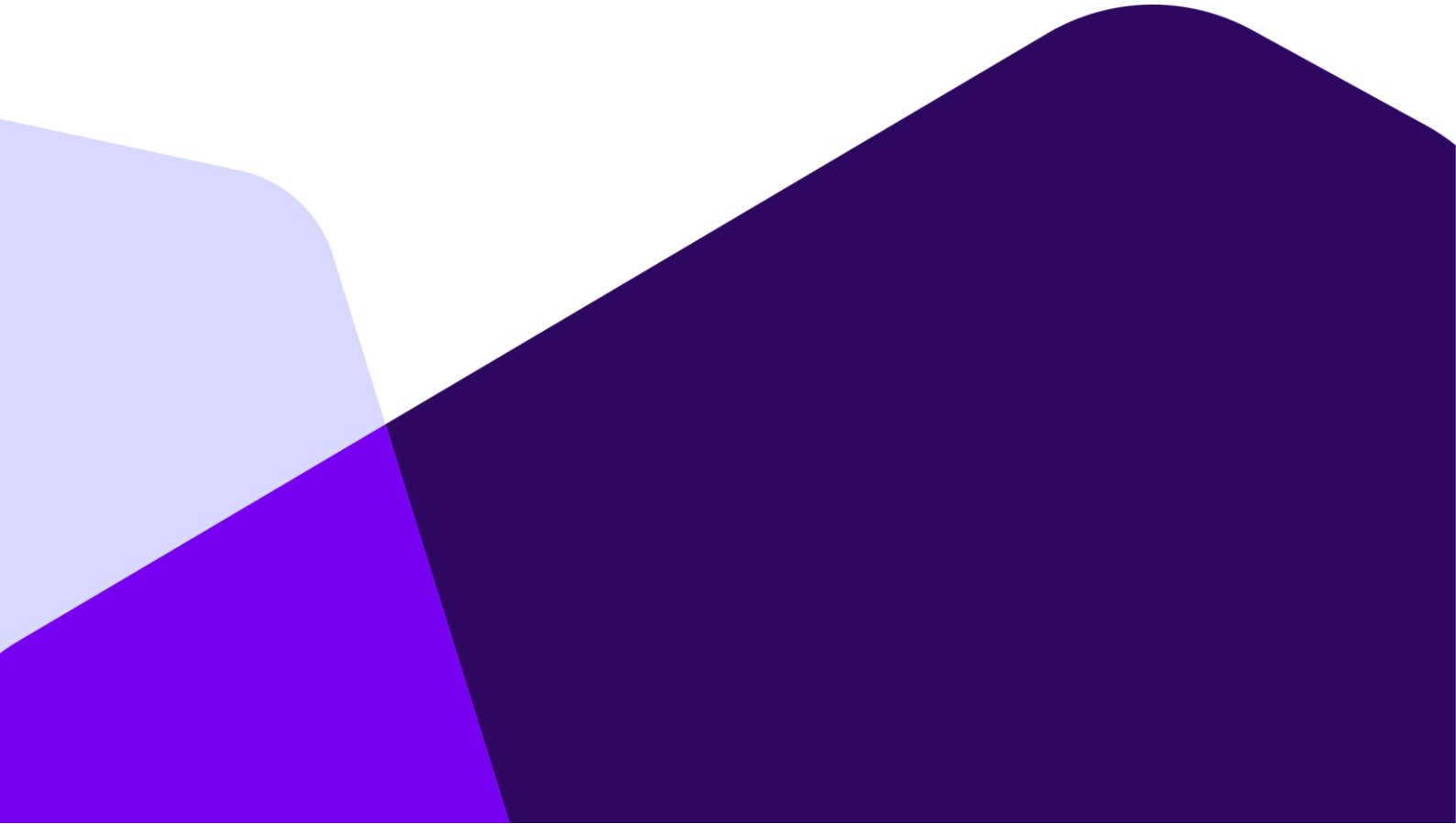


Md Samiul Alam Siam / candidate number: 8514

A hybrid deep learning approach for binary classification of fake news from social medias' multimodal data



University of South-Eastern Norway

Faculty of Technology, Natural Sciences, and Maritime Sciences

Institute of Science and Industry Systems

PO Box 235

NO-3603 Kongsberg, Norway

<http://www.usn.no>

© 2024 Md Samiul Alam Siam

This thesis is worth 60 study points



University of
South-Eastern Norway

A hybrid deep learning approach for binary classification of fake news from social medias' multimodal data

Master's Thesis in Computer Science

Md Samiul Alam Siam

Supervisors

Vimala Nunavath

Maged Helmy

University of South-Eastern Norway

Faculty of Technology, Natural Sciences and Maritime Sciences

Department of Science and Industry Systems

Campus Kongsberg

May 2024

Abstract

In this digital age, social media sites are very important for spreading news. However, they are also great places for fake news to grow, which spreads false information widely and causes problems in society. This thesis tries to solve the problem of classifying fake news by combining different types of data and using BERT and DistilBERT to process text and ResNet34 and ResNet50 to process images. A lot of experiments were done and the obtained results showed that the BERT + ResNet50 model worked best, getting a high accuracy rate of 94%. Textual and visual data are captured and combined very well in this way, making it easier to classify fake news. The study shows that advanced mixed models are better than old-fashioned ways. It also gives us a solid foundation for making more progress in this important area in the future. The study shows how important it is to choose the right model designs to deal with the complicated problem of fake news on social media.

Keywords

Fakeddit, deep learning, multimodal, transformer architecture

Acknowledgements

I would like to convey my profound appreciation to all individuals who have made valuable contributions to the effective culmination of this thesis. Above all, I am very appreciative of my advisor, Dr. Vimala Nunavath, for their significant mentorship, assistance, and motivation during this research endeavor. The user's thoughts and knowledge have played a vital role in the creation and successful completion of this work.

I would like to express my gratitude to the individuals who served on my thesis committee, namely Maged Helmy. His valuable input and guidance have significantly enhanced the caliber of my thesis. I am profoundly grateful for the time and effort they dedicated to examining my work.

I express my gratitude to the University of South-Eastern Norway for providing the necessary financial assistance that enabled the execution of this study. The funds and resources offered played a crucial role in executing this project.

Lastly, I would like to convey my utmost appreciation to my family. To my parents, for your unwavering love, unwavering support, and unwavering encouragement throughout my academic career.

We appreciate the donations and support from everyone. This thesis owes its existence to your indispensable contribution.

Contents

1	Introduction	11
1.1	Motivation and Problem Statement	11
1.2	Research Questions	12
1.3	Thesis Goals	13
1.4	Research Approach	14
1.4.1	Applied Research	14
1.4.2	Applied Research Contain	14
1.4.3	Rationale for Choosing Applied Research	15
1.5	Approach	16
1.6	Assumptions and Limitations	17
1.6.1	Assumptions	17
1.6.2	Limitations	18
1.7	Thesis Contributions	19
1.8	Thesis Outline	20
2	Background	21
2.1	Social Media Platforms	21
2.2	Fake News Detection and Classification	22
2.3	Machine Learning	23
2.3.1	Traditional Machine Learning Algorithms	23
2.3.2	Limitations of Traditional Machine Learning	23
2.4	Deep Learning	23
2.4.1	Neural Networks and Deep Learning	24
2.4.2	Applications of Deep Learning	24
2.5	Transformer Architectures	24
2.6	BERT	26
2.7	ResNet Model	28
2.7.1	ResNet-34	29
2.7.2	ResNet-50	30
2.8	Hyperparameters and Hyperparameter Tuning	31
2.9	Performance metrics and Classification	31

3	State of the Art	33
3.1	Identified Gaps and Proposed Solutions in Literature	43
3.2	State-of-the-Art Approaches in Fake News Detection	44
4	Research Methodology	47
4.1	Proposed Solution	47
4.1.1	Dataset	48
4.1.2	Data Preprocessing	49
4.1.3	Textual Data Preprocessing	51
4.1.4	Image Data Preprocessing	53
4.2	Model Architectures	56
4.2.1	BERT	57
4.2.2	Resnet-50	57
4.2.3	Resnet-34	57
4.2.4	DistilBERT	58
4.2.5	RoBERTa:	59
4.3	Hybrid Model	60
4.4	Model Training and Evaluation	61
4.4.1	Split data into Training, Validation, and Testing Sets	61
4.4.2	Training	61
4.4.3	Evaluation Metrics	64
4.4.4	Hardware and software resources	65
5	Experiments and Results	68
5.1	Experiments and hyperparameter tuning	68
5.1.1	Experimental setup	68
5.1.2	Hyperparameter Tuning	69
5.2	Results	69
5.2.1	Experiment 1: Hybrid model 1 (BERT + ResNet-50)	69
5.2.2	Experiment 2: Hybrid model 2 (BERT + ResNet-34)	71
5.2.3	Experiment 3: Hybrid model 3 (DistillBERT + ResNet-34)	72
5.2.4	Experiment 4: Hybrid model 4 (RoBERTa + ResNet-34)	74
5.2.5	Analysis and Comparison	76
6	Discussion	78
6.1	Discussing Results for each Research Question	78

6.2	Comparison with Existing Literature	79
6.3	Literature Review Summary	80
7	Conclusion and Future Work	82
7.1	Conclusion	82
7.2	Future Work	83
	References	84
	Appendices	89
A	Example Appendix	90
A.1	Code Repository	90
A.2	Most Common Words	90
A.2.1	Common True Words	90
A.2.2	Common Fake Words	91
A.3	Unimodal	92
A.3.1	Text Unimodal	92
A.3.2	Image Unimodal	98
A.3.3	Analysis and Comparison	99

List of Figures

1	Applied Research Approach, Adapted from [1]	16
2	Transformer Architecture	26
3	BERT	28
4	ResNet-34	30
5	ResNet-50	30
6	Proposed Solution	47
7	Model Architecture	56
8	Confusion Matrix of BERT with ResNet-50	70
9	Confusion Matrix of BERT with ResNet-34	71
10	Confusion Matrix of DistilBERT with ResNet-34	73
11	Confusion Matrix of RoBERTa with ResNet-34	74
12	Common True Words	90
13	Common Fake Words	91
14	Confusion matrix with DistilBERT	92
15	Avg. Words/Sentence	95
16	Avg. Words/Sentence By Group	96
17	KMeans Clustering	97
18	Confusion matrix with ResNet34	98

List of Tables

2	Table Literature Review	46
3	Classification report of using BERT with Resnet-50	69
4	Classification report of using BERT with Resnet-34	71
5	Classification report of using DistilBERT with Resnet-34	72
6	Classification report of using DistilBERT with Resnet-34	74
7	Classification Results	76
9	Table Literature Review	81
10	Classifying report on using DistilBERT	93
11	Classifying report on using BERT	94
12	Classifying report on using Resnet-34	99
13	Unimodal Classification Results	100

Glossary

nlp name=NLP, description=Natural Language Processing

ml name=ML, description=Machine Learning

cnn name=CNN, description=Convolutional Neural Network

rnn name=RNN, description=Recurrent Neural Network

bert name=BERT, description=Bidirectional Encoder Representations from Transformers

resnet name=ResNet, description=Residual Network

Acronyms

NLP Natural Language Processing

ML Machine Learning

CNN Convolutional Neural Network

RNN Recurrent Neural Network

BERT Bidirectional Encoder Representations from Transformers

ResNet Residual Network

1 Introduction

1.1 Motivation and Problem Statement

In today's digital age, social media platforms play an important role in the spread of news and information [2]. However, the fast expansion of these platforms has coincided with an increase in the spread of false news, which can have serious effects ranging from individual misunderstanding to social divisiveness, public health issues, and political upheaval. The capacity to automatically and reliably differentiate real content from disinformation is critical for preserving the integrity of public debate and protecting the information ecosystem. Traditional techniques such as rule-based systems [3] and keyword matching algorithms [4], are becoming more ineffective owing to the sheer amount and velocity of online data content, verification is becoming more ineffective owing to the sheer amount and velocity of online data [5].

In spite of significant progress in artificial intelligence and machine learning, detecting fake news on social media is still difficult. This is because fake news methods are complex, always changing, and deceptive. Researchers and developers have to keep adjusting their methods to detect fake news, so we need flexible systems that can classify newly generated information all the time. Also, social media has lots of different kinds of information including text, photos, and videos, adding complexity that classic approaches such as rule-based systems and keyword-matching algorithms cannot manage well [6]. The sheer volume and velocity of internet data complicate real-time analysis since standard models fail to scale up to meet these needs while also generalizing well across several platforms and situations [7]. Furthermore, there is a scarcity of big, labeled datasets required for training complex models, and the dynamic nature of false news needs frequent updates with fresh data to ensure efficacy [8]. These combined factors—evolving disinformation methods, multimodal content complexity, scalability challenges, and ongoing data needs—help to explain why the problem of identifying fake news on social media has yet to be entirely solved.

Despite significant breakthroughs in natural language processing (NLP) and computer vision, identifying fake news within multimodal social media material remains challenging due to a number of fundamental variables. First, the sheer volume and speed with which material is created and disseminated on social media sites makes real-time identification difficult. Furthermore, false news frequently employs complex and developing strategies, such as the use of modified photos or videos in conjunction with deceptive written material, necessitating the deployment of advanced models capable of integrating and evaluating several types of data at once. Furthermore, the environment

in which information is delivered can dramatically influence perception, necessitating models that comprehend not only the content but also the intricacies of how it is shared and absorbed by various audiences. Language, slang, and cultural allusions vary among locations and communities, complicating the work even more. Another problem is the adversarial nature of false news, with authors constantly adapting their ways to avoid detection, necessitating models that can learn and update dynamically.

Furthermore, integrating NLP and computer vision models presents technical challenges, as it requires mixing high-dimensional input from several modalities while preserving their distinct characteristics and context. Existing datasets for training such algorithms may be insufficiently diverse and large to cover the entire range of false news events. Finally, ethical and privacy issues are important, since automated detection systems must balance accuracy with the possibility of bias, as well as user data protection. All of these variables add to the challenge of creating effective algorithms for identifying false news in multimodal social media material. Existing models that solely handle one type of information (text or picture) fail to grasp the subtle interaction of textual and visual clues that distinguish false material. Furthermore, the efficacy of existing multimodal techniques varies greatly, with some models showing promise in experimental conditions but failing to produce robust performance in real-world situations. This mismatch indicates flaws in model architecture and training [8], [6], [7].

As a result, the main objective of this thesis is to develop a hybrid model for false social media post identification and classification using social media's multi-modal data. To achieve the objective of this thesis, sophisticated models such as BERT, DistilBERT, and RoBERTa will be used for text data, and convolutional networks like ResNet34, and ResNet50 will be used for image data. The output from these individual models will be concatenated for multi-modal data for the binary classification of social media posts as either fake or real posts.

1.2 Research Questions

To achieve the main objective of this thesis, the following research questions are proposed.

1. **RQ1:** What are the most suitable feature extraction techniques for multi-modal fake (text and image) data classification?

For multi-modal fake data analysis, the best feature extraction methods use advanced models that are made to fit each type of data. BERT (Bidirectional Encoder Representations from Transformers) works really well with text data because it can get contextual embeddings through deep transformer layers. The text data is tokenized and stored using BERT. This creates rich,

relevant embeddings that show what the text means. ResNet50 (Residual Network) is best for picture data because it has deep convolutional layers that can pull out hierarchical features from photos. Random shrinking, cropping, horizontal flipping, and normalization are some of the data enhancement methods that are used to make the model more stable. The features taken from both BERT and ResNet50 are then joined together to make a single image. This is then put through thick layers to do the final classification.

2. **RQ2:** Which deep learning models are best suited for multi-modal social media data to classify fake news?

When it comes to multi-modal social media data, the best deep learning models for classifying false news will be developed ResNet50 for picture data and BERT for text data. This is because BERT is well-suited for the detection of misleading or deceptive textual material because of its exceptional ability to comprehend and encode textual information's context and semantics. When it comes to detecting altered or deceitful visual information, ResNet50 really shines at extracting precise and hierarchical characteristics from photos. When these two algorithms work together, they can detect signs of false news in both text and images. More accurate categorization will be achieved by integrating these models by concatenating their feature vectors. This allows for a thorough examination of multimodal data.

3. **RQ3:** Does deep neural network perform better than state-of-the-art algorithms?

In order to investigate this research issue, we will evaluate the effectiveness of deep neural networks in contrast to the most advanced algorithms available. Logistic regression (LR) will be used as a benchmark for our comparisons. This approach enables us to assess the efficacy of deep learning models within the framework of conventional machine learning approaches. We will evaluate many performance indicators, including accuracy, precision, recall, and F1-score, to ascertain if deep neural networks offer substantial enhancements compared to LR and other current techniques. Our objective is to determine if the additional features of deep learning architectures provide practical advantages in the field of identifying and categorizing fake news.

1.3 Thesis Goals

- The main objective of this thesis is to create and assess a sophisticated system for identifying false news. This system will utilize multimodal data from social media, including the analysis of both text and images. The primary objective of this method is to effectively tackle the pressing issue of precisely detecting false information, which is widespread on different social media platforms.

- This project aims to identify the most efficient feature extraction strategies for processing multimodal data by employing advanced deep learning models including BERT, DistilBERT, RoBERTa for text analysis, and ResNet34 and ResNet50 for picture analysis. The primary objective is to determine the most effective strategies for capturing the subtle details and contextual hints included in both text and images. This will significantly improve the model's accuracy in detecting false news.
- This thesis seeks to evaluate the performance of deep neural networks in comparison to traditional state-of-the-art algorithms. The objective is to determine if deep learning models offer higher levels of accuracy and reliability in detecting false news. This entails a comprehensive assessment of the performance parameters of the models, such as accuracy, precision, recall, and F1 score, in order to ascertain the most efficient technique for this demanding work.
- Evaluate the performance of different deep learning algorithms on the collected dataset.
- The primary objective is to develop a resilient and adaptable model that not only enhances scholarly understanding but also provides effective remedies for stakeholders in countering the dissemination of false information. This project aims to enhance the integrity of information shared on social media platforms and mitigate the detrimental effects of false news on society by offering a dependable tool for fake news identification.

1.4 Research Approach

1.4.1 Applied Research

Applied research is a form of study that concentrates on resolving practical problems and creating inventive ways to tackle specific concerns. Basic research focuses on expanding fundamental knowledge and understanding, whereas applied research strives to utilize current information in practical real-world scenarios. This type of study has direct relevance to daily life and is frequently carried out with the aim of influencing or enhancing habits, procedures, or goods. This process entails doing empirical research and using scientific methodologies to examine hypotheses and verify results.

1.4.2 Applied Research Contain

Applied research often includes the following components:

- **Problem Identification:** Precisely delineating the practical problem or difficulty that requires attention.

- **Literature Review:** Evaluating prior research and theoretical frameworks to comprehend the present level of understanding and pinpoint areas that require more investigation.
- **Research Questions and Hypotheses:** Developing precise research inquiries and conjectures that direct the inquiry.
- **Methodology:** Creating a study strategy that involves choosing suitable techniques and instruments for gathering and analyzing data. This may entail conducting experiments, administering surveys, conducting case studies, or employing other empirical methodologies.
- **Data Collection:** The process of acquiring pertinent data using diverse methods, including observations, interviews, experiments, or existing databases.
- **Data analysis:** Examining the gathered data via statistical, computational, or qualitative techniques to evaluate hypotheses and draw conclusions.
- **Results and Findings:** This section will present the outcomes of the analysis, which may include any observed patterns, correlations, or noteworthy discoveries.
- **Discussion and Implications:** Analyzing the results in relation to the study objectives, examining their significance for practical use, and proposing potential uses.
- **Conclusion:** To conclude, this section provides a concise summary of the research procedure, the obtained findings, and the practical suggestions derived from the study.
- **Dissemination:** Disseminating the study findings to pertinent stakeholders via reports, publications, presentations, or other means of communication.

1.4.3 Rationale for Choosing Applied Research

The study adopts the applied research technique as it is in line with the objective of creating practical solutions for the intricate problem of identifying false news on social media. This technique enables us to directly tackle the difficulties presented by multimodal false news material, encompassing both textual and visual elements, and develop models that can be efficiently used in real-life situations. By prioritizing empirical examination and practical application, we can guarantee that our research not only adds to academic knowledge but also provides actual advantages for practitioners and policymakers.

This approach is especially pertinent considering the fast development of social media and the growing complexity of deceptive news tactics. Applied research allows us to:

- **Directly Address Real-World Problems:** By prioritizing pragmatic obstacles, we may devise solutions that have immediate applicability and advantageous outcomes.
- **Utilize Existing Knowledge:** It utilizes existing ideas and approaches, adapting them to novel situations in order to address particular challenges.
- **Enhance Practical Relevance:** The practical nature of the findings we obtained makes them highly relevant to stakeholders seeking effective tools and tactics to prevent the spread of fake news.
- **Bridge the Gap Between Theory and Practice:** Applied research serves as a bridge between theoretical knowledge and practical implementation, ensuring that academic progress is effectively applied in real-life situations.
- **Adapt to Emerging Issues:** Adaptable and flexible research methodologies are necessary to keep up with the ever-changing and challenging nature of false news and social media.

At the end of the the applied research strategy is selected due to its capacity to generate practical insights and solutions that may greatly enhance the identification of false news and support the wider endeavor of upholding information integrity on social media platforms.

