COMP9444 Project Summary

<Emotion Classification using Tweets>

z5270075: Jiangye Song, z5530515: Pan Lu, z5504217: Cong Du

z5471262: Jiayue Tang, z5465193: Xiangyu Zheng

I. Introduction

The aim of this project is to develop deep learning-based solutions for accurately identifying emotions presented in textual data, specifically tweets. Understanding emotions as conveyed through text is essential for numerous applications, including disaster response, marketing, security, and content creation. Emotions are expressed in nuanced ways, varying by individual experiences and cultural contexts, making this a challenging problem. Our contribution lies in employing advanced AI algorithms to improve performance on relevant metrics for emotion classification, leveraging various deep learning models such as LSTM, GRU, and CNN.

The dataset used for this project is the Hugging Face dair-ai Emotion dataset, which contains English Twitter messages labeled with the six basic emotions. This dataset provides a rich resource (split and unsplit) train, validation and test samples for training and evaluating machine learning models, enabling the development of a robust emotion classification system.

II. Related Work

Several studies have addressed sentiment analysis and emotion classification on social media platforms. For instance, Md Parvez Mollah's paper on using LSTM for Twitter sentiment analysis highlights the effectiveness of LSTM in handling sequential data but also notes its slower training time and potential overfitting issues. Federico Bianchi et al.'s work on emotion classification in Italian using BERT demonstrates high performance but requires substantial computational resources. Muhammad Abdul-Mageed and Lyle Ungar's EmoNet employs GRNN for fine-grained emotion detection, showcasing GRNN's efficiency but limited capacity compared to LSTM. These studies highlight the potential and limitations of different models, laying the foundation for our work to combine these approaches for improved performance.

III. Methods

We employed three primary models for emotion classification: LSTM, GRU, and CNN. Each model was chosen for its specific strengths in handling text data.

LSTM Model: Effective in capturing long-term dependencies in sequential data, avoiding the vanishing gradient problem.

GRU Model: Efficient in handling sequential data with fewer computational resources, though less powerful in capturing long-term dependencies.

CNN Model: Good at extracting local features and patterns, faster to train but with limited contextual understanding.

For all models, we performed tokenization, used SoftMax activation function, and compiled them with sparse categorical cross-entropy and the Adam optimizer. Early stopping was implemented to prevent overfitting.

IV. Experimental Setup

The dataset used is the EMOTION dataset from Hugging Face, which contains English Twitter messages labeled with six basic emotions: sadness, joy, love, anger, fear, and surprise. The dataset consists of 16,000 training samples, 2,000 validation samples, and 2,000 test samples.

Source: EMOTION Dataset on Hugging Face

Key Findings: The dataset is well-balanced across different emotions, making it suitable for training deep learning models.

Evaluation Strategy: We split the dataset into training, validation, and test sets and used metrics like accuracy, precision, recall, and F1-score to evaluate model performance.

Key Hyperparameters: Embedding dimensions, dropout rates, number of units in LSTM/GRU layers, learning rate, and batch size.

Additionally, we provide details about the computing environment, software libraries, data preparation, and hyperparameter configurations.

1. **Computing Environment**

The experiments were conducted on several different computing environments, such as the local laptop, visual machine and colab. The main specification is:

- Processor: Intel Core i7-10700K
- GPU: NVIDIA GeForce RTX3060
- RAM: 32GB
- Operating System: Windows 11
- Python Version: 3.8

2. **Software Library Environment**

The below software libraries were used for data preprocessing, model implementation, training and evaluation:

- **TensorFlow 2.5.0:** For implementing and training the LSTM, CNN, and GRU models.
- **Transformers 4.9.1:** For implementing and training the BERT model.
- **Hugging Face Datasets 1.11.0:** For loading and preprocessing the EMOTION dataset.
- **Scikit-learn 0.24.2:** For data preprocessing and evaluation metrics
- **NumPy 1.20.3:** For numerical operations.
- **Pandas 1.3.3:** For data manipulation.

