Motilal Nehru National Institute of Technology
Allahabad

A Report Submitted
in Partial Fulfillment of the Requirements
for the Degree of
Bachelor of Technology
in
Computer Science & Engineering

Internet Hinglish Memes Classification using Multimodal Learning

Under Supervision of:
Dr. Abhinav Kumar
Assistant Professor
Computer Science and Engineering Department

Submitted by:
Abhay Vishwakarma (20214279)
Abhishek Singh Dhruwanshi (20214322)
Amber Kumar Shakya (20214038)
Aditya Ashish Parmar (20214011)

Undertaking

We declare that the work presented in this report titled "**Internet Hinglish Memes Classification using Multimodal Learning**", submitted to the Computer Science and Engineering Department, Motilal Nehru National Institute of Technology Allahabad, Prayagraj, for the award of the Bachelor of Technology degree in Computer Science & Engineering, is our original work. We have not plagiarized or submitted the same work for the award of any other degree. In case this undertaking is found incorrect, we accept that our degree may be unconditionally withdrawn.

May 2024
Allahabad

Certificate

Certified that the work contained in the report titled "**Internet Hinglish Memes Classification using Multimodal Learning**", by Abhay Vishwakarma (20214279), Abhishek Singh Dhruwanshi (20214322), Amber Kumar Shakya (20214038), Aditya Ashish Parmar (20214011), has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

May 2024

Dr. Abhinav Kumar
Assistant Professor
Computer Science & Engineering
MNNIT Allahabad

Preface

In this exploration of internet memes classification, we delve into the realm of digital culture, seeking to unravel the intricate dynamics of humor and communication in the online world. Harnessing the power of multimodal learning, we navigate through a myriad of memes, dissecting their structure, context, and impact.

Amidst the ever-evolving landscape of internet culture, our endeavor aims to shed light on the underlying patterns and trends that govern the viral spread of memes. From image macros to viral videos, we embark on a journey through the vast expanse of internet humor, employing a diverse array of analytical tools and techniques.

At the heart of our exploration lies a deep appreciation for the creativity and ingenuity of meme creators and consumers. We recognize the importance of understanding the cultural and social contexts in which memes thrive, and the implications they hold for digital communication and identity.

This report stands as a testament to our dedication to advancing our understanding of internet memes and their role in contemporary society. It is our hope that this work will contribute to the ongoing dialogue surrounding digital culture and inform future research in this dynamic field.

Acknowledgements

Our deepest gratitude goes to **Dr. Abhinav Kumar** for guiding us through this project.

Their expertise in **Machine Learning, Deep Learning, Crisis Informatics, Natural Language Processing, Image Processing** and **Hate Speech Detection** has been invaluable.

His guidance has shaped our understanding and played a pivotal role in navigating the complexities of Hinglish memes classification.

We are sincerely thankful for their mentorship, which has broadened our skills and enriched our project.

May 2024

Abhay Vishwakarma (20214279)
Abhishek Singh Dhruwanshi (20214322)
Amber Kumar Shakya (20214038)
Aditya Ashish Parmar (20214011)

Contents

Chapter 1

Introduction

1.1 The Phenomenon of Internet Memes

Internet memes represent a unique form of cultural expression, blending humor, creativity, and intertextuality to convey complex ideas and emotions in a concise and accessible format. From image macros and reaction gifs to viral videos and catchphrases, memes encompass a diverse array of content genres, reflecting the ever-evolving nature of online discourse. What sets memes apart is their virality and adaptability, allowing them to spread rapidly across digital networks and undergo remixing and reinterpretation by users worldwide.

1.2 Challenges in Meme Analysis

Despite their ubiquitous presence in digital culture, memes present significant challenges for analysis and classification. Unlike traditional textual data, memes often rely on visual elements, cultural references, and contextual cues to convey meaning, making them inherently multimodal in nature. Moreover, memes exhibit a high degree of variability in terms of content, style, and intended audience, posing challenges for automated analysis and categorization.

1.3 The Need for multi-modal learning

To address the complexities of meme analysis, researchers have increasingly turned to multi-modal approaches that integrate both visual and textual data modalities. By leveraging advances in computer vision, natural language processing, and machine learning, multi-modal analysis techniques offer a more comprehensive understanding of meme content, capturing the intricate interplay between visual and semantic elements. Among these approaches, Memotion 3 stands out as a pioneering framework for internet memes classification, utilizing state-of-the-art deep learning models to unravel the emotional nuances embedded within meme content.

In this report, we present a comprehensive investigation into internet hinglish memes classification using multi-modal learning, with a focus on the Memotion 3 framework. We begin by providing a theoretical foundation for meme analysis, exploring the socio-cultural dynamics of internet memes and the challenges inherent in their classification. We then introduce the Memotion 3 framework, detailing its architecture, components, and methodologies for multi-modal analysis. Next, we

describe our experimental setup and dataset collection process, followed by a comprehensive evaluation of Memotion 3's performance in meme classification tasks.

Through this thesis, we aim to contribute to the advancement of meme analysis research and provide insights into the transformative potential of multi-modal analysis in understanding internet memes' cultural significance and impact on digital communication.

1.4 Motivation

The pervasive presence of internet memes in contemporary digital culture underscores the need for robust and comprehensive analysis techniques. Internet memes, characterized by their diverse formats and nuanced expressions, serve as dynamic artifacts reflecting the collective consciousness of online communities. Understanding the motivations behind meme creation, propagation, and evolution is essential for researchers, social scientists, and technologists seeking to unravel the intricacies of digital communication.

Motivated by the ever-growing significance of internet memes in shaping online discourse, this thesis embarks on a journey to explore the underlying motivations driving meme culture. By delving into the socio-cultural dynamics, psychological underpinnings, and technological innovations shaping the meme ecosystem, we aim to shed light on the multifaceted nature of meme phenomena.

Through an interdisciplinary lens, we seek to uncover the drivers behind meme creation, ranging from individual expression and social identity to collective humor and cultural commentary. By examining the motivations of meme creators, distributors, and consumers, we hope to gain insights into the complex interplay of factors influencing meme production and dissemination.

Furthermore, this section aims to elucidate the broader implications of meme analysis, beyond mere entertainment or amusement. Internet memes have emerged as powerful tools for political activism, social commentary, and cultural critique, challenging conventional notions of communication and expression. By understanding the motivations driving meme production and consumption, we can better appreciate their role in shaping public discourse and cultural narratives.

In essence, the motivation section of this thesis serves as a foundational exploration into the underlying drivers and implications of internet meme culture. By interrogating the motivations behind meme creation and dissemination, we seek to uncover the deeper meanings embedded within the seemingly whimsical world of online memes.

1.5 Some Wonderful Minds

The field of internet meme analysis owes much to the pioneering work of visionary researchers and scholars who have illuminated the path towards a deeper understanding of digital culture. In this section, we pay tribute to some of these remarkable minds whose insights and contributions have shaped the landscape of meme studies and computational humor.

Among the luminaries of meme research is Dr. Richard Dawkins[13], whose seminal work on the concept of the "meme" laid the groundwork for contemporary discus-

sions on cultural transmission and imitation in the digital age. Dawkins' notion of memes as "units of cultural evolution" sparked a paradigm shift in how we conceptualize the spread of ideas and behaviors in online communities.

Building upon Dawkins' foundational ideas, scholars like Susan Blackmore and Daniel Dennett have further explored the evolutionary dynamics of memes and their role in shaping human culture. Blackmore's research on "memetics" delves into the parallels between genetic and memetic evolution, offering insights into the mechanisms driving the proliferation of internet memes.

In the realm of computational humor and meme analysis, researchers such as Dr. Shlomo Argamon and Dr. Mark J. Lee have made significant strides in developing computational models for understanding humour and irony in textual data. Their work on sentiment analysis, humor detection, and sarcasm recognition has paved the way for more nuanced approaches to meme classification and interpretation.

Additionally, we cannot overlook the contributions of contemporary scholars like Dr. Limor Shifman[11], whose research on internet memes and digital culture has provided invaluable insights into the socio-cultural dynamics of meme production and circulation. Shifman's interdisciplinary approach, combining media studies, sociology, and communication theory, offers a holistic understanding of the complexities of meme culture in the age of social media.

As we embark on our own journey into meme analysis, we draw inspiration from these wonderful minds whose pioneering research and innovative thinking continue to shape the field. Their intellectual curiosity, interdisciplinary perspectives, and dedication to understanding the nuances of digital culture serve as guiding lights in our exploration of internet memes and their impact on society.

