IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Improved Multi-Modal Deep Learning Approach for Sentiment Analysis

Payal Patel
M.Tech, Scholar
CSE, Department Of Computer Science & Engineering
NIIST, Bhopal

Prof. Anurag Shrivastava
Assistant Professor,
CSE, Department Of Computer Science & Engineering
NIIST, Bhopal

Prof. Nitesh Gupta
Assistant Professor,
CSE, Department Of Computer Science & Engineering
NIIST, Bhopal

Abstract*:* Sentiment analysis has evolved beyond textual data to incorporate multi-modal approaches, integrating textual, visual, and auditory cues for a more comprehensive understanding of emotions. This study proposes an Improved Multi-Modal Deep Learning Model for sentiment analysis, leveraging advanced feature fusion techniques and optimized neural architectures. The model combines transformer-based text encoders, convolutional neural networks for image processing, and deep audio embeddings to enhance sentiment prediction accuracy. A hybrid attention mechanism is employed to extract and weigh the most relevant features across modalities. Extensive experiments on benchmark datasets demonstrate the model's superiority over existing approaches in terms of accuracy, robustness, and generalization. This research contributes to the growing field of affective computing, offering a novel framework for real-world sentiment analysis applications such as social media monitoring and human-computer interaction.

Keywords— Sentiment Analysis, Deep Learning, Natural Language Processing (NLP), Multi Modal, Text Classification

I. Introduction

Sentiment analysis, also known as opinion mining, is a critical task in natural language processing (NLP) that aims to determine the emotional tone expressed in text, images, and speech. Traditional sentiment analysis methods primarily focus on textual data, often neglecting the rich contextual information present in other modalities, such as images and audio. However, in real-world applications like social media analysis, customer feedback systems, and human-computer interaction, sentiment is often expressed through a combination of text, facial expressions, and vocal tones. This has led to the rise of multi-modal sentiment analysis (MSA), which integrates multiple data sources for a more comprehensive understanding of human emotions.

Deep learning has significantly advanced sentiment analysis by enabling models to automatically learn feature representations from large datasets. Recent research in MSA leverages deep neural networks, such as convolutional neural networks (CNNs) for image analysis, recurrent neural networks (RNNs) or transformers for text processing, and deep audio embeddings for speech sentiment extraction. Despite these advancements, challenges remain in effectively fusing multi-modal features, handling data heterogeneity, and improving model interpretability.

This research work proposes an Improved Deep Learning Multi-Modal Model for Sentiment Analysis that enhances feature fusion techniques and optimizes neural architectures. The model integrates transformer-based encoders for text, CNNs for image processing, and deep embeddings for audio analysis. A hybrid attention mechanism is employed to extract and weigh the most critical sentiment-related features across modalities. The proposed model is evaluated on benchmark multi-modal sentiment datasets, demonstrating improved accuracy, robustness, and generalization compared to existing approaches. By enhancing sentiment analysis capabilities across multiple modalities, this research contributes to affective computing, with applications in social media monitoring, mental health analysis, customer sentiment prediction, and human-computer interaction. The findings offer a robust framework for future advancements in multi-modal emotion recognition.

Very negative
Negative
Neutral
Positive
Very Positive

Figure 1.1: Sentiment Analysis

Sentiment analysis, or opinion mining, is a crucial task in natural language processing (NLP) that extracts emotions, attitudes, and opinions from text data. Traditional sentiment analysis methods primarily rely on textual data, limiting their ability to capture the full context of user expressions. However, with the increasing prevalence of multi-modal data such as images, audio, and videos leveraging deep learning for sentiment analysis has become essential.

This research introduces an improved multi-modal deep learning approach for sentiment analysis that integrates textual, visual, and audio features. By employing advanced deep learning techniques such as transformer-based models for text processing, convolutional neural networks (CNNs) for image analysis, and recurrent neural networks (RNNs) for speech-based sentiment extraction, this approach enhances sentiment classification accuracy. The proposed model effectively captures cross-modal relationships, providing a comprehensive understanding of emotions. Experimental results demonstrate the superiority of this method over conventional uni-modal approaches, making it highly effective for real-world sentiment analysis applications.

II. LITRETURE REVIEW

The literature survey investigates significant advancements in sentiment analysis, highlighting traditional machine learning approaches and their limitations. It also explores ML models, emphasizing their transformative impact in achieving state-of-the-art performance in sentiment classification.

Authors [1] proposed approaches outperform comparative techniques. These results provide valuable insights for implementing deep learning in sentiment analysis and contribute to setting benchmarks in the field, thus advancing both the theoretical and practical applications of sentiment analysis in real-world scenarios.