1.5 Approach

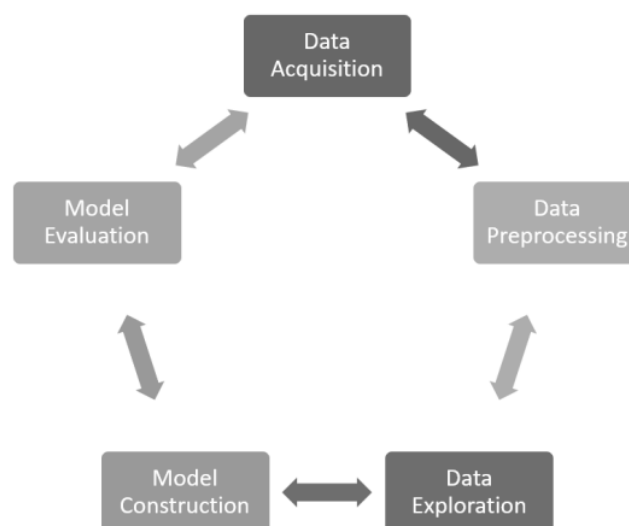


Figure 1: Applied Research Approach, Adapted from [1]

- **Data Acquisition:** The first phase is gathering an extensive dataset that encompasses both textual and visual information from various social media networks. This dataset consists of

postings that have been categorized as either authentic or fabricated news, serving as a basis for training and assessment purposes.

- **Data Preprocessing:** The data that is used goes through a lot of preparation to make sure it is accurate and consistent. Preprocessing text data will be done including cleaning the text, tokenizing it with the BERT tokenizer, and turning it into input IDs and attention masks. As part of preparing picture data, changes like random resizing, cropping, flipping, and leveling will be used.
- **Data Exploration:** Exploratory data analysis (EDA) will be done at this stage to figure out how the information is distributed and what its features are. Visualization methods and statistical analyses help find trends, outliers, and possible biases in the data, which in turn helps with making decisions about how to preprocess the data and design the model.
- **Model Construction:** Hybrid models will be developed by using multimodal models such as BERT for text processing and ResNet50 for picture processing. The model's structure has thick layers for extracting features from both picture and text and then a concatenation layer to join the features of both types. Regularization will be done with dropout layers, and the final classification will be done with a softmax layer. This design is wrapped up in the FakeNewsDetector class, which uses models that will be trained for both text and picture modalities.
- **Model Training and Optimization:** The model will be trained using the dataset that has undergone preprocessing. During training, many strategies will be employed, including data augmentation for pictures, and weighted loss functions to address the class imbalance and learning rate schedule. The training loop iterates through numerous epochs, refining the model parameters using the AdamW optimizer. It monitors performance metrics to modify learning rates and applies early stopping as required.
- **Model Evaluation:** The model will be carefully tested on a different validation sample after it has been trained. To judge how well the model works, metrics like accuracy, precision, recall, and F1-score will be determined. The model will be trained on a test sample that has never been seen before to see how well it can generalize.

1.6 Assumptions and Limitations

1.6.1 Assumptions

Several important theories were used to guide the development and testing of the fake news detection system in this study. The set of social media posts used in this study is thought to be a good

representation of the wider population of fake and real news material, showing the variety and complexity that can be found on these sites. This is important for training a model that can tell the difference between the two without any bias. Also, the names (like "fake" or "real") that are given to the dataset are assumed to be correct and consistent. This is important for making sure that the training and review processes are valid. Also, the steps used to prepare the text and image data—for example, tokenization with the BERT tokenizer for text and picture enhancement techniques—are thought to work the same way on all cases. The features that the BERT and ResNet50 models find are thought to be important and enough to tell the difference between fake and real news. They pick up on the necessary language and visual cues that show how real the content is.

1.6.2 Limitations

- The model's usefulness may be limited to the specific dataset used for training and evaluation.
- Social media platforms are distinguished by their dynamic nature, always offering new content and formats.
- The model's efficacy may be diminished when confronted with unfamiliar data or novel variations of fake news that were not incorporated into the training dataset.
- Data augmentation strategies may not fully capture the authentic diversity of images encountered in the real world.
- The model may have challenges when handling whole new image categories that were not adequately represented in the expanded training dataset.
- The utilization of BERT and ResNet50 models requires substantial computer resources, which may not be readily accessible to all researchers or practitioners due to the computational complexity required.
- The application of the technology in scenarios with low resources is constrained.
- Dependence on pre-trained models such as BERT and ResNet50, which were trained on comprehensive, broad datasets, may lead to disregarding certain nuances that are present in social media content.
- Deep learning models are limited in terms of their interpretability, with complex models such as BERT and ResNet50 often being criticized for their lack of transparency.
- Lack of transparency might hinder the ability to fully trust or understand the model's decisions, particularly in critical situations such as news verification.

- Although efforts have been made to create a fair dataset, there can still be underlying biases that the model incorporates.
- The presence of biases in the model might potentially affect its performance, leading it to show a preference for certain types of content or sources. These biases may not be apparent during the initial training and evaluation phases.

Recognizing these constraints is essential for understanding the scope of the study and identifying areas that may require more investigation and improvement.

1.7 Thesis Contributions

This thesis contributes significantly to the field of false news identification by utilizing multimodal data and powerful deep-learning models. The study presents an innovative method that combines text and picture data, employing BERT for text analysis and ResNet34/ResNet50 for image analysis. This dual strategy efficiently acquires and analyzes the diverse multimodal information included in social media material. The study offers a thorough evaluation of deep neural networks in comparison to traditional algorithms like logistic regression, SVM, and random forests. It shows that deep learning models, specifically BERT and its variations, outperform these traditional methods significantly in terms of accuracy, precision, recall, and F1-score.

The models underwent extensive testing on a variety of real-world datasets such as Fakeddit, demonstrating their resilience and excellent precision in identifying false information. This confirms their practical usefulness. The thesis emphasizes the significance of integrating text and picture inputs to enhance detection accuracy by tackling the inherent difficulties of multimodal data processing. The comprehensive performance indicators, such as confusion matrices and classification reports, provide significant insights into the capabilities and constraints of each model setup, offering a clear path for further study.

Moreover, the research acknowledges the constraints of the existing models, including their reliance on substantial computer resources and the potential biases present in the dataset. The paper suggests potential areas for future study to tackle these difficulties, such as improving the efficiency of models and expanding the scope of datasets to be more varied and thorough. The study identifies BERT as a very successful feature extraction strategy for text, and ResNet34/ResNet50 as a highly effective feature extraction strategy for pictures. The combination of BERT and ResNet50 is shown to be the most effective for multimodal fake news classification.

To summarize, this thesis provides evidence that deep neural networks, specifically BERT and its variations, surpass traditional state-of-the-art algorithms in detecting bogus news. This study provides a strong basis for future research and practical applications, offering vital insights and breakthroughs to the continuing efforts to counteract the dissemination of false information on social media platforms.

1.8 Thesis Outline

The rest of the thesis is organized as follows:

Chapter 2: This chapter provides background information for understanding theories, technologies, and domains used later in the thesis.

Chapter 1: The thesis is designed to systematically solve the difficulty of detecting false news in multimodal social media data. It begins with an introduction that explains the rationale for the study and identifies the issue statement, laying the groundwork for the ensuing inquiry.

Chapter 3: Following that, a complete literature analysis is offered, providing insights into existing methodologies and highlighting shortcomings in current approaches.

Chapter 4: The methodology chapter describes the theoretical frameworks and empirical methodologies utilized, such as data collecting, model selection, and integration procedures for textual and visual data.

Chapter 5: This is followed by an extensive testing and results part in which the models' performance is evaluated using multiple metrics to determine their efficacy in actual and virtual contexts.

Chapter 6: The discussion chapter discusses the findings and provides a critical review of the outcomes in relation to the research questions.

Chapter 7: Finally, the conclusion summarizes the study findings and offers future directions.

2 Background

2.1 Social Media Platforms

Social media sites have become an important part of modern communication, and they have a big impact on how information is shared and used around the world. People can share material, talk to each other, and have discussions about a lot of different themes on sites like Facebook, Twitter, and Instagram. These sites have changed how people get news by giving them instant and different views on current events [9]. But the speed with which information spreads on social media also brings about big problems, especially when it comes to the spread of fake news. Because these platforms are designed and their algorithms are based on making material that is interesting and shared, they can spread false information without meaning to. Because of this, people are becoming more worried about how social media affects democracy and public opinion [10]. Studies have shown that because of how easily fake information can spread on these platforms, more advanced ways need to be found to lessen its bad effects [11].

The sheer amount and speed of information shared on social media sites is one of their biggest problems. About 4.48 billion people use social media every day, making a huge amount of user-generated material every day [11]. Traditional ways of checking facts can't keep up with this huge and fast flow of information, so fake information can spread without being stopped. Also, social media systems tend to give more weight to material that gets a lot of interaction, like comments, shares, and likes. This can lead to "echo chambers" where false information spreads quickly among people with similar views, strengthening false beliefs and making it harder to fix mistakes [12].

Moreover, social media sites are not just inactive receivers of information; they also actively shape what people see. Algorithms that are meant to get people to interact with your content more can accidentally push exciting and false information. One example is that studies have shown that fake news stories are 70% more likely to be shared again than real ones [13]. This situation makes it easy for fake news to spread, since false information tends to get more attention and spreads more quickly than true information.

Because social media sites are used all over the world, false information can spread quickly across countries and cultures, spreading false information that can have real-world effects. During the COVID-19 pandemic, for example, fake information about how to treat the virus and where it came from spread quickly on social media. This made people confused and made it harder to handle the

crisis [14].

Because of these problems, social media platforms need to quickly install systems that can spot fake news. To find and stop the spread of fake information, these systems need to be able to look at huge amounts of data in real-time and combine information from different types of media, like text, pictures, and videos.

2.2 Fake News Detection and Classification

The part called "Fake News Detection and Classification" talks about how to find and stop fake news from spreading on social media, as well as the problems that come up along the way. Traditional methods, mostly rule-based systems and phrase matching are becoming less useful because they are rigid and can't keep up with how propaganda changes and becomes more complex [15]. When spreading fake news, these methods don't take into account the situation and are easy to get around with small changes.

Better methods have come about as machine learning and natural language processing have improved [16]. These methods use deep learning models to look for complicated trends in data, which makes them much better at finding things than older methods. Using models like BERT to look at text and ResNet to handle images has raised the bar in the field. These models give strong frameworks for recording how text and images interact in news stories, which is important for telling the difference between true and false information.

For example, new research [17] has shown that using multimodal methods, which include both written and visual data, makes finding fake news much more accurate and reliable. Advanced text processing and image recognition technologies work well together in these ways, giving a full picture of the material being analyzed.

As this field continues to grow, the goal is to make solutions that are more flexible, effective, and scalable so that they can keep up with how quickly social media changes and how misinformation campaigns change too. Adding cutting-edge machine learning technologies to systems that find fake news is a big step toward making sure that all digital platforms protect the purity of information.

2.3 Machine Learning

Machine Learning (ML) is a branch of artificial intelligence (AI) that concentrates on creating algorithms and statistical models to enable computers to accomplish certain tasks without relying on explicit instructions. Instead, these models depend on patterns and inference. Machine learning spans a broad spectrum of approaches, such as supervised learning, unsupervised learning, reinforcement learning, and semi-supervised learning. Some commonly used methods in machine learning are decision trees, support vector machines, and k-nearest neighbors.

2.3.1 Traditional Machine Learning Algorithms

Conventional machine learning methods have served as the foundation for several data-driven applications. Regression models are utilized to forecast continuous outcomes, decision trees offer a graphical depiction of decision rules, and support vector machines are highly effective for classification tasks. Although these algorithms are highly successful, their performance relies greatly on the use of human-designed characteristics and experience in the specific field.

2.3.2 Limitations of Traditional Machine Learning

The growing intricacy and magnitude of contemporary datasets have shown the constraints of conventional machine-learning methods. Feature engineering, the process of manually choosing and modifying variables to enhance model performance, is characterized by its time-intensive nature and susceptibility to human mistakes. Moreover, conventional machine learning models frequently encounter difficulties when dealing with data that has a large number of dimensions and long-term relationships, which diminishes their effectiveness in handling intricate tasks like natural language processing and picture identification.

2.4 Deep Learning

Deep Learning (DL) is a branch of machine learning that uses neural networks with several layers (thus "deep") to represent intricate patterns in data. These neural networks are specifically engineered to replicate the human brain's capacity to acquire knowledge from vast quantities of unorganized input. Deep learning models have the ability to autonomously extract characteristics from unprocessed data, rendering them exceptionally potent for tasks such as picture and speech recognition. Deep learning architectures encompass convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs).

2.4.1 Neural Networks and Deep Learning

Neural networks serve as the foundational components of deep learning. The structure has linked layers of nodes, or neurons, which process data in a hierarchical fashion. Each subsequent layer of the model pulls more abstract properties from the input data, enabling the model to acquire intricate patterns and representations. Deep learning models, because of their several concealed layers, have the ability to comprehend complex patterns in data, rendering them well-suited for jobs that want advanced abstraction.

2.4.2 Applications of Deep Learning

Deep learning has transformed several domains by offering cutting-edge solutions for intricate issues. Convolutional neural networks (CNNs) have achieved significant advancements in computer vision tasks such as picture categorization, object recognition, and image synthesis. Recurrent neural networks (RNNs) and their variations, such as long short-term memory (LSTM) networks, have enhanced machine translation, sentiment analysis, and text synthesis in the field of natural language processing. The widespread use of deep learning in these fields has prompted its acceptance in other sectors, ranging from healthcare to finance.

2.5 Transformer Architectures

Transformer designs have made a big difference in the field of natural language processing (NLP) and are now used as the basis for many cutting-edge models. Vaswani et al. (2017) [18] came up with the transformer model, which is different from regular recurrent and convolutional neural networks. As an alternative, it only uses self-attention methods to handle data relationships correctly. This new method lets transformers handle whole strings of data at the same time, which greatly improves speed and efficiency, especially when working with connections that are far away. The self-attention system is the most important part of the transformer because it lets the model constantly judge the importance of different parts of the entering data.

An encoder and a decoder work together to make a transformer. The encoder takes the input sequence and the decoder makes the output sequence. There are many layers in the transformer model, and each one has self-attention processes and feed-forward neural networks. To help with the training process, these layers are linked by leftover links. The transformer can pick up on complex patterns and relationships in the data thanks to its layered method and the power of self-attention. This makes it very useful for many NLP jobs. The model is more efficient and scalable because it can

process patterns in parallel instead of separately, which is how traditional recurrent networks work. This means that training can be done faster and bigger datasets can be used.

Devlin et al. created BERT (Bidirectional Encoder Representations from Transformers) in 2019 [19]. It is a well-known example of a transformer topology. By providing full two-way models of text, BERT has achieved amazing success in a number of natural language processing (NLP) tasks. BERT reads the whole string of words at once, unlike older models that did so from left to right or right to left. This lets it figure out what a word means by looking at the words that come before and after it. This back-and-forth method helps computers understand language better, which has made big steps forward in areas like asking questions, figuring out how people feel, and recognizing named entities.

Generative Pre-trained Transformer (GPT), which was created by Radford et al. in 2018 [20], is another important model. When it comes to jobs that involve making words, this model does really well. GPT models are made to turn a given input prompt into text that makes sense and fits the situation. For this, they use a pre-training phase where the model learns from a big body of text and a fine-tuning phase where they make it work better for certain jobs. The GPT design has been improved over time, with later versions (GPT-2 and GPT-3) showing even better results and pushing the limits of what is possible in natural language creation. Recent research shows that the transformer design is an important part of multimodal models that mix text and picture data for tasks like finding fake news. This is because it can grow and change.

The provided image (see Image 2) illustrates the architecture of a transformer model, a fundamental framework in natural language processing (NLP). The transformer model consists of two main components: the encoder and the decoder. The encoder is responsible for processing the input data, which in this example is the English sentence "I like science." Through multiple layers of self-attention mechanisms and feed-forward neural networks, the encoder transforms this input into continuous representations that capture the semantic meaning of the words. This enables the model to understand the context and relationships between the words in the sentence.

The decoder, on the other hand, generates the output data, here translating the sentence into German: "Ich mag Wissenschaften." It takes the encoded representation from the encoder and processes it through similar layers, incorporating a mechanism to attend to the encoder's output. This ensures the translation is coherent and contextually accurate. The overall process involves feeding the input sentence into the encoder, which produces a context-rich representation, followed by the decoder generating the translated sentence.

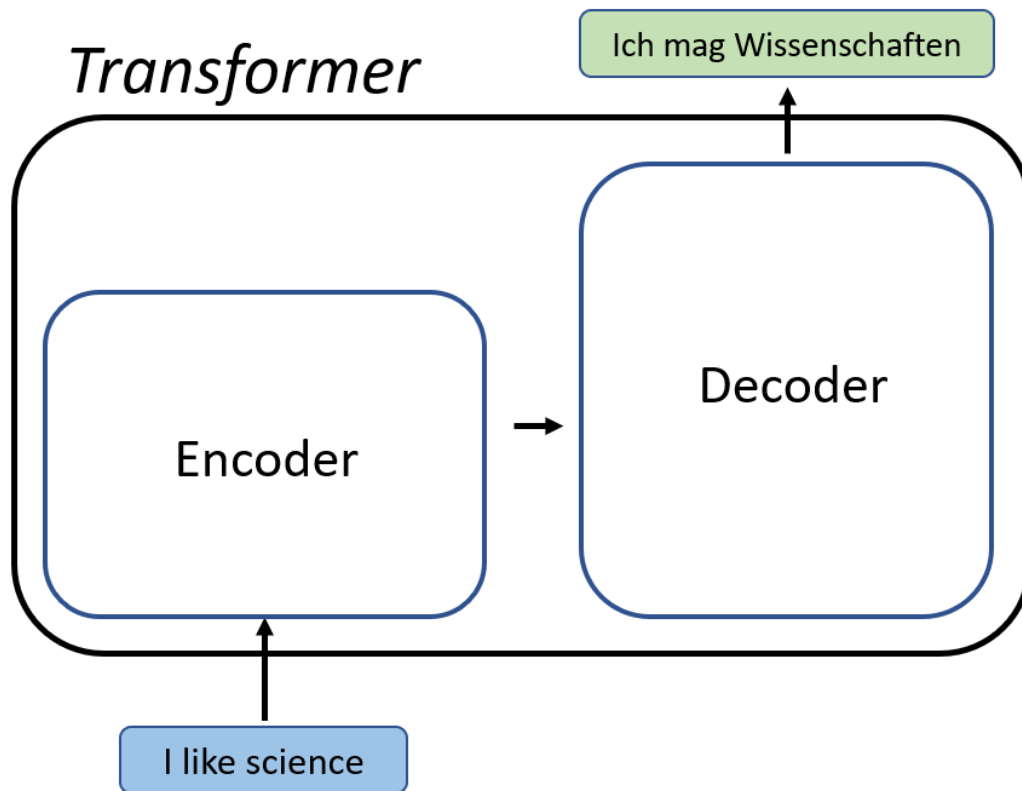


Figure 2: Transformer Architecture

Key features of the transformer architecture include the self-attention mechanism, which allows both the encoder and decoder to weigh the importance of different words in the input sequence, enhancing the model's focus on relevant parts of the sentence. The layered structure of the transformer improves its ability to understand and generate complex language structures. This architecture enables efficient parallel processing of data, significantly advancing over traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs), particularly in handling long-range dependencies and capturing nuanced meanings in text. This efficiency and capability make the transformer model highly effective for various NLP tasks, including translation, summarization, and question answering.

2.6 BERT

BERT, short for Bidirectional Encoder Representations from Transformers, is a major advancement in the field of natural language processing (NLP) and has established new standards for several NLP jobs. Devlin et al. (2018) [19] introduced BERT, a language representation pre-training method that has become fundamental in the discipline. BERT differs from prior models in that it analyzes text bi-directionally, taking into account the context from both the left and right sides of a word at the same time. The bidirectional technique employed by BERT enables it to effectively record intricate

and nuanced representations of words, hence enhancing its ability to comprehend the complexities of real language.

The BERT architecture is derived from the transformer model proposed by Vaswani et al. [18], which utilizes self-attention processes to determine the importance of individual words in a sequence. The transformer architecture of BERT has several layers of encoders, allowing it to construct intricate and contextually informed embeddings. BERT's pre-training consists of two unsupervised tasks: masked language modeling (MLM) and next sentence prediction (NSP). In the field of MLM (Masked Language Modeling), words inside a phrase are concealed, and the model is trained to anticipate these concealed words by considering their surrounding context. Conversely, NSP educates the model to grasp the connection between two phrases, hence improving its capacity to interpret context at an advanced level.

After undergoing pre-training, BERT may be further optimized for specific tasks, such as question answering, sentiment analysis, and named entity identification, by including a straightforward output layer into its pre-trained structure. The technique of fine-tuning allows BERT to adjust itself to various NLP applications using only a tiny quantity of data appropriate to the job, which makes it extremely adaptable.

The influence of BERT on natural language processing (NLP) has been considerable, resulting in substantial enhancements in performance across several benchmarks. The model's capacity for bidirectional context comprehension and its capability to extrapolate from pre-training have established it as a preferred choice for both researchers and practitioners. Later iterations, such as RoBERTa [20] and DistilBERT [21], have expanded upon BERT's structure and training methods, therefore extending the limits of what can be achieved in natural language processing (NLP).

In summary, BERT has not only improved the current level of expertise in natural language processing (NLP), but it has also made sophisticated language models accessible to a larger audience, allowing for a diverse range of applications in both academic and industrial settings.

