizing emotions, giving a precise description of emotional content within the feedback responses. By concentrating on the nine categories, we were able to capture a wide range of sentiments that are significant to the experiences of the respondents. Fig. 2 illustrates the correlations among the nine emotions based on Strapparava's findings [31] that emotions with similar sentiments (positive, negative, and ambiguous or neutral) tend to cluster together. This is based on the annotator's judgement. Accordingly, emotions that share similar intensity, such as admiration, gratitude, and relief, are closely correlated. It is worth noting that majority of the collected feedback conveys positive emotions classified as ambiguous or neutral and only disappointment as negative emotion.

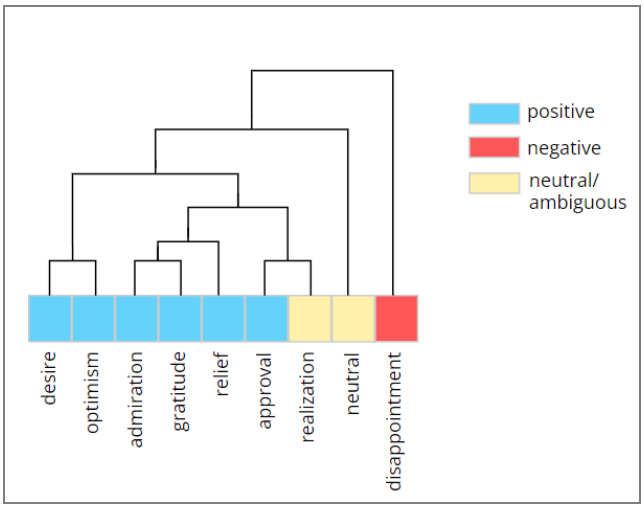

Fig. 2. Hierarchical clustering of emotion labels.

Similar to the source taxonomy, our emotion samples are not distributed uniformly in the corpus. This imbalance number only reflects the natural distribution of classes in the real-word scenario. While imbalance dataset is a common challenge in classification tasks, transfer learning method exhibits a reduced sensitivity to these imbalances. As demonstrated in previous studies [47, 48], transformer-based models surpassed the performance of deep learning models particularly in handling imbalanced data. This makes them more effective in scenarios where class distribution is uneven.

The boxplot in Fig. 3 shows the average number of words used to express different emotions about the UAQTE program. Desire has the highest average word count, suggesting that respondents tend to give more elaborate feedback when expressing aspirations and concerns. Neutral class shows the least number of words, possibly indicating a lack of strong opinion or engagement. The respondents typically use a moderate length of text to express feelings of approval, relief, optimism, realization, and gratitude. Feelings of disappointment and admiration both show a wide range of word counts with the median leaning towards longer responses, suggesting that these emotions evoke a more in-depth reaction. Overall, the varying lengths and distributions of responses highlight how different emotions influence the detail and depth of feedback on the program.

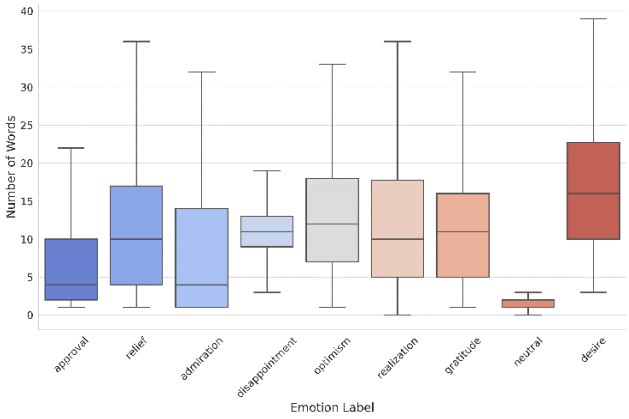

Fig. 3. Average number of words per emotion.

To enhance the quality of the corpus, after annotation, the dataset has undergone minimal data cleaning using the Natural Language Toolkit (NLTK). This process included tasks on lowercasing, removal of punctuations except for '/', lemmatization and expanding word contractions.

In our initial experiments, we attempted to remove stopwords from the texts. However, it reduced the performance metric scores of the model compared to training the corpus using the entire lexicon. This decrease in performance may be attributed to the nature of our collected text responses: they are concise, domain-specific, and provide direct feedback which often express crucial contextual information essential for accurate emotion extraction. To ensure consistency and reliability in our analysis, we performed these tasks to standardize the text data.

For the task of converting the feedback corpus to numerical representations that the model can understand, we used a pre-trained tokenizer essential for the fine-tuning of DistilBERT-based and RoBERTa-based models. As derivatives of BERT, DistilBERT and RoBERTa were pre-trained using the tokens generated by the Byte Pair Encoding (BPE) subword tokenization method. This tokenizer learns merging rules to construct a new symbol from two symbols in the base vocabulary and builds a base vocabulary with all the symbols that appear in the set of unique words. We incorporated all Special Tokens–Padding (PAD), Unknown Word (UNK), Special Classifier (CLS), Separator (SEP), and Masked or Hidden (MASK) tokens to make the model process and understand the inputs and the

task. Based on BPE, there are 2,359 tokens in the tokenizer vocabulary present in our dataset.

