Diverse Approaches to Emotion Detection

CSE 4120: Technical Writing and Seminar

Submitted To:

Dr. K. M. Azharul Hasan
Professor

Sunanda Das
Assistant Professor

Submitted By:
MST. Rubaiya Raktin Raha

Roll: 1907111

Department of Computer Science and Engineering
Khulna University of Engineering & Technology, Khulna

Abstract

This report attempts to look into various ways to detect emotion. It represents a comparative analysis of three research papers on emotion detection for text data or touch interaction using different techniques. The first paper applies transfer learning with BERT for textual emotion detection. The second paper integrates emoticons and text from micro-blogs, and the third paper uses touch interactions during text entry on smartphones. Considering their different methodology and implementation fields, their result has been analyzed and recommendations have been made.

Contents

List of Figures

List of Tables

1 Introduction

Emotion is a key in human interaction, impacting communication, decision, and well-being. There is no way to ignore or set aside emotions in human life because we, humans, use these emotions to communicate with each other, or make decisions. Emotions can be expressed in various ways, it could be by speech, text, or body gestures.

Emotion detection from text or bodily gestures is not that easy, even for humans themselves. However, with increasing interactions between humans and computers, have led to the necessity of detecting emotions to give humans better services in many aspects.

With the increase in digital communication availability, accurately interpreting user's emotions has become increasingly important. Various methodologies have even been proposed to tackle this problem, each with its unique approach and challenges. This report analyzes three distinct approaches to emotion detection to highlight their methodologies, effectiveness, and areas for improvement. By comparing different techniques, we aim to provide insights into their relative strengths and weaknesses and suggest possible methods for certain situations.

2 Problem Statement

The growing use of text-based communication and user's dependency on digital communication system has led to an increase interest in emotion detection. The significance of mention detection is huge. Detecting emotions can improve customer service, User experience, mental health support, marketing, personalized recommendations, etc.

3 Review of Literature

Various approaches have been explored in the field of emotion detection from text data, building on traditional machine learning and evolving towards more advanced deep learning models. Here are some of the approaches that have been mentioned in those papers.

3.1 Traditional Machine Learning Approaches

- **Lexicon-Based Methods**: Early works used predefined emotion lexicons to detect emotions in text. This often depended on the presence of specific words to infer emotions which was not very efficient.

- **Naive Bayes and SVM**: Machine learning algorithms such as Naive Bayes and Support Vector Machines(SVM)[8] were applied to text classification tasks, including emotion detection. They couldn't capture complex patterns in data and also required manual feature extractions. They also can't capture the sequential nature of text[1]

3.2 Deep Learning Models

- **LSTM and GRU**: Recurrent Neural Networks (RNNs)[3], particularly Long Short-Term Memory (LSTM)[7] and Gated Recurrent Units (GRU)[6] were introduced to capture the sequential nature of text. These models could learn contextual information but were computationally expensive and required large datasets for training. They are also prone to overfitting.

- **Bi-LSTM and SENN** Bi-LSTM is a powerful neural network architecture[10]. Instead of processing sequences only in one direction (e.g., left-to-right), they process them in both directions simultaneously. But it increased computational requirements. SENN includes both CNN and Bi-LSTM[2].

3.3 Multimodal Approaches

Emotion detection has also been explored with other data modalities such as audio, video, and touch interactions[9]. Multiple studies have aimed to determine emotion states based on smartphone activities (like SMS, call pattern, voice data)[4], additional sensor readings (like wrist sensor, skin conductor)

3.4 Transfer Learning

BERT and its variants, such as TinyBERT and DistilBERT, have been fine-tuned for various tasks, including emotion detection, demonstrating superior performance even with limited labeled data[5].

4 Methodology

Each of the three studies employs a unique methodology for detecting emotions from text data, leveraging different data sources, preprocessing techniques, and machine learning models. The methodology of each paper is discussed below:

4.1 Paper 1: Textual emotion detection utilizing a transfer learning approach

This paper utilizes the BERT model, a powerful pre-trained language model, for emotion detection. The methodology includes:

- **Data collection**: Collected Wang dataset (2.5 million tweets, 7 emotions) and MELD dataset (16,000 utterances from TV series Friends)

- **Data Preparation**: In this step, stopwords and punctuations were removed from texts and converted texts into lowercase. Tokenized the text using BERT's word piece tokenizer. Truncated sentences to a maximum length of 160 tokens.