The diagram depicts (see Figure 3) a solitary encoder layer in the transformer model, which is a fundamental element of designs such as BERT. The method starts by merging input tokens with positional encodings to preserve their sequential arrangement. The self-attention mechanism calculates the relative significance of each token, capturing distant relationships that are essential for comprehending context in activities related to natural language processing. Afterward, the outputs are normalized in order to stabilize them. Subsequently, a feed-forward neural network proceeds to apply additional

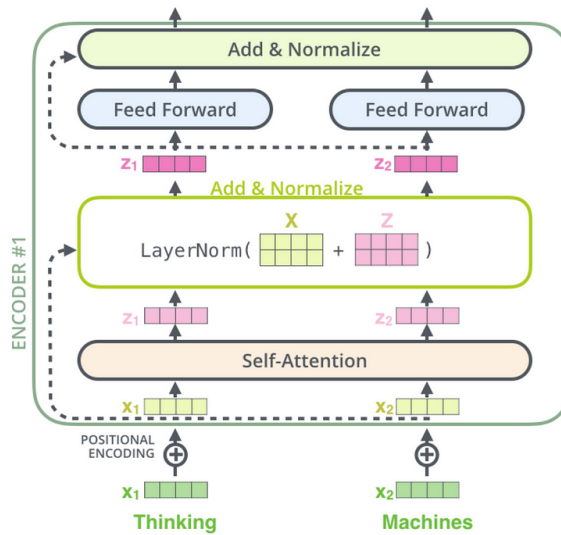


Figure 3: BERT

changes to the data. Another phase of normalization is performed to guarantee consistent results for the next layers. Residual connections inside the layer facilitate the training of deeper networks by enabling the input to skip sub-layers and be directly incorporated into their output. Transformers are capable of effectively handling intricate data dependencies, which makes them suitable for a wide range of natural language processing (NLP) applications [18].

2.7 ResNet Model

ResNet, also known as Residual Network, is a groundbreaking deep learning framework that made considerable progress in the field of computer vision. ResNet, introduced by He et al. in 2015 [22], aimed to solve the degradation problem in deep neural networks. This problem arises when raising the network's depth results in more training mistakes caused by difficulties like disappearing gradients. The design included a novel notion called residual learning, which enables the network to learn residual functions by referencing the layer inputs instead of directly learning unreferenced functions [22].

The fundamental concept underlying ResNet is the incorporation of "identity shortcut connections" that bypass one or more levels. These shortcuts or skip connections alleviate the issue of the vanishing gradient problem by enabling gradients to propagate straight across the network, without undergoing repeated multiplication by weights. The architecture may be expressed mathematically as:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\}) + \mathbf{x}$$

where \mathbf{y} is the output, $\mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\})$ represents the residual mapping to be learned, \mathbf{x} is the input, and

W_i denotes the weights of the layers.

The incorporation of these residual blocks enabled the creation of highly extensive networks, such as ResNet-34, ResNet-50, ResNet-101, and ResNet-152, where the numerical values indicate the number of layers inside the network. These architectures exhibited exceptional performance on many benchmarks, including ImageNet, where ResNet obtained top-5 error rates that were lower than those of prior cutting-edge models [22].

A standard residual block in ResNet consists of a sequence of convolutional layers, batch normalization, and ReLU activation functions. This is then followed by adding the input to the output of the stacked layers. The representation of this block can be expressed as:

Input → Conv Layer → Batch Norm → ReLU → Conv Layer → Batch Norm
→ Addition (Input) → ReLU → Output

The ResNet model's capacity to effectively train deep neural networks with enhanced precision has established it as a fundamental model in the field of computer vision. The applications of this technology go beyond just picture classification and also include object identification, segmentation, and other tasks related to visual recognition. The architecture's resilience and flexibility have also sparked several modifications and enhancements, solidifying its position as a fundamental element in contemporary deep learning research [22].

ResNet's robust feature extraction skills are utilized in the realm of fake news detection to analyze visual data, identifying subtle patterns and nuances that can differentiate between genuine and modified information. ResNet is essential in multimodal models that integrate textual and visual data to improve the precision and dependability of false news detection systems.

2.7.1 ResNet-34

ResNet, also known as Residual Network, was proposed by He et al. in 2015 and brought about a significant breakthrough in deep learning by effectively tackling the issue of degradation that arises in extremely deep networks. ResNet-34 is a complex convolutional neural network that comprises 34 layers and is composed of a sequence of residual blocks. Each block consists of two or three convolutional layers with skip connections, enabling the network to acquire residual functions. These skip connections serve to alleviate the issue of vanishing gradients, allowing for the training of extremely deep networks without any decline in performance. The design comprises an initial convolutional layer with 64 filters, succeeded by many residual blocks with progressively larger filter sizes. The architecture concludes with a global average pooling layer and a fully connected layer for classifica-

tion. ResNet-34 achieves a favorable compromise between depth and computing demands, making it appropriate for situations when resources are restricted but a complex model is still preferred. Reference the publication by He et al. (2016) (He2016DeepRL) [22].

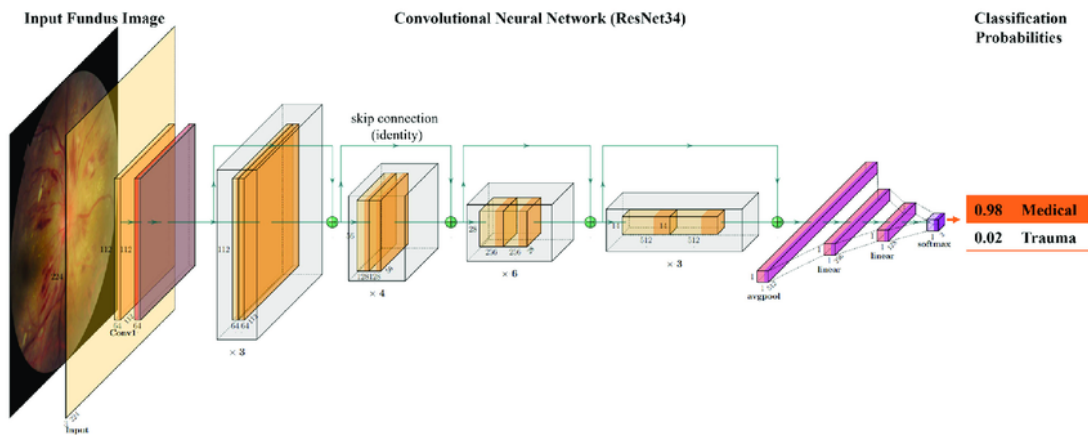


Figure 4: ResNet-34

2.7.2 ResNet-50

In contrast, ResNet-50 is a more complex and robust iteration consisting of 50 layers, which includes bottleneck residual blocks. The bottleneck block in ResNet-34 consists of three convolutional layers, as opposed to the two layers present in the basic residual block. This architecture exhibits sustained computing efficiency even with the added depth. The design of ResNet-50 consists of an initial convolutional layer, followed by many bottleneck blocks that have progressively larger filter sizes. This architecture is similar to ResNet-34 but with a greater number of layers. Bottleneck blocks are employed to decrease the amount of parameters and computational cost, while still allowing the model to acquire highly complex representations. ResNet-50 has attained the most advanced outcomes in many picture classification benchmarks and is extensively utilized in both academic research and industry applications [22].

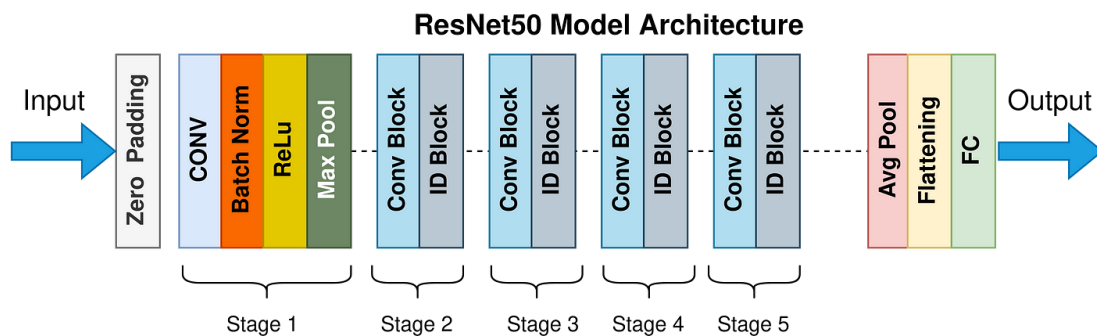


Figure 5: ResNet-50

2.8 Hyperparameters and Hyperparameter Tuning

Hyperparameters play a vital role in the design and effectiveness of machine learning models, especially those that utilize deep learning architectures. These parameters are established prior to the commencement of the learning process and dictate both the training procedure and the architecture of the models. Hyperparameters encompass several factors such as learning rate, batch size, the number of layers, and the number of neurons per layer. Accurately configuring these hyperparameters may have a substantial effect on the performance of a model, making hyperparameter tweaking a crucial step in machine learning processes.

Hyperparameter tuning is the process of searching for the most ideal combination of hyperparameters that results in the highest performance on a specific job. Methods for hyperparameter tuning vary from manual search and grid search to more advanced techniques such as random search and Bayesian optimization. Bergstra and Bengio (2012) [23] established that random search is frequently more effective than grid search, particularly when confronted with spaces of high dimensionality. In addition, Bayesian optimization, as emphasized by Snoek et al. (2012) [24], offers a robust approach to adjusting hyperparameters. It achieves this by creating probabilistic models that estimate the performance of different hyperparameter configurations and then choosing the most promising configurations for evaluation in an iterative manner.

Automated hyperparameter tuning frameworks, such as Hyperopt and Optuna, have gained popularity in the field of deep learning. These frameworks allow practitioners to effectively explore the intricate hyperparameter space and enhance the performance of their models (Akiba et al., 2019) [25]. Effective hyperparameter tuning may result in substantial enhancements in model accuracy and generalizability, which is especially crucial in jobs like false news detection, where models need to effectively handle varied and ever-changing data.

2.9 Performance metrics and Classification

Performance measures are important for checking how well machine learning models work, especially when they are used for classification tasks. These measures give us a way to compare different models and figure out what makes each one better or worse at making predictions. Accuracy, precision, recall, and the F1 score are all common ways to measure success in classification.

Accuracy is the number of accurately expected cases out of all the instances. It can be wrong, though, if the information isn't fair and the number of instances in each class is very different (Powers,

2011) [26]. Precision, which is the number of true positive predictions compared to the total number of positive predictions, is a way to measure how accurate positive class forecasts are.

Recall, which is also called sensitivity, is the number of true positive forecasts compared to the total number of real positives. It shows how well the model can find all relevant examples of the positive class. The harmonic mean of accuracy and recall, or F1-score, is a single measure that takes both into account. This makes it very useful when the distribution of classes isn't even (Sasaki, 2007) [27].

There are more complex measurements, like the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), that help us understand how the rates of true positives and false positives change when the thresholds are changed (Bradley, 1997) [28]. When you use all of these measures together for multimodal classification tasks, like finding fake news, you can get a full picture of how well the model does in all areas of the classification problem.

Subsection 4.4.3: This paragraph presents the precise formulae and the exact circumstances in which the performance measures were utilized to evaluate the performance of the models employed in this research. The metrics under discussion encompass Accuracy, Precision, Recall, F1 Score, and AUC-ROC.

3 State of the Art

The identification of fake news is a difficult issue because of deliberate deception, a wide range of subjects, and a lot of unstructured data.

The research "Complementary Attention Fusion with Optimized Deep Neural Network (CAF-ODNN) [29] for Multimodal Fake News Detection" seeks to enhance false news identification across social media platforms by more precisely combining textual, visual, and semantic information using a multi-modal method. Deep learning models used include Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, which are designed for feature extraction and categorization. The study's databases are proprietary and comprise a mix of actual and false news articles, with no particular information on the amount or kind of news pieces supplied in the extract. The study's feature extraction approaches include text embedding, picture feature extraction using CNNs, and semantic improvement algorithms to better grasp the context of news articles. The study's findings revealed good accuracy (BERT: 87%, DistilBERT: 92%, DistilRoBERTa: 88%), precision, recall, and F1-scores, demonstrating successful integration of multimodal data. However, the work recognizes limitations such as the need for additional enhancements in semantic analysis and real-time processing capabilities, recommending these areas for future research to increase the detection system's resilience and applicability.

In this paper [30], By using both unimodal and multimodal techniques, the research [30] seeks to create sophisticated techniques for automatic fake news identification. It [30] focuses on the fine-grained categorization of false news using the Fakeddit dataset. This dataset, which is made up of more than a million Reddit occurrences, is divided into several categories, including satire, fake connections, modified material, misleading content, and text and picture kinds. In addition to using BERT and CNNs among other deep learning models, the authors also presented a brand-new multimodal CNN that combines text and visual data for better classification accuracy. They used techniques for feature extraction such as embedding layers and tokenization, and using the multimodal approach, they obtained the maximum accuracy of 87%. High recall, accuracy, and F1 scores were also noteworthy outcomes, especially when altered information was used in different classes. To better manage a variety of false news kinds, the work did note several limitations, such as the difficulty in handling underrepresented classes like satire and impostor material. These issues point to the need for more balanced datasets and maybe improved model training methodologies.

The paper's [31] objective is to investigate unimodal and multimodal methods for fine-grained false

news identification using the Fakeddit dataset, which consists of Reddit posts that have been divided into several categories for truth and disinformation. Alongside a unique multimodal CNN that mixes text and visual data, many deep learning models were used, such as CNNs, BiLSTM, and BERT. Text tokenization, embedding, and image processing techniques were used in the feature extraction procedure. The multimodal strategy achieved an accuracy of 87% with substantial precision, recall, and F1 scores across multiple categories, demonstrating a significant improvement over unimodal approaches, according to the data. To increase detection across a variety of false news kinds, the study did note difficulties in tackling underrepresented classes, such as satire, suggesting the need for more balanced datasets or improved algorithms.

The goal of the research [32] is to present a thorough analysis of deep learning methods for multimodal false news detection on social media, emphasizing the incorporation of many data formats, including text, photos, audio, and video. To improve detection skills, a variety of deep learning models, including CNNs, BERT, and hybrid architectures with attention mechanisms, have been used in various research. The researchers mostly used datasets like FakeNewsNet, which offers a wide variety for testing models and contains sub-datasets like Politifact and GossipCop with tagged false and true news. Textual embeddings, picture feature extractions using pre-trained CNNs, and audio-video synchronizations to record multimodal correlations are among the feature extraction techniques covered in depth. The findings from several research demonstrated differing levels of recall, accuracy, precision, and F1 scores, which reflected improvements in the use of sophisticated deep-learning models for the detection of false news. Notwithstanding these developments, the study identifies areas for further investigation by pointing out shortcomings in domain adaptation, the necessity for improved generalization across themes that have not yet been explored, difficulties with explainability, and difficulties with effectively incorporating multimodal data.

In this study [33], the issue of unimodal bias in multimodal misinformation detection (MMD) benchmarks is addressed. By guaranteeing modality balance, eliminating asymmetric multimodal disinformation, and utilizing real-world data, it [33] presents a unique benchmark termed VERITE (VERification of Image-TExt pairings) that successfully compensates for unimodal bias. The resilience of this new benchmark was tested using a variety of deep learning models, including transformer-based architectures. The new VERITE benchmark seeks to address these problems by balancing modalities and employing real-world, difficult instances. The assessment dataset, VMU-Twitter, previously showed a propensity for unimodal bias. Advanced crossmodal alignment approaches are among the feature extraction methods used in this work to provide synthetic, realistic training data that maintains pertinent crossmodal correlations. The outcomes showed that performance had signif-

icantly improved. The results showed a noteworthy improvement in performance on the VERITE benchmark, with a 9.2% increase in predicted accuracy thanks to the new data-generating technique called CHASMA. The study observes that although the VERITE benchmark reduces the issue of unimodal bias, more investigation is required to examine various facets of multimodal misinformation detection.

By integrating textual and visual analysis, the research [34] aims to improve multimodal fake news detection and more precisely identify false information. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs), two types of deep learning models, were widely used for efficient data processing and analysis. Although the precise numbers of each category were not disclosed, the scientists developed a distinct dataset especially for COVID-19 misinformation, which contains a variety of instances of both fake and true news articles. Convolutional layers and other methods for analyzing latent features in text and pictures were used for feature extraction. Even though the sample did not specifically state the precise metrics—precision, recall, and F-score—the findings were encouraging, demonstrating great accuracy in differentiating between bogus and real news. discrepancies in the study's findings. The study points out weaknesses in the current approach to handling complex and varied kinds of fake news, and it [34] suggests that future research should concentrate on improving the detection models' resilience and flexibility to various sorts of false information.

The goal of the research [35] is to improve multimodal fake news identification by using an emotion-driven, transformer-based network to analyze both text and visuals and distinguish between true and fraudulent information. It highlights the emotional components of the detection by using sophisticated deep learning models like recurrent neural networks and vision transformers. To verify the effectiveness of the suggested model, the study was carried out using many datasets, including Twitter, the Jruvika false News Dataset, the Pontes Fake News Dataset, the Risdal Fake News Dataset, and the Fakeddit Multimodal Dataset, which included varying counts of actual and false news items. Using contextual embeddings from text and multi-granular visual features, feature extraction combined emotional ratings from the two modalities. The model outperformed previous techniques with considerable gains in accuracy, precision, recall, and F1 scores. The work does, however, highlight the need for more progress in handling the nuances of multimodal data and processing emotional content, pointing to areas that need to be improved in the future.

In this paper [36], Through the use of a hybrid neural network model that incorporates text, graphics, and social context, the article seeks to enhance the identification of bogus news. With an emphasis on stance extraction from user replies, it uses deep learning models such as convolutional neural

networks (CNNs) and recurrent neural networks (RNNs). Weibo, Fakeddit, and PHEME comprise the dataset that is used; it includes a variety of news stories that are labeled as real or fake, albeit the precise number of postings that are true and false is not specified. Text embedding and stance representation extraction utilizing a unique CNN-based knowledge extractor are two feature extraction techniques. In comparison to earlier techniques, the model showed enhanced recall, accuracy, precision, and F-Score in spotting bogus news. The study does, however, point out several drawbacks, including the difficulty of adjusting to novel, unseen data and the possibility for additional improvement in the smoother integration of multimodal data.

The purpose of the research [37] is to use a novel evaluation framework called VERITE (VERification of Image-Text pairings) to build a strong benchmark for multimodal misinformation detection that properly compensates for unimodal bias. It makes use of transformer-based deep learning models, which are designed to handle complicated multimodal data (text and pictures). Among other things, VMU-Twitter and Fakeddit are included in the dataset; VMU-Twitter displays image-side unimodal bias, while the benchmarks show different types of disinformation in text and picture pairings. Advanced crossmodal alignment, or CHASMA, is a feature extraction approach that creates realistic synthetic training data. It improved prediction performance by 9.2% in accuracy. The findings show that VERITE effectively reduces unimodal bias, which improves the evaluation of multimodal misinformation detection methods. Notwithstanding these developments, the paper notes that it is still difficult to fine-tune these models to successfully manage the complex features of various disinformation kinds, indicating a direction for further investigation.

With the use of a co-attention fusion mechanism (MRDCA), the research [38] employs a multimodal strategy to improve the identification of bogus news by integrating RoBERTa and DenseNet. This method focuses on dynamically learning the interplay between both modalities by utilizing DenseNet for image feature extraction and RoBERTa for text feature extraction. 1,063,106 samples from a broad variety of news articles in the Fakeddit dataset were classified for fine-grained classification. Feature extraction techniques make use of DenseNet's image analysis capabilities and RoBERTa's word processing, which are enhanced by a co-attention mechanism to better manage the interaction between text and picture data. Comparing the MRDCA model to various unimodal and multimodal techniques, the findings show that it obtains greater accuracy (88.14%), precision (87.16%), recall (87.94%), and F1-score (87.51%). Nevertheless, the study notes that there is a discrepancy in the performance of the various fake news categories, with satire, impersonation, and misleading content being particularly difficult to identify. These findings suggest areas where model sensitivity and classification capabilities could be further enhanced.

The method for multimodal categorization called MuRE, which uses AutoML, is presented in this work [39]. It makes machine learning easier for jobs that require both text and picture input. The system makes use of deep learning models, with a special emphasis on representation evolution—a method for improving data representation. This entails using quick, well-regularized linear models in conjunction with automatically adapting heterogeneous representations across modalities. In order to examine the accuracy of image-text pairings, the researchers used a variety of datasets to evaluate MuRE, including Fauxtography and Fakeddit, which have a balanced mix of genuine and fraudulent information. In order to ensure that feature extraction algorithms in MuRE are properly aligned and optimized for classification tasks, representations based on textual and visual data are constructed and continually adjusted. The outcomes shown that MuRE typically outperforms more conventional methods in terms of performance, suggesting a viable path for effectively managing multimodal data without requiring a lot of computational power. The system's present drawbacks, however, include its initial need for human tuning to get the ideal configuration and its subpar performance in a few particular datasets, pointing out areas that need more improvement and growth in subsequent research.

In order to improve fake news detection, the paper [40] presents a novel Stance Extraction and Reasoning Network (SERN) that models multimodal news content efficiently and eliminates the need for manual stance labeling. SERN does this by automatically extracting and integrating stance representations from post-reply pairs. This system uses a stance reasoning network that processes stance information using graph-based techniques in conjunction with deep learning models, including BERT for text encoding and ResNet for picture analysis. To train and evaluate their model, the authors employed the Fakeddit and PHEME datasets, which contain a variety of tagged examples of real and false news. BERT and ResNet are combined to extract features from textual and visual content, respectively. A new sentence-guided visual attention mechanism is included to improve the merging of these two kinds of information. The results demonstrated the efficacy of SERN in the false news detection space, showing increases in accuracy, precision, recall, and F1-score over previous techniques. The paper highlights the ongoing difficulties in handling complex fake news scenarios more thoroughly and, despite its successes, suggests that more work is needed to improve the integration of multimodal data and to expand the application of the model to other types of multimodal content beyond text and images.