B. Fine-Tuning Process

The adoption of transfer learning techniques through the pre-trained models was instrumental in achieving our research objective. Fine-tuning these models on our dataset allowed us to customize their capabilities to the context of educational feedback, enhancing their accuracy in emotion detection. This methodology directly supports the research goal on deploying an effective emotion classifier by combining the strengths of advanced NLP models with the specific requirements of the emotion analysis task.

Two emotion detection models fitted on the GoEmotions dataset were selected for the transfer learning experiments. The first model, EmoRoberta [33] was trained on a more generalized variant of BERT. It adapted the benchmark approach in transfer learning that integrated a pre-trained and fine-tuned RoBERTa model with the GoEmotions dataset. This model was tested across three distinct emotion taxonomies, surpassing the current state-of-the-art model with a higher macro-F1 score.

On the other hand, the second model employed the zero-shot learning pipeline utilizing the same emotion dataset. This model employed a distillation process to generate a more efficient and faster classifier. Fig. 4 illustrates an overview of how the distilled student model for emotion detection was trained. Initially, a Natural Language Inference (NLI) teacher model was trained on RoBERTa using a set of candidate emotions and sample unlabeled data. This process generated probability distribution for each sample across all possible classes, with the class having the highest probability score chosen as the predicted class by the teacher model. The student model was then trained using a combination of unseen and unlabeled GoEmotions dataset, a list of emotion labels and the predictions made by the teacher model. To generate a faster classifier, the student was trained on DistilBERT – a smaller and faster version of BERT, while still retaining comparable performance. This process combines the strength of zero-shot learning and distillation, resulting in a more capable emotion detection.

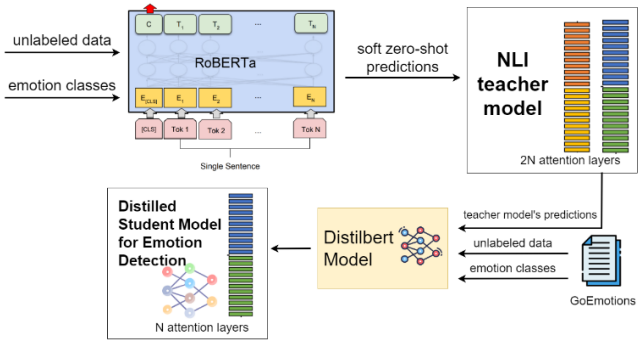

Fig. 4. Zero-shot learning pipeline for training a distilled student model for emotion detection.

The pre-trained EmoRoBerta and distilled zero-shot student models were fine-tuned on 80% of the feedback dataset to adapt their pre-trained weights to our emotion detection target task. In total, 15 finetuning experiments using the two knowledge-source models were completed.

Pytorch and Tensorflow deep learning libraries were utilized in the transfer learning process. The models were evaluated under various hyperparameter configurations–batch sizes of 8 and 16, epochs of 3, 5 and 10, and learning rate of 2e−5 with Adam optimizer. Cross-entropy loss was employed and training and validation losses were monitored. This was done to check if each model was improving over epochs and was generalizing well on new data.

C. Model Evaluation and Testing

The held-out 20% of the dataset was equally divided between validation and test sets. The validation set's purpose was to monitor the model's performance throughout the training process as well as to adjust the hyperparameter tuning. This serves as an independent dataset to help prevent overfitting and help choose the best-performing model during training. On the other hand, the test set was employed to assess the final performance of the model, providing an impartial estimate of its generalization ability on new, unseen data.

Performance of the models were described by the confusion matrices that properly present the models' predictions. We focused on four key classification metrics--accuracy, precision, recall and F1 score. Precision measures the accuracy of the positive predictions, while recall measures the ability of the model to find all relevant cases in the dataset. A higher precision signifies less False Positives (FP), whereas lower precision implies more FPs. Additionally, a higher recall correlates with fewer False Negatives (FN), reflecting the classifier's ability to return more relevant and accurate results. The harmonic mean of these two metrics is measured by F1-Score.

The model that obtained the best scores across all metrics was chosen for implementation of the emotion detection classifier.

D. Deployment of the Emotion Classifier

The optimal model was utilized in a practical setting to analyze emotions and sentiments in feedback from recipients of the UAQTE program. A simple Python code in Jupyter Notebook was designed for this purpose. The feedback responses underwent standardized cleaning procedures: lowercasing, removal of punctuations except for '/', lemmatization and expanding word contractions in the text. The cleaned texts were tokenized using AutoTokenizer and then automatically classified to identify the corresponding emotions. The result of the emotion classification was further analyzed to serve as an aid in the continuous evaluation of the free tertiary education, further improvement of the current program implementation as well as for the development of policy recommendations.

The methodology, from data preparation to model deployment, was designed to directly contribute to the evaluation and improvement of the UAQTE program. By providing a detailed account of our methodological choices and their validation, we aim to highlight the significance of our research in the broader NLP field. This demonstrates the potential of transfer learning and zero-shot models in enhancing the understanding of feedback responses within educational settings.

