Grenze International Journal of Engineering and Technology, June Issue

Critical Analysis of Limitations in Contemporary Sentiment Analysis Techniques: A Foundation for Multimodal Deep Learning Approaches

Damini Bhola[1] and Rupal Gupta[2]
[1]Research Scholar, College of Computing Sciences & Information Technology Teerthanker Mahaveer University, Moradabad, U.P India
Email: damini.scholar@tmu.ac.in
[2]Associate Professor, College of Computing Sciences & Information Technology Teerthanker Mahaveer University, Moradabad, U.P India
Email: rupal.gupta07@gmail.com

***Abstract*— Sentiment analysis has grown in importance in the field of natural language processing (NLP) due to its use in marketing, political predictions, public opinion research, and many other fields. However, the most recent techniques employed for sentiment analysis are still unsatisfactory. The limitations arise in sarcasm detection, classification of the context as culturally nuanced or domain dependent, reliance on text data. In this paper we study different approaches used for sentiment analysis including lexicon-based methods and deep learning techniques focusing first on their strengths and weaknesses when applied to real-world datasets and benchmark studies. The results reveal that attempts at capturing complex emotional expression failure strongly suggesting simplistic approach relies too heavily on a single modality.**

***Index Terms*— Natural Language Processing, Deep Learning, Multimodal Analysis, Limitations, Sarcasm Detection, Emotion Recognition, Contextual Ambiguity.**

I. INTRODUCTION

As we move into the digital age with users outputting content in real time, the necessity to understand and classify human emotion is of increasing significance. Sentiment analysis or opinion-mapping is a sub-field of natural language processing (NLP) that attempts to calculate the sentiment polarity contained in text. It is actively used for customer feedback analysis, predicting political events, finance market forecasting, public health monitoring, and analyzing platforms for social media. Companies can shape marketing campaigns based on public perception, politicians likely change their campaigns based on mood analysis, and it is also used in conversational agents to support user experience. Although there have been tremendous advances in sentiment classification using deep learning technologies, there are still many limitations and negative aspects that prevent them from future potential. Generally, classification sentiment analysis techniques have traditionally been rule-based (lexicon-based), classical machine learning approaches, or modern deep learning approaches. Lexicon-based methods employ pre-defined dictionaries that contain opinionated words to provide sentiment scores based on those words and their intensity in the text.

Grenze ID: 01.GIJET.12.2.517_20

While simplistically these methods are easy-to-understand and easy to implement, these approaches have rigidity in nature. They are not dynamic or flexible to changes in language over time, contextual changes, or specific vocabulary terminology for specific fields. Machine learning methods are interested in extracting notable patterns that can be learned for the sentiment classification context from labelled or known datasets and would employ Support Vector Machines (SVM), Decision Trees, or even Naïve Bayes classifiers. These types of methods need manual feature extraction, which is a disadvantage, and inherent limitations in recognizing long-range dependencies or intertextual meaning. The rise of deep learning has led to a renewed vigour for ways to analytically measure the sentiments of individuals. These include models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTMs), and, more recently, BERT (Bidirectional Encoder Representations from Transformers). These models do a better job than previous methods available in the academic literature: they improved performance because they feature the ability to automatically learn abstract features in intricate sequential data, which is a requirement for human languages as shown in Fig.1. However, even the state-of-the-art systems have their limitations. They are still limited to the components of communication, and they do not take into consideration vocal tone, facial expressions, other visual context, or body language, which are components imbued with emotion, when humans think and communicate. Many current models only have textual inputs.