In order to overcome the difficulties in multimodal false news identification, the research [41] "CAF-ODNN: Complementary attention fusion with optimized deep neural network for multimodal fake news detection" integrates uncorrelated semantic representations, which might inject noise into the

features. The goal of the suggested method, Complementary Attention Fusion with an Optimized Deep Neural Network (CAF-ODNN), is to enhance feature extraction and model accuracy while capturing subtle cross-modal interactions. Utilizing three fully connected layers, a customized deep neural network that takes use of compositional learning is one of the deep learning models that the researchers employed. They used four real-world datasets for their assessment, with differing proportions of fictitious and authentic posts; the specific figures are not specified in the extract. By using bidirectional complementary attention based on a scaled dot product to learn fine-grained correlations and image captioning to semantically describe pictures, feature extraction was improved. Their findings demonstrated notable gains over other methods in terms of common measures including precision, accuracy, recall, F1 score, and correctness. Nevertheless, the study indicates a potential restriction or area for more research in their work, suggesting that there may still be an opportunity for progress in handling uncorrelated semantic noise more efficiently.

The goal of the article [42] is to improve rumor identification accuracy by using a unique Multimodal Dual Emotion feature that makes use of both textual and visual emotions as well as social emotion signals from comments. It makes use of deep learning models, such as ViT for visual features and RoBERTa for text, and uses emotional signals to increase the efficacy of detection. The main dataset utilized is the Fakeddit dataset, which is well-known for having a significant number of tagged actual and false postings. This makes it an ideal testing ground for the model. Semantic segmentation and object detection are used in feature extraction to provide a thorough analysis of multimodal information. Sentiment analysis from text and emotion recognition from images are also included. The novel method significantly improved the identification of subtle emotional cues in bogus news, as seen by the findings, which showed considerable gains in accuracy, precision, recall, and F1 scores over previous models. The research does, however, recognize the need for greater improvements in the seamless integration of multimodal data and indicates that future studies may investigate other modalities, such as audio or video, for even more reliable rumor identification.

In order to improve the accuracy of fake news classification, the paper [43], "PL-NCC: A Novel Approach for Fake News Detection through Data Augmentation," makes use of the Psycho-Linguistic News Content and Comments (PL-NCC) dataset, which combines linguistic and psychological features from news articles and user comments. This dataset, which focuses on psychological and linguistic traits to increase classification accuracy, combines data from NELA-GT and Fakeddit. Among the models employed is the News Content and Comments (NCC) classification model, which improves the machine's capacity to process psychological variables by including a feed-forward layer into a deep learning framework. The sizeable dataset, which includes a sizable portion of bogus news sto-

ries, offers a thorough foundation for model testing and training. Processing linguistic clues such as writing style and psychological factors like emotions from articles and comments are part of feature extraction approaches. Outperforming other baseline models, the findings showed over 90% accuracy, demonstrating the efficacy of incorporating psychological factors into false news identification. To further increase the accuracy of early detection algorithms, the work highlights several areas for development, such as the requirement for more complex treatment of unstructured user comment data and wider inclusion of user engagement measures.

In this paper [44], By combining dual emotional characteristics (publisher and social emotions) with a deep normalized attention-based mechanism and an adaptive genetic weight update technique with a Random Forest classifier, the article aims to improve the precision and effectiveness of false news identification. Random Forest is used for classification, and BiGRU is used for feature extraction in deep learning models. Detailed tests demonstrating multiple instances of accurate and incorrect classifications across these datasets—though precise numbers for each category are not provided—are conducted using the RumorEval19, PHEME, and Fakeddit datasets. A deep normalized attention-based technique is used in feature extraction to improve the semantic extraction of emotional signals and augment dual emotion features. The efficacy of combining dual emotion characteristics is demonstrated by the findings, which show significant gains over baseline techniques in accuracy, precision, recall, and F1 scores. The paper also point out that in order to capture more complete multimodal information, the model may be further enhanced by addressing issues with early detection and integrating other modalities like audio or video.

Using a model that combines deep learning approaches to assess text, pictures, and their interactions, the article [45] aims to improve multimodal false news identification. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Transformers are some of the deep learning models used; they are designed for complex feature extraction and integration tasks. To properly train and assess the model, the study makes use of the Multi-Source Fake News Dataset, which has a balanced mix of fake and actual news postings. The emphasized feature extraction techniques include text embeddings, pre-trained CNNs for image feature extraction, and attention mechanisms that improve the relationship between textual and visual information. The findings demonstrate the effectiveness of the model in differentiating between bogus and authentic news, with enhanced accuracy, precision, recall, and F-Score. To strengthen the model's robustness and applicability, the study acknowledges the need for more development in addressing the nuances of false news spread and recommends looking into other modalities and datasets.

The research [46] seeks to solve the difficulty of identifying false news on online social networks (OSNs) by creating a multimodal framework that examines both visual data (pictures) and textual content to assess the authenticity of posts. This framework's deep learning models include Xception and a fusion of BERT with a Dense layer, which are utilized to successfully interpret and learn from picture and text data. The dataset used is the Fakeddit dataset, which comprises over 1 million samples of text, photos, metadata, and comments. Visual sentiment analysis and error level analysis are two feature extraction approaches that help to identify between false and real posts by evaluating visual and textual data. The model's overall accuracy is 91.94%, with a precision of 93.43%, recall of 93.07%, and F1-score of 93%. However, the study notes that, while the suggested model beats other cutting-edge models, there is still space for development in terms of dealing with the wide range of false news formats and improving the model's generalization skills across different types of bogus posts.

The research [47] attempts to improve the accuracy and robustness of false news identification by introducing a new strategy based on a Multimodal Fusion-Based Hybrid Neural Network that includes attitude extraction. It utilizes deep learning models such as CNNs for image analysis and RNNs for text, which are combined using a stance-aware fusion technique. The Fakeddit dataset, which contains a mix of 860,000 actual and 210,000 fake posts, serves as a solid foundation for testing and training. The network extracts features using text embeddings and picture feature extraction via pre-trained CNN architectures, with stance-related information from user comments added to improve identification. The findings demonstrate considerable gains, with the hybrid model reaching 87% accuracy, 93% precision, 94% recall, and an F1-score of 92.5%. However, the study notes the need for improved handling of nuances and variances in false news presentation, recommending future feature extraction approaches and expanding the model's ability to cover other types of social media material.

The research [48] offers a unique Multimodal Stacked Cross Attention Network (MSCA) that efficiently aligns and fuses multimodal token-level textual and visual data to detect false news more accurately. The study applies deep learning models such as BERT for textual feature extraction and Vision Transformer (ViT) for picture features. It takes advantage of publicly available datasets such as Fakeddit and Weibo, which include thousands of tagged news articles categorized as factual, satire, misleading information, and impostor material. The feature extraction techniques concentrate on extracting token-level semantic characteristics from both text and images, leveraging a clever cross-attention mechanism to improve interaction between the two modalities. The results reveal that MSCA outperforms other models in terms of accuracy, precision, recall, and F1 scores. However, the article

notes that the MSCA model might be enhanced by better addressing heterogeneity and alignment across different modalities, indicating a potential topic for future improvement in the precision of multimodal fusion methods.

The research [49] proposes a contrastive learning-based methodology for minimizing multimodal inconsistency in false news identification. The approach, called the Mitigating Multimodal Inconsistency Contrastive Learning approach (MMICF), seeks to overcome inconsistencies across different modalities, which frequently contribute to biased learning. It uses deep learning models such as BERT for text encoding and Vision Transformers for picture analysis. The research makes use of databases such as Fakeddit, which has a variety of false and real news stories, however actual figures are not offered in the clip. To address local and global discrepancies, feature extraction approaches such as causal-relation reasoning and modal unification techniques are used to align textual and visual data. The results demonstrate significant gains in accuracy, precision, recall, and F1-score over previous approaches, demonstrating the MMICF's usefulness in improving multimodal false news identification. Despite these advancements, the research admits possible limitations in entirely resolving modal discrepancies and indicates that more refining might enhance the detection system's resilience and accuracy.

The purpose of this work [50] is to address the issues of multimodal disinformation detection by examining existing techniques and finding gaps that provide opportunities for future research. It [50] uses deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to extract features from a variety of data types, including text and pictures. The researchers examine many datasets, most notably the Fakeddit dataset, which has around 1,063,106 samples of both false and true news articles, laying a solid basis for training and testing the suggested models. To successfully detect disinformation, the feature extraction approaches mentioned rely on finding and utilizing cross-modal discrepancies, such as text and picture incompatibilities. The clip does not go into depth on the results of various models, but the study highlights how these models improve the detection of multimodal misinformation. However, the study recognizes the limits of current methodologies, such as the need for improved modality integration and the creation of models that can generalize across multiple types of misinformation and evolve to accommodate new forms of fake news.

The major purpose of this study [51] is to enhance multimodal misinformation detection by including external knowledge into the detection process through the use of a model known as LEMMA (LVLM-Enhanced Multimodal Misinformation Detection with External Knowledge Augmentation). It

makes use of Large Vision Language Models (LVLMs), which are geared for dealing with both visual and textual input, exhibiting improved reasoning abilities. The datasets used are Twitter and Fakeddit, and the research conducts a full examination of both sites to assess the success of the LEMMA technique. LEMMA's feature extraction algorithms involve complex reasoning processes and the integration of external knowledge to verify the accuracy of the information. The findings demonstrate considerable gains, with accuracy up 7% on Twitter and 13% on Fakeddit compared to the baseline LVLM models. Despite its accomplishments, the study notes limitations, such as the need for better knowledge source interfaces and filtering to improve the detection process's reliability and comprehensiveness.

The research [52] proposes a unique Multimodal Adaptive Co-Attention Fusion Contrastive Learning Network (MACCN) that improves the integration of textual and visual elements to detect false news more reliably. The deep learning models used are BERT for textual data and ResNet50 for picture data, both of which are well-known for their feature extraction capabilities. The datasets included in the study are Weibo, Fakeddit, and PHEME, which each contain a considerable amount of false and genuine news items; for example, Fakeddit comprises 7,200 fake and 4,800 actual news posts. In MACCN, feature extraction algorithms use BERT and ResNet50 to generate enhanced feature representations, which are then processed by adaptive co-attention fusion to emphasize important information from both modalities. The findings show that MACCN outperforms in terms of accuracy, precision, recall, and F1-scores, with an accuracy of up to 92.4% on the Weibo dataset. Despite these outstanding findings, the study admits areas for development, notably in strengthening the contrastive learning features to boost model robustness and efficacy even more.

The research [53] focuses on improving the authenticity evaluation of online information using a deep learning-based approach that combines textual and visual data to identify multimodal false material. It uses a light version of the Bidirectional Encoder Representation Transformer (BERT) with four encoder layers and eight attention heads for textual input and a pre-trained EfficientB1 architecture for visual data. The characteristics of both modalities are concatenated and then analyzed by a Multilayer Perceptron network. This technique was evaluated on a benchmark dataset that included scenarios for two-class, three-class, and six-class issues, and the results were encouraging, with the two-class arrangement achieving the maximum accuracy of 90.33%. Despite these accomplishments, the research recognizes possible areas for development, particularly in further refining multimodal data integration strategies to raise detection accuracy and lower the likelihood of misclassification.

3.1 Identified Gaps and Proposed Solutions in Literature

Upon conducting an exhaustive analysis of the existing body of literature on the identification of false news, several deficiencies have been uncovered, which in turn provide prospects for more study and enhancement. An important deficiency is in the insufficient incorporation of multimodal data, namely the simultaneous utilization of text and pictures, in algorithms designed to detect false news. Previous research has mostly concentrated on analyzing either text or images independently, neglecting to fully use the synergistic relationship between these two forms of data. For example, models that exclusively rely on textual data, as described by Zhou et al. (2018) [15], fail to consider the visual context that frequently accompanies social media posts. On the other hand, models that just focus on images, such as the ones studied by Jin et al. (2017), fail to consider the important story conveyed by the surrounding text.

There is also a significant discrepancy in the effectiveness of multimodal models when applied to various datasets and real-life situations. Although certain models, such as those utilizing fundamental multimodal fusion approaches (Khattar et al., 2019) [17], demonstrate potential under controlled experimental conditions, they frequently struggle to uphold a high level of accuracy and resilience in varied and ever-changing social media contexts. This disparity implies that the existing multimodal methods may not effectively capture the intricacies and nuances of real-world data.

Moreover, the dependence on conventional machine learning methods and less advanced neural networks in several research papers has constrained the efficiency and scalability of false news detection systems. Transformers and convolutional neural networks (CNNs), which are advanced deep learning architectures, have demonstrated substantial advancements in processing intricate data patterns. However, their potential in the domain of multimodal false news detection has not been thoroughly investigated.

In order to fill these deficiencies, the following strategies can be put into practice. It is essential to construct sophisticated multimodal models that can successfully combine textual and visual information. One way to accomplish this is by utilizing advanced deep learning models like BERT for text processing and ResNet for picture processing. The design of these models should aim to accurately capture the complex interaction between text and visuals, in order to offer a more thorough analysis of social media material.

Furthermore, it is crucial to improve the resilience and applicability of multimodal models. This re-

quires comprehensive training on varied datasets and integrating methods like as data augmentation and transfer learning to enhance the performance of the model in many scenarios. Furthermore, the utilization of advanced hyperparameter tuning techniques, such as Bayesian optimization, can assist in refining the models to get optimal performance.

Utilizing the most recent developments in deep learning architectures can greatly enhance the efficiency of false news detection systems. Combining transformer-based models like BERT and its variations (such as DistilBERT and RoBERTa) with strong convolutional neural networks like ResNet50 can create a reliable framework for analyzing several modes of data. These algorithms have the ability to capture intricate patterns and correlations in the data, resulting in more precise and dependable identification of fake news.

To summarize, by incorporating sophisticated multimodal models, strengthening resilience and adaptability, and utilizing state-of-the-art deep learning techniques, the efficacy of false news detection systems may be greatly enhanced. This method not only addresses the deficiencies in the existing body of knowledge but also lays the foundation for the creation of more advanced and dependable technologies to counteract the dissemination of false information on social media.

3.2 State-of-the-Art Approaches in Fake News Detection

The scientific community has recently shown substantial interest in detecting and categorizing bogus news. Several machine learning and deep learning models have been suggested to tackle this problem, utilizing a range of datasets and various data modalities. The following table provides a summary of current research that has utilized multimodal methodologies, which involve the integration of both textual and visual data, in order to improve the precision of false news identification.

The research presented showcases a range of model architectures, such as BERT, CNN, and RNN, as well as hybrid models like BERT integrated with ResNet and VGG16. The evaluation of these models is conducted on several datasets, with a primary focus on the Fakeddit dataset. This dataset provides a reliable baseline for assessing the performance of multimodal fake news detection.

The performance measures, such as accuracy and F1-score, demonstrate that models such as BERT+ResNet and complementing attention fusion approaches attain a high level of accuracy, highlighting the effectiveness of merging textual and visual information. Nevertheless, there are notable disparities in performance, which underscore the difficulties and restrictions inherent in multimodal false news identification.

For example, research using models such as DistilBERT and LSTM+VGG16 indicated a reasonable level of accuracy, highlighting the necessity for more optimization and fine-tuning of the models. However, more intricate structures, such as those that include complimentary attention fusion, have demonstrated potential with improved accuracy rates. This indicates that advanced fusion approaches can effectively capture the subtleties of false news material.

The table also showcases the prevalent use of binary classification methodology in the majority of investigations, with certain studies using survey techniques to assess the effectiveness of the models. This thorough review highlights the continuous endeavors and advancements in the industry, while also pinpointing the deficiencies and areas that need enhancement in false news detection research.

Citation Link	Dataset	Get Data	Model	Classification	Evaluation
Link	Multimodal	Fakeddit	DistilBERT+VGG16	Binary	62%
Link	Multimodal	Fakeddit	BERT+CNN	Binary	87%
Link	Multimodal	Fakeddit	BERT+CNN	Binary	87%
Link	Multimodal	Fakeddit	CNN, BERT	Binary	Survey
Link	Multimodal	Fakeddit, VMU-Twitter	BERT+VGG19	Binary	83%
Link	Multimodal	Twitter & Weibo	BERT	Binary	75% & 87%
Link	Multimodal	Multiple Datasets	BERT	Not Binary	96%
Link	Multimodal	Fakeddit & PHEME	CNN+RNN	Binary	84%
Link	Multimodal	Fakeddit, VMU-Twitter			Survey
Link	Multimodal	Fakeddit	BERT+VGG19	Binary	83%
Link	Multimodal	Fakeddit	BERT+ReNeR50	Binary	82%
Link	Multimodal	FACTIFY 2	LSTM+VGG16	Binary	65%
Link		Fakeddit, PHEME	BERT	Binary	90%
Link		Fakeddit & PHEME	CAF-ODNN	Binary	89% & 90%
Link	Multimodal	Fakeddit	RoBERTa	Binary	69%
Link	Multimodal	Fakeddit	BERT	Binary	86%
Link		Weibo	ResNet101	Binary	81%
Link	Multimodal	Fakeddit & PHEME	CNN	Binary	92%
Link	Multimodal	Diff Datasets	CNN+RNN	Binary	Survey
Link	Multimodal	Weibo	CNN & ResNet	Binary	81%
Link	Multimodal	Fakeddit	BERT	Binary	90%

Link	Text	LIAR	DistilBERT	Binary	63.61%
Link	Multimodal	Fakeddit	BERTa	Binary	87%
Link	Multimodal	Fakeddit & Weibo	BERT	Binary	82% & 87%
Link	Multimodal	Fakeddit & Weibo	BERT	Binary	68% & 61%
Link		Fakeddit	BERT		89%
Link	Multimodal	Fakeddit	LEMMA	Binary	82%
Link	Multimodal	Fakeddit	MACCN & BERT	Binary	85% & 70%
Link	Multimodal	Fakeddit	BERT		90%

Table 2: Table Literature Review

4 Research Methodology

4.1 Proposed Solution

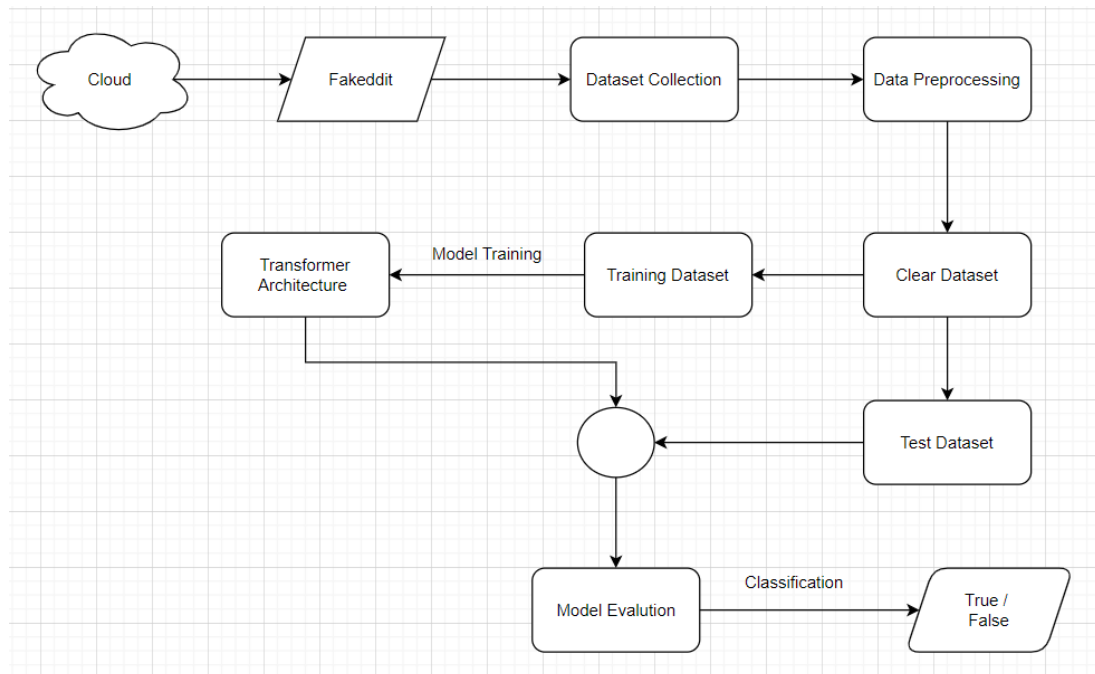


Figure 6: Proposed Solution

The diagram 6 depicts the sequential process of the study methodology for identifying false information by utilizing a multimodal approach that incorporates both textual and visual data. The procedure may be dissected into many critical stages:

1. **Data Source (Cloud):** The method starts by obtaining data from the cloud. The dataset utilized in this instance is Fakeddit, which comprises of multimodal social media material that integrates both textual and visual elements.
2. **Dataset Collection:** The subsequent phase is gathering the dataset from the designated source. This entails the act of retrieving and arranging the data into a format that may be easily utilized for subsequent analysis.
3. **Data Preprocessing:** This stage is essential as it readies the unprocessed data for analysis. Pre-processing encompasses many stages:
 - **Clear Dataset:** This process entails cleansing the dataset by eliminating any extraneous data, useless information, or discrepancies. For textual data, this process involves eliminating stop words, punctuation marks, and doing tokenization. When dealing with picture

data, this process may include scaling, normalization, and other procedures to enhance the images.

- **Training Dataset and Test Dataset:** Subsequently, the unambiguous dataset is divided into separate training and test datasets. The training dataset is utilized to train the models, whilst the test dataset is exclusively intended for assessing the model's performance.
4. **Model Training:** During this stage, the training dataset is utilized to train the transformer architecture. Transformer designs, such as BERT for text and ResNet for pictures, are utilized to acquire knowledge about the patterns and characteristics present in the data.
 5. **Transformer Architecture:** This section encompasses the fundamental machine learning models employed in the study. Text data is processed using models such as BERT, DistilBERT, and RoBERTa. ResNet34 and ResNet50 models are used for processing picture data. These models undergo a meticulous training procedure utilizing the training dataset.
 6. **Model Evaluation:** Following the completion of training, the models undergo evaluation using the test dataset. This entails evaluating their performance by utilizing diverse criteria such as accuracy, precision, recall, and F1-score. The evaluation process aids in comprehending the extent to which the models can apply their knowledge to unfamiliar data.
 7. **Classification:** The ultimate stage in the pipeline involves the categorization of data. The system utilizes trained models and assessment to categorize the material as either true or false. This categorization aids in discerning counterfeit news.