Chapter 2

Related Work

Sentiment and Emotion Analysis: Extensive research has been conducted on analyzing sentiment in text, spanning various methodologies including machine learning (ML) techniques like SVM, logistic regression, and deep learning (DL) methods such as neural networks. Surveys on sentiment analysis in social media provide comprehensive insights into this field.

For humor detection on social media, the HaHa[3] shared task offers a dataset, while Öhman et al. released an English dataset to identify eight emotions. Textual emotion detection is extensively surveyed in literature.

Hatespeech Detection: Detecting hatespeech is crucial for maintaining a safe environment on social media platforms, particularly for marginalized groups. Various datasets have been curated and released for this purpose, including those focused on offensive language and hate speech targeting specific groups such as women and immigrants. Methods for hatespeech detection include convolutional and recurrent neural networks, Bert-like models, and linguistic feature incorporation.

Codemixed Language Processing: Processing codemixed language, characterized by the informal mixing of multiple languages, presents significant challenges. Tasks like sentiment analysis and offense detection in codemixed languages have gained attention, with methods employing graph convolutional networks, BERT-based models, and adaptations in loss functions or positional embeddings.

Multimodal Analysis: While much research has concentrated on unimodal text analysis, there is a growing interest in multimodal content analysis involving text, images, and videos. Multimodal datasets facilitate various tasks like image captioning, hatespeech detection, and sentiment analysis. However, there's a scarcity of studies focusing on codemixed meme analysis, with limited works exploring methods like image-text joint embeddings and transformer-based[7] models.

Previous Iterations of Memotion: Previous iterations[1] of the Memotion shared task provided datasets of English memes, whereas Memotion 3 emphasizes Hinglish codemixed memes. Common methods across Memotion iterations include ensembling and bert-like models.

Chapter 3

Proposed Work

This project aims to develop a robust classification system for internet memes utilizing multimodal learning techniques. Leveraging the combined information from textual and visual modalities inherent in memes, the system will seek to accurately categorize memes into predefined classes or themes.

The methodology involves collecting a diverse dataset of internet memes from various online platforms, ensuring representation across different genres, formats, and cultural contexts. Textual content will undergo preprocessing, including tokenization and normalization, while image data will be subjected to techniques like resizing and feature extraction.

3.1 Dataset

The dataset utilized in this project is sourced from the Memotion 3.0[2] competition, which is a part of the De-Factify 2 workshop held in AAAI-22. This dataset is specifically curated for Hate Speech Detection and comprises a collection of internet memes. It consists of 10,000 memes, including a subset of Hinglish memes, which adds linguistic diversity to the dataset.

The dataset is divided into three subsets: training images, validation images, and test images. The training set contains 7,000 memes, while both the validation and test sets consist of 1,500 memes each. This division allows for the training, validation, and evaluation of the classification model.

Each subset of the dataset is accompanied by three CSV files providing additional metadata. These files include the image IDs, corresponding labels indicating the presence of humor, sarcasm, offensiveness, and motivational content, as well as the overall sentiment conveyed by the memes. Additionally, optical character recognition (OCR) text extracted from the memes is provided in the CSV files, facilitating textual analysis alongside visual content processing.

The dataset's composition and annotations offer a rich and varied resource for training and evaluating the multimodal classification model. The inclusion of Hinglish memes reflects the dataset's awareness of linguistic and cultural diversity, which is essential for robust meme classification in diverse online communities.

3.2 Tasks

Within the Memotion 3.0 competition, which encompasses three distinct tasks (A, B, and C), our project primarily focused on addressing Tasks A and B.

Task A - Sentiment Analysis: This task involves classifying internet memes into positive, negative, or neutral categories based on the underlying sentiment conveyed by the meme content.

Task B - Emotion Classification: In this task, the system identifies various emotional nuances within memes, including humor, sarcasm, offensiveness, and motivational content. Memes may belong to more than one emotion category.

As for **Task C**, which centers on quantifying the intensity or scale of emotional expression within memes, our project did not directly address it. Instead, we concentrated our efforts on sentiment analysis and emotion classification within internet memes.

3.3 Flow Chart

The classification process initially involves using a text-only model and an image-only model separately. Then, both text and images, including memes, are integrated for a comprehensive analysis.

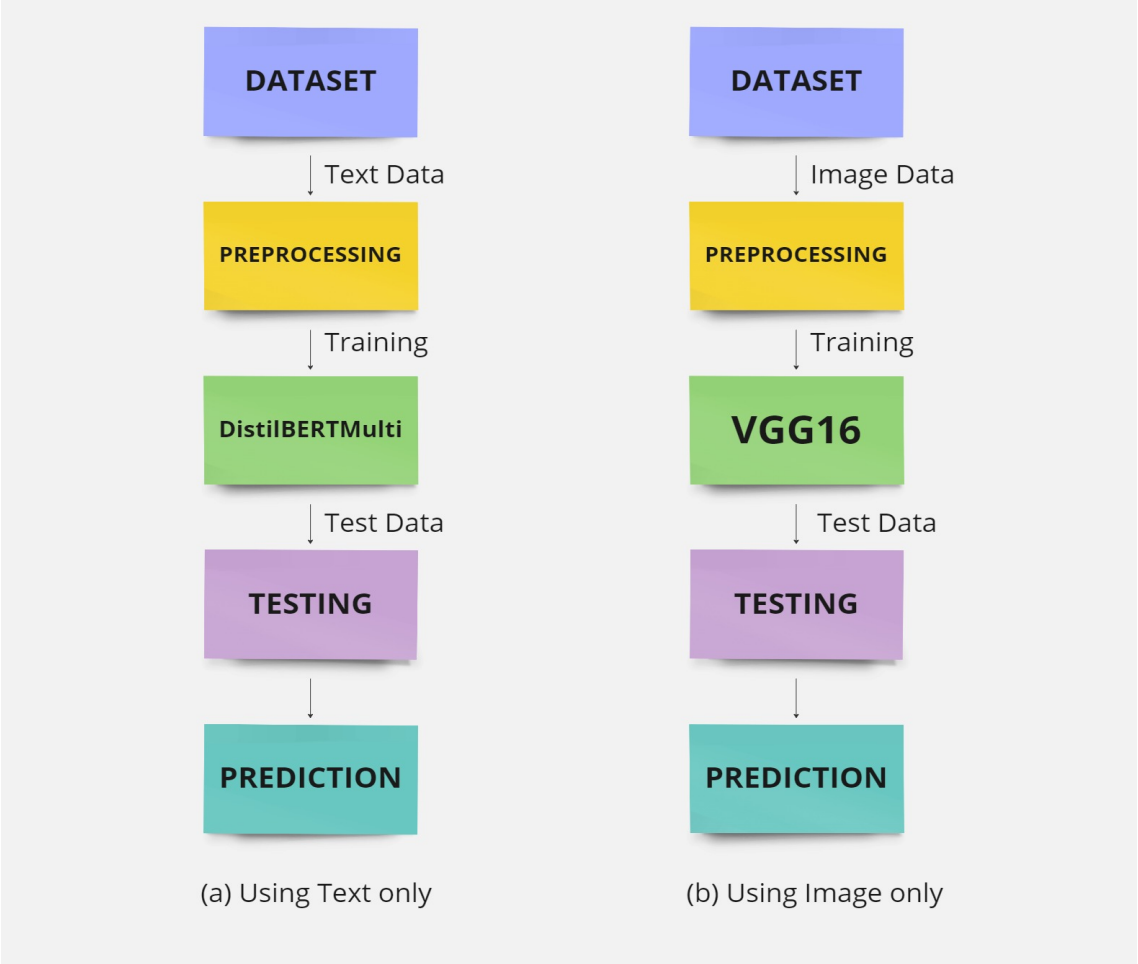


Fig. 1 : Separate utilization of text and images for predictions

In Figure 1, VGG16[5] is trained on images, while DistilBERTMulti[10] is trained on texts for classification.