V. Results and Discussion

The results presented here answered the research objectives on identifying the best emotion classification model from the transfer learning experiments and deploying the emotion classifier to analyze scholar's feedback. Particularly, this section comprehensively outlines the results of experiments done in the fine-tuning stage, the best emotion classification model, and the result and analysis of the labeled feedback using the developed emotion classifier.

A. Fine-Tuning Experimental Results

Table 1 presents the comparative results of the fine-tuning experiments conducted with the EmoRoberta model and the distilled zero-shot student model. These models were evaluated using the 10% unseen and labeled test set.

Table 1. Results of the fine-tuning experiments

Model	Stop-words Removal	Epoch	Batch size	Acc (%)	Pre (%)	Rec (%)	F1-Score (%)
EmoRo-Berta Model	No	5	8	74.30	75.14	74.3	74.72
	No	10	8	76.55	75.89	76.74	76.31
	No	5	16	78.54	78.12	79.32	78.72
	No	10	16	77.08	77.45	77.08	77.26
	Yes	5	8	74.10	73.23	73.78	73.51
	Yes	10	8	74.45	74.67	74.36	74.51
Distilled Zero-Shot Student Model	No	3	8	82.97	83.36	82.97	82.77
	No	5	8	82.35	82.52	82.35	82.36
	No	10	8	80.8	81.35	80.80	80.80
	No	5	16	84.21	84.93	84.21	84.23
	No	10	16	83.59	83.94	83.59	83.53
	Yes	5	8	79.12	78.67	78.89	78.76
	Yes	10	8	78.64	79.6	78.64	79.12
	Yes	5	16	77.81	79.28	78.14	78.71
	Yes	10	16	79.57	80.46	79.57	80.01

1) Distilled zero-shot student model outperforms EmoRoberta

Across all configurations, the distilled zero-shot learned model generally demonstrates superior performance compared to the EmoRoberta model especially in terms of accuracy and F1 score. The best performing EmoRoberta model gained 78.54% accuracy and 78.96% F1 score, while the best distilled zero-shot student model received 84.21% accuracy and 84.23% F1-Score.

2) Optimal results achieved without removing stopwords

The distilled zero-shot student model retaining the stopwords in the training corpus performs better compared to the same model with stopwords removal across all metrics. Removing stopwords yields the highest F1 score of 79.46%, while keeping them boosts the score to 84.22%. The removal of stopwords seems to negatively impact the performance of the model, an implication that stopwords are significant in providing valuable context for emotion detection task. Previous findings [49] have demonstrated that stopwords, often deemed irrelevant in many NLP tasks, actually hold significant contextual information that is crucial for understanding the subtleties of emotional expression in text. This validates our observation that stopwords provide the necessary linguistic context which enhances the model's capacity for accurate emotion detection.

3) Effect of epochs and batch size

Increasing epochs from 5 to 10 does not consistently improve performance in either model. For instance, the EmoRoberta model shows a slight decrease in performance when increasing epochs from 5 to 10 with a batch size of 16. The lack of consistent improvement with increased epochs indicates a complex interaction between model training duration and performance. This suggests that longer training does not always equate to better performance, possibly due to overfitting issues or diminishing returns after a certain point in training.

Accordingly, in the experiments, a larger batch size (16 *vs.* 8) tends to improve performance for both models. This is particularly noticeable in the distilled zero-shot model. The general improvement in performance with a larger batch size for both models underscores the role of batch size in effective learning. Larger batches provide more diverse data per update, potentially leading to better generalization and robustness in the model's predictions.

B. Best Fine-tuned Distilled Zero-Shot Emotion Classification Model

The best model was trained on the corpus without removing the stopwords and has been configured with 5 epochs, 2e−5 learning rate and a batch size of 16. Fig. 5 illustrates the training progress during the experiment. The pattern shows consistent improvement across all epochs in terms of both reducing losses and improving performance metrics on accuracy, precision, recall, and F1 score. The relatively stable validation loss and the consistent improvement in the performance metrics suggest that the model is learning effectively, not overfitting, and is likely to generalize well to new, unseen data.

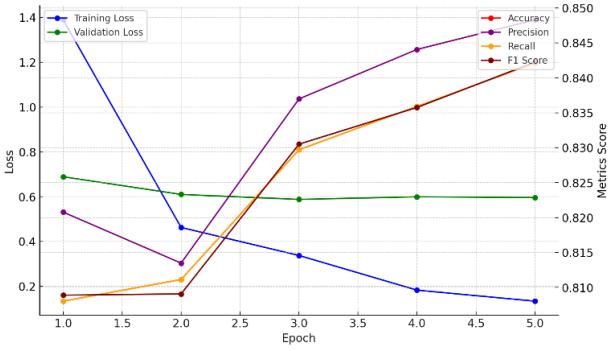

Fig. 5. Training progress metrics of the distilled zero-shot student model, fine-tuned with 5 epochs, 2e−5 learning rate and 16 batch size.