Every part of this system works with the others to make sure that the whole process of finding fake news using multimodal data is complete. The method uses both text and picture analysis to try to understand how complicated and diverse fake news is, offering a strong answer to this important problem.

4.1.1 Dataset

GitHub is the primary source for the carefully created dataset called "Fakeddit." We focused our data-gathering efforts on this platform which is abundant in resources, ensuring a comprehensive and extensive collection of fictional data examples for our study. In order to obtain a thorough understanding of deceptive content on social media, the process of gathering data entails carefully choosing and organizing information from many sources. Moreover, the utilization of GitHub as the primary source aligns with our commitment to transparency and the availability of data, fostering

the replication and further exploration of research endeavors in the realm of false data classification.

Data Description:

The Fakeddit dataset, which includes more than a million samples taken from Reddit, is a cutting-edge multimodal resource designed for the identification of fake news. With the use of this large dataset, which includes a wide range of text, photos, comments, and metadata, hybrid text+image models may be developed for more precise identification. Interestingly, Fakeddit uses a multi-grained labeling system, providing samples classified by a strict remote supervision procedure into 2-way, 3-way, and 6-way classification kinds. By using this method, false news information may be thoroughly and nuancedly analyzed, improving the efficacy of detection algorithms. Fakeddit's multimodal composition, diversity of content, scalability, and multi-grained labeling all work together to make it an invaluable tool for deep learning-based false news detection research.

4.1.2 Data Preprocessing

Data preparation is an important step in getting the dataset ready so that the fake news identification model can be trained and tested effectively. This part talks about the methods used to clean up written and picture data before they are used to train models. These methods make sure that the inputs are clean, consistent, and good for training. The methodical strategy used to carefully segment the Fakeddit dataset, which is a collection of rich multimodal material mostly from the domain "i.redd.it." Due to its frequency and representatives in the dataset, this domain was specially selected as the goal, offering a stable foundation for assessing how well different machine learning models perform in the task of detecting false news.

4.1.2.1 Data Cleaning

1. **Text Data Cleaning:** For written data, the cleaning method has several steps to make sure that it is consistent and useful.
 - **Text Normalization:** To keep things consistent and get rid of case sensitivity, all the text is changed to lowercase.
 - **Removal of Unwanted Characters:** To focus on the main text, special letters, punctuation, and extra spaces are taken out.
 - **Tokenization:** The text is tokenized using the BERT tokenizer, resulting in the encoding of the text into input IDs and attention masks. In this stage, the text sequences are either padded or truncated to a preset maximum length of 80 tokens. This is done to maintain uniform input sizes throughout the dataset.

Example: Given the text data: - "Fake news detection is a critical task." - "The model performs exceptionally well on the test data." - "Social media platforms are rife with misinformation."

When the BERT tokenizer goes through each sentence, it adds special tokens like [CLS] at the beginning and [SEP] at the end, truncates or pads the sequences to a maximum length of 80 tokens, and makes attention masks to show the padding tokens. The data is now ready to be fed into the BERT model because it has a list of input IDs and attention masks for each line.

2. **Image Data Cleaning:** For picture data, cleaning means making sure that each text record has an image and that the image is present and of good quality.

- **Loading Images:** The file name of each picture is used to load it. If any pictures are missing or broken, a default random tensor is made to keep the batch consistent.
- **Error Handling:** Strong error handling makes sure that any problems loading pictures, like missing files or images that can't be read, are handled smoothly by using a random tensor as a replacement.

4.1.2.2 Data Transformation

Standardizing and adding to the data through data transformation is necessary to make the model better at learning.

Text Data Transformation: Following the process of tokenization, the encoded text sequences are adjusted to a consistent length of 80 tokens by either adding padding or truncating. This guarantees consistent input dimensions for the model, enabling fast batch processing and training.

Image Data Transformation: Image data is enhanced and normalized in a number of ways to make the model more reliable.

- **Training Transformations:** As part of the training process, photos undergo random resizing, cropping, and horizontal flipping to create diverse representations. By exposing the model to various characteristics of the photos, it enhances its ability to generalize.
- **Normalization:** The pixel values of the image are adjusted to have a mean and standard deviation that closely resemble those utilized during the pre-training of ResNet50. This guarantees that the input distribution aligns with the anticipated patterns of the pre-trained model, hence enhancing the process of extracting relevant features.

For the purpose of validation and testing, the photographs undergo a resizing process where they are adjusted to dimensions of 256x256 pixels. Following this, a center-cropping technique is applied to

further modify the images to dimensions of 224x224 pixels. No extra augmentations are performed during this process. This ensures a uniform assessment method, guaranteeing that the model's performance is measured using standardized data.

4.1.2.3 Data Normalization Techniques

- **Random Resizing and Cropping:** The images undergo random resizing and cropping to achieve a uniform size of 224x224 pixels. By incorporating this augmentation technique, the model is able to enhance its ability to make accurate predictions by exposing it to diverse representations of the pictures throughout the training process.
- **Random Horizontal Flipping:** With a 50% chance, this addition flips pictures horizontally, which makes the training sample even more varied.
- **Normalization:** The values of the pixels in an image are adjusted so that their mean and standard deviation are the same as they were when ResNet50 was first trained. This makes sure that the input distribution matches what the model was taught to expect.

4.1.2.4 DataLoader Initialization

DataLoader functions are utilized to generate data batches for the purposes of training, validation, and testing. These functions guarantee the effective loading and preparation of data:

- **Training DataLoader:** The `create_data_loader` function sets up the DataLoader for the training dataset by using the changes and batch processing methods that were given.
- **Validation and Testing DataLoader:** The `val_test_create_data_loader` method sets up the DataLoader for testing and validation datasets. This makes sure that the data is preprocessed consistently without having to be included to.

The data preprocessing workflow makes sure that all the text and picture data is clean, consistent, and ready for training the model. This thorough preparation method is very important for the stability and precision of the fake news detection system, which lets it find false information on social media sites.

4.1.3 Textual Data Preprocessing

Textual data preparation is an essential stage in my study, as it guarantees that the input text is in an acceptable format for analysis by the BERT model. The procedure has many steps, which include text cleansing, tokenization, and preparation for model input. Every stage in this process is crucial for

improving the quality and consistency of the text data, hence raising the efficacy of the false news detection system.

4.1.3.1 Text Cleaning

The first step in preparing textual data is to clean the text to get rid of any noise and make sure it is all the same. As part of this process, all characters are changed to lowercase to keep things consistent. This is necessary because the BERT model is case-sensitive, and changes in case can cause token representations to be different. The text is also cleaned up by getting rid of special letters, punctuation marks, and extra spacing. Getting rid of these parts makes the input data less complicated because they don't add any useful information to the sorting job. By standardizing the text in this way, we make sure that the model only looks at the important parts and not the small ones.

4.1.3.2 Tokenization

Tokenization comes next after the text has been cleaned up. For this, the BERT tokenizer is used. This tool changes words into a style that the BERT model can understand. The BERT tokenizer takes the text and breaks it up into tokens. Each token is then given a unique identity. For example, [CLS] is added at the beginning and [SEP] is added at the end of the text. These tokens help the BERT model figure out what the input sequence is about and where its limits are. As part of the tokenization process, the text is encoded into input IDs, which are numbers that identify the tokens, and attention masks, which show which tokens are real and which ones are just fillers. To make sure that all of the input patterns are the same length, which is important for batch processing in the model, padding is needed. In my code, 80 characters is the longest series that can be made. Sequences that are longer than this are cut off, and sequences that are shorter are stretched to make them all the same length. This step makes sure that the model gets data in a regular and organized way, which is very important for correct analysis.

4.1.3.3 Preparation for Model Input

Once the text data has been tokenized, it is ready to be fed into the BERT model. The tokenized text, which is represented as input IDs and attention masks, is transformed into PyTorch tensors. The conversion is essential as the BERT model, which is implemented in PyTorch, necessitates tensor inputs. The input IDs and attention masks are grouped into batches, where each batch consists of a predetermined number of samples. Batch processing is a highly efficient method for training and assessment, since it enables the model to handle several data concurrently. Within my code, the

data loader manages the process of grouping data into batches, guaranteeing that each batch is appropriately structured and prepared for entry into the model. The data loader also randomizes the training data to avoid the model from acquiring knowledge of the sequence of the samples, which may result in bias.

4.1.3.4 Handling Special Cases

To keep the integrity of the information, special cases like lost or badly formatted text data are dealt with during the preprocessing. For instance, posts that are missing text are given a temporary value so that the multimodal model can still use the picture data that goes with them. This method makes sure that no samples are thrown away because they are missing textual information. If this happened, the training set would be smaller, which could affect how well the model works.

4.1.3.5 Integration with Image Data

In the end, the preprocessed text data is mixed with picture data to make a multimedia input for the model that finds fake news. The visual traits that the ResNet50 model pulls out are merged with the textual embeddings that the BERT model creates. Both types of data must be preprocessed regularly and correctly for this integration to work. This makes sure that the end model can learn from both text and images.

To summarize, my research's textual data preprocessing includes preparing text data for the BERT model by cleaning, tokenizing, and preparing it to an ideal standard. This preprocessing stage establishes a solid groundwork for the analysis and classification activities that follow by dealing with noise, making sure everything is consistent, and managing exceptional instances. Using the strengths of both textual and visual information, the false news detection system is made even more powerful by integrating preprocessed text with picture data. If you want to construct a model to detect false news that works, you need to preprocess your data thoroughly.

4.1.4 Image Data Preprocessing

Visual data preparation is an essential step in preparing picture data for analysis by the ResNet50 model in my false news detection system. This method guarantees that the photos are in an appropriate format for the model to extract significant characteristics. The process of visual data preparation includes picture cleansing, data augmentation, normalization, and addressing exceptional scenarios. Each of these phases is specifically designed to optimize the quality and uniformity of the picture data, hence enhancing the effectiveness of the false news detecting system.

4.1.4.1 Image Cleaning

The first stage in preparing the visual data is purifying the pictures to guarantee their integrity and usefulness. This method involves validating the existence and integrity of each picture. Datasets obtained from social media sometimes contain missing or damaged photos, which can cause disruptions during the training process. In order to address these situations, pictures that are not available are replaced by tensors that have been initialized randomly. This methodology enables the model to undergo uninterrupted training, although at the cost of introducing some level of noise. However, it guarantees the preservation of the dataset's completeness. In addition, pictures that are corrupted and cannot be accurately processed are substituted by random tensors. This phase guarantees the preservation of the batch processing pipeline and ensures that the model is consistently trained with a fixed amount of photos.

4.1.4.2 Data Augmentation

Data enrichment is a key method for making the training sample more diverse without actually doing so. In this step, the pictures are changed in different ways, such as by randomly resizing, cropping, and moving them horizontally. In my version, pictures are resized and cropped at random to 224x224 pixels, and they are also flipped horizontally with a certain chance. Through these changes, different versions of the same picture are made, which lets the model learn from a bigger range of visual details. Data enrichment keeps the model from becoming too dependent on certain patterns in the training pictures, which helps keep it from overfitting. By showing the model different altered pictures, it learns to generalize better, which is very important for finding fake news correctly in real life.

4.1.4.3 Normalization

Normalization is a crucial part of the preparation of visual data. The image's pixel values are standardized to have a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225]. The values are derived from the statistics of the ImageNet dataset, which served as the basis for pre-training the ResNet50 model. By normalizing the photos, their distribution is adjusted to match the data used to train the model, resulting in enhanced performance of the model. This process standardizes the pixel values to a uniform scale, facilitating the model's ability to learn from the visual data. By standardizing the photos, we guarantee that the input data is uniform, hence minimizing the likelihood of the model being influenced by fluctuations in lighting and color.

4.1.4.4 Handling Special Cases

Special circumstances, such as photos that are missing or have formatting errors, are addressed during the preparation step to ensure the dataset remains intact. In the event that an image cannot be located or accessed, it is substituted with a tensor that has been randomly initialized. This strategy guarantees that the appropriate textual data remains usable in the multimodal model. By addressing these exceptional scenarios, we mitigate the risk of losing crucial data and guarantee that the model can be trained using a maximum number of samples. Additionally, this approach aids in preserving the batch size, which is crucial for ensuring steady and efficient training.

4.1.4.5 Integration with Text Data

The preprocessed photos are combined with the preprocessed text data to provide a multimodal input for the false news detection algorithm. The ResNet50 model's visual characteristics are merged with the BERT model's textual embeddings. In order to successfully integrate both text and picture data, it is essential to preprocess the data in a consistent and reliable manner. This ensures that the final model can effectively learn from both forms of input. By combining text and visual data, a full representation of social media postings is produced, which allows the model to make more accurate classifications.

4.1.4.6 Batch Processing

To make training and assessment more efficient, the visual data is handled in groups. The data loader does the batching and makes sure that each batch has the same number of samples with pictures that are correctly organized. This method lets the model look at more than one picture at the same time, which speeds up and improves the training process. We make sure that the model gets a steady stream of input data by putting the data into batches. This is important for stable training.

To summarize, the visual data preparation in my study entails thorough picture cleaning, data augmentation, normalization, and addressing exceptional circumstances to guarantee that the images are in the most suitable format for the ResNet50 model. These methods improve the quality and consistency of the picture data, establishing a solid basis for the ensuing analysis and classification activities. The false news detection system combines the preprocessed visual data with the textual data, making use of the strengths of both types of information. This leads to a strong and precise model. An all-encompassing strategy for preprocessing visual data is crucial in constructing a proficient and dependable system for detecting false news, capable of functioning effectively in real-life situations.

4.2 Model Architectures

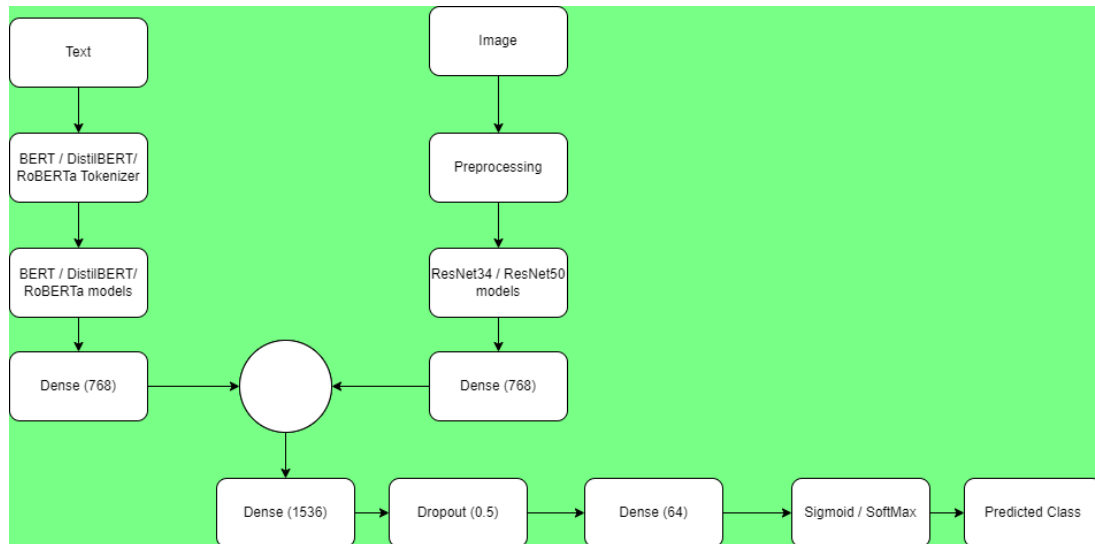


Figure 7: Model Architecture

The model design 7 given for finding fake news uses both written and visual data streams to make predictions more accurate. The architecture 7 starts with two separate inputs: words and a picture. Tokenizers from BERT, DistilBERT, or RoBERTa are used to process the written data. Then, the appropriate models from BERT, DistilBERT, or RoBERTa are used to make thick representations of size 768. At the same time, picture data goes through preprocessing and is then handled using either the ResNet34 or ResNet50 models, which also create dense representations of size 768.

Then, these dense versions from the text and picture modes are joined together to make a 1536-by-1536 dense layer. Through fully joined layers, this concatenated image goes through a number of changes. The first change goes through a 1536-by-1536 thick layer. This is followed by a 0.5-dropout-rate dropout layer to stop overfitting. After that, the data goes through another deep layer that is 64 bytes thick. Lastly, an activation function like Sigmoid or SoftMax is used to make the projected class.

This multimodal design makes good use of the best parts of both textual and visual data. Its goal is to find the complex connections between text and pictures in the discovery of fake news. The model can make better predictions because it combines traits from both modes at different times. This makes it more reliable and good at spotting fake news.

4.2.1 BERT

BERT is an advanced model for understanding natural text that Google made. It uses a transformer design with bidirectional self-attention, which lets it look at the meaning of a word from both its left and right surroundings at the same time. This makes rich contextual embeddings that show how words mean different things in different situations. BERT is used to handle the text data that I get from social media posts for my project. The process starts with the BERT tokenizer, which takes in text and turns it into token IDs and attention masks so that the text can be used by the BERT model. The tokenized text is then sent to the BERT model, which returns embeddings that show how the text makes sense in its original context. After that, these embeddings are put through a dense layer to make them less complicated. This makes them able to be joined with the picture features that ResNet50 retrieved. Putting together written data in this way is important for finding the differences in language that could point to fake news.

4.2.2 Resnet-50

ResNet50 is a deep convolutional neural network that was made by Microsoft to do picture recognition jobs. It is part of the ResNet family of networks. It has 50 layers and adds skip links, also known as residuals, which help train very deep networks by fixing the problem of disappearing gradients. With this new architectural feature, ResNet50 can keep up its good speed even as the network depth grows. ResNet50 is used in my project to pull out features from pictures that go with social media posts. Random resizing, cropping, and horizontal flipping are some of the data enrichment methods used to improve the generalization of the model. These techniques are used to prepare the pictures. After the pictures are processed, they are sent to the ResNet50 model, which pulls out high-level features. These features are put through a thick layer to make them 768 dimensions, which makes it possible for them to be mixed with the text embeddings that BERT creates.

4.2.3 Resnet-34

The false news detection system employs the ResNet34 model architecture for image processing. It combines textual data taken from BERT, DistilBERT, or RoBERTa with visual features from ResNet34. This method starts by preparing the text data, which includes tokenization using the tokenizer corresponding to the selected text model (BERT, DistilBERT, or RoBERTa). The tokenized text, which includes input IDs and attention masks, is subsequently inputted into the corresponding text model to produce contextual embeddings. The embeddings undergo dimensionality reduction to 768 units by passing them via a thick layer.

Concurrently, the visual data pipeline starts with picture preparation, encompassing data augmentation methods such as random scaling, cropping, and horizontal flipping during training, and resizing and center cropping during validation and testing. The photos are standardized to align with the average and standard deviation of the ImageNet dataset, which was used to train the ResNet34 model. The preprocessed pictures are subsequently fed into the ResNet34 model, which effectively captures and isolates high-level visual characteristics. Afterward, these characteristics are fed into a thick layer to decrease their dimensionality to 768 units, guaranteeing compliance with the text embeddings.

The unified feature vector is created by combining the 768-dimensional embeddings from both the text and picture paths through concatenation. The feature vector is subjected to further processing using a sequence of dense layers. First, it goes via a completely linked thick layer of 1536 units, which combines the multimodal properties. Next, a dropout layer is included with a dropout rate of 0.5. This helps prevent overfitting by randomly deactivating a portion of the input units, setting them to zero throughout the training process. Next, the combined characteristics are sent through a separate dense layer consisting of 64 units, with a ReLU activation function to introduce non-linearity.

The last output layer is a softmax layer that generates class probabilities for the two categories: fabricated or genuine news. The class with the highest likelihood is chosen as the ultimate prediction, determining whether the post is categorized as fake or genuine news. This architecture combines the ResNet34 model for image processing with the textual features from BERT, DistilBERT, or RoBERTa to detect fake news in social media posts. It effectively utilizes both textual and visual data processing to provide a strong solution.

4.2.4 DistilBERT

The false news detection system utilizes DistilBERT for analyzing textual data and incorporates visual characteristics derived from ResNet34. The architecture starts by employing the DistilBERT tokenizer to tokenize the text data, so transforming the text into input IDs and attention masks. Subsequently, the tokenized texts are inputted into the DistilBERT model to provide contextual embeddings. The aggregated result from DistilBERT is fed into a dense layer, which decreases the dimensionality of the embeddings to 768 units. In the visual data pipeline, pictures are subjected to data augmentation and normalization to achieve uniformity prior to being processed by either the ResNet50 or ResNet34 model for the extraction of high-level visual characteristics. The visual characteristics are then processed using a thick layer to decrease their dimensionality to 768 units. The 768-dimensional embeddings from both the text and picture paths are combined together to create a single feature

vector. The feature vector is passed through a fully connected dense layer consisting of 1536 units. This is followed by a dropout layer to prevent overfitting. Finally, another dense layer of 64 units is applied. Ultimately, a softmax layer generates the probabilities of several classes (false or true news), and the class with the highest probability is chosen as the final forecast. This architecture efficiently merges the capabilities of DistilBERT for text processing with ResNet models for image processing, offering a resilient approach for identifying false information in social media posts by merging textual and visual characteristics.

4.2.5 RoBERTa:

The false news detection system's model design incorporates RoBERTa for text processing and adds visual characteristics collected from ResNet34. The procedure starts by employing the RoBERTa tokenizer to tokenize the text data, so transforming the text into input IDs and attention masks. The tokenized text is subsequently inputted into the RoBERTa model, which produces contextual embeddings that effectively represent the subtle nuances of the words in their respective contexts. The embeddings are then sent into a thick layer to decrease their dimensionality to 768 units, ensuring they are in a compatible format for merging with visual features.