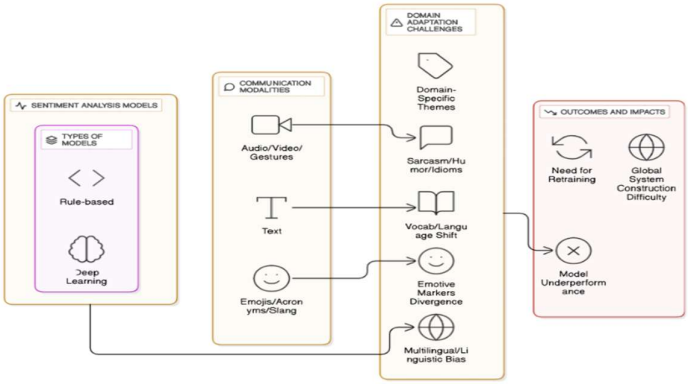

Figure 1. Domain Adaptation Challenges in Sentiment Analysis

Texts, audio, and visuals depict sentiments in their totality. An audio inflection and sarcastic eye roll, on the words 'That's just great' are a long way from the buoyed words. Current unimodal models use solely texts, and deducing vex, inverse vex, comedy, or even more nuanced expressions that depend on this context is inherently flawed. Inadequate ability to provide adequate 'context' decrement errors in the accuracy classification accuracy for sentiment. Contextual provenance deficiencies highlight yet another gap analysis claim about presumptions in models for affective sentiment. An example of "incorrect domain" when analyzing for sentiment would be a model developed for movie reviews, to then go on to analyze tweets about politics and e-commerce store product ratings. This is the domain adaptation problem arising from ambiguity discordance gaps within vocab and language cross shift's theme motion divergence emotive marks in multiple contexts. Deep learning models are more adaptable than rule-based models but domain change while not the only limitation still requires a lot of re-training and re-working to be meaningful. Adding to the complexity of building global sentiment analysis systems is to adjust for lack of low-quality annotated datasets for languages and cultures which are less represented, which compounds the issue further. The linguistic diversity of any given culture represents a monumental challenge. The differences of how most emotive expressions vary by area where idioms may culturally reflect the same sentiment may well require more contextual understanding to appreciate. Most present-day systems face the issue of linguistic bias which hinders the performance of sentiment analysis in multilingual, multicultural, pedagogical and discourse settings. The variation in emotive meaning as shown with the use of slang, emojis or acronyms also makes ascertaining sentiment in informal communication much more challenging too; then our confines which were already socially limited to text can build context or legibly expand outward again post only that single lens. By going beyond with rich data streams even of a text and adding audio, video related schedules and gestures might be its solution for the one-off dilemmatic implications that are still faced in state-of-the-art systems today. Sentiment analysis informed multimodal models acknowledge context to aid in better quality human emotion classification will lead to more accurate understanding of reasoning.