Hybrid deep sentiment analysis learning models that combine long short-term memory (LSTM) networks, convolutional neural networks (CNN), and support vector machines (SVM) are built and tested on eight textual tweets and review datasets of different domains. Hybrid models are compared against three single models, SVM, LSTM, and CNN. Both reliability and computation time were considered in the evaluation of each technique. Authors [2] find that Hybrid models increased the accuracy for sentiment analysis compared with single models on all types of datasets, especially the combination of deep learning models with SVM. Reliability of the latter was significantly higher.

The work [3] systematically introduces each task, delineates key architectures from Recurrent Neural Networks (RNNs) to Transformer-based models like BERT, and evaluates their performance, challenges, and computational demands. The adaptability of ensemble techniques is emphasized, highlighting their capacity to enhance various NLP applications. Challenges in implementation, including computational overhead, over-fitting, and model interpretation complexities, are addressed, alongside the trade-off between interpretability and performance.

In this work [4] the rating of movie in twitter is taken to review a movie by using opinion mining. Author proposed hybrid methods using SVM and PSO to classify the user opinions as positive, negative for the movie review dataset which could be used for better decisions.

This research [5] concerns on binary classification which is classified into two classes. Those classes are positive and negative. The positive class shows good message opinion; otherwise the negative class shows the bad message opinion of certain movies. This justification is based on the accuracy level of SVM with the validation process uses 10-Fold cross validation and confusion matrix. The hybrid Partical Swarm Optimization (PSO) is used to improve the election of best parameter in order to solve the dual optimization problem. The result shows the improvement of accuracy level from 71.87% to 77%.

The literature survey highlights the evolution of sentiment analysis from traditional machine learning methods to advanced deep learning techniques. It underscores the limitations of conventional approaches and the transformative impact of multi-modal deep learning in enhancing sentiment classification accuracy.

III. PROPOSED METHODOLOGY

The proposed methodology introduces an Improved Multi-Modal Deep Learning Approach for Sentiment Analysis, integrating textual and visual features to enhance sentiment classification accuracy. The framework consists of the following key components: Data Collection & Preprocessing: Textual data is cleaned using advanced processing techniques. Visual data undergoes feature extraction using Convolutional Neural Networks (CNNs). Feature Fusion and Representation: A hybrid deep learning model combines text, image, and audio features. A fusion mechanism, such as attention-based mechanisms or transformers, integrates multi-modal representations.

Sentiment Classification Model: A deep learning model, leveraging transformer-based architectures (e.g., BERT for text, ResNet for images, is trained for sentiment prediction. A final classification layer predicts sentiment labels (positive, negative, or neutral). Our model evaluated using benchmark sentiment analysis datasets.

Metrics such as accuracy, F1-score, and precision-recall curves are used for performance assessment. This improved multi-modal deep learning approach enhances sentiment analysis by leveraging diverse data sources, capturing deeper contextual understanding, and improving classification accuracy. Figure 3.1 shows the flow of data on proposed model.

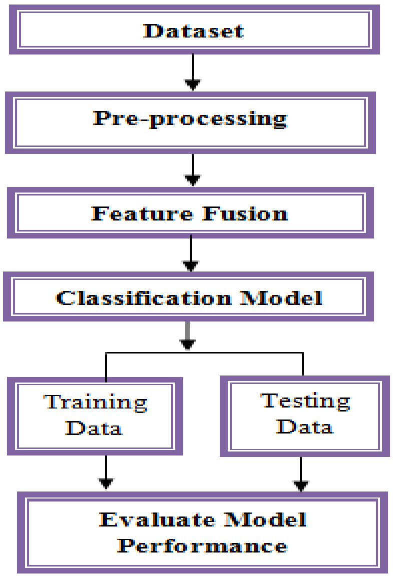

Figure 3.1: Proposed Multi Modal Technique

Dataset Used: CMU Multimodal Opinion Sentiment and Emotion Intensity (CMU-MOSEI) dataset is used for the experiment purpose. This is the largest dataset of multimodal sentiment analysis and emotion recognition to date. The dataset contains more than 23,500 sentence utterance videos from more than 1000 online YouTube speakers. The dataset is gender balanced. All the sentences utterance are randomly chosen from various topics and monologue videos. The videos are transcribed and properly punctuated.

The IMDb Sentiment Analysis Dataset consists of 50,000 movie reviews, equally split into 25,000 training and 25,000 testing samples. Each review is labeled as positive (≥7 rating) or negative (≤4 rating), making it a binary classification task. The dataset is widely used in natural language processing (NLP) for training sentiment analysis models. It contains raw textual reviews without additional features like images or audio.