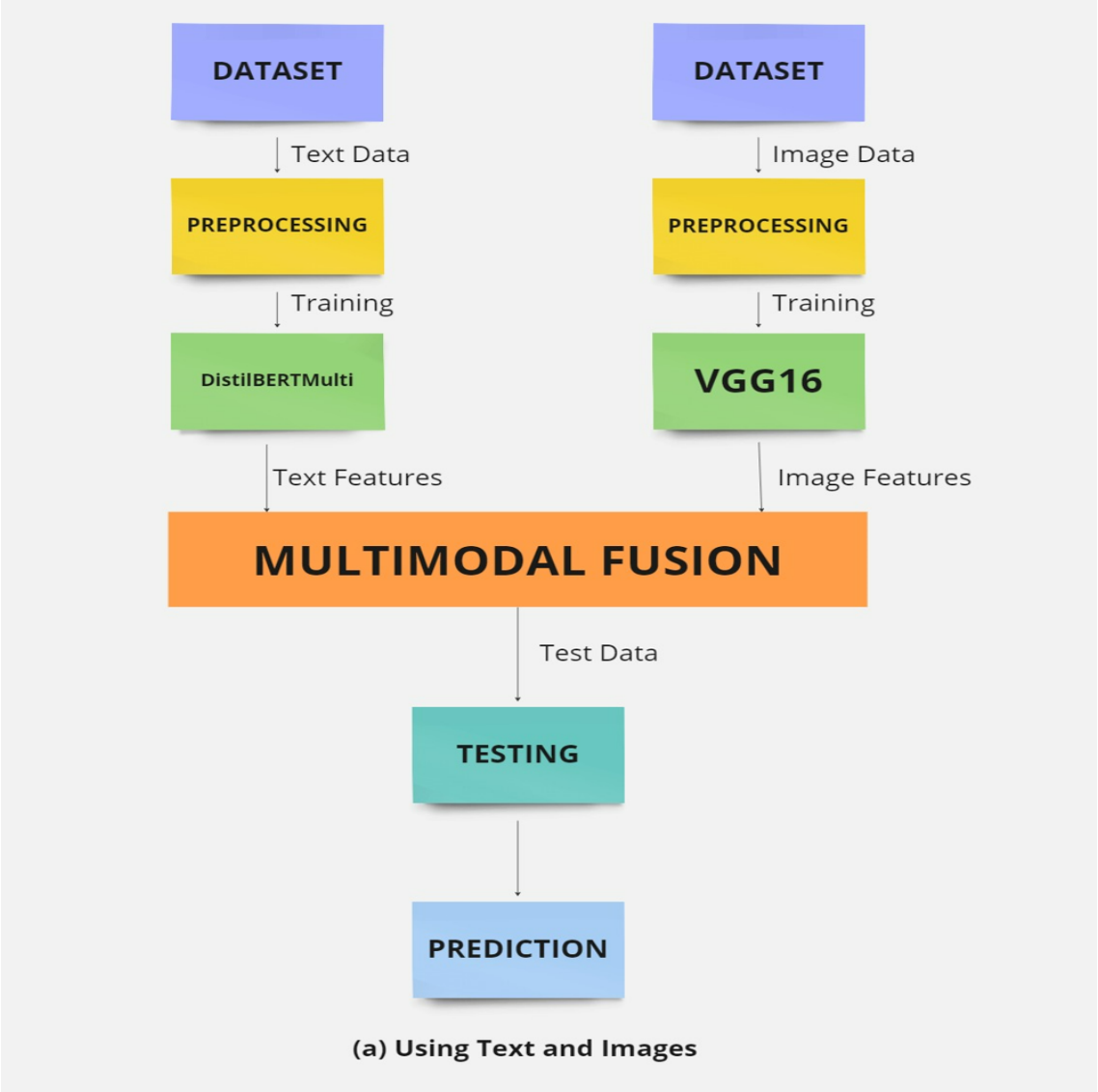


Fig. 2 : Combined utilization of text and images for predictions

In Figure 2, features extracted from images by VGG16 and from text by DistilBERTMulti are fused in the feature space for predictions.

3.4 Model Architecture

DistilBERTMulti - It stands out as a condensed iteration of BERT[4], honed through training on a vast array of languages. This diverse linguistic exposure equips it for various tasks across language barriers, from machine translation to cross-lingual document classification and multilingual sentiment analysis.

DistilBERTMulti shares the transformer architecture with DistilBERT and employs self-attention mechanisms for processing input sequences. However, it stands out by training on a multilingual corpus encompassing 104 languages. This extensive training equips it to recognize language-agnostic patterns, enabling seamless analy-

sis across diverse linguistic landscapes. With text tokenization, position embedding, semantic representation generation, and feature extraction, DistilBERTMulti efficiently deciphers input text, making it adept at tasks like sentiment analysis even in languages like Hinglish.

The advantages of employing DistilBERTMulti are manifold:

Efficiency: Much like its predecessor, DistilBERT, DistilBERTMulti boasts computational efficiency, requiring fewer resources for training and inference while maintaining competitive performance levels.

Multilingual Proficiency: With exposure to a broad spectrum of languages, including Hindi and English, DistilBERTMulti exhibits remarkable adaptability in comprehending and generating text across multiple linguistic contexts.

Transfer Learning Facilitation: Pretrained versions of DistilBERTMulti can swiftly adapt to specific tasks with minimal amounts of task-specific data. This streamlined process leverages transfer learning to efficiently tailor the model to new applications.

Cross-Lingual Applications: DistilBERTMulti's multilingual prowess makes it particularly adept at cross-lingual tasks, where input and output texts may vary across different languages. This versatility extends to the nuanced realm of Hinglish, enabling the analysis of text sentiments embedded within Hinglish memes.

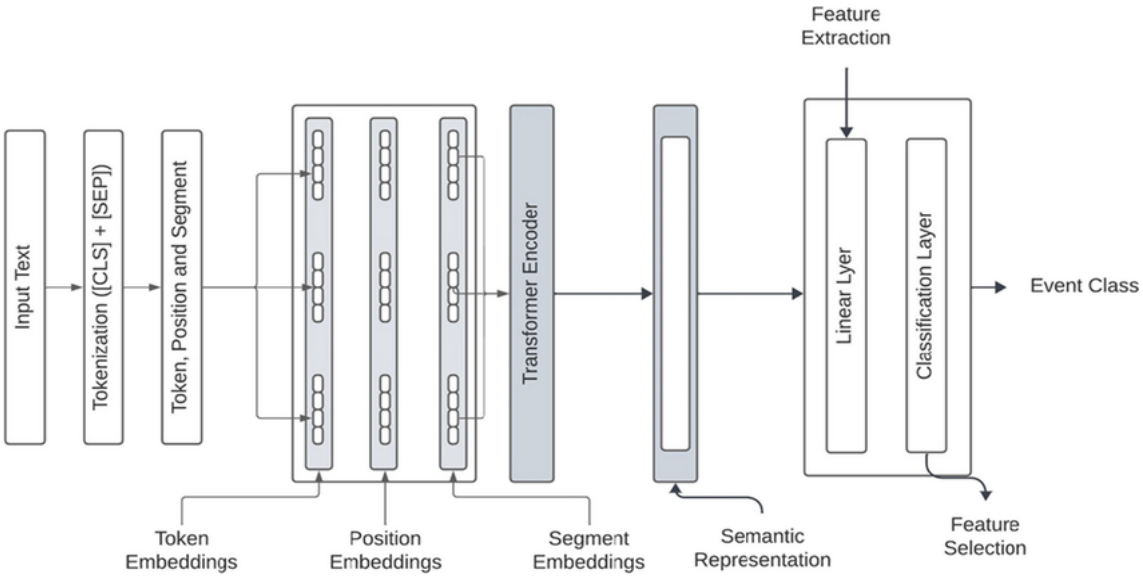


Fig. 3 : Architecture of DistilBERT[6]

VGG16 - It is a convolutional neural network (CNN)[12] architecture that gained prominence for its simplicity and effectiveness in image classification tasks. It is composed of 16 layers, including 13 convolutional layers and 3 fully connected layers.

The structural framework of VGG16 follows a straightforward design, with stacked convolutional layers followed by max-pooling layers. The convolutional layers are responsible for extracting features from input images through the application of filters, which detect patterns at different spatial scales.

After each set of convolutional layers, max-pooling layers are applied to downsample the feature maps, reducing their spatial dimensions while retaining important fea-

tures. This hierarchical process allows VGG16 to progressively learn and abstract features from input images.

Following the convolutional layers, VGG16 includes three fully connected layers, which serve as a classifier. These layers take the high-level features extracted by the convolutional layers and map them to the corresponding class labels.

VGG16's architecture is characterized by its deep and uniform structure, where convolutional layers have small 3x3 filters and are stacked one after the other. This uniformity contributes to its effectiveness in learning hierarchical features.

The advantages of using VGG16 include:

Simplicity: VGG16's straightforward architecture makes it easy to understand and implement, even for those new to deep learning.