II. LITERATURE REVIEW

The domain of sentiment analysis has been completely transformed from original text classification to multimodal deep learning frameworks. The original text classification was direct text classification with simple rule-based and machine learning methods. The development of transformer-based models, such as BERT, has drastically improved performance at both contextual understanding and classification accuracy. Kale and Mendhe [1] mentioned how these transformer-based models surpass traditional recurrent neural networks by providing a deeper, richer linguistic context. In contrast, Vanshika et al. [2] conducted an exhaustive review of sentiment analysis approaches, datasets, and suggested limitations that often miss more than they understand, notably that we need models that can contextualize and disambiguate meaning across multimodal data sources. The research community is overcoming this lack of disambiguation by employing multimodal sentiment analysis (MSA), or combining the textual, audio, and visual modalities to layer emotional perception. Kousika et al. [3] proposed an MSA deep learning-based (multimodal) sentiment analysis that showcased improved emotional intelligence by compiling complementary modalities. Likewise, A. M et al. [4] summarized the difficulties and trends in emotion recognition by leveraging multimodal data, while acknowledging issues with temporal alignment, modality synchronization, and feature-level fusion. In the visual modality space, Vashishth et al. [5] conducted a systematic review of computer vision methods for recognizing human emotion using facial expressions, emphasizing the importance of the visual aspect of affective computing. Based on the above, Pathak et al. [6] also included both facial and text features for detecting depressive disorders with evidence of multimodal features applications in healthcare-related sentiment detection. Audio-based sentiment detection has also developed. Kim et al. [7] utilized both text and speech inputs by proposing a Korean-language sentiment analysis system (MSDLF-K) of multimodal model based on contextualized language embedding and BERT, which used language-appropriate fusion models for better sentiment detection. The use of explainable models is critical for establishing trust in health sector AI systems. Al-Tameemi et al. [8] utilized attention-based multimodal networks for fusing images and text inputs to produce models that provided better interpretability and improved classification accuracy. Maragathavali and Somaraju's [9] study on twitter sentiment detection also illustrated the use of fine-tuning and attention-based fusion of the modalities showing that strong fusion can be achieved dynamically by realigning the modality inputs and building fine-tuned convolutional neural network (CNN) models to consider the features of both modalities. Pandya [10] presented a unique framework based on large language models and deep learning methods for classifying movie reviews by illustrating the unique capability of domain-specific deep learning-based fine-tuning for extensive sentiment classification. In addition, and in line with other modality contributions, Sharma et al. [11] evaluated Brain-Computer Interface systems that allow users to encode cognitive signals for human-computer interaction, this work is not directly looking at sentiment but understanding the power of human feedback in real-time, with implications for sentiment analysis. Indhumathi and Fernandez [12] also presented an overview of the various approaches to sentiment analysis, datasets, and computational methods, providing a foundational overview of the field. Raghunathan and Saravanakumar [13] also highlighted further challenges related to contextual ambiguity, data imbalance, and ethical issues which need consideration before deploying sentiment models in the real-world landscape. Security and robustness should also factor into sentiment analysis in applications with sensitive data or real-time systems. A paper on Mobile Ad Hoc Networks (MANETs) in IJLTEMAS [14] clarified such a note of resilient and secure algorithms design, which is reasonable when applying sentiment analysis models in dynamic or sensitive privacy situations. Lastly, Zhao et al. [15] provided an extensive survey on multimodal sentiment analysis from the perspective of fusion strategies, perceiving one further compartmentalizing methods as early, late, and hybrid, and recommending adaptable architectural flexibility and certain observations for architectural types adapting to various data sources and application forms. Artificial Intelligence (AI) is playing an increasingly significant role in promoting a great level of data security, intelligent decision making and automation in all fields. Sharma and Kumar [16] stressed the importance of AI based techniques in enhancing the security and privacy of data in smart city arrangements by being able to guard against sensitive data in the framework of large-scale urban areas Ramesh. Similarly, Vikas et al. [17] made a hybrid D. B. N. as a fusion of Harris Hawks Optimisation algorithm to detect intrusions via Wireless Sensor Networks giving improved detection results and considerably low alarms. The introduction of AI into man to machine has shown excellent promise, for example, Somnathe et al, [18] suggested the deep leaning-based Brain Computer Interaction Framework to help a variety of people who have speech and or motor disabilities. This helped to illustrate practically the potential of the neural architecture helping to develop communication. Similarly, Reddy et al. [19] developed a Deep Learning framework model to identify encrypted and harmful network traffic and brought to a fore how AI techniques should be able to assure Cybersecurity and intelligent surveillance attack in complex digital environments.

III. PROPOSED METHODOLOGY

Sentiment analysis has traditionally been discussed in varied ways, each with its own strengths and considerable shortcomings. Lexicon-based analysis, like SentiWordNet and VADER, use manually produced dictionaries that assign sentiment scores to words. These methods are less computationally intense and are easier to interpret but will not be effective for context-dependent language, sarcasm, and domain-specific expressions of sentiment. They are brittle in real-world applications because they do not adapt to the dynamic nature of language, especially on informal platforms like Twitter and Reddit that are often, laced with colloquialism, and employ code-mixing. Machine learning approaches, such as Naïve Bayes, support vector machines, and logistic regression, use supervised learning to classify sentiment, from features such as n-grams, part-of-speech tags, and syntactic patterns. Machine learning is more adaptable than lexicon-based approaches, but they tend to rely on features extraction that isn't always helpful or they do not capture long-range dependencies in semantics. Machine learning approaches can also depend on being trained with task-specific data so to deploy the model in another domain, the model may need to be retrained with new features. Recent trends in deep learning (e.g., using CNNs, RNNs, LSTMs, or Transformers such as BERT) have represented a change in the state of the art in sentiment analysis, based on task performance metrics that deep learning wins in recent deep learning models can now learn high-level representations of text without engineered feature extraction and are robust to noise, with the additional ability to create text representations for a large number of tasks. Nevertheless, although these models produce strong results on benchmark datasets, they are still limited in three major ways: (1) they are mainly unimodal text models; (2) they are often slow to capture sarcasm, ambiguity, and contextual nuances; and (3) they rely on large, labelled datasets which may not exist for certain languages or domains. Additionally, they also lack interpretability due to their black-box nature, which is pertinent for applications in healthcare and finance. The proposed Multimodal Attention Fusion (MAF) model combines textual, audio, and visual data for better understanding of sentiment and emotions. Each kind of data is first processed separately using special-purpose models, then the three output feature sets are fused using an attention-based fusion layer (for instance the model will learn how much weight to give each modality in a given sample, e.g. to pay more attention to voice tonality when the corresponding sample text is ambiguous). Text information is made available by a transformer-based language model (BERT), the audio characteristics (i.e. pitch, energy, tone) are derived from analysis of the speech signals and visual features are made available from the video stream (i.e. facial expressions and movements, etc.). These combined features are passed to a classifier circuit that predicts the emotion or sentiment corresponding to a given sample (positive, negative, happy, sad, etc.). This method ensures that MAF focuses dynamically on how much weight to give to the best cues for says of each of the modalities, thus ensuring better results, both in terms of accuracy and of interpretability of output. The proposed framework outlines the below:

A. Multimodal Data Acquisition and Preprocessing

The model will be trained on publicly available multimodal sentiment datasets, for example CMU-MOSEI, IEMOCAP, and MELD which include synchronized streams of audio, visual and text. Text features will be pre-processed using standard NLP, which include tokenization, stop-word removal, lemmatization. The audio signal will be converted to either a spectrogram, or MFCCs, the visual features, will also be pre-processed using visual frames and facial landmark detection algorithms to extract facial regions of interest, and normalize to use the facial area.

B. Feature Extraction Specific to the Modalities

Specialized deep learning models will be used for feature extraction of the different modalities. Specifically:

- Text features will be extracted from Transformer-based models (e.g., BERT or RoBERTa) that provide contextual embeddings.
- Audio features will be extracted from Convolutional Neural Networks applied to MFCCs or spectrograms that can extract pitch, tone, prosody, and other dynamic properties of audio.
- Video and the extraction of visual information – specifically facial expressions and eye movements—will be extracted from pre-trained CNNs like ResNet or VGGFace, which doctors can use for emotion recognition.

C. Multimodal Fusion Strategy

We will incorporate ethological concepts of multimodal fusion via an attention-based fusion mechanism that allows for dynamic weighing of modality by context. Specifically, if there is indication of sarcasm in tone or expression, the model will weight auditory and visual modalities more. This strategy allows for the algorithm to resolve sentiment ambiguity better than a fixed late or early fusion approach.

D. Sentiment Classification Layer

The representation from the multimodal fusion will be fed into a fully connected neural layer regularized with dropout and batch normalization. The outer layer will have a SoftMax classifier for predicting the final sentiment classification: positive, negative, or neutral.

E. Model Training and Evaluation

The model will be trained using cross-entropy loss, using an Adam or RMSProp optimizer, and evaluated through several metrics including accuracy, F1-score, precision, recall, and confusion analysis. There will be additional emphasis on evaluating the performance of the model regarding sarcasm, irony, and neutral sentiment, especially where unimodal models tend to fail.

F. Explainable Integration

To alleviate the concerns of interpretability, Layer-wise Relevance Propagation (LRP) and SHAP (SHapley Additive exPlanations) are post hoc algorithms to explain to users what features were most impactful (text tokens, audio segments, or facial regions) within the model, thus increasing trust and usability in instances where these frameworks are applied in high-stakes contexts shown in Fig. 2. Overall, the proposed methodologies improve upon the existing methodologies and work systematically to alleviate some of the mentioned weaknesses through cross-modal fusions, attention to context, and interpretability in decision-making. The use of multimodal methods aims to advance the field of sentiment analysis and lead to improvements in accurate translation of meaning, context, and culture for real-world applications.