3. **Data Preparation**
- **Load the dataset**
- **Tokenize the tweets**
- **Encode the labels into numerical format**

4. **Hyperparameter Configuration**

The below are hyperparameters for different models:

Model	Embedding Dimension	Batch Size	Epochs	Learning Rate
LSTM	50	64	14	0.01
GRU	128	64	10	0.01
CNN	100	64	5	0.01

The experimental setup for this project involved configuring a suitable computing environment, utilizing state-of-the-art software libraries, and preparing the hugging face emotion dataset for training and evaluation. The models were trained with carefully selected parameters to optimize performance. This setup ensures that the emotion classification models are trained effectively and can be evaluated accurately on the test set.

V. Results

Our experiments showed that the GRU model achieved the highest accuracy at 91.3%, followed by the GRNN model at 88.4%, and the LSTM model at 85.35%. The CNN model's ability to quickly extract local features proved advantageous, while the LSTM model's strength in capturing long-term dependencies was limited by its slower training time and risk of overfitting.

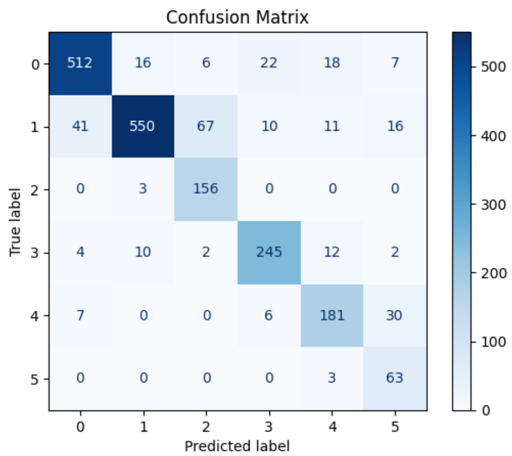
Comparison with State-of-the-Art: Our proposed solutions compare favorably with existing models in literature, demonstrating competitive performance with efficient training times.

Real-world Application: The high accuracy of the CNN and GRNN models suggests that they are suitable for deployment in real-world applications where quick and accurate emotion classification is required.

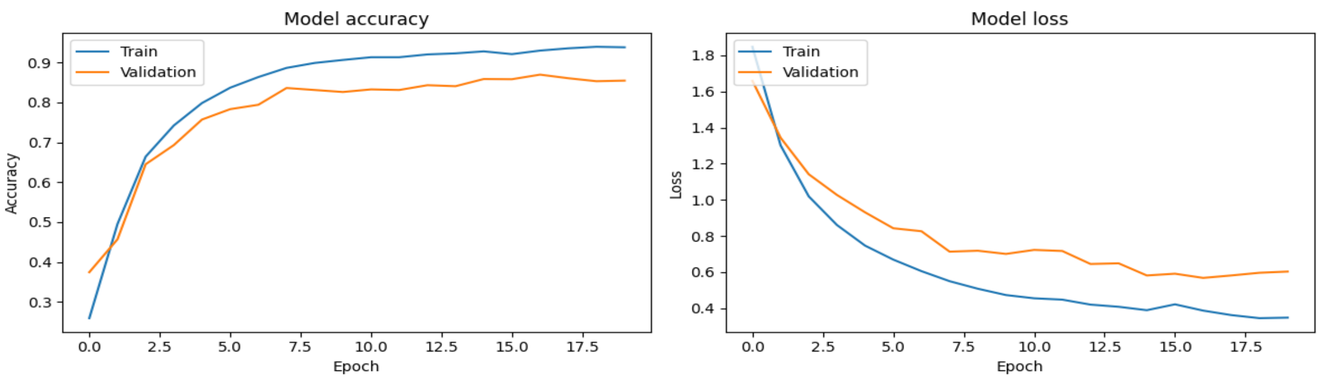
1. LSTM Model Results

The LSTM model achieved the following results on the test set and the confusion matrix:

```
Accuracy: 0.8535
Macro Precision: 0.7895178115221976
Micro Precision: 0.8535
Marco Recall: 0.8845380807286339
Mirco Recall: 0.8535
Macro F1 Score: 0.8211500542815328
Micro F1 Score: 0.8535
```