- **Model Architecture**: EmotionalBERT is a model based on the pre-trained BERT, which utilizes only the encoder part of the transformer architecture. BERT is pre-trained on a large corpus of text. EmotionalBERT incorporates this pre-trained BERT and fine-tunes it for emotion detection tasks.

  The input sequence includes a special token [CLS] at the beginning for classification purposes. The model consists of 12 transformer blocks, each containing a 768-dimensional hidden layer and a 12-head self-attention mechanism. The final hidden state of the [CLS] token is passed through a feed-forward neural network, followed by a softmax layer to classify the emotions in the text.

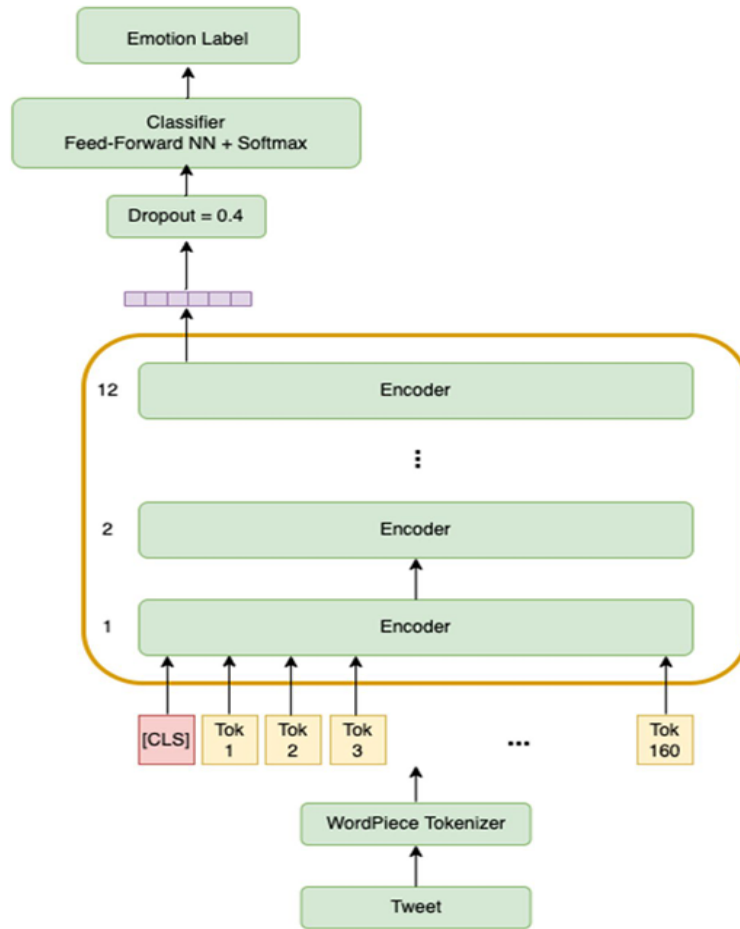Figure 4.1: The EmotionalBERT model architecture

- **Model Training**: Three sets of data amounts are used for training: 100,000, 250,000, and 500,000 tweets from Wang dataset.