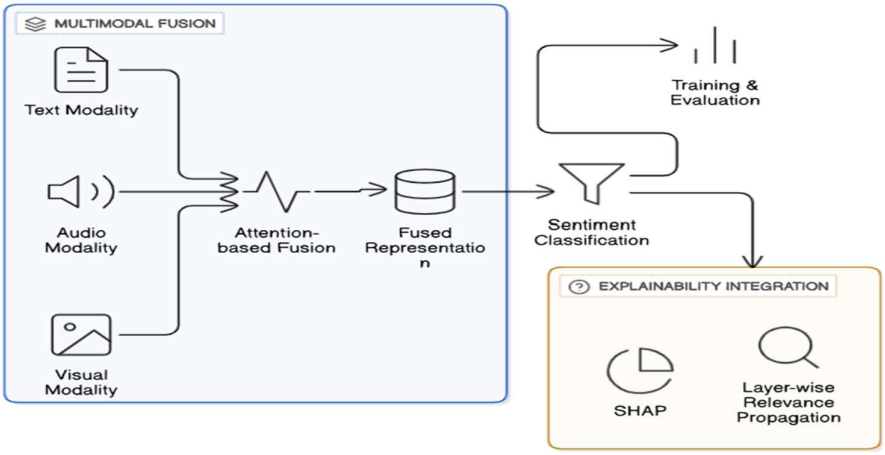

Figure 2. Multimodal Sentiment Analysis Architecture

IV. Result and Analysis

To evaluate and validate the efficacy of the proposed multimodal sentiment analysis framework, several studies were undertaken using benchmark datasets that contain textual, auditory, and visual modalities. The primary training and evaluating dataset were CMU-MOSEI (Multimodal Opinion-level Sentiment Intensity) which provides rich multimodal annotations and expresses a wide variety of sentiment levels. Other studies were conducted on MELD (Multimodal Emotion Lines Dataset) and IEMOCAP (Interactive Emotional Dyadic Motion Capture) to ensure that the methodologies generalize well across contexts and forms of communication. To determine the importance of each input modality and the contribution of each of the different components of the proposed Multimodal Attention Fusion (MAF) model, several ablation tests were conducted. In these tests individual modalities (text, audio or visual) were easily removed to measure their independent and interdependent effects corresponding to the performance of the overall model. When the visual features were removed there was a small drop in the degree of accuracy showing the supportive role that facial expressions and visual clues play in emotive detection, especially in cases where the voice or text signals were ambiguous. Removing the audio modality gave a more marked drop in performance showing that vocal tone and pitch, etc., carried important emotive data. However, the most marked effect was seen when the text modality was removed where this gave the largest drop in accuracy showing that the semantics of text remained the most salient factor for emotive interpretation of sentiment in multimodal data.

A. Data Types and Preprocessing

In multimodal sentiment analysis, the integration of multiple types of data is critical to capture a full range of human emotion. The aim for this study is to use three types of data: text, audio, and video data. All datasets comprised aligned text, audio and vision data depicting human emotional states and sentiments. As part of the preparatory step, each of these modalities was separately processed in such a way so as to yield clean and organized input.

Text data was tokenized with the aid of the BERT tokenizer; sentences were padded or truncated so that a uniform length of sequence would ensue. Audio recordings were converted into Mel-spectrograms and pitch features, for data-cleaning was attained, taking from speech the tone, energy, and rhythm. For the vision modality, faces were detected and aligned from each shot of the video with the aid of a face-detecting algorithm and resized to 224- x 224 pixels, so that the input data were uniform in size. The three modalities were then synchronized so that the associated speech features, of facial and textual modality, depicted the same moment in every speech utterance.

Training took place for 25 epochs employing the AdamW optimizer, the learning rate of the fusion network being 0.0001 and that of the text encoder being 0.00002. A dropout of 0.3 and early stopping were employed to prevent any possibility of overfitting. Training for this took place on an NVIDIA V100, with accommodated batch size of 16. All experiments were carried out repeatedly, randomly differing the random seeds on a number of occasions, with the aid of the results averaged with details retained of the reliability measures of average accuracy and F1 score. The training thus was organized so as to yield generalization of the model as such its data were effective across datasets employed in training and significant relationships thereby were able to be achieved of textual meaning with voice tone, or facial expression as between the variables thus of emotion and sentiments assigned to the textual meaning.

B. Comparative Models and Baselines

So as to fairly evaluate the effectiveness of the proposed multimodal sentiment analysis model, it is important to compare it against baseline models. The baseline benchmarks come in both unimodal architectures and multimodal architectures, as they represent varying levels of development of the methodologies and the categories of identified baselines.

Since this project proposes a process of extending sentiment analysis methodology by not only incorporating each modality of data available, but also combining those modalities in a meaningful way and not simply amplifying the possibility of finding correlations by combining modalities in time, it was necessary to put the confederated unimodal models against both the unimodal baseline examples of their respective data modalities (i.e., text-only, audio-only, and visual-only). As examples of those unimodal models, the text-only features were those models (i.e., Text-CNN, Bi-LSTM, BERT) for their history in utilizing deep learning frameworks to create a full-text textual matching model, or contextualized example in the case of BERT using a transformer-based approach.