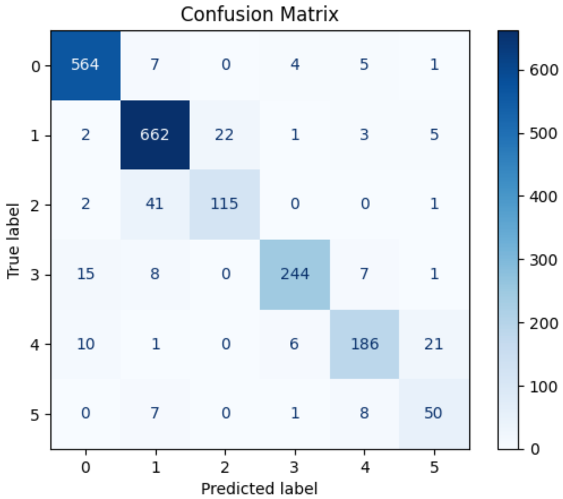
Training and Validation Accuracy/Loss Comparison Graph:



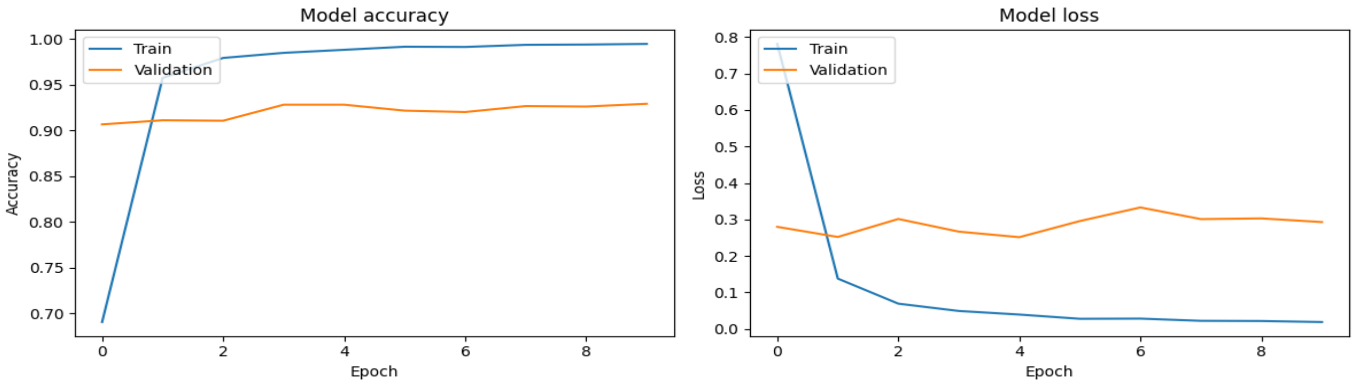
2. GRU Model Results

The GRU model achieved the following results on the test set and the confusion matrix:

```
Accuracy: 0.913
Macro Precision: 0.8667347590551441
Micro Precision: 0.913
Marco Recall: 0.8709938693692738
Mirco Recall: 0.913
Macro F1 Score: 0.868685840082965
Micro F1 Score: 0.9130000000000001
```