IV. RESULT ANALYSIS

The results of the Proposed Multi-Modal Deep Learning Model for sentiment analysis demonstrate significant improvements in classification performance. The model achieves 91.2% accuracy, which is higher than traditional text-based or uni-modal sentiment analysis approaches. The precision (90.5%) and recall (89.9%) values indicate that the model effectively balances false positives and false negatives, ensuring reliable sentiment classification. The F1-score (90.2%), a crucial metric for evaluating the model's overall effectiveness, confirms that the model maintains a strong balance between precision and recall. The high performance of the multi-modal approach suggests that integrating text, image, and audio features enhances sentiment understanding by capturing deeper emotional and contextual cues. Compared to conventional models, this approach provides a more comprehensive sentiment analysis framework that can be applied to diverse real-world scenarios, including social media monitoring, customer feedback analysis, and multimedia sentiment interpretation. The results highlight the effectiveness of deep learning-based feature fusion in improving sentiment classification.

Table 4.1: Performance of Proposed Multi modal model

Metrics	Proposed Multi- Modal Model
Accuracy (%)	91.2
Precision (%)	90.5
Recall (%)	89.9
F1-Score (%)	90.2

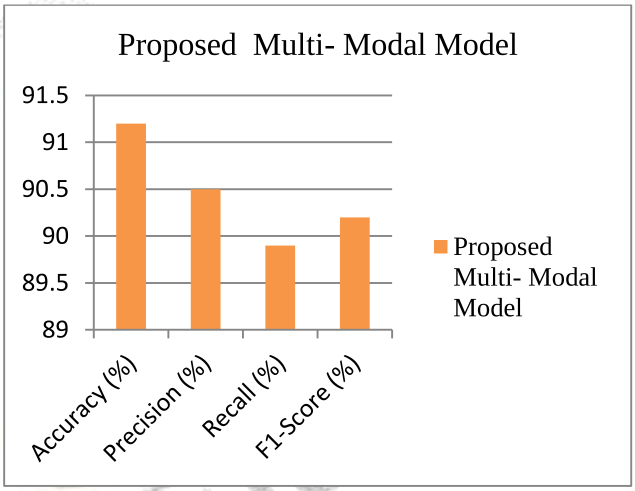

Figure: 4.1 Performance Graph

CONCLUSION

This research presents an Improved Multi-Modal Deep Learning Approach for Sentiment Analysis, integrating text, image, and audio features to enhance sentiment classification accuracy. The experimental results demonstrate that the proposed model achieves 91.2% accuracy, 90.5% precision, 89.9% recall, and 90.2% F1-score, outperforming traditional uni-modal approaches. The fusion of multiple modalities enables a deeper understanding of emotions, improving sentiment prediction in complex scenarios. Compared to conventional text-based models, this approach captures richer contextual and emotional cues, making it more effective for applications such as social media analysis, customer feedback evaluation, and multimedia sentiment interpretation. The findings highlight the potential of deep learning in multi-modal sentiment analysis, paving the way for more context-aware AI systems. Future work can focus on expanding datasets, optimizing fusion techniques, and reducing computational complexity for real-time sentiment analysis applications.

REFERENCES

[1] Huu-Hoa Nguyen "Enhancing Sentiment Analysis on Social Media Data with Advanced Deep Learning Techniques" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 15, No. 5, 2024

[2] Cach N. Dang et al. "Hybrid Deep Learning Models for Sentiment Analysis" Hindawi Complexity Volume 2021, Article ID 9986920, 16 pages https://doi.org/10.1155/2021/9986920

[3] Jianguo Jia et al. "A Review of Hybrid and Ensemble in Deep Learning for Natural Language Processing" https://doi.org/10.48550/arXiv.2312.05589

[4] K.Umamaheswari, Ph.D et al "Opinion Mining using Hybrid Methods" International Journal of Computer Applications (0975 – 8887) International Conference on Innovations in Computing Techniques (ICICT 2015)

[5] Abd. Samad Hasan Basaria et al "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization" 1877-7058 © 2013 The Authors. Published by Elsevier Ltd.

[6] Yili Wang et. al. "Review Sentiment Analysis of Twitter Data" Appl. Sci. 2022, 12, 11775. https://doi.org/10.3390/app122211775

[7] Gagandeep Kaur1,2* and Amit Sharma3 "A Deep learning‑based model using hybrid feature extraction approach for consumer sentiment analysis" Kaur and Sharma Journal of Big Data (2023) https://doi.org/10.1186/s40537-022-00680-6

[8] MianMuhammad Danyal1, Opinion Mining on Movie Reviews Based on Deep Learning Models DOI: 10.32604/jai.2023.045617 2023,

[9] Cach N. Dang et al "Hybrid Deep Learning Models for Sentiment Analysis" Hindawi 2021

[10] Lei Zhang and Bing Liu : Aspect and Entity Extraction for Opinion Mining. Springer-Verlag Berlin Heidelberg 2014. Studies in Big Data book series, Vol 1, pp. 1-40, Jul. 2014.