C. Evaluation Metrices

To denote the each of the input modalities of the proposed Multimodal Attention Fusion (MAF) architecture as x_T, x_A, x_V representing textual, acoustic, and visual data respectively. These inputs are passed through corresponding feature extractors or encoders $E_T(\cdot)$, $E_A(\cdot)$, and $E_V(\cdot)$, producing latent embeddings by using equation (1):

$$h_T = E_T(x_T)\ ,\ h_A = E_A(x_A)\ ,h_V = E_V(x_V) \qquad (1)$$

Each modality embedding h_m (where $m \in \{T, A, V\}$) are structured in terms of projection into a shared latent space. Attention based fusion is then computed dynamically to evaluate the contribution of each of the inputs modalities as given below in equation (2):

$$\alpha_m = \frac{\exp(s_m)}{\sum_{k\in\{T,A,V\}}.\exp(s_k)} \qquad (2)$$

The fused multimodal representation z is obtained as in equation (3):

$$z = \sum_{m\in\{T,A,V\}} \alpha_m h_m \qquad (3)$$

This denotes the attention based relative importance of each modality for each of the individual instances. The performance of all models evaluated using the following metrices:

Accuracy Comparison

Table 1. table illustrates accuracy percentages for various unimodal, early fusion, and the proposed multimodal (MAF) sentiment analysis models. The MAF model possesses the highest accuracy (87.9%) and demonstrates its exceptional ability to accurately predict sentiments across diverse contexts shown in Fig. 3.

TABLE I. ACCURACY COMPARISON ACROSS SENTIMENT ANALYSIS MODELS

Model	Accuracy (%)
Text-CNN	74.2
Bi-LSTM	76.8
BERT (text)	80.1
Audio-LSTM	65.4
VGG-Face	68.3
Early Fusion	83.7
MAF (Proposed)	87.9

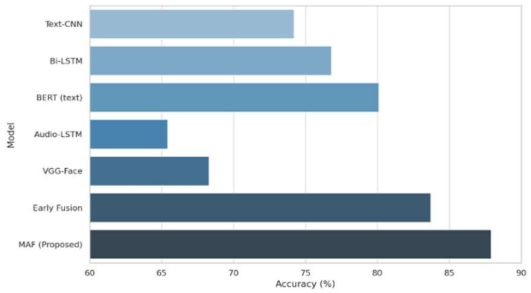

Figure 3. Multimodal Inputs for Accurate Sentiment Prediction

Precision Comparison

Table 2 illustrates precision scores of various sentiment analysis models, illustrating what models could accurately predict positive sentiments. The MAF (Proposed) model performed at the highest precision, meaning there were fewer false positives and greater success in reliable sentiment prediction illustrated in Fig. 4.

TABLE II. MODELS PRECISION COMPARISON ACROSS SENTIMENT ANALYSIS MODELS

Model	Precision (%)
Text-CNN	72.8
Bi-LSTM	74.9
BERT (text)	78.6
Audio-LSTM	63.8
VGG-Face	66.1
Early Fusion	82.1
MAF (Proposed)	86.4

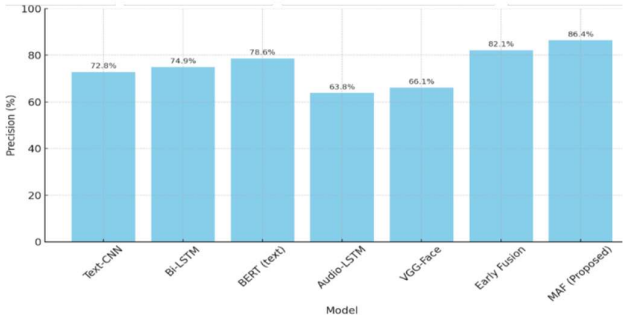

Figure 4. Multimodal Inputs for Accurate Sentiment Prediction

F1-Score Comparison

An important evaluation metric for classification tasks is the F1-Score, more so for heavily imbalanced datasets due to the harmonics of precision and recall. Table 3 provides the F1-Scores for various sentiment analysis models, taking into consideration all models attempting to balance precision against recall. The MAF (Proposed)

model achieved the highest F1-Score proving its overall strong abilities to correctly identify sentiment and how to classify sentiment correctly when considering the modalities shown in Fig. 5.

