Name: Ayush Himmatsinghka **EE798R** Roll no: 210245

1 Improved Model Architecture

Our enhanced model captures both sentence-level and dialog-level context. We integrated the following components to improve accuracy and contextual understanding:

1.1 Sentence-Level Model

The sentence-level model processes individual sentences and classifies them into one of the four emotion categories. The model architecture is as follows:

- **Word-Level Encoding:** Each word in a sentence is embedded with a randomly initialized embedding layer of size 100. A Bidirectional LSTM with 128 hidden units processes the embedded words to capture forward and backward dependencies, producing a sentence representation that captures contextual meaning within each sentence.
- **Classification Layer:** The final sentence representation is passed through a dense layer with a softmax activation function to classify the sentence into one of the four emotions: *angry*, *happy*, *neutral*, and *sad*.

1.2 Dialog-Level Model

The dialog-level model incorporates several enhancements to capture interactions across multiple sentences within a conversation. Key improvements in this model include:

- **Bidirectional LSTM (BiLSTM):** Replacing the standard LSTM with BiLSTM enables the model to capture information flowing in both directions within a sentence, leading to richer sentence representations that provide a better basis for dialog-level context.
- **Multi-Head Attention Mechanism:** This layer helps the model focus on different parts of the dialog, attending to relevant utterances based on the context. This multi-headed approach allows the model to analyze complex interactions between sentences, especially beneficial in multi-turn conversations.
- **Attention-Based Pooling:** This mechanism aggregates the dialog information by focusing on the most critical parts, allowing the model to enhance its representation of the overall conversation context. Attention-based pooling ensures that the model prioritizes key dialog elements for classification.

2 Implementation Details

Our implementation, using TensorFlow and Keras, includes:

- **Embedding Layer:** Size 100 with a vocabulary size of 5000. This layer transforms word indices into dense vector representations, serving as the initial input to the model.
- **BiLSTM and GRU Layers:** The BiLSTM captures word-level context, while the GRU layer captures relationships across sentences in the dialog-level model.
- **Multi-Head Attention and Pooling Layers:** Enhances focus on relevant dialog sections, allowing the model to weigh specific utterances based on their contribution to the emotional tone.

The dataset was split 80% for training, 10% for validation, and 10% for testing, with class weights applied to handle imbalance.

3 Results and Observations

3.1 Sentence-Level Model Results

The improved sentence-level model achieved an accuracy of 63%. Key metrics are shown below:

Class	Precision	Recall	F1-Score
Angry	0.70	0.66	0.68
Happy	0.69	0.57	0.62
Neutral	0.56	0.54	0.55
Sad	0.50	0.68	0.57

Figure 1: Enhanced Sentence-level classification report.

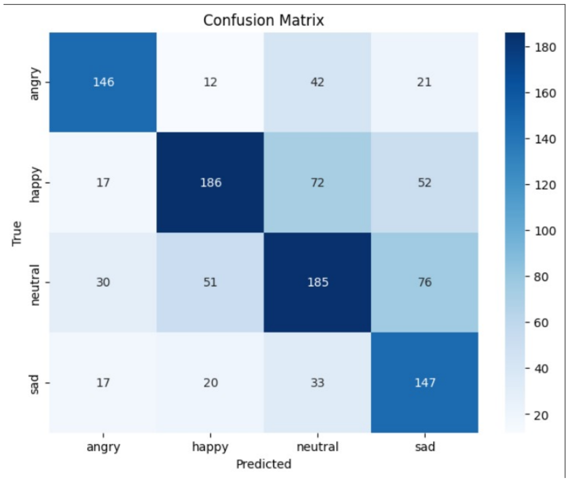

Figure 2: Sentence-level Model Performance

3.2 Dialog-Level Model Results

Using dialog context with the aforementioned enhancements improved accuracy. Results with two window sizes are summarized below:

- Window Size 5: Achieved a classification accuracy of 67%, with improvements in precision for *angry* and *happy* classes. The model benefits from the context provided by a smaller window, enhancing emotion differentiation across sentences.
- Window Size 10: Achieved best results with 72% accuracy, leveraging broader dialog context to capture subtle emotion shifts. This larger window allows the model to better understand the flow and changes in emotional tone across conversations, significantly improving recall for complex emotions.

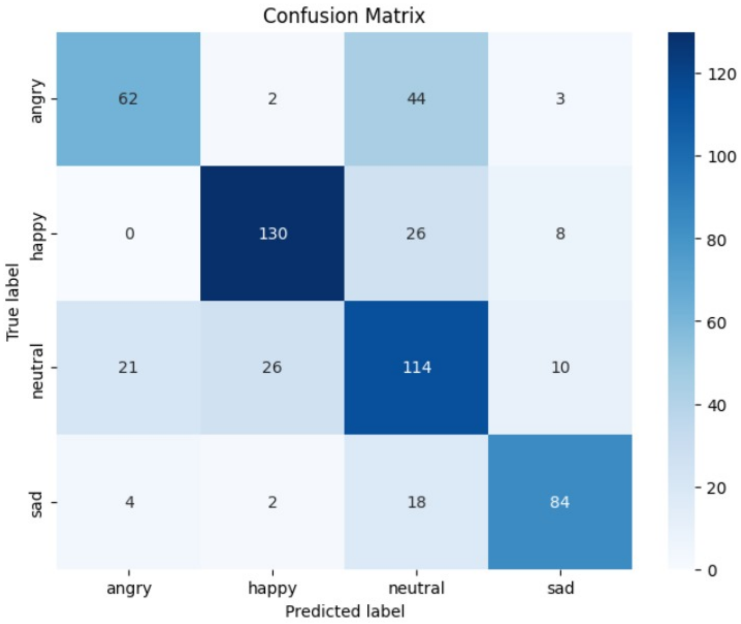

Figure 3: Dialog-level Model Performance (Window Size 10)

4 Conclusion

The improved Self-Attentive Emotion Recognition Network (SERN) effectively leverages self-attention and hierarchical encoding to capture nuanced emotional context across dialogs. The model demonstrated significant improvements, particularly in dialog-level context, achieving 72% accuracy with a window size of 10. These enhancements make SERN suitable for real-time sentiment analysis applications in online social networks, enabling a deeper understanding of user emotions.