Effectiveness: Despite its simplicity, VGG16 achieves competitive performance on image classification tasks, often outperforming more complex architectures.

Transfer Learning: Pretrained versions of VGG16 on large-scale image datasets like ImageNet can be fine-tuned for specific image recognition tasks with relatively small amounts of task-specific data.

Robustness: VGG16's deep architecture allows it to learn intricate features from images, making it robust to variations in input data.

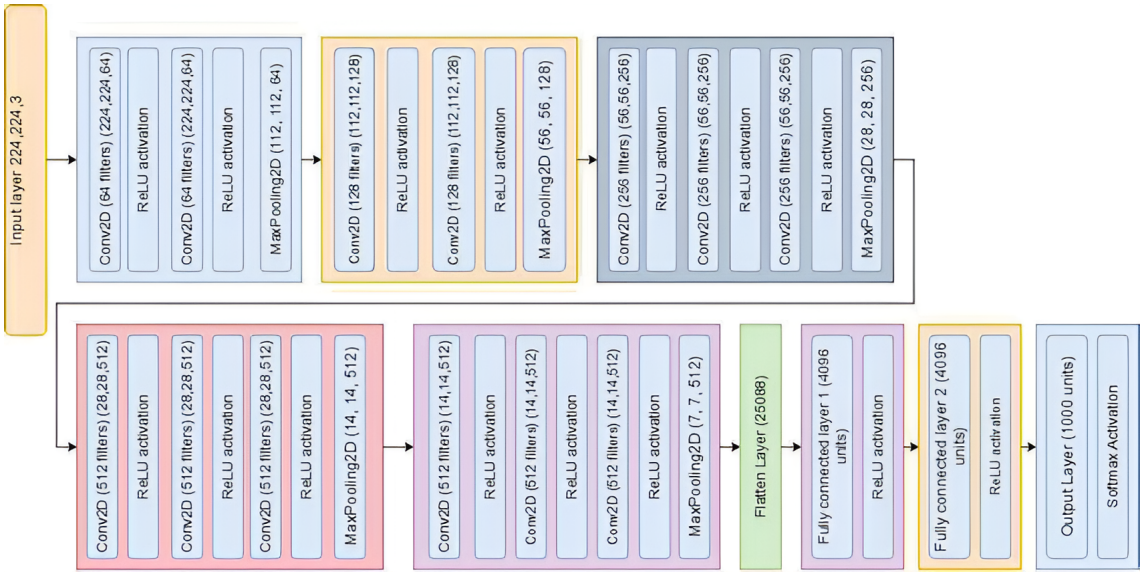


Fig. 4 : Architecture of VGG16[5]

Multimodal(DistilBERTMulti + VGG16) - It combines the strengths of both DistilBERTMulti and VGG16 architectures to tackle multimodal[9] tasks, particularly in areas like image-text fusion and cross-modal retrieval.

The structural framework of this multimodal model integrates the feature extraction capabilities of VGG16 with the contextual understanding of text provided by DistilBERTMulti. By leveraging DistilBERTMulti, the model can process textual information efficiently, capturing semantic nuances and contextual cues. Meanwhile, VGG16's convolutional layers excel at extracting visual features from images, detecting patterns at different spatial scales.

In this architecture, input data from different modalities (text and images) are processed separately by DistilBERTMulti and VGG16, respectively. The extracted

features are then fused at a later stage, enabling the model to learn joint representations that capture both visual and textual semantics.

The advantages of Multimodal(DistilBERTMulti + VGG16) include:

Comprehensive Feature Extraction: By combining DistilBERTMulti and VGG16, the model can capture rich semantic information from both text and images, enhancing its understanding of multimodal data.

Synergistic Fusion: The fusion of features from different modalities allows the model to exploit complementary information sources, improving overall performance in tasks like image-text matching and retrieval.

Transfer Learning: Pretrained versions of both DistilBERTMulti and VGG16 can be fine-tuned on multimodal datasets, leveraging transfer learning to adapt the model to specific tasks with minimal data requirements.

Robustness and Generalization: The joint representation learning enables the model to generalize well to new multimodal data, making it robust to variations and noise across different modalities.