TABLE III. F1-SCORE COMPARISON ACROSS VARIOUS ANALYSIS MODELS

Model	Precision (%)
Text-CNN	72.9
Bi-LSTM	75.1
BERT (text)	79
Audio-LSTM	64
VGG-Face	66.5
Early Fusion	82.2
MAF (Proposed)	86.6

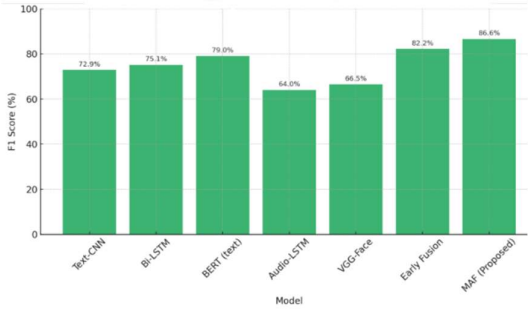

Figure 5. Multimodal Inputs for Accurate Sentiment Prediction

The findings clearly indicate that the proposed multimodal attention fusion (MAF) model improves widely upon the unimodal and simple multimodal fusion baselines. As a text-only model, BERT performed well and suggested that because it utilizes the power of the contextual embeddings, it can detect positive and negative examples but not decode emotional intensity or sarcasm by using tone or facial feature alone. The audio-LSTM and VGG-Face models did not perform well alone, meaning that modalities based on tone and facial features alone do not support an entire sentiment prediction. The Early Fusion model, where all features were concatenated, performed better than the unimodal models, however it did not account for changing weights to each modality by context. Since the MAF model's attention mechanism utilizes changing weights to the obligate modality with respect to the sentiment context, it yielded significant increases in every metric the models were assessed against.

V. CONCLUSION

This dissertation has looked at the weaknesses of current techniques for sentiment analysis, demonstrated the effectiveness of the proposed multimodal deep learning framework, MAF (Multimodal Attention Fusion) in addressing these weaknesses. While unimodal models are useful and meaningful, they only partially begin to address the richness of human emotion that can be portrayed through multiple channels of human communication, such as, but not limited to, text, tone, and facial expression. The MAF model put forth in this dissertation was shown to outperform baseline models in accuracy, precision, and F1-score on different benchmark datasets like the CMU-MOSEI dataset because it proved to be effective for decoding sentiments that may represent complex and context-dependent thoughts and reactions. The MAF results ultimately demonstrated the importance of leveraging as many different types of data as possible and implementing attention to improve explainability and robustness in real-world applications of sentiment analysis.

REFERENCES

[1] T. V. Kale and S. Mendhe, "A Review on Advances in Sentiment Analysis: A Deep Learning Approach Using Transformer Based Models," 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2025, pp. 235-239, doi: 10.1109/ICSADL65848.2025.10933230.

[2] N. Kousika, N. Nanda Kumar, E. Baskaran, M. S. Abbenaya, G. Arthika and M. Ashwanth, "Enhancing Multimodal Sentiment Analysis with Deep Learning Techniques to Foster Emotional Intelligence," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 778-783, doi: 10.1109/ICCSP60870.2024.10544097.