- **Evaluation**: Comparing the performance of EmotionalBERT with traditional RNN-based models like LSTM and GRU on the same datasets. The model demonstrates superior performance, especially with smaller training datasets.
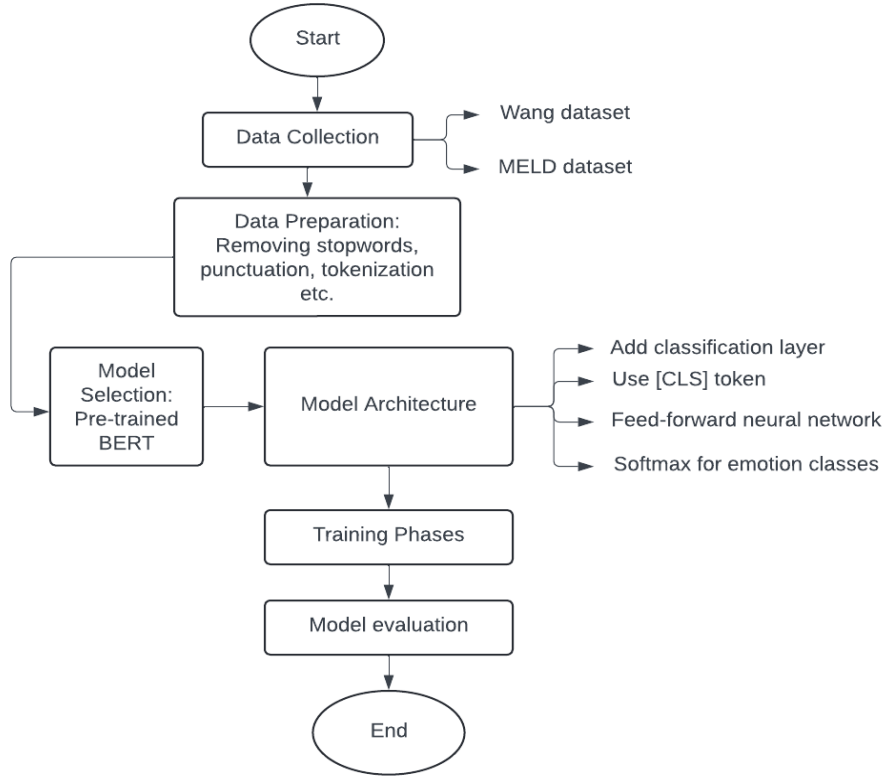
Figure 4.2: Methodology of paper 1

4.2 Paper 2: Emotion Detection from Touch Interactions During Text Entry on Smartphones

- **Data collection**: A diverse group of participants was recruited to ensure a wide range of typing behaviors and emotional responses. Participants included individuals from various age groups, professions, and typing proficiency levels.A custom application running on smartphones (Touchsense) was used to record the touch inputs during text entry that we the authors developed to record the detailed touch interactions without interfering with the natural typing behavior of the participants. We had them perform predefined typing and free text while experiencing induced emotional states with methods such as watching emotional videos, recalling personal experiences, and engaging with emotion-eliciting content. The study used features of touch interaction (such as typing speed– the time to press between two successive keys, touch pressure, touch duration, and swipe dynamics) relentlessly recorded by the app.

- **Feature extraction**: Key features such as inter-key delay(time interval between consecutive key presses), pressure(duration of key press and time taken to release the key), and swipe dynamics were extracted from the raw touch data.

- **Machine learning Model**: The processed data was split into training, validation, and test sets to ensure robust model evaluation. Then a personalized Random Forest-based emotion detection model was used to identify four emotion states(happy, sad, stressed, and relaxed). The models were trained using touch interaction features such as typing
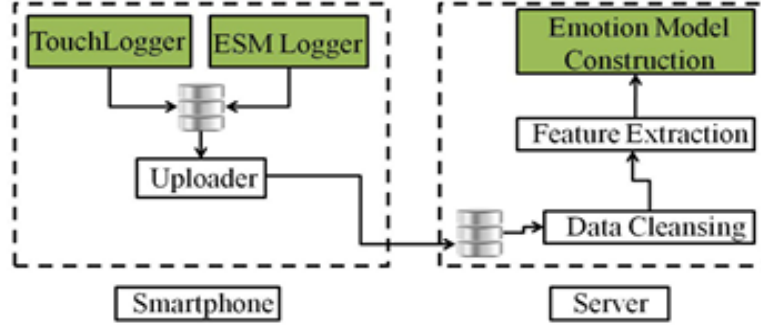
8

Figure 4.3: TouchSense architecture

(a) TouchSense Keyboard.
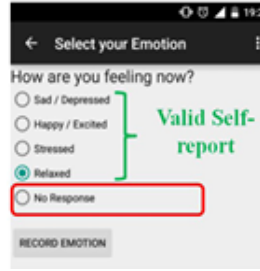
(b) TouchSense self-report collection interface.

Figure 4.4: TouchSense prompt

speed, touch pressure, touch duration, and swipe dynamics. Hyperparameter tuning was performed using grid search and cross-validation to optimize the model parameters for the best performance. During training, the models were continuously monitored using the validation set to prevent overfitting.