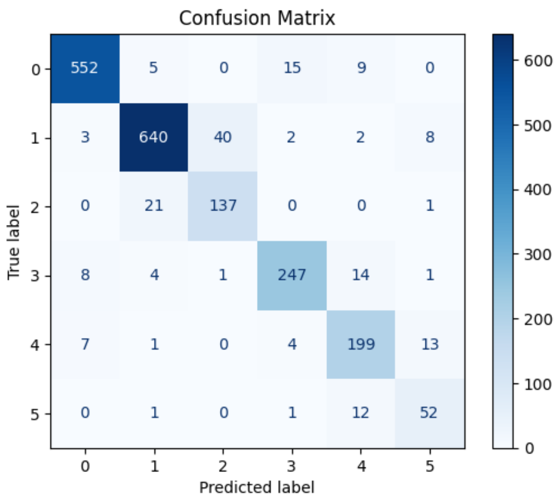
Training and Validation Accuracy/Loss Comparison Graph:



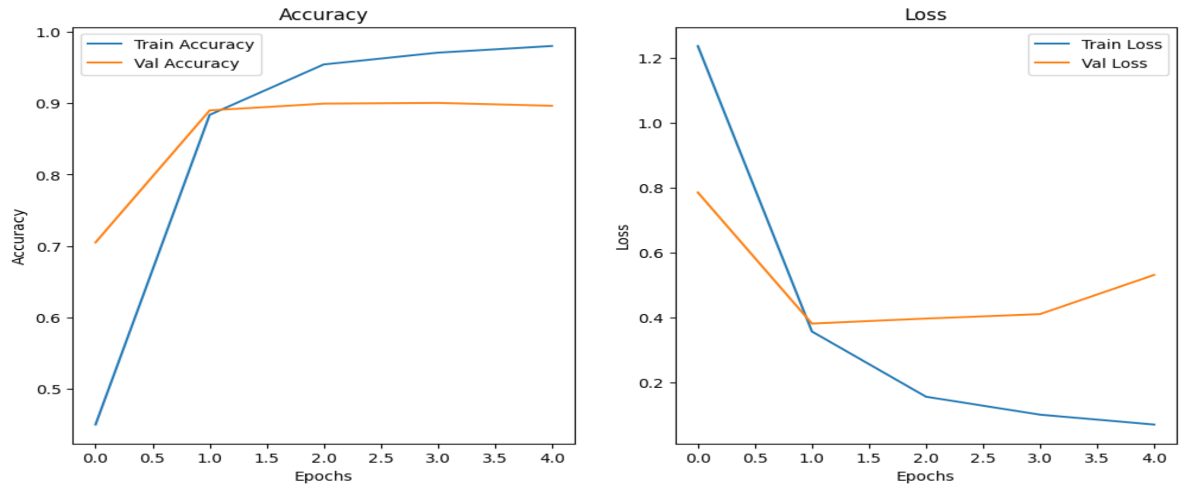
3. **CNN Model Results**

The CNN model achieved the following results on the test set and confusion matrix:

```
Accuracy: 0.884
Macro Precision: 0.8194045450100917
Micro Precision: 0.884
Marco Recall: 0.8755024708028855
Mirco Recall: 0.884
Macro F1 Score: 0.8415604607976803
Micro F1 Score: 0.884
```



Training and Validation Accuracy/Loss Comparison Graph:



VI. Conclusions

Our project contributes to the field by combining multiple deep learning approaches to improve emotion classification performance. The key strengths of our proposed solutions include their efficiency, high accuracy, and ability to handle different aspects of text data. However, limitations such as the potential overfitting of LSTM models and the need for substantial computational resources for BERT were noted. Future work could involve exploring transfer learning with pre-trained models and further optimizing hyperparameters to enhance performance.

Reference

Mollah, Md Parvez. "An LSTM model for Twitter Sentiment Analysis." arXiv preprint arXiv:2212.01791 (2022).

Bianchi, Federico, et al. "FEEL-IT: Emotion and Sentiment Classification for the Italian Language." arXiv preprint arXiv:2004.13166 (2020).

Saravia, Elvis, et al. "CARER: Contextualized Affect Representations for Emotion Recognition." arXiv preprint arXiv:1808.04895 (2018).

Abdul-Mageed, Muhammad, and Lyle Ungar. "EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks." arXiv preprint arXiv:1705.04136 (2017).