The confusion matrix visualized as heatmap in Fig. 6

displays the proportions of each predicted label against the true labels in the test dataset. The color intensity indicates the magnitude of percentages, with darker shades signifying higher percentages. Diagonal values are the correct predictions for each class, whereas the lighter shades indicate the model's difficulties in accurately classifying emotions. Gratitude, desire, admiration, and neutral classes with darker colors, show high percentages, suggesting that the model demonstrates strong predictive ability in identifying these emotions. This result can be attributed to the distinct textual patterns or linguistic markers in the dataset that make it easier for the model to recognize. For instance, expressions of desire often contain phrases indicating wants or needs, such as *"I hope it to be improved"* and *"we want to benefit"*. The feeling of gratitude is typically associated with words expressing appreciation, like "*thank you"*, *"grateful"* and *"blessed"*. Admiration signifies a strong positive sentiment towards the program expressed in the statements like *"*excellent*"*, *"great program"*, and *"*commendable*"*. Neutral expressions, lacking strong emotional markers, might be easier to distinguish from more emotive states. Conversely, the lighter shades for realization and optimism indicate the model's difficulties in accurately classifying these emotions.

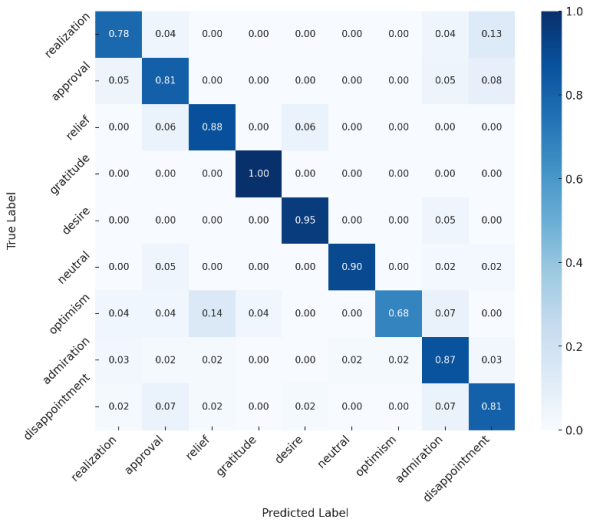

Fig. 6. Confusion matrix of the best model showing the relative frequency of predictions in each emotion category.

On the other hand, off-diagonal values show the distribution of incorrect predictions. Well-performing classes obtained no to a smaller number of misclassifications, while some emotion classes tend to be confused with other classes. For example, realization is sometimes confused with disappointment (at 13%), and optimism with relief (at 14%). This pattern of confusion may stem from the inherent overlap in the expression of these emotions within the text. Both realization and disappointment can be associated with expressions of understanding, although with different emotional polarities. Optimism and relief are both forward-looking emotions that can share a context of positive anticipation or the resolution of burden.

Moreover, the similarities on how these emotions are expressed or represented in the data possibly arise from the use of common keywords or phrases that are context-dependent. For instance, phrase like "good, but can be…" could be linked to both realization and disappointment, depending on the subsequent context. The writer might has come to an understanding or recognized the need for something more. Similarly, expressions of hope or expectation might be found in both statements of optimism and relief. The distinction here lies in the surrounding words and the overall sentiment of the text, which can be subtle and thus more challenging for the model to understand.

Some instances of the misclassified text responses are provided in Table 2. The first three examples were annotated as realization. However, the model predicted them as disappointment, likely due to the inherent negative connotation in the statements implying an issue or dissatisfaction. The third example is intended to express admiration for the program's effectiveness. However, the model's prediction of relief might be influenced by the phrase, "helped a lot of students," which could be construed as alleviating a problem or concern, a common context for relief.

Table 2. Sample misclassified data points

	Text (Feedback)	Actual Emotion	Predicted Emotion
1	*"I think that students from wealthy or students from family that are financially stable should not be able to avail it"*	realization	disappointment
2	*"it needs to be better"*	realization	disappointment
3	*"there are a lot of areas that could be improved"*	realization	disappointment
4	*"it was a great implementation and helped a lot of students"*	admiration	relief
5	*"I hope it will continue as it will help a lot of students in their college life"*	optimism	relief
6	*"great help for me and my family. I wish other students would still benefit from it in the future.*	relief	desire
7	*"great program. I hope it will continue to provide quality education"*	optimism	admiration
8	*"it is very helpful and very supporting"*	relief	approval

The fifth example expresses optimism about the program's future impact, but inaccurately tagged as relief–this could be influenced by the model's focus on the immediate benefit mentioned. Describing the program as "very helpful and very supporting" in the last example evokes sense of approval, but also expresses a sense of relief. There is an overlap of recognizing the program's value and expressing ease of burden.

The spread of misclassifications suggests either a lack of distinctive features learned by the model to differentiate these emotions or the existence of more than one emotion in a single feedback. A notable observation for most of the misclassified instances is the presence of two emotions in a statement. The fine-tuned emotion detection model operates on the assumption of a single dominant emotion per statement, but human expression is rarely so straightforward.

The nuanced interplay of emotions in language reflects the complexity of how the respondents communicate their thoughts, thereby challenging the ability of the emotion detection model.