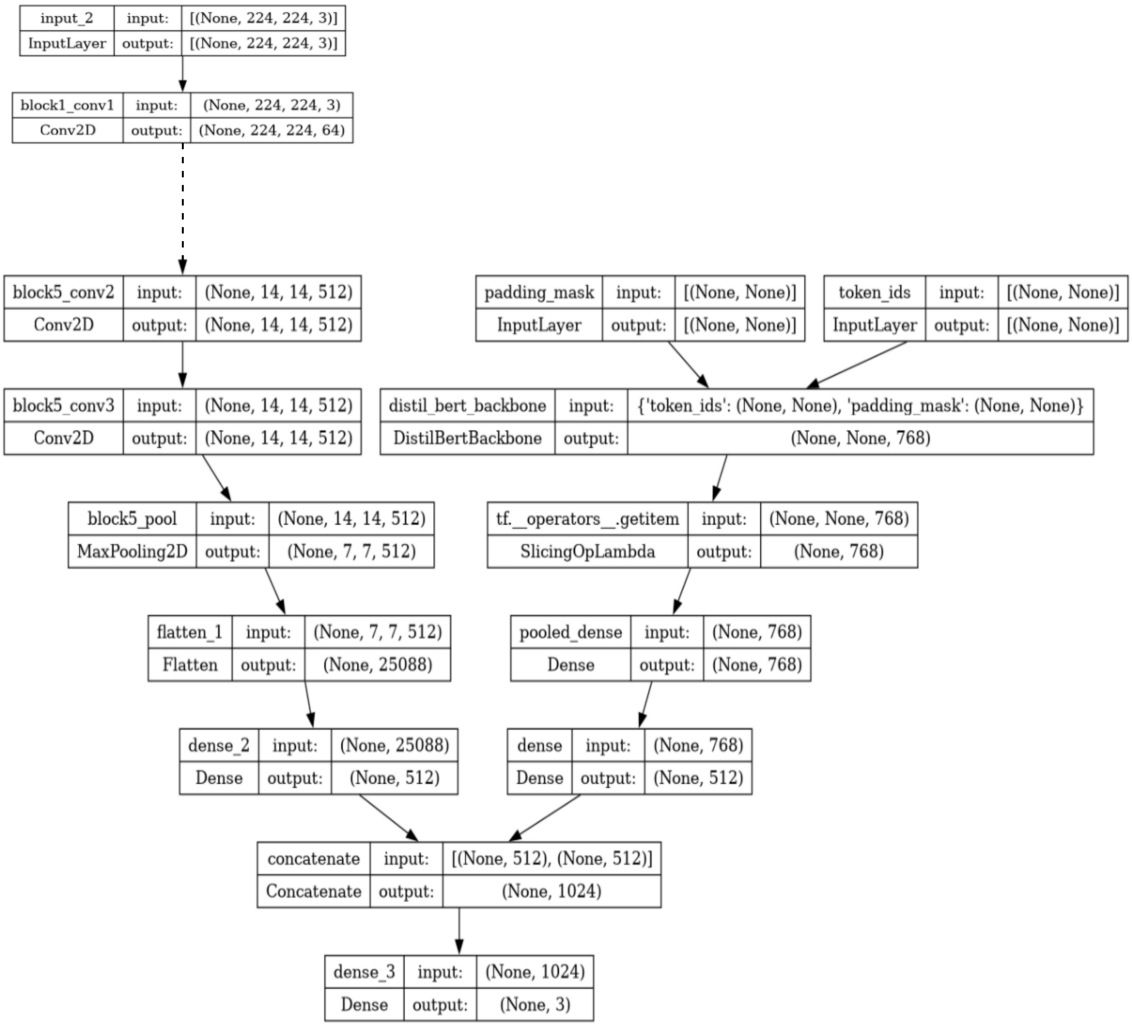


Fig. 5 : Architecture of multimodal for task A

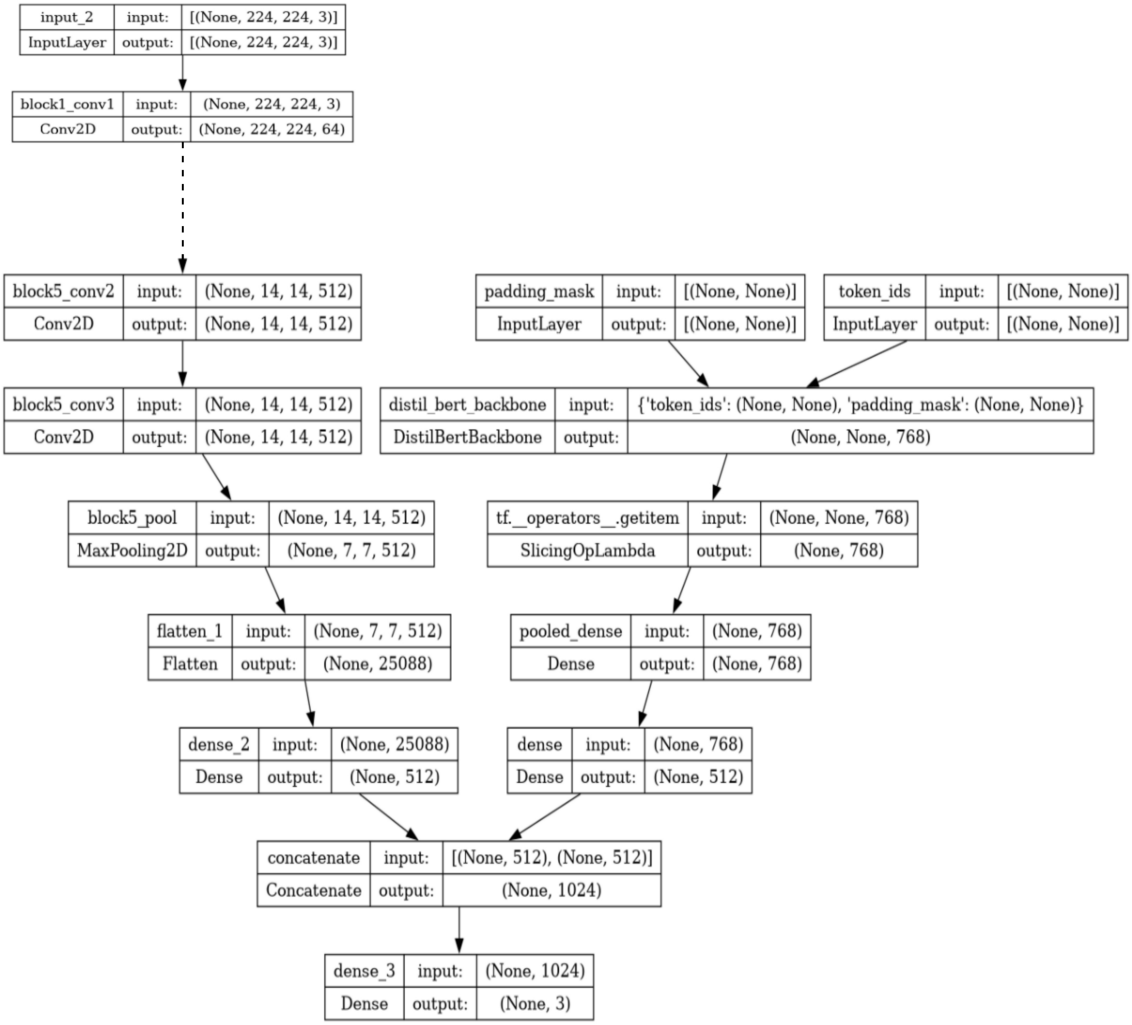


Fig. 6 : Architecture of multimodal for task B

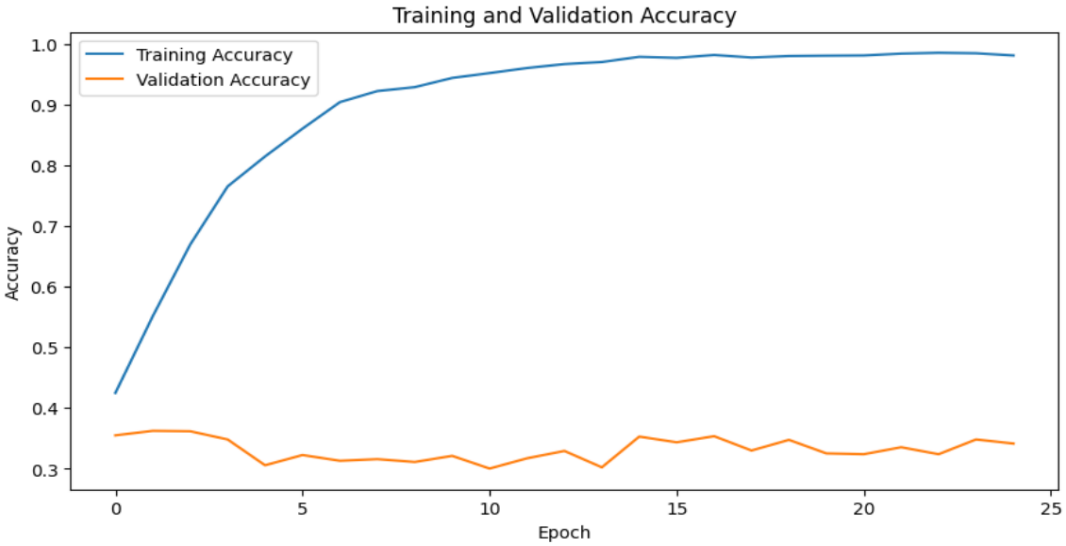
Chapter 4

Results Analysis

4.1 Task A

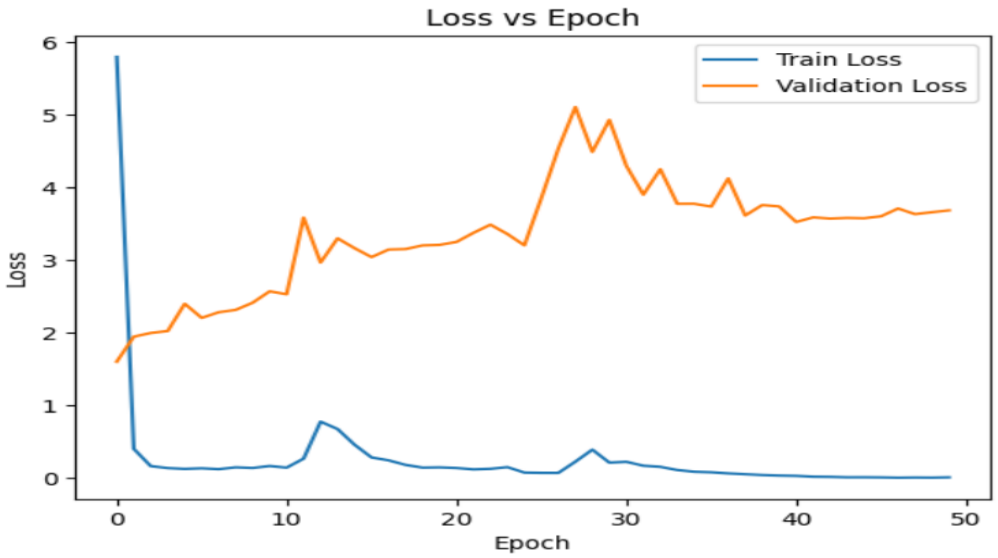
Plots of Text only Model(DistilBERTMulti) -