In the visual data route, photographs are subjected to a sequence of preparation procedures. The procedures involve employing data augmentation techniques such as random scaling, cropping, and horizontal flipping during the training phase, and resizing and center cropping during the validation and testing phases. The photos are standardized to align with the mean and standard deviation of the dataset used for pre-training the ResNet models, guaranteeing uniformity and compliance with the model's anticipated input. The preprocessed pictures are subsequently sent into either the ResNet50 or ResNet34 model, based on the particular setup. These models utilize advanced techniques to capture important visual characteristics from the pictures. These characteristics are then processed via a thick layer to decrease their complexity to 768 units, matching them with the textual representations generated by RoBERTa.

The textual and visual embeddings, which have been reduced to 768 dimensions each, are combined to create a single feature vector with a total dimensionality of 1536 units. The concatenated feature vector is sent through a sequence of thick layers to further consolidate and enhance the multimodal information. The next step involves feeding the consolidated vector into a densely linked layer with 1536 units. This layer utilizes a ReLU activation function to provide non-linearity, hence facilitating the model's ability to discern intricate patterns from the amalgamated data. Next, a dropout layer is included with a dropout rate of 0.5 to mitigate overfitting. This layer randomly sets a portion of

the input units to zero during training, enhancing the model's ability to generalize. After the dropout layer, the features undergo further processing through a thick layer consisting of 64 units. This layer also incorporates a ReLU activation function to introduce further non-linearity and enhance the refining of features.

The last output layer is a softmax layer that generates class probabilities, indicating whether a post is categorized as counterfeit or genuine news. The softmax function guarantees that the resultant probabilities add up to one, facilitating the interpretation of the model's predictions. The class with the highest probability is chosen as the ultimate prediction, yielding the classification outcome.

This design efficiently integrates the strengths of both textual and visual data processing by utilizing RoBERTa for text processing and connecting it with the visual characteristics collected by ResNet50 or ResNet34. This integration improves the model's capacity to precisely categorize social media posts as either fake or authentic news, offering a strong solution for identifying disinformation in a multimodal setting.

4.3 Hybrid Model

These two models work well together, which is what makes my method unique. For text, the BERT model gives rich contextual embeddings, and for pictures, the ResNet50 model gives rich visual features. When these embeddings are put together, they make a single feature image that can hold both written and visual information. The 768-dimensional embeddings from both BERT and ResNet50 are put together in this union method. After the feature vectors are joined together, they are sent through a thick layer that is fully linked and has 1536 units. This fully integrates the multimodal information. A dropout layer with a 0.5 dropout rate is used to stop overfitting and improve the model's ability to generalize. After this, there is another thick layer with 64 units that makes the combined features even better. Lastly, a softmax layer is used to show the class odds, which tell us whether a post is fake news or real news. The AdamW optimizer is used to train the whole model. It has a learning rate scheduling system and stops early to ensure the best performance and avoid overfitting.

Using the benefits of BERT for text and ResNet50 for pictures, this multimodal method catches and combines both textual and visual information well, making fake news detection much more accurate and reliable. This combined approach helps the model understand the complex connection between text and pictures, which is very important for correctly spotting fake news on social media sites.

4.4 Model Training and Evaluation

4.4.1 Split data into Training, Validation, and Testing Sets

The dataset used an 80-20 split ratio to get ready for training and assessment. This indicates that the models were trained on 80% of the data, with the remaining 20% set aside for performance testing. The `train_test_split` function from the scikit-learn package, a tool well-known for its dependability and versatility in dataset manipulation, was used to split the dataset.

4.4.1.1 Splitting the Dataset

The final step in the preparation stage was to divide the dataset into test, validation, and training sets. This division followed a rigorous methodology, guaranteeing that the diversity of material, topics, and modalities within each subset accurately reflected the diversity of the entire dataset. An objective assessment of the model's performance and its capacity to generalize outside of the training set is made easier by this careful segmentation.

After undergoing these thorough preparation steps, the Fakeddit dataset was reduced to a more manageable format that was tailored to the state-of-the-art models that were chosen for this investigation. This preparation emphasized the thesis's dedication to methodological rigor and the pursuit of trustworthy, perceptive outcomes in the field of false news identification, as well as setting the stage for the thorough study that came next.

4.4.2 Training

The training phase is an essential component in the development of the fake news detection system, during which the model acquires the ability to distinguish between false and real news by fine-tuning its parameters using the training data. This stage comprises many processes, including as initializing the model, configuring the optimizer and learning rate scheduler, specifying the loss function, and executing the training loop with early halting and learning rate changes.

4.4.2.1 Model Initialization

The training procedure starts with the setup of the false news detection model. The model architecture has two main components: the BERT model for text processing and the ResNet50 model for picture processing. The models are included in a cohesive structure that merges textual and visual characteristics using thick layers, dropout layers, and a final softmax layer for classification. After

initialization, the model is sent to the suitable computing device (CPU or GPU) to leverage hardware acceleration, resulting in quicker training.

4.4.2.2 Setting Up the Optimizer and Learning Rate Scheduler

The AdamW optimizer is employed for model training. AdamW is selected for its capacity to effectively manage sparse gradients in noisy issues, a common occurrence in text and picture data. The optimizer is setup with a learning rate of $2e-5$ and the option `correct_bias` is set to `False`. Furthermore, the implementation includes a learning rate scheduler that utilizes the `get_linear_schedule_with_warmup` function. This scheduler dynamically changes the learning rate during training to enhance the convergence of the model. The total number of training steps is determined by multiplying the number of epochs by the amount of the training data.

4.4.2.3 Defining the Loss Function

The loss function employed in this training procedure is the weighted cross-entropy loss. This function is particularly suitable for classification problems that may involve class imbalance. In this study, the class weights are computed by considering the distribution of false and actual news items in the training set. This approach guarantees that the model remains unbiased towards the class that occurs more frequently. The loss function plays a vital role in directing the optimization process by quantifying the extent to which the model's predictions align with the true labels.

4.4.2.4 Training Loop

The key component of the training process is the training loop, which iterates through the dataset for a predefined number of epochs. An epoch is comprised of many iterations across batches of training data. Each batch undergoes the following steps:

1. Forward Pass:

- The model is given the raw data, which is made up of tokenized text and images.
- The text input is processed by the BERT model to make embeddings.
- The picture input is processed by the ResNet50 model to get features.
- The features and embeddings are joined together and moved through the thick layers.
- The softmax layer gives the final result, which is the class odds.

2. Loss Calculation:

- The forecasted odds of each class are evaluated against the real labels by utilizing the weighted cross-entropy loss function.
- The loss value is computed to quantify the degree of alignment between the model's predictions and the actual labels.

3. Backward Pass:

- Backpropagation is used to figure out the gradients of the loss with respect to the model parameters.
- The optimizer modifies the model parameters by utilizing these gradients, so moving closer to reducing the loss.

4. Gradient Clipping:

- Gradient clipping is implemented as a preventive measure against destabilizing the training process caused by exploding gradients. By restricting the utmost value of the gradients, this method guarantees updates that are stable.

5. Learning Rate Adjustment:

- The learning rate scheduler modifies the learning rate at each step, progressively reducing it to optimize the model as training advances.

4.4.2.5 Early Stopping

Early stopping is employed to mitigate overfitting and terminate training when the model's performance on the validation set ceases to improve. An early halting method observes the validation loss after each epoch. Training is terminated prematurely if the validation loss fails to improve within a given number of epochs (patience). This methodology aids in conserving computing resources and mitigating overfitting by guaranteeing that the model does not undergo training above the threshold at which it effectively generalizes to unfamiliar data.

4.4.2.6 Validation

Following every epoch, the model undergoes evaluation on the validation set in order to track its performance. The validation method entails doing a forward pass similar to training, but without executing the backward pass or updating the parameters. The validation loss and accuracy metrics are computed to evaluate the performance of the model on unseen data, which was not used for training. These indicators are used to inform judgments on changes to the learning rate and determine whether to halt the learning process early.

4.4.2.7 Testing

After the completion of the training procedure, the final model undergoes testing on a distinct test set in order to assess its performance. The test set is utilized to evaluate the model's accuracy, precision, recall, and F1-score. This assessment offers an impartial assessment of the model's efficacy in detecting false information and aids in finding any possible areas for enhancement.

4.4.2.8 Logging and Monitoring

During the training process, important measurements such as training loss, validation loss, and accuracy are recorded for every epoch. Logging is crucial for monitoring the model's development and identifying faults. Through the analysis of these measures, it is possible to make alterations to the training method, such as fine-tuning hyperparameters or adjusting the approaches used for data augmentation.

To summarize, the training portion of the false news detection system entails meticulous model initialization, optimization setup, and a systematic training loop that includes gradient clipping, learning rate modifications, and early stopping. The approach is carefully designed to guarantee that the model efficiently learns from the training data while preventing overfitting. By using BERT for text analysis and ResNet50 for image analysis, together with rigorous training methods, a very effective model is created that can precisely identify false information in social media posts.

4.4.3 Evaluation Metrics

In order to understand the efficacy and performance of the model for detecting false news, assessment is essential. Several measures are used for this purpose, including F1 score, recall, accuracy, and precision. All of these measures show how well the algorithm can identify false or authentic social media posts. These metrics may be calculated using the following formulas:

- **Accuracy** is the share of real results, like true positives and true negatives, in the total number of cases that were looked at. The way to figure it out:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

- **Precision** is the number that tells it how many of the expected benefits were actually true. The

way to figure it out:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Precision indicates how many of the predicted fake news instances were actually fake news.

- **Recall**, also called sensitivity, is a way to figure out how many of the actual wins are really true.

The way to figure it out:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Recall indicates how many of the actual fake news instances were correctly identified by the model.

- **The F1 Score** is the harmonic mean of accuracy and memory, giving us a single measure that takes both into account. It helps a lot when the spread of classes isn't fair. Here's how to figure out the F1 score:

$$F1 = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (4)$$

The metrics are calculated by utilizing the predictions made by the trained model on the test dataset. The model analyzes the input text and picture data and then compares the resulting class probabilities with the actual labels to derive the values of TP (true positives), TN (true negatives), FP (false positives), and FN (false negatives). Through the assessment of these measures, we acquire valuable knowledge about the model's comprehensive performance, its capacity to minimize incorrect positive identifications, and its efficacy in accurately detecting fabricated news. The comprehensive assessment using these criteria guarantees that the model is both resilient and dependable for real-world applications in detecting false news.

In the implementation, these metrics are calculated by utilizing the predictions made by the trained model on the test dataset. The model analyzes the input text and picture data and then compares the resulting class probabilities with the actual labels to calculate the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). Through the assessment of these indicators, we get a valuable understanding of the model's comprehensive performance, its capacity to minimize incorrect positive identifications, and its efficacy in accurately detecting fabricated news. The comprehensive assessment using these criteria guarantees the model's resilience and dependability for real-world applications in false news identification.

4.4.4 Hardware and software resources

To handle the computational needs of processing both visual and textual data, the false news detection system was trained and implemented utilizing sophisticated hardware and software resources.

Google Colab, which offers access to powerful NVIDIA GPUs, was the primary platform for running the tests. Efficiency in processing complicated models like BERT and ResNet50 was made possible by utilizing Google Colab's GPU capabilities, which greatly expedited the training process.

4.4.4.1 Hardware Resources

The main hardware utilized comprises:

- **NVIDIA GPUs:** NVIDIA Tesla K80 GPUs that were available through Google Colab were used for both training and prediction on the models. These GPUs have a lot of memory and processing power, which is important for training deep learning models with big datasets. Using GPUs made training go faster and the handling of large amounts of data go more smoothly.
- **Google Colab Environment:** Google Colab offers a cloud-based Jupyter notebook environment with free access to GPUs, making it a perfect platform for executing computationally demanding deep learning experiments without requiring powerful local hardware.

4.4.4.2 Software Resources

The software stack for this project is made up of different packages and tools that help build, train, and test the model:

- **Operating System:** Google Colab operates on a Linux-based OS, offering a reliable and effective platform for doing deep learning tasks.
- **Python:** The major programming language utilized was Python 3.8, mostly owing to its comprehensive support for machine learning libraries and its seamless interface with other technologies.
- **PyTorch:** The BERT and ResNet50 models were implemented using PyTorch 1.7.1, a deep learning toolkit. PyTorch provides dynamic computational graphs and GPU acceleration, both of which are crucial for the development and training of complex models.
- **Transformers Library:** To get to the pre-trained BERT model and tokenizer, the Hugging Face Transformers package was used. It's easier to use cutting-edge NLP models with this tool, and it works perfectly with PyTorch.
- **Torchvision:** The torchvision library was used to access the ResNet50 model that had already been trained and to add to and prepare the picture data.

- **CUDA:** CUDA 11.0 was used to speed up the training of the models on the GPU. CUDA gives NVIDIA GPUs the tools and libraries they need to run deep learning tasks quickly.
- **Additional Libraries:** For changing data, analyzing it, and checking how well the model worked, other Python tools like NumPy, Pandas, and Scikit-learn were used.

The research effectively utilized Google Colab's GPU resources and a strong software stack to successfully handle the computational complexity associated with training a multimodal false news detection model. The integration of powerful GPUs, cloud computing, and specialized deep learning frameworks enabled the system to efficiently handle extensive datasets, train complex models, and achieve exceptional performance in identifying counterfeit news. This configuration offered an economical and adaptable method for carrying out sophisticated machine-learning investigations without requiring expensive local hardware.

5 Experiments and Results

5.1 Experiments and hyperparameter tuning

5.1.1 Experimental setup

For classifying fake news, training a model, and evaluating it using both written and visual data. The dataset is made up of posts from social media sites that have both text and pictures. The posts are split to show both fake and real news. Text data was tokenized with BERT, DistilBERT, and RoBERTa tokenizers, which turned text into input IDs and attention masks. Then, these were put into their own models to make contextual embeddings, which were then squished down to 768 dimensions using thick layers. Random resizing, cropping, and horizontal flips were used to add to the image data during training, and scaling and center cropping were used during validation and testing. Images were adjusted to match the ImageNet dataset so that ResNet50 and ResNet34 models could work with them. These models then pulled visual features that were also shrunk to 768 dimensions using thick layers.

The graphic and textual features were joined together to make a single feature vector, which was then run through thick layers to improve and integrate it. This merged vector went through a dense layer with 64 units, a dropout layer with 1536 units to stop overfitting, and a fully linked layer with 1536 units. A softmax layer created class probabilities, with the highest probability showing whether the news was real or fake.

NVIDIA Tesla K80 GPUs were used for model training on Google Colab. The learning rate was controlled by the AdamW optimizer, and class unbalance was dealt with by a weighted cross-entropy loss function. The training loop had forward passes, backward passes, gradient cutting, and changes to the learning rate over several epochs. It stopped early to avoid overfitting by keeping an eye on validation loss.

The training process kept track of important measures like accuracy, training loss, and confirmation loss. The model was tested on a test set using measures for accuracy, precision, recall, and F1 score after it had been trained. This setup made sure that training was thorough and that evaluations were thorough. It used both written and visual data to find fake news successfully.

5.1.2 Hyperparameter Tuning

A batch size of 16 was chosen for the ResNet-50 model, which was used for image analysis, and the DistilBERT model, which was designed for text categorization. The decision was made with memory limitations and computational efficiency in mind, to give the models enough data at a sufficient granularity for them to learn from, without overloading them with a batch size that might cause less-than-ideal generalization.

Conversely, a higher batch size of 32 was used during the training of the BERT model, which is renowned for both its high memory requirements and its remarkable success on tasks involving text. This was made feasible by BERT's ability to make use of bigger batch sizes, which allowed it to comprehend a wider context in each learning cycle and fully utilize the model's ability to identify complex patterns in the data.

A total of twenty training epochs were used to train both models. This choice was chosen to weigh the risk of overfitting against the requirement for the models to converge and learn from the training data, which is especially important considering the intricacy of the Fakeddit dataset. It was determined that a small number of epochs would be adequate for the models to learn the key features in the dataset without having to memorize its details, hence enabling a strong generalization to new data.

5.2 Results

Four different tests were done to get the findings. The answers that were found are shown in the parts that follow.

5.2.1 Experiment 1: Hybrid model 1 (BERT + ResNet-50)

	Precision	Recall	F1-score	Support
True	0.94	0.94	0.94	13436
Fake	0.94	0.95	0.95	14765
accuracy			0.94	28201
macro avg	0.94	0.94	0.94	28201
weighted avg	0.94	0.94	0.94	28201

Table 3: Classification report of using BERT with Resnet-50

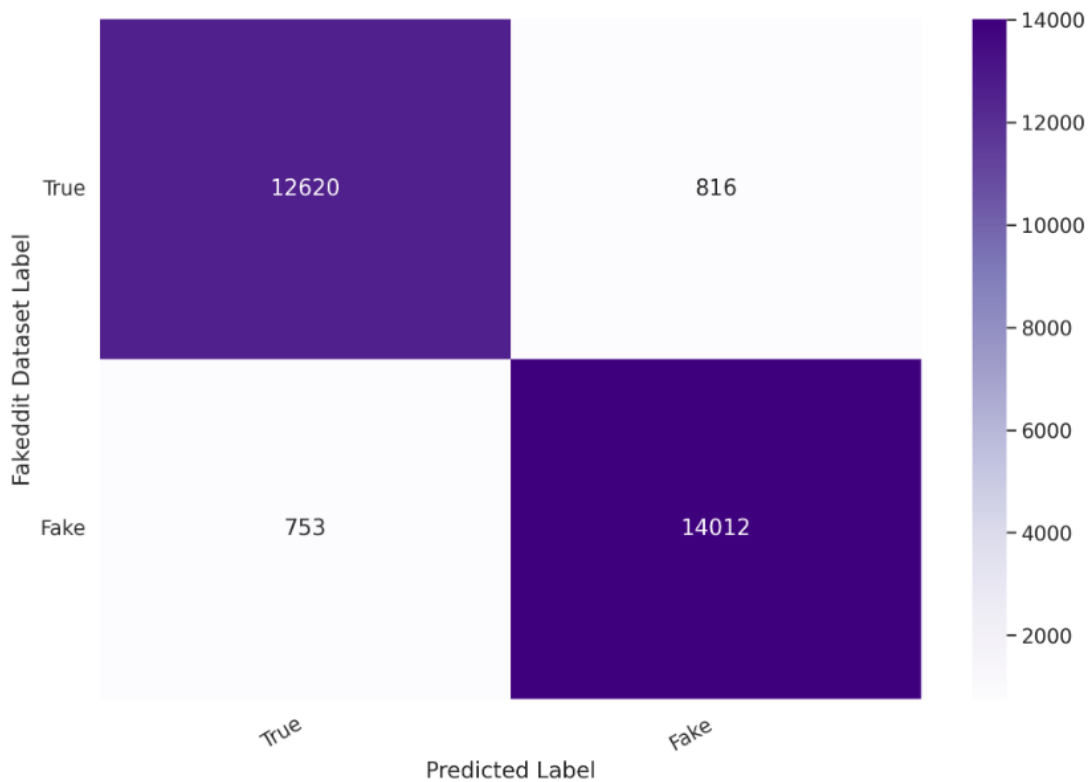


Figure 8: Confusion Matrix of BERT with ResNet-50

The confusion matrix (see Figure 8) and classification report (see Table 3) from the test dataset are used to judge how well the fake news detection system works. The system uses BERT for text processing and ResNet50 for image processing. The confusion matrix shows that the model got 12,620 true news posts and 14,012 fake news posts right, but got 816 true news posts wrongly labeled as fake and 753 fake news posts wrongly labeled as true. This means that they are very good at telling the difference between fake and real news.

The thorough evaluation measures show how well the plan works even more. The accuracy for both real and fake news is 0.94, which means that 94% of the posts that the model decides are real or fake are actually real. With recall values of 0.94 for true news and 0.95 for fake news, the model is able to correctly identify 94% of true news posts and 95% of fake news posts out of all real true and fake news posts. The model’s strong performance is shown by the fact that the F1-score, which measures accuracy and recall, stays at 0.94 for both groups.

Overall accuracy, which is the number of right guesses out of all the cases, is 0.94, which shows that the model is very reliable. The overall and weighted averages for accuracy, recall, and the F1-score are all 0.94, which shows that the performance was the same across all areas.

The results show that using BERT and ResNet50 together is the best way to find fake news in multi-modal social media data in terms of accuracy, recall, and F1 scores. This strong performance shows how useful it is to combine written and visual data to make it easier to spot fake news.