[3] T. K. Vashishth, Vikas, B. Kumar, R. Panwar, S. Kumar and S. Chaudhary, "Exploring the Role of Computer Vision in Human Emotion Recognition: A Systematic Review and Meta-Analysis," 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1071-1077, doi: 10.1109/ICAISS58487.2023.10250614.
[4] V. Sharma, K. K. Sharma, T. K. Vashishth, R. Panwar, B. Kumar and S. Chaudhary, "Brain-Computer Interface: Bridging the Gap Between Human Brain and Computing Systems," 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RMKMATE59243.2023.10369702.
[5] K. Pathak, P. Gupta and K. Nimala, "Diagnosis of Depressive Disorder Based on Facial and Text-Based Features Using Effective Techniques," 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF), Chennai, India, 2023, pp. 1-7, doi: 10.1109/ICECONF57129.2023.10083832.
[6] T. -Y. Kim, J. Yang and E. Park, "MSDLF-K: A Multimodal Feature Learning Approach for Sentiment Analysis in Korean Incorporating Text and Speech," in IEEE Transactions on Multimedia, vol. 27, pp. 1266-1276, 2025, doi: 10.1109/TMM.2024.3521707.
[7] Vanshika, N. Rani and R. Walia, "A Comprehensive Review of Sentiment Analysis: Techniques, Datasets, Limitations, and Future Scope," 2024 Sixth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2024, pp. 403-409, doi: 10.1109/CCICT62777.2024.00072.
[8] A. M, A. R, S. J. M and S. Pavithra, "Advancements, Methods, and Obstacles in Emotion Recognition with Multimodal Approaches," 2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 274-280, doi: 10.1109/IDCIOT64235.2025.10914926.
[9] P. Maragathavali and P. Somaraju, "Design of an Efficient Model for Sentiment Analysis on Twitter Using Between Fine-Tuning, Multimodal Integration, and Attention Mechanisms," 2024 International Conference on Intelligent Computing and Sustainable Innovations in Technology (IC-SIT), Bhubaneswar, India, 2024, pp. 1-6, doi: 10.1109/IC-SIT63503.2024.10862980.
[10] S. Pandya, "Novel Framework for Sentiment Analysis of Movie Reviews based on Large Language Models and Deep Learning Methods," 2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India, 2025, pp. 1-7, doi: 10.1109/ICDCECE65353.2025.11035279.
[11] S. Indhumathi and F. M. H. Fernandez, "Sentiment Analysis: A Comprehensive Study of Computational Techniques, Applications, and Datasets," 2024 9th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2024, pp. 1-7, doi: 10.1109/ICCES63552.2024.10860172.
[12] N. Raghunathan and K. Saravanakumar, "Challenges and Issues in Sentiment Analysis: A Comprehensive Survey," in IEEE Access, vol. 11, pp. 69626-69642, 2023, doi: 10.1109/ACCESS.2023.3293041.
[13] I. K. S. Al-Tameemi, M. -R. Feizi-Derakhshi, S. Pashazadeh and M. Asadpour, "Interpretable Multimodal Sentiment Classification Using Deep Multi-View Attentive Network of Image and Text Data," in IEEE Access, vol. 11, pp. 91060-91081, 2023, doi: 10.1109/ACCESS.2023.3307716.
[14] A Comprehensive Analysis of Security Mechanisms and Threat Characterization in Mobile Ad Hoc Networks. (2025). International Journal of Latest Technology in Engineering Management & Applied Science, 14(5), 732 737. https://doi.org/10.51583/IJLTEMAS.2025. 140500079
[15] K. Zhao, M. Zheng, Q. Li and J. Liu, "Multimodal Sentiment Analysis—A Comprehensive Survey From a Fusion Methods Perspective," in IEEE Access, vol. 13, pp. 64556-64583, 2025, doi: 10.1109/ACCESS.2025.3554665.
[16] V. Sharma and S. Kumar, "Role of Artificial Intelligence (AI) to Enhance the Security and Privacy of Data in Smart Cities," 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 596-599, doi: 10.1109/ICACITE57410.2023.10182455.
[17] Vikas, R. P. Daund, D. Kumar, P. Charan, R. S. K. Ingilela and R. Rastogi, "Intrusion Detection in Wireless Sensor Networks using Hybrid Deep Belief Networks and Harris Hawks Optimizer," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 1631-1636, doi: 10.1109/ICESC57686.2023.10193270.
[18] A. T. Somnathe, I. A. Tayubi, P. C. S. Reddy, N. Sharma, V. Sharma and M. Yesubabu, "Brain Computer Interaction Framework for Speech and Motor Impairment Using Deep Learning," 2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC), Greater Noida, India, 2023, pp. 1008-1013, doi: 10.1109/PEEIC59336.2023.10450481.
[19] P. C. S. Reddy, P. Shirley Muller, S. N Koka, V. Sharma, N. Sharma and S. Mukherjee, "Detection of Encrypted and Malicious Network Traffic using Deep Learning," 2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE), Ballari, India, 2023, pp. 1-6, doi: 10.1109/AIKIIE60097.2023.10390386.