In addition to the confusion matrix, performance metric scores for each emotion are shown in Table 3 to provide a more comprehensive context on how the model performed across all classes. The model excels in identifying desire, gratitude and neutral, evidenced by high precision and recall rates. These emotions are not only accurately identified but also rarely misclassified as other emotions (low false positives). However, there are notable challenges with optimism and relief. In the case of optimism, there is a significant disparity between the precision and recall, indicating that while the model is accurate when it predicts optimism, it often fails to identify it. For the case of relief, while the recall is high, the lower precision suggests a tendency to misclassify other emotions as relief. This variance could stem from various factors, including the distribution of sample sizes in the dataset or the inherent complexity in classifying certain emotions.

The balanced macro and weighted averages on all emotions suggest a good overall performance of the model, effective in predicting most common instances and can be deployed in actual application.

Table 3. Performance metric scores per emotion

	Emotion Label	Precision (%)	Recall (%)	F1 score (%)
1	admiration	81.54	86.89	84.13
2	approval	82.76	81.36	82.05
3	desire	90.48	95.00	92.68
4	disappointment	81.03	81.03	81.03
5	gratitude	94.12	100.00	96.97
6	neutral	97.44	90.48	93.83
7	optimism	95.00	67.86	79.17
8	realization	72.00	78.26	75.00
9	relief	70.00	87.50	77.78

The foregoing analyses not only shed light on the model's current capabilities and limitations, but also underscore areas for improvement. The limitations of the proposed model include potential biases in the training data and the model's ability to understand context and differences in emotional expression. Another limitation is that the model's designed to be only single-label classifier, which assumes that each response is associated with a single dominant emotion. This simplification may not accurately reflect the complexity of human emotions, where multiple feelings can be experienced and expressed simultaneously in the feedback.

To address the challenges and limitations identified, future efforts should focus on mitigating data biases. This can be done by adding data samples for minority classes and resampling or adjusting the weights of data points to balance the representation of various emotional expressions. Also, incorporating enhanced feature extraction techniques and context-aware models for improved contextual understanding could make the models better differentiate emotions with closely related expressions. Lastly, developing multi-label emotion classifier could address the complexity of human emotional responses more effectively, recognizing that feedback can carry multiple sentiments at once.

Our proposed model demonstrates superiority and uniqueness compared to earlier works in the related domains. Previous efforts in transfer learning to improve emotion detection, such as the works [12, 35] utilized models which were trained on labeled dataset. In this research, we capitalized on the inherent knowledge of a zero-shot learned model. This strategy significantly enhances the model's adaptability and performance in detecting nuanced emotions specific to the educational context, making it particularly effective for analyzing the diverse expressions found in feedback on the UAQTE program. While previous research has demonstrated the effectiveness of pre-trained language models across formal and informal text types [23, 38, 39], our approach extends this versatility to the educational feedback domain, characterized by its unique mix of expressions. This broad applicability, facilitated by the fine-tuning of a zero-shot learned model, illustrates our method's uniqueness in adapting to the specific challenges of educational sentiment analysis. Within the educational context, previous research that fine-tuned an mBERT model for sentiment analysis using code-mixed data [5] demonstrated the potential of NLP in analyzing student feedback. Our study extends these findings by not only focusing on sentiment polarities, but by identifying specific emotions present in the student feedback.

Ultimately, the findings of the study contribute to the ongoing dialogue in the NLP community about the importance of developing adaptable models. The research's exploration of zero-shot model for emotion detection not only provides valuable insights for educational application, but also marks a significant step forward in NLP field. It demonstrates the feasibility and effectiveness of fine-tuning such model in achieving a context-aware understanding of language that can be adapted to specialized domains like education.

C. Emotion Classification in Academic-Related Feedback

The model was deployed to perform automatic emotion labeling of feedback responses. We analyzed 362 newly collected survey responses answering the question, "What is your overall feedback on the implementation of the UAQTE program?". These feedback responses were cleaned, pre-processed, and automatically labeled with their corresponding emotions. Presented in Fig. 7 is the summary of the results. The most frequent emotions identified are relief and approval. The presence of relief as the top emotion indicates that the program has significantly alleviates financial burden from students and their families. Students likely appreciate the financial support provided by the program, allowing them to pursue education without the added stress of financial constraints. The prevalence of approval indicates that students are generally satisfied with the implementation of the UAQTE program. This may be attributed to scholars' experiences to access tertiary

education, career opportunities and equality and inclusivity. Accordingly, emotions like gratitude, admiration, and optimism further underscore positive experiences. Scholars' responses reflect the program's positive impact, which is expressed through appreciation, satisfaction, as well as hope for its continued success and improvement.

The frequency of the positive emotions strongly suggests a positive reception of the UAQTE program. This supports the conclusion that the program has successfully achieved its objective of providing deserving Filipino students access to quality education and has alleviated financial burdens. Moreover, these results reinforce the need for program sustainability to support the next generations of students, increasing its economic and social impact.

However, the presence of disappointment in the feedback shows negative sentiment towards certain aspects of the program, which fail to meet the expectations of the recipients. These responses often critique the program's limited scope, conveying that it favors certain group, and point to issues like corruption or inadequate measures. They highlight various negative aspects that scholars have encountered or observed, including administrative challenges, delays in the release of funds and other educational needs not provided by the program. Also, there is a substantial number of neutral responses observed. This implies that the recipients recognize the program without clear satisfaction or dissatisfaction. Such responses acknowledge the benefits and effectiveness of the program but also indicate areas needing improvement or further enhancement.