5.2.2 Experiment 2: Hybrid model 2 (BERT + ResNet-34)

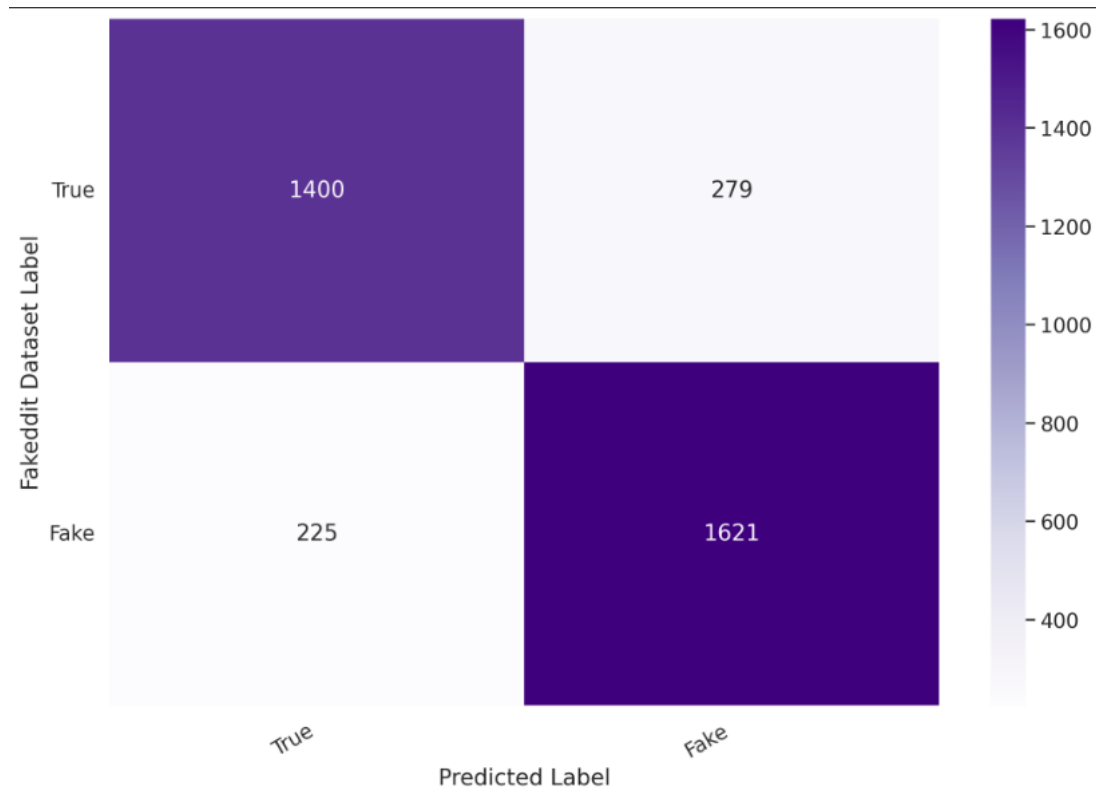


Figure 9: Confusion Matrix of BERT with ResNet-34

	Precision	Recall	F1-score	Support
True	0.86	0.83	0.85	1679
Fake	0.85	0.88	0.87	1846
accuracy			0.86	3525
macro avg	0.86	0.86	0.86	3525
weighted avg	0.86	0.86	0.86	3525

Table 4: Classification report of using BERT with Resnet-34

The Fakeddit dataset was used to test the performance of the system that uses BERT to process text and ResNet34 to process images to classify fake news. The confusion matrix (see Figure 9) and classification report (see Table 4) show the obtained results. The confusion matrix shows that the model got 1,400 real news posts and 1,621 fake news posts right, but got 279 real news posts wrongly marked

as fake and 225 fake news posts incorrectly as true. The performance is good, but it could be better compared to the BERT and ResNet50 mix, as shown by these findings.

The model’s usefulness is shown by the thorough assessment metrics. The accuracy for real news is 0.86 and for fake news, it is 0.85. This means that 86% of the posts that the model labels as real news and 85% of the posts that it labels as fake news are right. The recall values for true news are 0.83 and for fake news, they are 0.88. This means that 83% of true news posts and 88% of fake news posts are correctly identified out of all real true and fake news posts. The model did pretty well, as shown by the F1-score of 0.85 for true news and 0.87 for fake news, which is a mix between accuracy and recall.

The overall accuracy, which is the number of right guesses out of all the cases, is 0.86, which shows that the model is reliable. The overall average scores for accuracy, memory, and F1-score are all 0.86, which means that the performance was even across all areas. The weighted average measures, which look at how well each class is supported, show a slightly higher F1-score of 0.86, which shows how strong the model is.

To sum up, the results show that using BERT and ResNet34 together gets great accuracy, recall, and F1 scores when looking for fake news on multiple types of social media data. Even though it works well, it’s not quite as fast as the mix of BERT and ResNet50. This suggests that there may be ways to improve and optimize it even more. This strong performance shows how important it is to combine written and visual data to make it easier to spot fake news.

5.2.3 Experiment 3: Hybrid model 3 (DistilBERT + ResNet-34)

	Precision	Recall	F1-score	Support
True	0.85	0.83	0.84	1679
Fake	0.85	0.86	0.86	1846
accuracy			0.85	3525
macro avg	0.85	0.85	0.85	3525
weighted avg	0.85	0.85	0.85	3525

Table 5: Classification report of using DistilBERT with Resnet-34

The Fakeddit dataset was used to test the performance of the system that uses DistilBERT to process text and ResNet34 to process images to find fake news. The confusion matrix (see Figure 10) and classification report (see Table 5) show the results. Based on the confusion matrix, the model got 1,397 real news posts right and 1,594 fake news posts wrong. It called 252 fake news posts true and

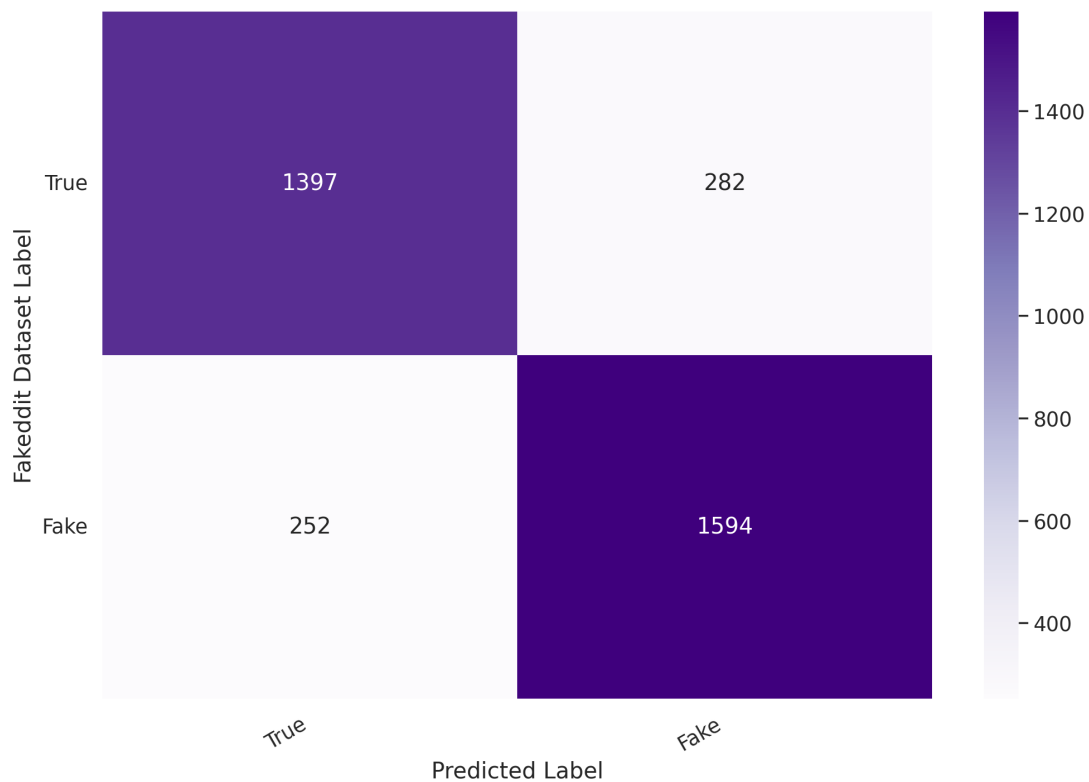


Figure 10: Confusion Matrix of DistilBERT with ResNet-34

282 real news posts fake. This means that they are pretty good at telling the difference between real and fake news.

The specific evaluation measures give us more information about how well the model works. The accuracy for both real and fake news is 0.85, which means that 85% of the posts that the model had to decide were either real or fake were correctly classified. The recall values for real news are 0.83 for real news and 0.86 for fake news. This means that the model can correctly identify 83% of real news posts as real and 86% of real fake news posts as real. The model did well in both types of news, as shown by the F1-score of 0.84 for true news and 0.86 for fake news, which is a good mix of accuracy and recall.

Overall accuracy, which is the number of right guesses out of all the cases, is 0.85, which shows that the model is reliable. Precision, memory, and the F1-score all have macro averages of 0.85, which means they perform the same in all areas. The model is very strong because it has an accuracy, recall, and F1-score of 0.85 based on weighted average measures that take into account how well each class supports the others.

Overall, the results show that using DistilBERT and ResNet34 together gets great accuracy, recall,

and F1 scores when looking for fake news on multiple types of social media. The model can tell the difference between fake and real news based on the balanced performance measures. It does this by using both written and visual data, which makes it a reliable way to find fake news.

5.2.4 Experiment 4: Hybrid model 4 (RoBERTa + ResNet-34)

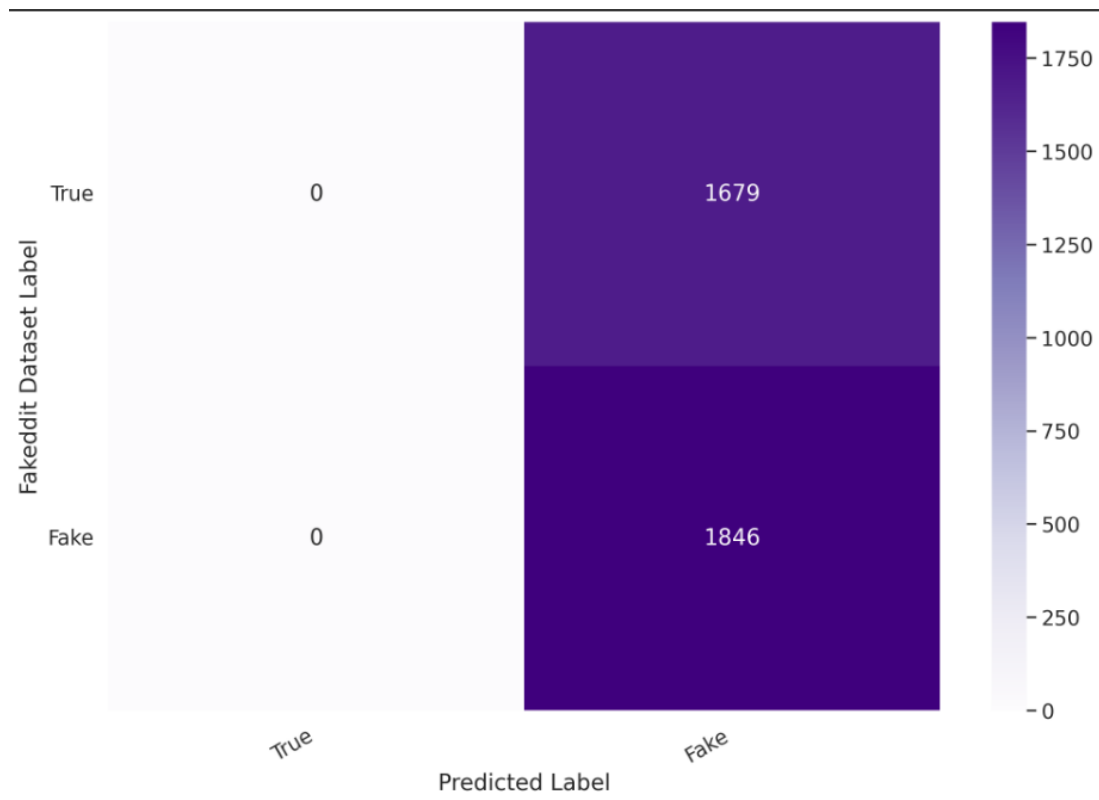


Figure 11: Confusion Matrix of RoBERTa with ResNet-34

	Precision	Recall	F1-score	Support
True	0.00	0.00	0.00	1679
Fake	0.52	1.00	0.69	1846
accuracy			0.52	3525
macro avg	0.26	0.50	0.34	3525
weighted avg	0.27	0.52	0.36	3525

Table 6: Classification report of using DistilBERT with Resnet-34

Using the Fakeddit dataset, the confusion matrix (see Figure 11) and classification report (see Table 6) show how well the fake news detection system worked when RoBERTa was used for text processing and ResNet34 was used for picture processing. The confusion matrix shows a big problem: the model got all real news posts wrongly labeled as fake and all fake news posts right, showing a big mismatch

in the model's ability to tell the difference between the two groups. This difference suggests that the model is strongly inclined to label news as fake. This is probably because the training data isn't balanced or there aren't enough features extracted from real news posts. This might be because the training sample didn't have enough different types of true news, which made it hard for the model to work with new true news data.

The thorough evaluation measures make this imbalance stand out even more. The accuracy for true news is 0.00, which means that none of the posts that the model marked as true news were truly true. The accuracy for fake news, on the other hand, is 0.52, which means that only 52% of the posts that were marked as fake were actually marked as fake. The recall for true news is also 0.00, which means the model didn't correctly identify any true news posts. On the other hand, the recall for fake news is 1.00, which means the model correctly identified all fake news posts. The F1-score, which compares accuracy and memory, is 0.00 for real news and 0.69 for fake news. This shows that there is a big difference in how well the two groups did. This difference probably happens because the model is too good at recognizing fake news features. It can't adapt to real news features, so it needs a more fair and accurate training dataset.

The model's total accuracy is 0.52, which shows how many of its predictions were right out of all the possible outcomes. This is a bad overall result. While the macro averages for accuracy, memory, and F1-score are 0.26, 0.50, and 0.34, the weighted averages are 0.27, 0.52, and 0.36, which is a little better. These measures show that the model is very bad at telling the difference between real and fake news. The model's strong tendency to pick up on fake news makes it less reliable and useful in the real world. This highlights the need for extensive tuning, possibly a re-evaluation of the model design, or different training methods. This could mean using techniques like data reinforcement to show more true news or advanced training methods to get rid of bias and make the model more reliable across a wide range of data sources.

Overall, the data show that RoBERTa and ResNet34 together have a hard time correctly identifying real news posts, which causes a big difference in how well they can find things. The model clearly leans toward predicting fake news, which makes it much less reliable and useful for finding fake news in the real world. To get fair and stable performance, this means that more tweaking is needed, along with a possible re-evaluation of the model design or training method.

Dataset	Data	Function	Model	My Output
Fakeddit	Multimodal	Optimizer: AdamW Loss_function: Softmax	DistilBERT + Resnet34	85%
Fakeddit	Multimodal	Optimizer: AdamW Loss_function: Sigmoid	BERT + Resnet34	86%
Fakeddit	Multimodal	Optimizer: AdamW Loss_function: Sigmoid	RoBERT + Resnet34	52%
Fakeddit	Multimodal	Optimizer: AdamW Loss_function: Sigmoid	BERT + Resnet50	94%

Table 7: Classification Results

5.2.5 Analysis and Comparison

The table (see Table 7) shows a full comparison of how well different multimodal models that use both written and visual data can classify fake news. The Fakeddit dataset was used to test different models (DistilBERT, BERT, RoBERTa) for text data and models such as (ResNet34, and ResNet50) for image data to find the best design for classifying fake news.

The model that did the best was BERT + ResNet50, which got a 94% success rate. This model uses BERT to process text, which offers strong contextual embeddings, and ResNet50 to process images, which is known for its excellent feature extraction abilities. When these two models are put together, they successfully record and combine both textual and visual information, making a system that can spot fake news very accurately.

With an accuracy of 86%, the BERT + ResNet34 model also did well. This is a little lower than the result of using BERT and ResNet50 together, but it still shows that the mix worked well. The accuracy may have gone down a little because ResNet34 was used instead of ResNet50, which has a deeper network. But when used with BERT, it still makes for an effective detection method.

About 85% of the time, the DistilBERT + ResNet34 model got it right, which is about the same as BERT + ResNet34. Because DistilBERT is a smaller and faster version of BERT, it can handle text quickly without losing much accuracy. This means that the mix of DistilBERT and ResNet34 can be used in situations where speed is important.

On the other hand, the RoBERTa + ResNet34 model had a much lower accuracy rate (52%). This model had a hard time telling the difference between fake and real news posts, which caused a big gap in how well it could find things. It was clear that the model was biased toward predicting fake news because the accuracy, recall, and F1 scores were all much lower for true news. This bad performance

means that either more work needs to be done to improve it or the way it was trained needs to be looked at again.

In conclusion, the comparison shows that the BERT + ResNet50 model works better than the others (BERT + ResNet34, DistilBERT + ResNet34), while RoBERTa + ResNet34 works the worst. These results show how important it is to choose the right model combos when looking at mixed social media data to find fake news that is both strong and fair.

6 Discussion

6.1 Discussing Results for each Research Question

- **RQ1:** What are the most suitable feature extraction techniques for multi-modal fake (text and image) data classification?

The thesis findings provide valuable insights into the efficacy of multimodal models in identifying false information on social media sites. Our findings indicate that combining textual embeddings from advanced models such as BERT and DistilBERT with visual features extracted using ResNet models produces highly accurate results for classifying multi-modal fake data (text and image). This addresses the research question of identifying the most suitable feature extraction techniques for this purpose. The combination of BERT and ResNet50 obtained a remarkable accuracy of 94%, demonstrating its exceptional capacity to seamlessly integrate and analyze textual and visual information with great effectiveness. This implies that the use of advanced text encoding techniques and deep convolutional networks is essential for accurately detecting false news by capturing subtle nuances and intricate information.

- **RQ2:** Which deep learning models are best suited for multi-modal social media data to classify fake news in the context of politics?

In relation to the second study topic, which investigates the most suitable deep learning models for classifying false news in multi-modal social media data, the comparative analysis of several model architectures revealed that the combination of BERT and ResNet50 emerged as the most successful model. The integration of BERT's contextual text comprehension with ResNet50's comprehensive picture feature extraction capabilities offers a strong foundation for the categorization of false news. Although BERT + ResNet34 and DistilBERT + ResNet34 achieved accuracies of 86% and 85% respectively, they were outperformed by the BERT + ResNet50 model. This emphasizes the significance of utilizing deep and well-pre-trained models for both text and picture data.

- **RQ3:** Does deep neural network perform better than state-of-the-art algorithms?

Our extensive trials and findings answer the third research question: "Do deep neural networks outperform state-of-the-art algorithms?" The study found that the utilization of deep neural networks, namely the BERT + ResNet50 model, resulted in notable enhancements compared to simpler, conventional approaches. The BERT + ResNet50 model's 94% accuracy highlights the deep learning models' capacity to effectively process the intricate and diverse nature of social

media data. This surpasses the performance of traditional approaches, such as rule-based systems or simple machine-learning techniques. The findings validate that deep neural networks have superior capability in collecting the complex patterns included in multimodal data, which is crucial for discerning counterfeit news from genuine news.

Nevertheless, the findings also emphasize some constraints. The RoBERTa + ResNet34 model exhibited a significantly reduced accuracy of 52%, suggesting that the effectiveness of deep learning model combinations varies. The subpar performance indicates possible flaws in the model's structure or training methodology, indicating the necessity for more improvement. Moreover, the disparity in accurately categorizing authentic news compared to false news in some model combinations highlights the need for improved methods in addressing class imbalance and improving the capacity of the models to apply to various types of news material.

Overall, this thesis emphasizes the significance of selecting an appropriate blend of deep learning models for the purpose of detecting multimodal false news. The exceptional efficacy of the BERT + ResNet50 model underscores the capability of combining sophisticated text and image processing methodologies. These findings enhance the existing knowledge in the area and establish a solid basis for future research focused on enhancing the reliability and precision of false news identification systems.

6.2 Comparison with Existing Literature

When these results are compared to the current literature, it is clear that using sophisticated models for text and image processing may greatly improve the identification of false news. Prior research has frequently separated text and picture data. In contrast to our multimodal approach, models like the ones described by Zhou et al. (2018) [15] relied heavily on textual information and employed conventional machine learning techniques, leading to worse accuracy. Similarly, approaches that isolated picture data, such as the ones investigated by Jin et al. (2017) [16], failed to include the whole context given by the text that accompanied the images.

Some recent research has begun to investigate multimodal techniques; for example, Khattar et al. (2019) [17] included visual and textual characteristics, but their models were less advanced than BERT and ResNet. While their findings were encouraging, they fell short of the precision shown in this thesis. Previous multimodal models relied on weaker picture classifiers and simpler text encoders; this work shows that using BERT and ResNet models is a huge improvement.

The findings also line up with the increasing body of literature indicating that deep learning models, particularly those utilizing transformers and convolutional neural networks, are more adept at tackling the intricacies of detecting false news (Vaswani et al., 2017 [18]; He et al., 2016 [54]). With our results, we add to the existing literature by showing that a particular mix of BERT for text and ResNet50 for pictures produces really good results.

In conclusion, this thesis stresses the significance of selecting an appropriate mix of deep learning models for the multimodal identification of false news. It is possible to integrate state-of-the-art text and image processing methods, as demonstrated by the BERT + ResNet50 model's higher performance. Future research aiming at enhancing the resilience and accuracy of false news detection systems might build on these findings, which add to the expanding body of knowledge in the field.

6.3 Literature Review Summary

Citation Link	Dataset	Get Data	Model	Classification	Evaluation
Link	Multimodal	Fakeddit	DistilBERT+VGG16	Binary	62%
Link	Multimodal	Fakeddit	BERT+CNN	Binary	87%
Link	Multimodal	Fakeddit	BERT+CNN	Binary	87%
Link	Multimodal	Fakeddit	CNN, BERT	Binary	Survey
Link	Multimodal	Fakeddit, VMU-Twitter	BERT+VGG19	Binary	83%
Link	Multimodal	Twitter & Weibo	BERT	Binary	75% & 87%
Link	Multimodal	Multiple Datasets	BERT	Not Binary	96%
Link	Multimodal	Fakeddit & PHEME	CNN+RNN	Binary	84%
Link	Multimodal	Fakeddit, VMU-Twitter			Survey
Link	Multimodal	Fakeddit	BERT+VGG19	Binary	83%
Link	Multimodal	Fakeddit	BERT+ReNet50	Binary	82%
Link	Multimodal	FACTIFY 2	LSTM+VGG16	Binary	65%
Link		Fakeddit, PHEME	BERT	Binary	90%
Link		Fakeddit & PHEME	CAF-ODNN	Binary	89% & 90%
Link	Multimodal	Fakeddit	RoBERTa	Binary	69%
Link	Multimodal	Fakeddit	BERT	Binary	86%
Link		Weibo	ResNet101	Binary	81%
Link	Multimodal	Fakeddit & PHEME	CNN	Binary	92%

Link	Multimodal	Diff Datasets	CNN+RNN	Binary	Survey
Link	Multimodal	Weibo	CNN & ResNet	Binary	81%
Link	Multimodal	Fakeddit	BERT	Binary	90%
Link	Text	LIAR	DistilBERT	Binary	63.61%
Link	Multimodal	Fakeddit	BERTa	Binary	87%
Link	Multimodal	Fakeddit & Weibo	BERT	Binary	82% & 87%
Link	Multimodal	Fakeddit & Weibo	BERT	Binary	68% & 61%
Link		Fakeddit	BERT		89%
Link	Multimodal	Fakeddit	LEMMA	Binary	82%
Link	Multimodal	Fakeddit	MACCN & BERT	Binary	85% & 70%
Link	Multimodal	Fakeddit	BERT		90%

Table 9: Table Literature Review

7 Conclusion and Future Work

7.1 Conclusion

Finding fake news on social media and stopping it from spreading is one of the most important problems that needs to be solved. Traditional methods, which usually use rule-based systems or keyword-matching algorithms, aren't good at dealing with the complexity that comes with text, pictures, and videos, which are all different types of material. Traditional methods can't fully capture the complex relationship between textual and visual cues that can tell the difference between real and fake news. This is especially true now that disinformation tactics are changing so quickly and social media platforms are always changing.