Plot 1 : Loss vs Epochs of DistilBERTMulti

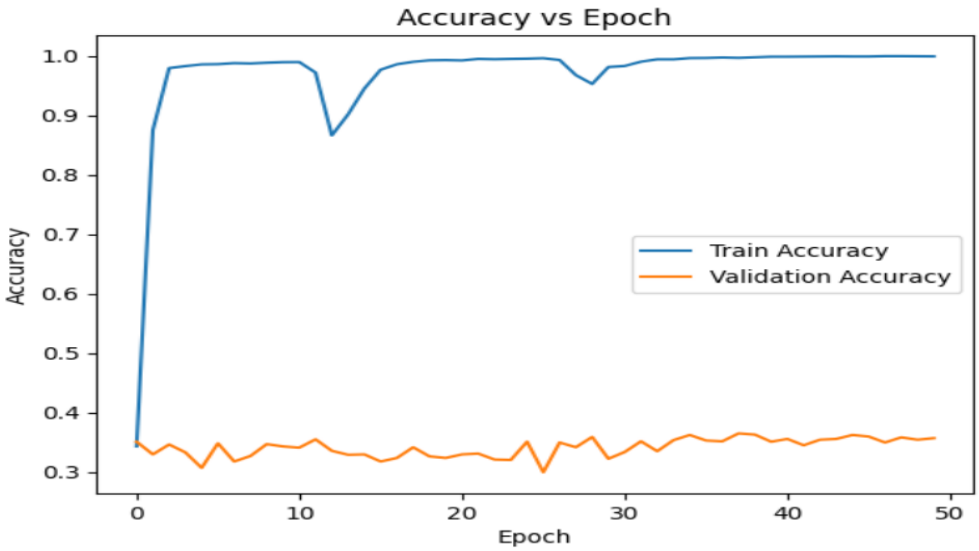


Plot 2 : Accuracy vs Epochs of DistilBERTMulti

Plots of Image only Model(VGG16) -

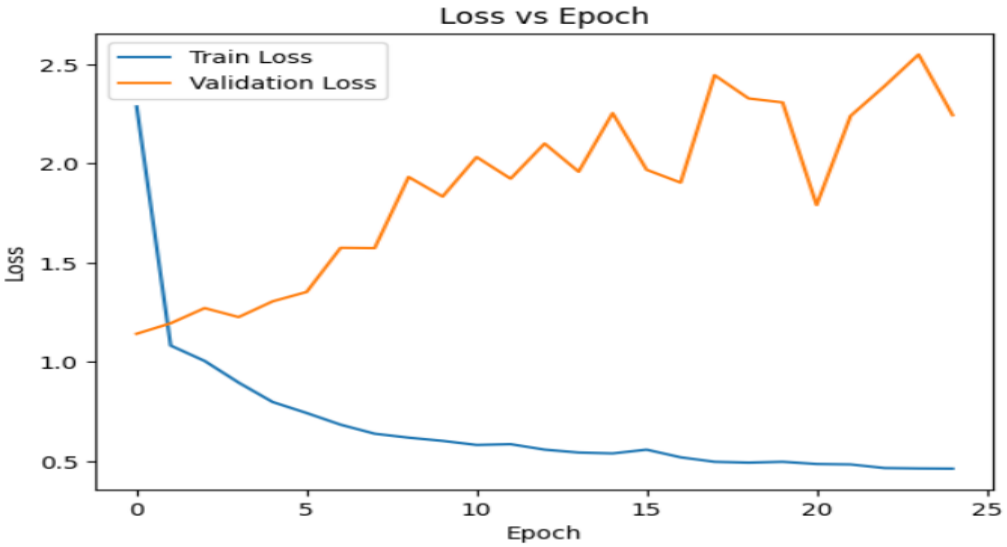


Plot 3 : Loss vs Epochs of VGG16

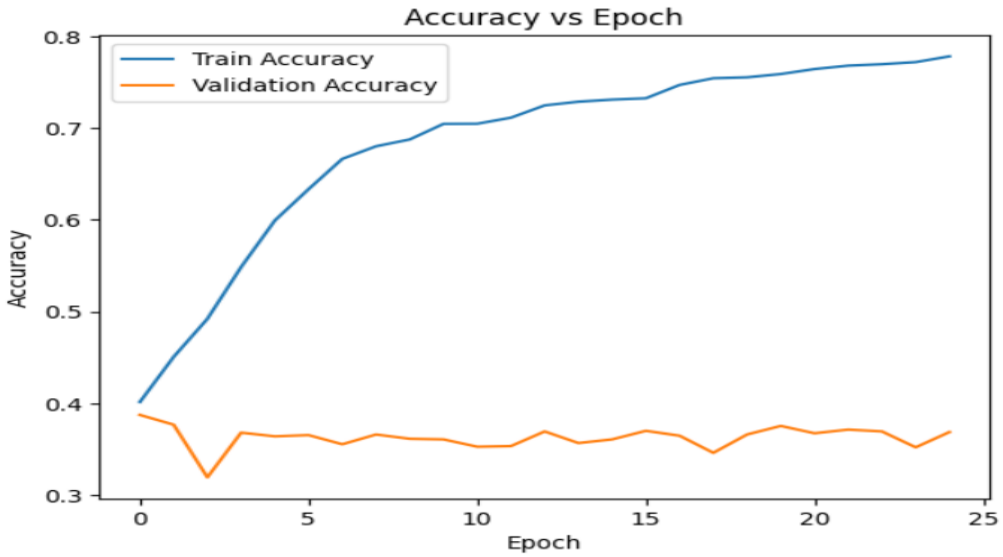


Plot 4 : Accuracy vs Epochs of VGG16

Plots of multimodal(DistilBERTMulti + VGG16) -



Plot 5 : Loss vs Epochs of DistilBERTMulti + VGG16



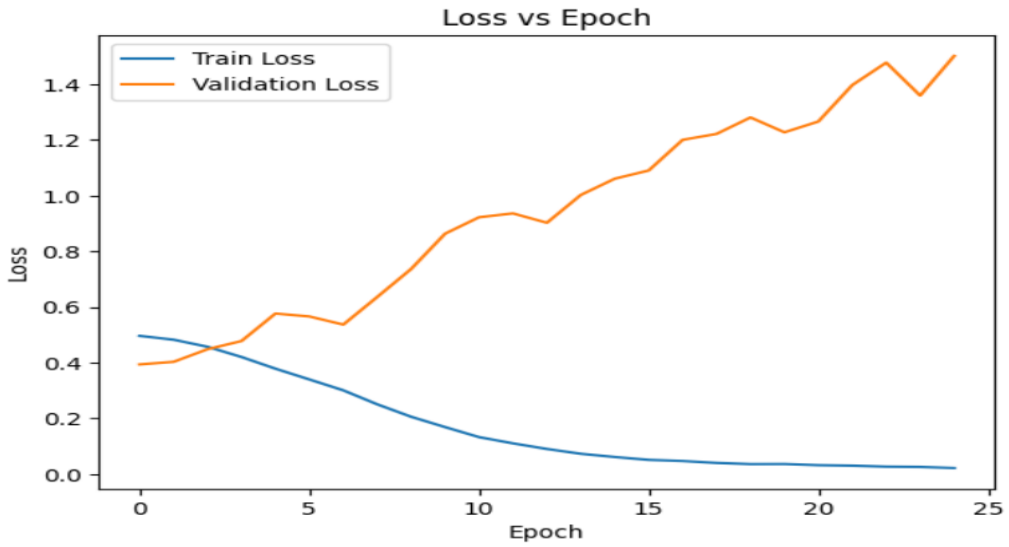
Plot 6 : Accuracy vs Epochs of DistilBERTMulti + VGG16

Table 1. Accuracy, Precision, Recall, F1-weighted

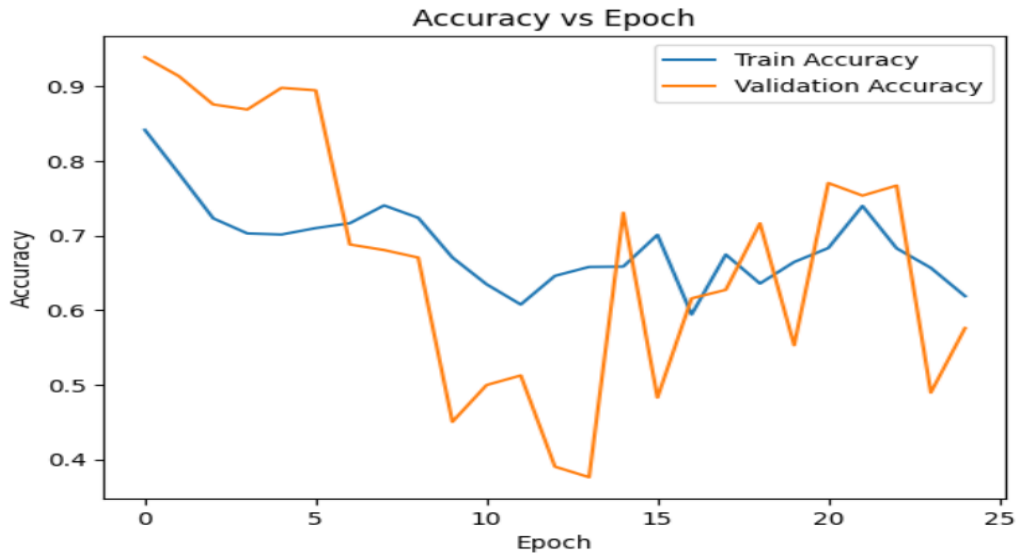
Model	Accuracy	Precision	Recall	F1-score
Text Only (DistilBERTMulti)	0.3458	0.3353	0.3373	0.3339
Image Only (VGG16)	0.3567	**0.3723**	0.3567	**0.3600**
Multimodal (DistilBERTMulti + VGG16)	**0.3687**	0.3548	**0.3687**	0.3563