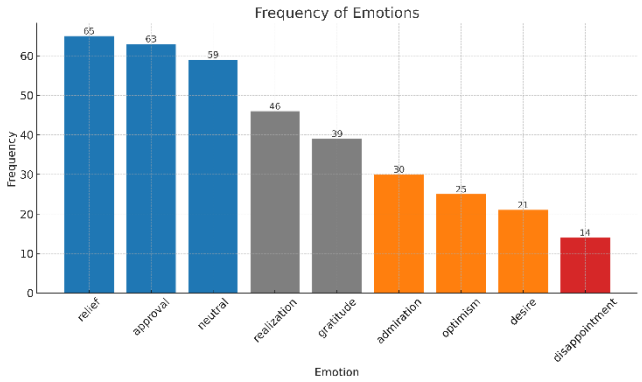

Fig. 7. Frequency of each predicted emotion in the actual application of the emotion classifier.

Based on the foregoing, while the UAQTE program is making a positive difference among its stakeholders, there is still room for refinement. Identifying key emotions in scholars' feedback can aid the program administrators and policymakers to understand the impact of the free education program and make data-driven decisions for its improvement and targeted interventions. For example, policy recommendations may include resource allocation based on identified emotions like approval and disappointment. This will emphasize the need to expand the program's scope to address specific concerns, ensure fair distribution of benefits, and implement transparency measures. Targeted interventions may focus on academic assistance, skill development workshops, streamlining of administrative processes, and community engagement. The presence of the identified emotions also points to the necessity of long-term strategies, like continuous assessment, research for longitudinal studies, and data-driven decision-making to proactively meet the stakeholders' evolving needs. By enhancing the components of the UAQTE program that receive positive reactions or fixing issues highlighted by negative feedback can help ensure that educational initiatives are not only effective academically, but also resonate positively with their target recipients.

VI. Conclusion

While the fine-tuning of pre-trained models, a standard architecture in transfer learning, has consistently delivered state-of-the-art results in emotion detection tasks, the potential of zero-shot learned model in this area remains largely unexplored. This paper introduces a novel method of fine-tuning a distilled zero-shot student model to a domain-specific task of classifying emotions present in the responses related to the implementation of the Universal Access to Quality Education (UAQTE) program in the Philippines. Select emotion labels from the GoEmotions taxonomy were utilized to annotate the domain-specific corpus. The proposed approach excludes the removal of stopwords in the cleaning and preprocessing pipeline. As observed, removing the stopwords in the training set negatively impacts the performance of the model, implying that these words play significant role in the context and understanding of emotions in the present task.

The experimental results revealed that the distilled zero-shot student model outperformed the EmoRoberta model across all performance metric scores, achieving 84.21% accuracy, 84.93% precision, 84.21% recall, and 84.23% F1 score. This model demonstrates excellent predictive ability to distinguish between various emotions, particularly those with distinct expressions. It is proved to be effective in accurately identifying most emotions, particularly excelling at desire, gratitude, and neutral classes. However, the model encountered challenges in classifying emotions like optimism and relief, suggesting future improvements on how to better represent them in the corpus. The presence of few misclassifications suggests either a lack of distinctive features learned by the model to differentiate these emotions or the existence of more than one emotion in single feedback.

In practical application, the model was successfully deployed to automatically label feedback from UAQTE scholars. The analysis revealed a predominantly positive reception of the program among its beneficiaries, with feelings of relief, approval, and gratitude being the most prominent. However, the presence of neutral and negative emotions like disappointment also highlights areas where the free education might need improvement. These insights can be valuable for policymakers and program administrators to understand the impact of the program and to make data-driven decisions for its improvement and targeted interventions.

To address the challenges and limitations identified, future works may focus on enhancing the model to reduce confusion between certain emotions. This can be achieved by adding data samples to minority classes and adjusting the weights of data points to balance the representation of

various emotional expressions. Also, incorporating enhanced feature extraction techniques and context-aware models for improved contextual understanding could make the models better differentiate emotions with closely related expressions. Exploring multi-label emotion classification to capture not just the dominant emotion, but all emotions present in text offers another direction for further research. Additionally, investigating the model's cross-domain adaptability beyond academic context, such as social media dataset.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

M.J.P.C. proposed the approach in fine-tuning a zero-shot model for emotion classification, performed the experiments, conducted formal analysis of results, and prepared the final draft. T.D.P. checked the soundness of the proposal and validity of the methodology. L.L.M. was in-charge in the data collection and analysis of experimental results. M.B.A. acquired the funding and contributed in the analysis of classification results. All authors contributed in completing the paper and had approved its final version.

FUNDING

This work is supported by Philippine Commission on Higher Education (CHED) Leading the Advancement of Knowledge in Agriculture and Science (LAKAS) Project No. 2021-007, eParticipation 2.1: Harnessing Natural Language Processing (NLP) for Community Participation.

ACKNOWLEDGMENT

The authors wish to thank the CHED-LAKAS Project.

REFERENCES

[1] P. Fung, D. Bertero, Y. Wan, A. Dey, R. H. Y. Chan, F. B. Siddique, Y. Yang, C.-S. Wu, and R. Lin, "Towards empathetic human-robot interactions," *Computational Linguistics and Intelligent Text Processing*, pp. 173–193, 2018.