- **Evaluation**: The trained models were evaluated using various performance metrics, including accuracy, precision, recall, and F1-score, to assess their effectiveness in emotion detection. The performance of the proposed models was compared with baseline models, such as logistic regression and naive Bayes classifiers, demonstrating significant improvements. The evaluation showed that the models could accurately classify emotions based on touch interaction data, validating the effectiveness of the proposed approach.

Figure 4.5: Methodology of paper 2

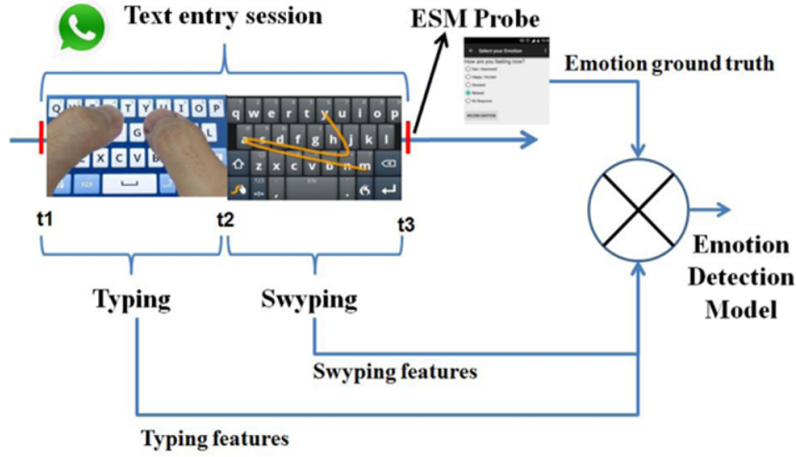## 4.3 Paper 3: Emotion Recognition from Microblog Managing Emoticon with Text and Classifying using 1D CNN

This paper integrates text and emoticons from microblogs to improve emotion recognition accuracy. The methodology includes:

- **Data Collection**: The Twitter API was used to collect a large dataset of tweets containing both text and emoticons. The dataset contained tweets in multiple languages and diverse geographical location, only tweets of English language was sorted and taken.

- **Data Preprocessing**: The raw tweets were cleaned to remove noise and irrelevant content. This included Extracting and removing hyperlinks to prevent them from influencing emotion detection and cleaning the tweets by removing hashtags, mentions, and other non-essential symbols. Emoticons in the tweets were converted to their corresponding words using a predefined lookup table. This ensured that emoticons contributed well enough in the classification as it as a whole dealt with text data. The overall text was normalized as well by converting all text to lowercase, removing punctuation, etc.

Emoticon	Word meaning	Emoticon	Word meaning
😀	Grinning face	😠	Angry face
😄	Grinning face with smiling eyes	😡	Pouting face
😁	Beaming face with smiling eyes	😤	Face with steam from nose
☺	Smiling face	🤬	Face with symbols on the mouth
😭	Loudly crying face	😍	Smiling face with heart-eyes
😢	Crying face	🥰	Smiling face with hearts
🥺	Pleading face	😘	Face blowing a kiss
☹	Frowning face	😚	Kissing face with closed eyes

Figure 4.6: Emoticons with corresponding word meanings

- **Feature Representation** The text and emoticons were converted into sequences of integers using integer encoding. Each unique word and emoticon was assigned a unique integer value. The sequences were padded to a fixed length to ensure uniform input size for the CNN model. Padding was done using zeros for shorter sequences and truncating longer sequences.