4.2 Task B

Plots of Text only Model(DistilBERTMulti) -

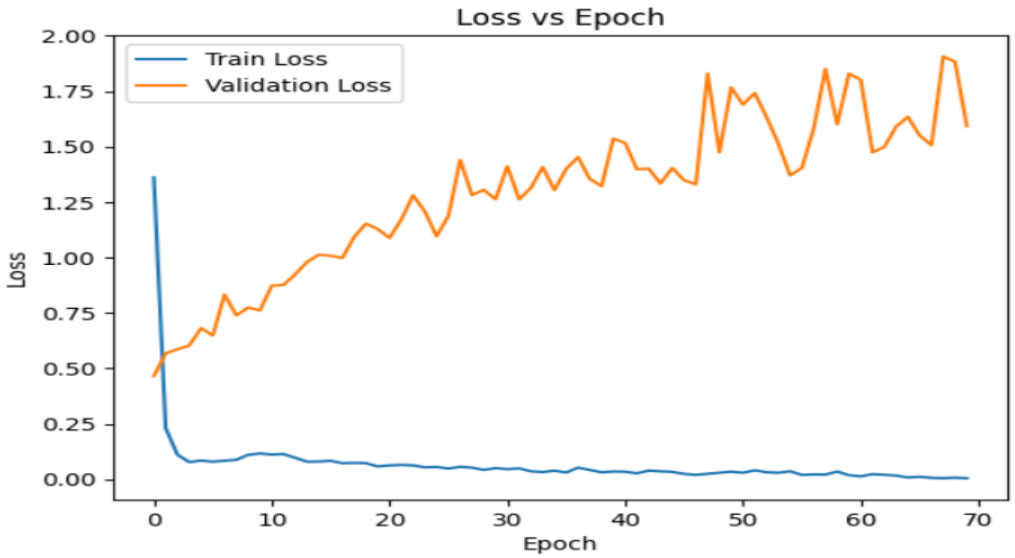


Plot 7 : Loss vs Epochs of DistilBERTMulti

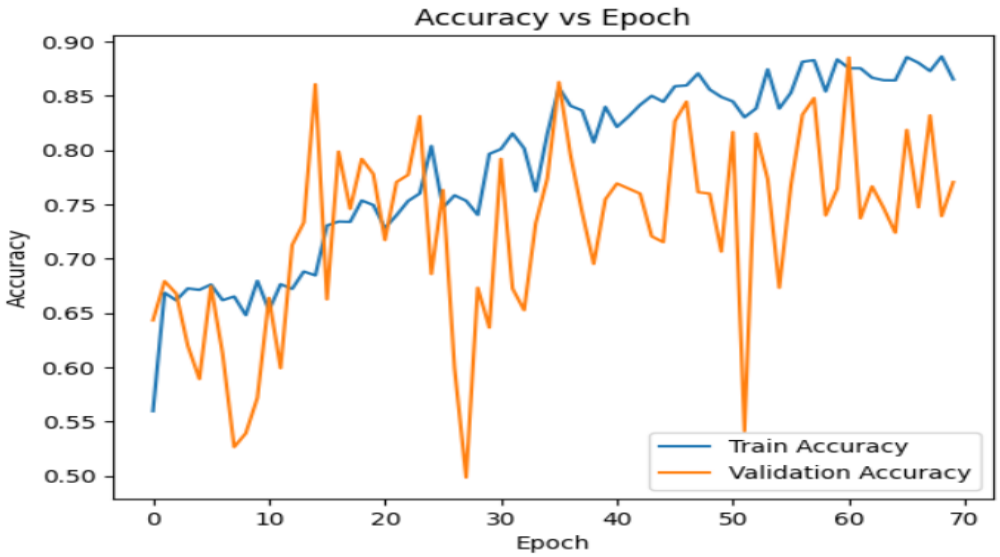


Plot 8 : Accuracy vs Epochs of DistilBERTMulti

Plots of Image only Model(VGG16) -

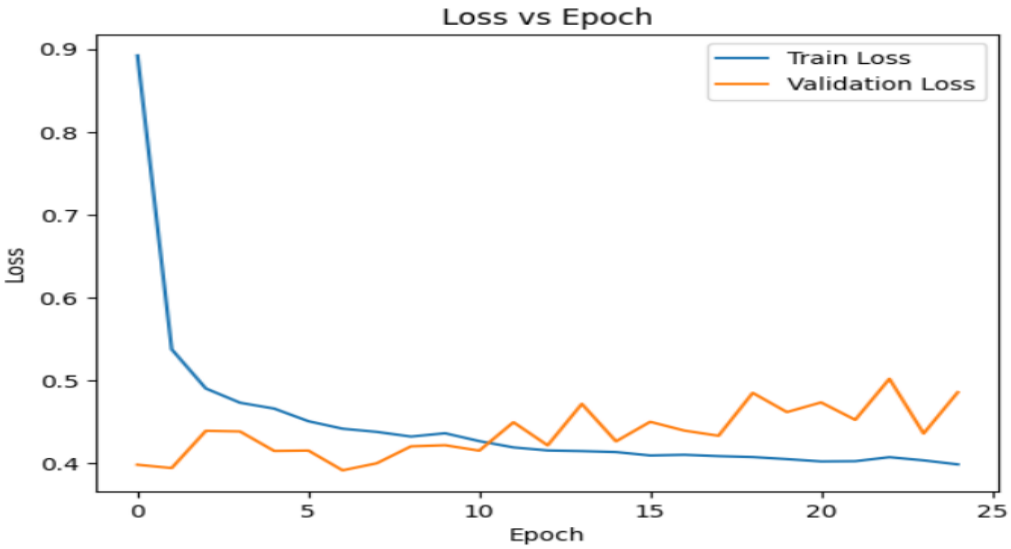


Plot 9 : Loss vs Epochs of VGG16

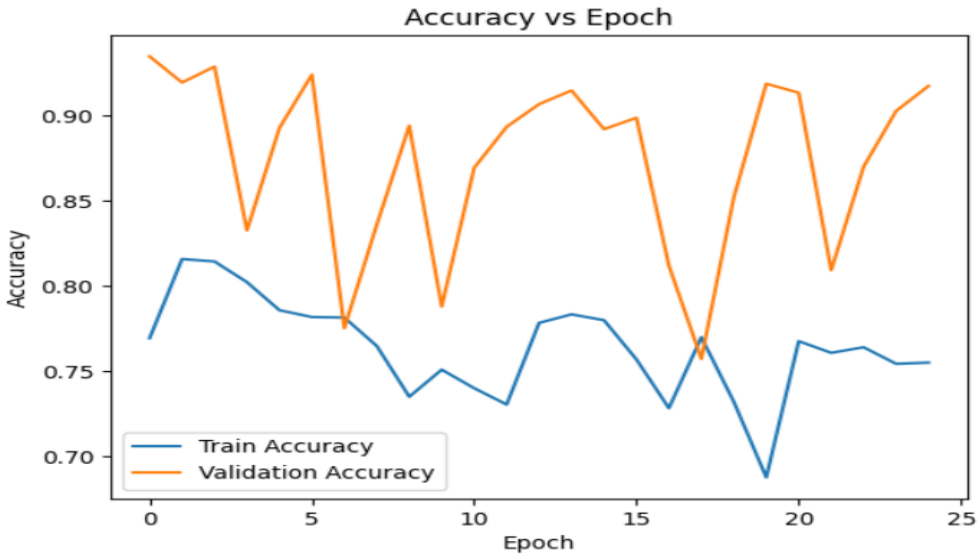


Plot 10 : Accuracy vs Epochs of VGG16

Plots of multimodal(DistilBERTMulti + VGG16) -



Plot 11 : Loss vs Epochs of DistilBERTMulti + VGG16



Plot 12 : Accuracy vs Epochs of DistilBERTMulti + VGG16

Accuracy, Precision, Recall, F1-weighted

(H = humour, S = sarcasm, O = offensive, M = motivational)

Table 2. F1-weighted of different classes

Model	H	S	O	M
Text Only (DistilBERTMulti)	0.8659	0.8323	**0.4788**	0.9326
Image Only (VGG16)	0.8996	0.8720	0.4388	0.9475
Multimodal (DistilBERTMulti + VGG16)	**0.9004**	**0.8764**	0.4198	**0.9562**