[2] B. Ribeiro, G. Oliveira, A. Laranjeira, and J. P. Arrais, "Deep learning in digital marketing: Brand detection and emotion recognition," *International Journal of Machine Intelligence and Sensory Signal Processing*, vol. 2, no. 1, pp. 32–50, 2017. https://doi.org/10.1504/IJMISSP.2017.088173

[3] A. Suciati and I. Budi, "Aspect-based sentiment analysis and emotion detection for code-mixed review," *International Journal of Advanced Computer Science and Applications*, vol. 11, pp. 179–186, 2020. http://dx.doi.org/10.14569/IJACSA.2020.0110921

[4] N. Gupta, M. Gilbert, and G. D. Fabbrizio, "Emotion detection in email customer care," *Computational Intelligence*, vol. 29, no. 3, pp. 489–505, 2013.

[5] L. L. Maceda, A. A. Satuito, and M. B. Abisado, "Sentiment analysis of code-mixed social media data on Philippine UAQTE using fine-tuned mBERT Model," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 7, 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140777

[6] R. Kamal, M. A. Shah, C. Maple, M. Masood, A. Wahid, and A. Mehmood, "Emotion classification and crowd source sensing: A lexicon based approach," *IEEE Access*, vol. 7, pp. 27124–27134, 2019. doi: 10.1109/ACCESS.2019.2892624

[7] A. Al Maruf, Z. M. Ziyad, M. M. Haque, and F. Khanam, "Emotion detection from text and sentiment analysis of ukraine russia war using machine learning technique," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 12, 2022. http://dx.doi.org/10.14569/IJACSA.2022.01312101

[8] D. Xu, Z. Tian, R. Lai, X. Kong, Z. Tan, and W. Shi, "Deep learning based emotion analysis of microblog texts," *Information Fusion*, vol. 64, pp. 1–11, 2020.

[9] D. Haryadi and G. P. Kusuma, "Emotion detection in text using nested long short-term memory," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 10, no. 6, 2019. http://dx.doi.org/10.14569/IJACSA.2019.0100645

[10] H. Binali, C. Wu, and V. Potdar, "Computational approaches for emotion detection in text," in *Proc. 4th IEEE International Conference on Digital Ecosystems and Technologies, Dubai, United Arab Emirates*, 2010, pp. 172–177. doi: 10.1109/DEST.2010.5610650

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171–4186, 2019.

[12] Y. K, D. Lee, T. Whang, S. Lee, and H. Lim, "Emotionx-ku: Bert-max based contextual emotion classifier," *CoRR*, arXiv:abs/1906.11565, p 6, 2019.

[13] Y.-H. Huang, S.-R. Lee, M.-Y. Ma, Y.-H. Chen, Y.-W. Yu, and Y.-S. Chen, "Emotionx-idea: Emotion bert–an affectional model for conversation," arXiv preprint arXiv:1908.06264, p 6, 2019.

[14] C. Huang, A. Trabelsi, and O. R. Zaiane, "Ana at semeval-2019 task 3: Contextual emotion detection in conversations through hierarchical lstms and bert," in *Proc. the 13th International Workshop on Semantic Evaluation*, pp. 49–53, 2019.

[15] S. J. Pan, "Transfer learning," *Data Classification: Algorithms and Applications*, pp. 537–570, 2014.

[16] Hugging Face. (2023). Zero-shot classification. *Hugging Face* [Online]. Available: https://huggingface.co/tasks/zero-shot-classification.

[17] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," *Journal of Personality and Social Psychology*, vol. 17, no. 2, pp. 124–129, 1971.

[18] R. E. Plutchik and H. R. Conte, "Circumplex models of personality and emotions," *American Psychological Association*, 1997.

[19] A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, 1990.

[20] E. Cambria, A. Livingstone, and A. Hussain, "The hourglass of emotions," *Cognitive Behavioural Systems*, pp. 144–157, 2012.

[21] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proc. the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 70–74, June 2007.

[22] B. Nojavanasghari, T. Baltrušaitis, C. E. Hughes, and L. P. Morency, "Emoreact: A multimodal approach and dataset for recognizing emotional responses in children," in *Proc. the 18th ACM International Conference on Multimodal Interaction*, pp. 137–144, October 2016.

[23] H. Y. Chen and J. D. Choi, "Character identification on multiparty conversation: Identifying mentions of characters in TV shows," in *Proc. the 17th Annual SIGdial Meeting on Discourse and Dialogue*, SIGDIAL'16, 2016.

[24] S. Mohammad, F. Bravo-Marquez, M. Salameh, and S. Kiritchenko, "Semeval-2018 task 1: Affect in tweets," in *Proc. the 12th International Workshop on Semantic Evaluation*, pp. 1–17, June 2018.

[25] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and S. Ravi, "GoEmotions: A dataset of fine-grained emotions," *ACL 2020*, pp. 4040–4054. doi: 10.18653/v1/2020.acl-main.372

[26] Z. Ahanin, M. A. Ismail, N. S. Singh, and A. AL-Ashmori, "Hybrid feature extraction for multi-label emotion classification in English text messages," *Sustainability*, 2023.