- **Model Training**: The 1D CNN architecture was designed to process the sequential text data. It begins with a Input layer that accepts padded sequences of integer-encoded text. Then it converts integer sequences to dense vectors using a pre-trained word embedding layer, which captures the semantic meaning of the words.Multiple convolutional layers are applied to detect local patterns in the text. Max-pooling layers reduce the dimensionality of the feature maps while retaining the most important features.The output from the convolutional and pooling layers is then fed into fully connected layers that perform high-level reasoning on the extracted features. Finally, the model includes an output layer with a softmax activation function to classify the text into different emotion categories, providing a probability distribution over the possible emotions.Hyperparameters such as the number of filters, kernel size, pooling size, and dropout rate were optimized using grid search and cross-validation.
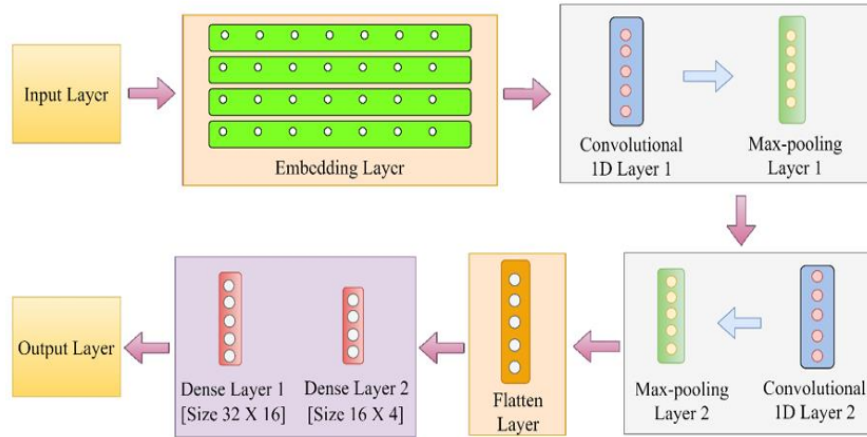
Figure 4.7: 1D CNN architecture of the proposed ER

- **Evaluation**: The 1D CNN model was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score to assess its performance. We tested the model on a held-out test set and it achieved a high classification accuracy of 88.0%, significantly outperforming baseline models that used only textual data without emoticons. According to the confusion matrix, the model accurately classified most emotions, with minor misclassifications primarily occurring between similar emotions. To further validate the model's effectiveness and generalizability in emotion recognition from microblogs, we conducted cross-validation and ablation studies which confirmed its robustness

Figure 4.8: Methodology of paper 3

4.4 Comparative Methodology Overview

Here is a comparative discussion of the three papers methodology

Table 4.1: Comparison of methodologies used in different papers.
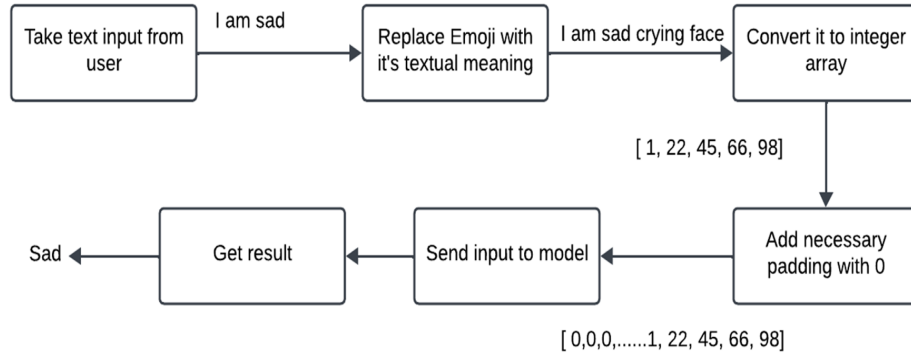
Component	Paper1	Paper2	Paper3
Data Source	Text Data from tweets	Touch interaction data from smartphone's keyboard	Microblog data, including text and emoticon
Data Processing	Stop words removal, Tokenization, Lowercasing	Typing sessions and swiping sessions (e.g., typing speed, touch pressure)	Cleaned text, Usernames removed, Hashtags converted to text
Model Construction	EmotionalBERT with BERT's pre-trained model and softmax layer	Personalized machine learning model using touch-based data	1D Convolutional Neural Network (CNN) with multiple layers
Experiments	Experiments with varying training data amounts	3-week study with 22 participants	Data split into training and test sets, followed by CNN training

5 Result Analysis

A comparative analysis of the results of these three papers has been done on the basis of a common parameter and that is accuracy.

Paper 1: Transfer Learning for Textual Emotion Detection

- Achieves 60.87% accuracy using the IEMOCAP dataset.

- Significantly outperforms current state-of-the-art models in real-time emotion prediction.