Table 3. Precision of different classes

Model	H	S	O	M
Text Only (DistilBERTMulti)	0.8776	**0.8520**	0.4965	**0.9474**
Image Only (VGG16)	**0.8819**	0.8498	**0.5351**	0.9443
Multimodal (DistilBERTMulti + VGG16)	0.8721	0.8424	0.5309	0.9434

Table 4. Recall of different classes

Model	H	S	O	M
Text Only (DistilBERTMulti)	0.8548	0.8143	**0.4780**	0.9187
Image Only (VGG16)	0.9226	0.8986	0.4766	0.9506
Multimodal (DistilBERTMulti + VGG16)	**0.9306**	**0.9133**	0.4680	**0.9693**

Table 5. Accuracy of different classes

Model	H	S	O	M
Text Only (DistilBERTMulti)	0.8548	0.8143	**0.4780**	0.9189
Image Only (VGG16)	0.9226	0.8986	0.4766	0.9506
Multimodal (DistilBERTMulti + VGG16)	**0.9306**	**0.9133**	0.4680	**0.9693**

Table 6. Overall Average Scores

Model	Accuracy	Precision	Recall	F1-score
Text Only (DistilBERTMulti)	0.7665	0.7934	0.7665	0.7774
Image Only (VGG16)	0.8121	**0.8029**	0.8121	**0.7894**
Multimodal (DistilBERTMulti + VGG16)	**0.8203**	0.7972	**0.8203**	0.7882

4.3 Overview with works of others

In the tables provided below, you can see an overview of our work alongside the works of others.

Table 7. Leaderboard of teams on Task A[8]:

Rank	Team	F1-scores
1.	NUAA-QMUL-AIIT	**0.3441**
2.	NYCU_TWO	0.3420
3.	CUFE	33.77
4.	CSECU-DSG	0.3334
5.	Baseline	0.3328
6.	wentaorub	0.3288

Our Model (Multimodal)	F1-scores
DistilBERTMulti + VGG16	**0.3563**

Table 8. Leaderboard of teams on Task B[8]:

Rank	Team	F1 scores				
		H	S	O	M	Overall
1.	Wentrob	0.8891	0.8674	0.4899	**0.9444**	**0.7977**
2.	NYC_TWO	**0.8984**	**0.8691**	0.4317	**0.9444**	0.7834
3.	NUAA-QMUL-AIIT	0.8706	0.7797	0.5072	**0.9444**	0.7755
4.	BASELINE	0.8455	0.7482	0.4884	0.9078	0.7474
5.	CUFE	0.8226	**0.8691**	**0.5078**	0.7760	0.7439
6.	CSECU-DSG	0.8864	0.8630	0.4932	0.6479	0.7226

Our Model	F1 scores				
Multimodal	H	S	O	M	Overall
DistilBERTMulti + VGG16	**0.9004**	**0.8764**	0.4198	**0.9562**	**0.7882**

We believe this comprehensive overview demonstrates the effectiveness of our approach, particularly in achieving better scores in Task A and Task B, excluding offensiveness detection. We hope this comparative analysis inspires further innovation and advancements in the domain.

Chapter 5

Conclusion and Future Work

Our exploration into internet meme classification using multimodal learning techniques has revealed significant insights into the complexities of digital culture and the challenges of analyzing and categorizing memes. Through a comprehensive investigation spanning theoretical foundations, dataset collection, model architectures, and results analysis, we have gained a deeper understanding of the dynamics underlying internet memes and their classification.

5.1 Key Findings

Internet Meme Phenomenon: Internet memes represent a unique form of cultural expression, blending humor, creativity, and intertextuality to convey complex ideas and emotions in a concise and accessible format. Understanding the phenomenon of internet memes is essential for deciphering digital culture and online discourse.

Challenges in Meme Analysis: Despite their widespread popularity, memes present significant challenges for analysis and classification due to their multimodal nature, variability in content, and cultural context dependence. Traditional analytical methods often fall short in capturing the nuances of meme content and context.

Multi-Modal Analysis Approach: Multi-Modal Analysis Approach: To address the complexities of meme analysis, we adopted a multi-modal analysis approach that integrates both textual and visual modalities. Leveraging advances in computer vision and natural language processing, our approach offers a comprehensive understanding of meme content, capturing both visual and semantic elements.

Model Architectures: We explored the architectures of DistilBERTMulti, VGG16, and a multimodal fusion model, combining the strengths of both architectures to tackle multimodal classification tasks. DistilBERTMulti excels in text analysis, while VGG16 specializes in image feature extraction, enabling the multimodal model to capture rich semantic and visual information.

Results Analysis: Our experimental results demonstrate the effectiveness of multimodal learning in meme classification tasks. The multimodal fusion model outperforms text-only and image-only models, achieving higher accuracy, precision, recall, and F1-score across sentiment analysis and emotion classification tasks.

5.2 Future Works

1. Enhanced Multi-Modal Fusion Techniques While our current approach combines textual and visual modalities for meme classification, future research could explore more sophisticated fusion techniques. Techniques such as attention mechanisms and graph-based fusion could be investigated to better capture the interactions between textual and visual elements in memes.

2. Fine-Tuning on Specialized Domains The Memotion 3 dataset provides a diverse collection of internet memes, but specific domains, such as political memes or memes in niche communities, may require specialized models. Future work could involve fine-tuning our classification system on domain-specific datasets to improve performance in targeted meme classification tasks.

3. Incremental Learning and Adaptation As internet meme culture evolves rapidly, our classification system should be capable of adapting to new trends and emerging meme formats. Incremental learning techniques could be explored to continuously update our model with fresh data, ensuring its relevance and accuracy over time.

4. Exploration of Cross-Lingual and Cross-Cultural Analysis Given the global nature of internet memes, there is significant potential in exploring cross-lingual and cross-cultural meme analysis. Future research could focus on developing models capable of understanding memes in multiple languages and cultural contexts, facilitating more inclusive and comprehensive meme classification.

5. Interdisciplinary Collaboration Collaboration with experts from diverse fields, including linguistics, psychology, and sociology, could enrich our understanding of meme culture and improve the interpretability of our classification system. Interdisciplinary research endeavors could uncover deeper insights into the socio-cultural dynamics driving meme creation and dissemination.

6. Ethical Considerations and Bias Mitigation As with any automated classification system, it's crucial to address ethical concerns and mitigate potential biases. Future works should prioritize fairness, transparency, and accountability in meme analysis, ensuring that our models do not perpetuate harmful stereotypes or misinformation.

7. Deployment in Real-World Applications Ultimately, the success of our meme classification system lies in its practical applications. Future efforts should focus on deploying our model in real-world scenarios, such as content moderation platforms or social media analytics tools, to assist users in navigating and understanding the vast landscape of internet memes.

By pursuing these avenues of research, we can continue to push the boundaries of meme analysis and unlock new insights into the ever-evolving world of digital culture.

Bibliography

[1] defactify.com. https://defactify.com.

[2] AIISC. Memotion Dataset, 2023.

[3] Chiruzzo, L., Castro, S., Góngora, S., Rosa, A., Meaney, J., and Mihalcea, R. Overview of haha at iberlef 2021: Detecting, rating and analyzing humor in spanish. *Procesamiento de Lenguaje Natural 67* (09 2021), 257–268.

[4] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding, 10 2018.

[5] Habibullah, Ullah, Md Kamal, M., and Sohag, Md Rahman, M. S. Improved convolutional neural network and transfer learning with vgg16 approach for image classification. pp. 389–396.

[6] Kinger, S., Kinger, D., Thakkar, S., and Bhake, D. Towards smarter hiring: resume parsing and ranking with yolov5 and distilbert. *Multimedia Tools and Applications* (03 2024), 1–19.

[7] Kokab, S., Asghar, S., and Naz, S. Transformer-based deep learning models for the sentiment analysis of social media data. *Array 14* (04 2022), 100157.

[8] Mishra, S., S, S., Chakraborty, M., Patwa, P., Rani, A., Chadha, A., Reganti, A., Das, A., Sheth, A., Chinnakotla, M., Ekbal, A., and Kumar, S. Overview of memotion 3: Sentiment and emotion analysis of codemixed hinglish memes, 09 2023.

[9] Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Multimodal deep learning. pp. 689–696.

[10] Sanh, V., Debut, L., Chaumond, J., and Wolf, T. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter, 10 2019.

[11] Shifman, L. *Memes in digital culture*. MIT press, 2013.

[12] Upreti, A. Convolutional neural network (cnn). a comprehensive overview, 08 2022.

[13] Wechsberg, M. *Darwin & Dawkins–Gene und Meme*. na, 2009.