[27] R. Kamath, A. Ghoshal, S. Eswaran, and P. Honnavalli, "An enhanced context-based emotion detection model using RoBERTa," in *Proc. 2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, 2022, pp. 1–6. doi: 10.1109/CONECCT55679.2022.9865796

[28] N. Alvarez-Gonzalez, A. Kaltenbrunner, and V. Gómez, "Uncovering the limits of text-based emotion detection," arXiv:2109.01900 [cs.CL], 2021.

[29] C. Strapparava and R. Mihalcea, "Learning to identify emotions in text," in *Proc. the 2008 ACM symposium on Applied computing*, 2008, pp. 1556–1560.

[30] A. Balahur, J. M. Hermida, and A. Montoyo, "Detecting implicit expressions of sentiment in text based on commonsense knowledge," in *Proc. the 2nd workshop on computational approaches to subjectivity and sentiment analysis (WASSA 2.011)*, 2011. pp. 53–60.

[31] C. Strapparava and R. Mihalcea, "Semeval-2007 task 14: Affective text," in *Proc. the Fourth International Workshop on Semantic Evaluations*, 2007, pp. 70–74.

[32] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends® in information retrieval*, vol. 2, no. 1, pp. 1–135, 2008.
[33] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1631–1642.
[34] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, and D. Amodei, "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.
[35] I. Ameer, N. Bölücü, M. H. F. Siddiqui, B. Can, G. Sidorov, and A. Gelbukh, "Multi-label emotion classification in texts using transfer learning," *Expert Systems with Applications*, vol. 213, 118534, 2023.
[36] C. Huang, A. Trabelsi, and O. R. Zaiane, "Ana at semeval-2019 task 3: Contextual emotion detection in conversations through hierarchical lstms and bert," in *Proc. the 13th International Workshop on Semantic Evaluation*, pp. 49–53, 2019.
[37] M. Hasan, E. Rundensteiner, and E. Agu, "DeepEmotex: Classifying Emotion in Text Messages using Deep Transfer Learning," in *Proc. 2021 IEEE International Conference on Big Data (Big Data)*, Orlando, FL, USA, 2021, pp. 5143–5152. doi: 10.1109/BigData52589.2021.9671803
[38] D. Kodati and R. Tene, "Identifying suicidal emotions on social media through transformer-based deep learning," *Applied Intelligence*, vol. 53, pp. 11885–11917, 2023. https://doi.org/10.1007/s10489-022-04060-8
[39] A. Avvaru, S. Vobilisetty, and R. Mamidi, "Detecting sarcasm in conversation context using transformer-based models," in *Proc. the Second Workhop on Figurative Language Processing*, pp. 98–103, 2020.
[40] A. Kane, S. Patankar, S. Khose, and N. Kirtane, "Transformer based ensemble for emotion detection," arXiv preprint, arXiv:2203.11899, 2022.
[41] A. C. T. Mary and A. L. P. J. Rose, "Multifaceted Sentiment Detection System (MSDS) to avoid dropout in virtual learning environment using multi-class classifiers," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no. 4, 2023. http://dx.doi.org/10.14569/IJACSA.2023.0140440
[42] A. A. Tzacheva and A. Easwaran, "Emotion detection and opinion mining from student comments for teaching innovation assessment," *International Journal of Education*, vol. 9, 2021.
[43] M. L. Barron-Estrada, R. Zatarain-Cabada, and R. O. Bustillos, "Emotion recognition for education using sentiment analysis," *Res. Comput. Sci*., vol. 148, no. 5, pp. 71–80, 2019.
[44] Z. Ye, Y. Geng, J. Chen, J. Chen, X. Xu, S. Zheng, and H. Chen, "Zero-shot text classification via reinforced self-training," in *Proc. the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 3014–3024.
[45] Q. Chen, W. Wang, K. Huang, and F. Coenen, "Zero-Shot text classification via knowledge graph embedding for social media data," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9205–9213, June 15, 2022. doi: 10.1109/JIOT.2021.3093065
[46] A. Gera, A. Halfon, E. Shnarch, Y. Perlitz, L. Ein-Dor, and N. Slonim, "Zero-shot text classification with self-training," arXiv preprint, arXiv:2210.17541, 2022.
[47] F. Demirkıran, A. Çayır, U. Ünal, and H. Dağ, "An ensemble of pre-trained transformer models for imbalanced multiclass malware classification," *Computers & Security*, vol. 121, 102846, 2022.
[48] N. E. Erciyes, A. K. Görür, "Deep learning methods with pre-trained word embeddings and pre-trained transformers for extreme multi-label text classification," in *Proc. 2021 6th International Conference on Computer Science and Engineering (UBMK)*, IEEE, 2021, pp. 50–55.
[49] U. Rashid, M. W. Iqbal, M. A. Skiandar, M. Q. Raiz, M. R. Naqvi, and S. K. Shahzad, "Emotion detection of contextual text using deep learning," in *Proc. 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, Istanbul, Turkey, 2020, pp. 1–5. doi: 10.1109/ISMSIT50672.2020.9255279