Paper 2: Emotion Detection from Touch Interactions

- Obtains an average accuracy (AUCROC) of 73% (std dev. 6%) and a maximum of 87% in a 3-week in-the-wild study.

Paper 3: Emotion Recognition from Microblogs with 1D CNN

- Achieves 88.00% accuracy on the Twitter dataset and outperforms every other model that has been tested with it.

6 Findings and Recommendations

Here the strength and weaknesses of each of the paper's approaches is highlighted, along with some other differences

Table 6.1: Comparative discussion

Aspect	Paper1	Paper2	Paper3
Strength	Transfer learning: using a pre-trained model, allows better performance, Very efficiency, fewer training epochs, Multilingual potential	Non-Intrusive and natural approach Doesn't require explicit user input High accuracy: 73%, Gives realistic insights	Combined Text and Emotion, High accuracy: 88%, Robustness and scalable, Outperforms every existing model
Weaknesses	Model complexity: Difficult to understand, Dataset Dependency, Emoticon Limitation	Limited emotional state(4), Small sample size, Not generalized	Emoticon Ambiguity: can have multiple meaning, Limited Emotional state(4)
Performance Metrics	F1-score, accuracy, and comparisons with RNN-based models	Average accuracy (AU-CROC), F-score	Accuracy metrics, comparisons with other emotion detection methods
Applications	Text-based sentiment analysis, Social media monitoring	Emotion detection in smartphone-based applications	Social media sentiment analysis, Microblog emotion detection

This comparative analysis provides valuable insights into the strengths and weaknesses of three distinct approaches to emotion detection. Paper1 uses transfer learning that results in better performance and it is resource efficient as well by using a pre-trained model. That said, the complexity of its model and the requirement for specific datasets are problematic. What made Paper2 different was that it took a completely non-intrusive, natural approach and got very close to the speed (and faster in a few cases) without requiring any explicit user input. Nonetheless, it is limited in representation of emotional state and suffers from a small sample size, and is therefore of questionable generalizability. In contrast, we use this together using text with emotion incorporated in Paper 3 which achieves the highest in terms of accuracy and scalability compared to other models. However, it still struggles with unclear emoticon meanings and a limited emotional state scope.

Recommendation: Based on the strengths and weaknesses, the recommended approach for emotion detection from text data is the use of transfer learning with the BERT model as demonstrated in Paper 1. This approach shows significant advantages in terms of accuracy, efficiency, and scalability, making it a robust solution for emotion detection tasks across various datasets and contexts.

The third paper uses a 1D CNN and shows better accuracy and effectiveness, especially in recognizing emoticons. However, for smartphone-based applications that need real-time emotion detection, the second paper's method of using touch interactions is more efficient. On the other hand, if the task involves analyzing text data with computational resources or working with multiple languages, the transfer learning approach from the first paper is more promising. Each paper has its own strengths and is designed for different application

7 Addressing Course Outcomes and Program Outcomes

Table 7.1 represents the course outcome(CO).

Table 7.1: Course Outcomes (CO)

CO No.	CO Statement	Corresponding PO No.
CO1	Analyze, and summarize technical papers, demonstrating comprehension of the main ideas, key concepts, supporting evidence, and logical arguments presented.	PO2
CO2	Demonstrate proficiency in employing appropriate citation methodologies and referencing techniques, ensuring accurate and consistent citation of sources, and effectively avoiding plagiarism.	PO8
CO3	Develop effective presentation skills, applying guidelines and techniques to deliver clear, organized, and engaging technical presentations, incorporating visual aids and adapting to different audience needs.	PO10
CO4	Write scientific reports in a clear, concise, and structured manner, following the conventions of scientific writing, and effectively communicating technical information.	PO9

Table 7.2 represents the program outcome(PO).

Table 7.2: Program Outcomes (POs)

PO	Description
PO2:Problem Analysis	Analyze and summarize research papers on Emotion Detection, demonstrating comprehension of key concepts, methodologies, and findings.
PO8: Ethics	Demonstrate proficiency in ethical research practices, ensuring accurate citation and referencing, and avoiding plagiarism in the context of summarizing and analyzing Emotion Detection.
PO9:Individual and Teamwork	Collaborate effectively within a class to analyze multiple aspects of Emotion Detection, contributing individual insights and integrating them into a cohesive report.
PO10:Communication	Develop and deliver clear, organized, and engaging presentations on Emotion Detection. Write scientific reports that effectively communicate technical information in a structured manner.