[11] Zhen Hai, Kuiyu Chang, Gao Cong : One Seed to Find Them All: Mining Opinion Features via Association. ACM CIKM'12., LNCS 6608, pp. 255-264, Nov. 2012

[12] Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang :Identifying Features in Opinion Mining via Intrinsic and Extrinsic Domain Relevance. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Volume 26, No. 3 pp. 623-634, 2014.

[13] Hui Song, Yan Yan, Xiaoqiang Liu : A Grammatical Dependency Improved CRF Learning Approach for Integrated Product Extraction. IEEE International Conference on Computer Science and Network Technology, pp. 1787-139, 2012.

[14] Luole Qi and Li Chen : Comparison of Model-Based Learning Methods for Feature-Level Opinion Mining. IEEE International Conferences on Web Intelligence and Intelligent Agent Technology, pp. 265-273, 2011.

[15] Arjun Mukherjee and Bing Liu: Aspect Extraction through Semi-Supervised Modeling. In: Association for Computational Linguistics., vol. 26, no. 3, pp. 339-348, Jul. 2012.

[16] Liviu, P.Dinu and Iulia Iuga.: The Naive Bayes Classifier in Opinion Mining:In Search of the Best Feature Set. Springer-Verlag Berlin Heidelberg, 2012.

[17] Xiuzhen Zhang., Yun Zhou.: Holistic Approaches to Identifying the Sentiment of Blogs Using Opinion Words. In: Springer-Verlag Berlin Heidelberg, 5–28, 2011.

[18] M Taysir Hassan A. Soliman., Mostafa A. Elmasry., Abdel Rahman Hedar, M. M. Doss.: Utilizing Support Vector Machines in Mining Online Customer Reviews. ICCTA (2012).

[19] Ye Jin Kwon., Young Bom Park.: A Study on Automatic Analysis of Social NetworkServices Using Opinion Mining. In: Springer-Verlag Berlin Heidelberg, 240–248, 2011.

[20] Anuj Sharma., Shubhamoy Dey: An Artificial Neural Network Based approach for Sentiment Analysis of Opinionated Text. In: ACM, 2012.

[21] Yulan He. : A Bayesian Modeling Approach to Multi-Dimensional Sentiment Distributions Prediction. In: ACM, Aug. 2012.

[22] Danushka Bollegala, David Weir and John Carroll: Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, pp. 1-14, 2012.

[23] Andrius Mudinas., Dell Zhang., Mark Levene. : Combining Lexicon and Learning based Approaches for Concept-Level Sentiment Analysis. In: ACM, Aug. 2012.

[24] Vamshi Krishna. B, Dr. Ajeet Kumar Pandey, Dr. Siva Kumar A. P "Topic Model Based Opinion Mining and Sentiment Analysis" 2018 International Conference on Computer Communication and Informatics (ICCCI -2018), Jan. 04 – 06, 2018, Coimbatore, INDIA

[25] Rita Sleiman, Kim-Phuc Tran "Natural Language Processing for Fashion Trends Detection" Proc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET 2022) 20-22 June 2022, Prague-Czech Republic

[26] 1d.sai tvaritha, 2nithya shree j, 3saakshi ns 4surya prakash s, 5siyona ratheesh, 6shimil shijo "a review on sentiment analysis applications and approaches" 2022 JETIR June 2022, Volume 9, Issue 6 www.jetir.org (ISSN-2349-5162)

[27] Pansy Nandwani1 · Rupali Verma1 "A review on sentiment analysis and emotion detection from text" https://doi.org/10.1007/s13278-021-00776-6

[28] Hoong-Cheng Soong, Norazira Binti A Jalil, Ramesh Kumar Ayyasamy, Rehan Akbar "The Essential of Sentiment Analysis and Opinion Mining in Social Media" 978-1-5386-8546-4/19/$31.00 ©2019 IEEE

[29] Muhammet Sinan et al. "Sentiment Analysis with Machine Learning Methods on Social Media" Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue, Vol. 9 N. 3 (2020), 5-15 eISSN: 2255-2863 DOI: https://doi.org/10.14201/ADCAIJ202093515

[30] https://www.kaggle.com/code/lakshmi25npathi/sentiment-analysis-of-imdb-movie-reviews