8 Addressing Complex Engineering Activities

Table 8.1 represents the activities regarding complex engineering.

Table 8.1: Characteristics of Complex Engineering Activities (A1 to A5) and their specific applications to Emotion Detection

Attribute	Characteristics of Complex Engineering Activities
Range of resources (A1)	Managing and integrating diverse resources such as research papers, datasets, computational tools to analyze Emotion Detection in various methods
Level of interaction (A2)	Addressing and resolving conflicts between different approaches and methodologies in Emotion detection, and synthesizing these into a coherent analysis.
Innovation (A3)	Studying innovative techniques and advanced ED models creatively to enhance the understanding and performance of Emotion Detection, such as leveraging deep learning architectures like Transformers or CNN.
Consequences for society and the environment (A4)	Evaluating the broader implications of Emotion Detection on society, such as ethical considerations, data privacy etc.
Familiarity (A5)	Venturing beyond established methods by employing new, principle-based approaches to tackle novel challenges in Emotion Detection.

9 Conclusion

Detecting emotions from text is crucial across many fields, from customer service to mental health support. It improves interactions, aids in sentiment analysis for marketing, and even helps assess student engagement in education. Social media, it enables better understanding of public opinion and enhances content moderation. Ultimately, accurate emotion detection enhances empathy and fosters safer and more effective communication online and offline.Emotion detection from text data is a challenging yet crucial task in NLP. This report focuses on three papers and how they apply different ways to detect emotions and how effective the methodology is representing human emotions accurately. Incorporating multimodal data, using emoticons and transfer learning are all promising methods, that could largely improve the performance of mood detection.

References

[1] Francis A Acheampong, Henry Nunoo-Mensah, and Wei Chen. Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829, 2021.

[2] Erdenebileg Batbaatar, Meijing Li, and Keun Ho Ryu. Semantic-emotion neural network for emotion recognition from text. *IEEE Access*, 7:111866–111878, 2019.

[3] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.

[4] Matteo Ciman and Katarzyna Wac. Individuals' stress assessment using human-smartphone interaction analysis. *IEEE Transactions on Affective Computing*, PP(99), 2016.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: pre-training of deep bidirectional transformers for language understanding. In *Association for Computational Linguistics*, 2018.

[6] Rahul Dey and Fathi M Salem. Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, 2017.

[7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[8] Y. Tang, Y.Q. Zhang, N.V. Chawla, and S. Krasser. Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1):281–288, 2009.

[9] Matthias Trojahn, Florian Arndt, Martin Weinmann, and Frank Ortmeier. Emotion recognition through keystroke dynamics on touchscreen keyboards. In *ICEIS (3)*, pages 31–37, 2013.

[10] Wenjie Wei et al. Microblog sentiment classification method based on dual attention mechanism and bidirectional lstm. In *Chinese Lexical Semantics: 20th Workshop, CLSW 2019, Beijing, China, June 28–30, 2019, Revised Selected Papers*, volume 20. Springer International Publishing, 2020.

Publication Details

Table 9.1: Publication Details

Serial No	Title	Author Details	Source of Publication	Published year
1	Textual emotion detection utilizing a transfer learning approach	Mahsa Hadikhah Mozhdehi, AmirMasoud Eftekhari Moghadam, Hossein Habib M. Aghdam	The Journal of Supercomputing, Volume 79, Pages 13075–13089, Springer	2023
2	Emotion Recognition from Microblog Managing Emoticon with Text and Classifying using 1D CNN	Md. Ahsan Habib, M. A. H. Akhand and Md. Abdus Samad Kamal	Journal of Computer Science (JCS), Pages 1170-1178, International Journal of Human-Computer Studies, Elsevier	2022
3	Emotion detection from touch interactions during text entry on smartphones	Surjya Ghosh, Kaustubh Hiwarea, Niloy Gangulya, Bivas Mitraa, Pradipta De	International Journal of Human-Computer Studies Volume 130, Pages 47-57, Elsevier	2019