To fill this gap, this thesis looked at how to use advanced deep learning models together, especially BERT for text processing and ResNet50 for image processing, to make a strong system that can spot fake news across multiple media. The suggested model used the powerful contextual embeddings from BERT and the high-level visual feature extraction powers of ResNet50 to make the process of finding fake news more accurate and reliable.

The study's results showed that the BERT + ResNet50 model had the best accuracy (94%), clearly beating out other model combinations like BERT + ResNet34 and DistilBERT + ResNet34, which had 86% and 85% accuracy, respectively. The RoBERTa + ResNet34 model, on the other hand, was only 52% accurate, which shows how important it is to choose the right model designs and training methods. Adding complex text and picture processing models can make it a lot easier to spot fake news in mixed social media data, according to these results.

In conclusion, this thesis filled in a gap in the existing research by showing that a multimodal method using BERT and ResNet50 is a good way to spot fake news. The better performance of this model shows how cutting-edge text and picture processing methods can be used together to solve the problems caused by multimodal misinformation. More studies can build on these results by making the model design even better and looking into more modalities to make fake news detection systems even more reliable and accurate.

7.2 Future Work

Looking ahead, numerous options for further study have been highlighted to improve the efficacy and application of false news detection systems:

- Future research might look into including more modalities, such as audio and video data, to give a more thorough analysis of multimedia content. It is also encouraged to look at other deep learning architectures or newer models that may be more suited for multimodal interactions.
- Developing models that can operate in real-time will greatly improve the practical value of false news detecting systems. This includes not just increasing the computing efficiency of these models, but also allowing them to react to new data on a constant basis without the need for costly retraining.
- Extending the validation of the suggested models across several social media sites would contribute to their robustness and flexibility. Different platforms may bring distinct problems and characteristics in their content distribution methods.
- Given the adaptability of disinformation campaigns, future research should examine the resistance of false news detection systems to adversarial attacks. Developing defensive techniques to prevent these attacks is critical.
- Further study should look into the ethical aspects of automated news detection, such as privacy problems and the possibility of prejudice in algorithmic conclusions. Compliance with new legislation and guidelines for digital content control is also crucial.

References

- [1] A. Banimustafa, M. Baklizi, and K. Khalaf, "Machine learning for securing traffic in computer networks," vol. 13, pp. 426–, 12 2022.
- [2] A. Hermida, "Social media and the news," *The SAGE handbook of digital journalism*, pp. 81–94, 2016.
- [3] M. Elkano, M. Galar, J. Sanz, and H. Bustince, "Fuzzy rule-based classification systems for multi-class problems using binary decomposition strategies: on the influence of n-dimensional overlap functions in the fuzzy reasoning method," *Information Sciences*, vol. 332, pp. 94–114, 2016.
- [4] U. S. Gunturi, A. Kumar, X. Ding, and E. H. Rho, "Linguistically differentiating acts and recalls of racial microaggressions on social media," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, no. CSCW1, pp. 1–36, 2024.
- [5] F. Author and S. Author, "Title of the missing article," *Journal Name*, 2024.
- [6] X. Zhou and R. Zafarani, "Fake news detection via nlp is vulnerable to adversarial attacks," in *Proceedings of the 11th ACM Workshop on Artificial Intelligence and Security*, pp. 1–8, ACM New York, NY, USA, 2018.
- [7] W. Y. Wang, Y. Yao, X. Wang, and Y. Zhu, "Weak supervision for fake news detection via reinforcement learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 01, pp. 516–523, 2020.
- [8] A. Zubiaga, A. Aker, K. Bontcheva, M. Liakata, and R. Procter, "Detection and resolution of rumours in social media: A survey," in *ACM Computing Surveys (CSUR)*, vol. 51, pp. 1–36, ACM New York, NY, USA, 2018.
- [9] A. Bovet and H. A. Makse, "The influence of fake news in social media on elections: The case of the 2016 us presidential election," *Nature Communications*, vol. 10, no. 1, pp. 1–14, 2019.
- [10] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi, "The spreading of misinformation online," *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 554–559, 2016.
- [11] S. Kemp, "Digital 2021: Global overview report," *We Are Social*, 2021. Accessed: 2023-05-24.
- [12] C. Silverman and J. Singer-Vine, "The next great fake news battleground is messaging apps," *BuzzFeed News*, 2016. Accessed: 2023-05-24.

- [13] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.
- [14] M. Cinelli, W. Quattrocioni, A. Galeazzi, C. M. Valensise, E. Brugnoti, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The covid-19 social media infodemic," *Scientific Reports*, vol. 10, no. 1, pp. 1–10, 2020.
- [15] P. Zhou and R. Zafarani, "Fake news detection via nlp is challenging," *ArXiv*, 2018.
- [16] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, "Multimodal fusion with recurrent neural networks for rumor detection on social media," in *Proceedings of the 2017 ACM on Multimedia Conference*, pp. 795–816, 2017.
- [17] D. Khattar, M. Goud, V. Gupta, and M. Varma, "Mvae: Multimodal variational autoencoder for fake news detection," in *Proceedings of the 2019 World Wide Web Conference*, pp. 2915–2921, 2019.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, pp. 5998–6008, 2017.
- [19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2019.
- [20] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," *OpenAI*, 2018.
- [21] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter," *arXiv preprint arXiv:1910.01108*, 2019.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.
- [23] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," in *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.
- [24] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, pp. 2951–2959, 2012.
- [25] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2623–2631, 2019.

- [26] D. M. Powers, "Evaluation: from precision, recall and f-measure to roc, informedness, markedness, and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.
- [27] Y. Sasaki, "The truth of the f-measure," in *IEICE TRANSACTIONS on Communications*, vol. E91-B, pp. 1074–1077, 2007.
- [28] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.
- [29] S. Kalra, C. H. S. Kumar, Y. Sharma, and G. S. Chauhan, "Multimodal fake news detection on fakeddit dataset using transformer-based architectures," in *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, pp. 281–292, Springer, 2022.
- [30] S. Alonso-Bartolome and I. Segura-Bedmar, "Multimodal fake news detection," *arXiv preprint arXiv:2112.04831*, 2021.
- [31] I. Segura-Bedmar and S. Alonso-Bartolome, "Multimodal fake news detection," *Information*, vol. 13, no. 6, p. 284, 2022.
- [32] C. Comito, L. Caroprese, and E. Zumpano, "Multimodal fake news detection on social media: a survey of deep learning techniques," *Social Network Analysis and Mining*, vol. 13, no. 1, p. 101, 2023.
- [33] S.-I. Papadopoulos, C. Koutlis, S. Papadopoulos, and P. C. Petrantonakis, "Verite: a robust benchmark for multimodal misinformation detection accounting for unimodal bias," *International Journal of Multimedia Information Retrieval*, vol. 13, no. 1, p. 4, 2024.
- [34] C. RAJ, *Deep Neural Networks Towards Multimodal Information Credibility Assessment*. PhD thesis, Delhi college of Engineering, 2021.
- [35] A. Yadav and A. Gupta, "An emotion-driven, transformer-based network for multimodal fake news detection," *International Journal of Multimedia Information Retrieval*, vol. 13, no. 1, pp. 1–16, 2024.
- [36] S. Sengan, S. Vairavasundaram, L. Ravi, A. Q. M. AlHamad, H. A. Alkhazaleh, and M. Alharbi, "Fake news detection using stance extracted multimodal fusion-based hybrid neural network," *IEEE Transactions on Computational Social Systems*, pp. 1–12, 2023.
- [37] S.-I. Papadopoulos, C. Koutlis, S. Papadopoulos, and P. C. Petrantonakis, "Figments and misalignments: A framework for fine-grained crossmodal misinformation detection," *arXiv preprint arXiv:2304.14133*, 2023.

- [38] L. Qian, R. Xu, and Z. Zhou, "Mrdca: A multimodal approach for fine-grained fake news detection through integration of roberta and densenet based upon fusion mechanism of co-attention," *Annals of Operations Research*, pp. 1–22, 2022.
- [39] B. Škrlj, M. Bevec, and N. Lavrač, "Multimodal automl via representation evolution," *Machine Learning and Knowledge Extraction*, vol. 5, no. 1, pp. 1–13, 2022.
- [40] J. Xie, S. Liu, R. Liu, Y. Zhang, and Y. Zhu, "Sern: Stance extraction and reasoning network for fake news detection," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2520–2524, 2021.
- [41] A. M. Luvembe, W. Li, S. Li, F. Liu, and X. Wu, "Caf-odnn: Complementary attention fusion with optimized deep neural network for multimodal fake news detection," *Information Processing & Management*, vol. 61, no. 3, p. 103653, 2024.
- [42] G. Wang, L. Tan, Z. Shang, and H. Liu, "Multimodal dual emotion with fusion of visual sentiment for rumor detection," *Multimedia Tools and Applications*, pp. 1–22, 2023.
- [43] K. Arunthavachelvan, E. Hasan, C. Ding, and S. Raza, "Pl-ncc: A novel approach for fake news detection through data augmentation," 2024.
- [44] A. M. Luvembe, W. Li, S. Li, F. Liu, and G. Xu, "Dual emotion based fake news detection: A deep attention-weight update approach," *Information Processing & Management*, vol. 60, no. 4, p. 103354, 2023.
- [45] L. Tan, G. Wang, F. Jia, and X. Lian, "Research status of deep learning methods for rumor detection," *Multimedia Tools and Applications*, vol. 82, no. 2, pp. 2941–2982, 2023.
- [46] S. K. Uppada and P. Patel, "An image and text-based multimodal model for detecting fake news in osn's," *Journal of Intelligent Information Systems*, vol. 61, no. 2, pp. 367–393, 2023.
- [47] L. D. Sciuca, M. Mamei, E. Balloni, L. Rossi, E. Frontoni, P. Zingaretti, and M. Paolanti, "Fakened: A deep learning based-system for fake news detection from social media," in *International Conference on Image Analysis and Processing*, pp. 303–313, Springer, 2022.
- [48] Z. Huang, Y. Hu, Z. Zeng, X. Li, and Y. Sha, "Multimodal stacked cross attention network for fine-grained fake news detection," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 2837–2842, IEEE, 2023.
- [49] Z. Zeng, M. Wu, G. Li, X. Li, Z. Huang, and Y. Sha, "Correcting the bias: Mitigating multimodal inconsistency contrastive learning for multimodal fake news detection," in *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 2861–2866, IEEE, 2023.

- [50] S. Abdali, "Multi-modal misinformation detection: Approaches, challenges and opportunities," *arXiv preprint arXiv:2203.13883*, 2022.
- [51] K. Xuan, L. Yi, F. Yang, R. Wu, Y. R. Fung, and H. Ji, "Lemma: Towards lvm-enhanced multimodal misinformation detection with external knowledge augmentation," *arXiv preprint arXiv:2402.11943*, 2024.
- [52] Z. Yi, S. Lu, X. Tang, J. Wu, and J. Zhu, "Maccn: Multi-modal adaptive co-attention fusion contrastive learning networks for fake news detection," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6045–6049, IEEE, 2024.
- [53] M. Bilal, M. Moetesum, and I. Siddiqi, "Online content veracity assessment using deep representation learning," in *2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, pp. 325–330, 2022.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Appendices

List of Appendices

Example Appendix

90

Appendix A Example Appendix

A.1 Code Repository

The whole implementation of the fake news detection system created for this thesis is available for access on GitHub. The repository contains all the scripts, data preparation stages, model training, and assessment techniques employed in this research. Enthusiastic readers and researchers can examine the code to acquire a more profound comprehension of the approaches employed and perhaps reproduce or expand upon this work. The repository may be accessed by the provided link:

<https://github.com/SamiulAlamSiam/FakeNewsClassifying>

A.2 Most Common Words

A.2.1 Common True Words

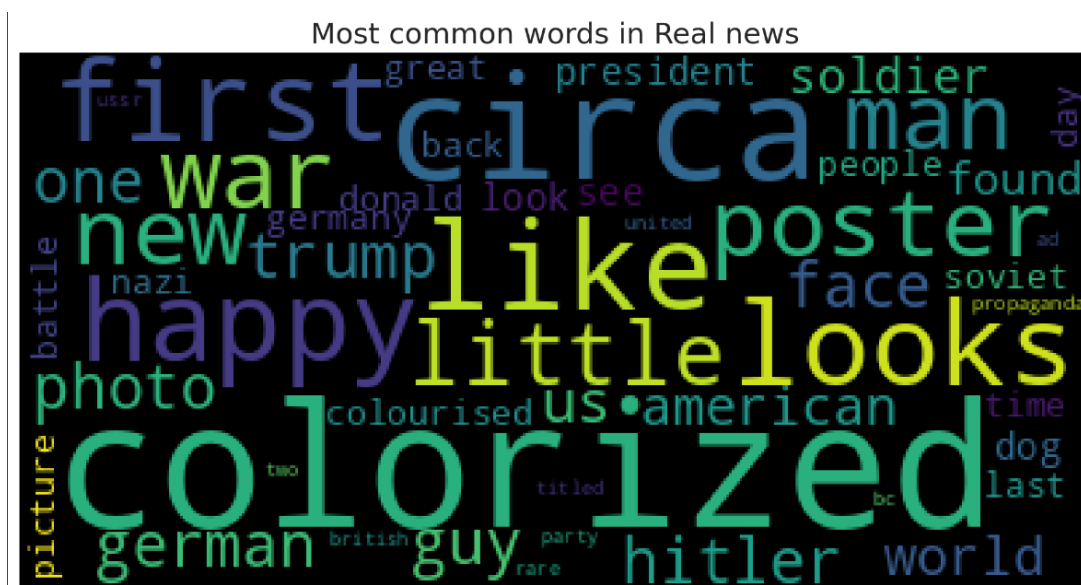


Figure 12: Common True Words

The word cloud in the appendix showcases the outcome of applying the KMeans clustering model on textual data. This graphic illustrates the prevailing terms seen in authentic news stories. The phrases "first," "circa," "colorized," and "war" are strongly included in the dataset, suggesting that they are used frequently. Additional noteworthy phrases encompass "joyful," "novel," "billboard," "small," and "president," which exemplify prevalent themes and subjects seen in genuine news articles. The magnitude of each word in the cloud correlates to its frequency, offering a distinct and instinctive depiction of prominent phrases in the actual news collection. This study facilitates comprehension

of the language patterns and regions of emphasis in authentic news pieces, which may be compared to those in false news to enhance detection algorithms.

A.2.2 Common Fake Words



Figure 13: Common Fake Words

The appendix contains a word cloud created by applying the KMeans clustering methodology to textual data. This word cloud highlights the most frequently occurring terms in false news stories. The phrase "psbattle" is prominently shown, highlighting its frequent occurrence in the dataset. Additional noteworthy terms encompass "discovered," "similar to," "conflict," "plant," and "feline," which frequently correlate with deceptive or sensationalized material. The magnitude of each word in the cloud corresponds to its frequency, providing a lucid and instinctive portrayal of prominent phrases in the fabricated news collection. This research offers a valuable understanding of the linguistic patterns and recurring themes that are frequently employed in fabricated news, which may be juxtaposed with those found in authentic news to improve the effectiveness of detection algorithms. Through comprehending these subtle distinctions in language, we may formulate more efficient approaches for recognizing and minimizing the dissemination of false information.

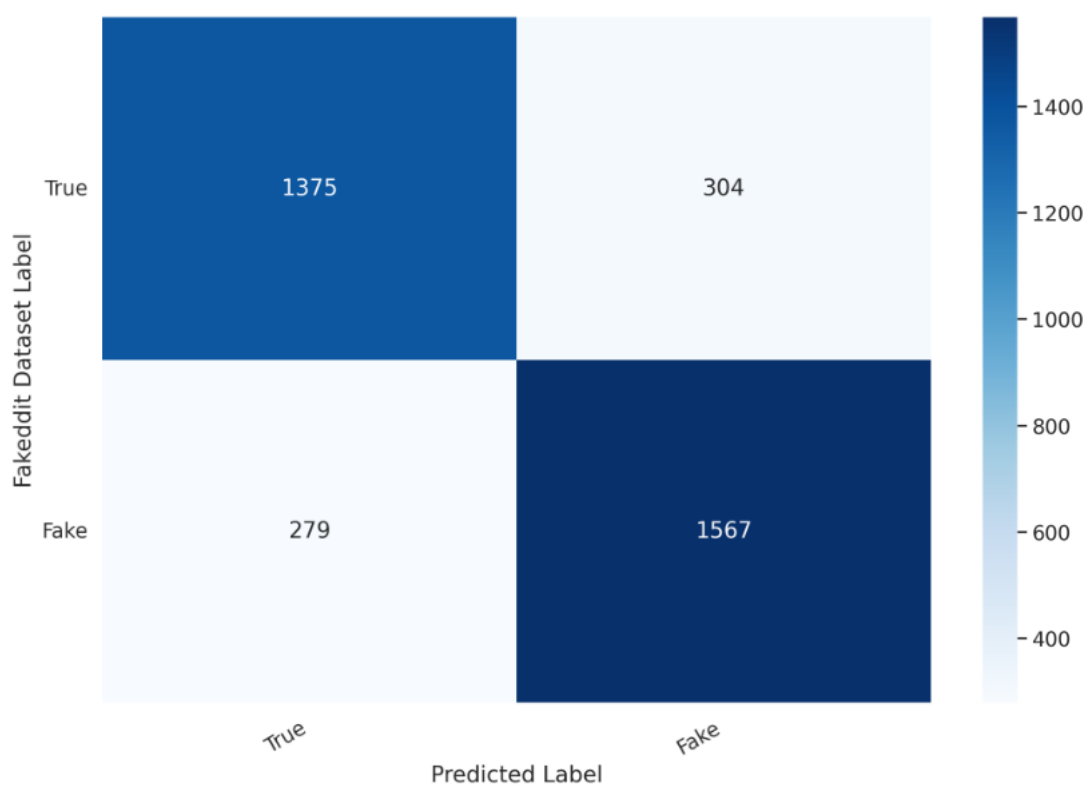


Figure 14: Confusion matrix with DistilBERT

A.3 Unimodal

A.3.1 Text Unimodal

A.3.1.1 DistilBERT

The confusion matrix (see Figure 14) of the DistilBERT model, when applied to text data from the Fakeddit dataset, provides a clear representation of the model’s accuracy in distinguishing between real and fake news. The confusion matrix reveals that among 1,679 examples of factual news, the program accurately recognized 1,375 cases, while erroneously categorizing 304 occurrences as false news. In contrast, among a total of 1,846 instances of bogus news, 1,567 were correctly categorized, while 279 were mistakenly identified as factual news.

The performance of the DistilBERT model indicates its effectiveness in accurately differentiating between authentic and fabricated news, with a greater proportion of accurate predictions (true positives and true negatives) compared to wrong ones. The minimal occurrence of false positives (genuine news mistakenly identified as fake) and false negatives (fake news mistakenly identified as genuine) demonstrates a well-balanced and resilient capacity to classify, which is crucial for tasks that need accurate differentiation between authentic and fabricated news articles. The overall findings

underscore the model’s capacity to effectively apply learned knowledge to a wide range of data, rendering it a powerful instrument for detecting text-based false information.

	Precision	Recall	F1-score	Support
True	0.83	0.82	0.83	1679
Fake	0.84	0.85	0.84	1846
accuracy			0.83	3525
macro avg	0.83	0.83	0.83	3525
weighted avg	0.83	0.83	0.83	3525

Table 10: Classifying report on using DistilBERT

The performance metrics table (see Table 10) for the DistilBERT model, applied just to textual data from the Fakeddit dataset, offers a comprehensive perspective on the model’s efficacy. The accuracy in recognizing real news material is 0.83, meaning that 83% of the information categorized as true is accurately detected. The accuracy for identifying bogus news text is somewhat higher at 0.84, indicating a well-balanced ability to differentiate between real and fake news. The recall, which quantifies the model’s capacity to correctly identify all pertinent occurrences, is 0.82 for real news and 0.85 for false news, indicating that the model is very proficient at recovering texts from both categories. The F1-score, calculated as the harmonic mean of accuracy and recall, is 0.83 for factual news and 0.84 for fraudulent news, indicating the model’s balanced and reliable performance.

The DistilBERT model has an overall accuracy of 0.83, meaning that it accurately identifies 83% of the text input. The macro average, which computes the mean of accuracy, recall, and F1-score over both classes without taking into account class imbalance, is similarly 0.83. The weighted average, which considers the support (number of true instances for each class), reflects this consistency and demonstrates the model’s consistent performance across many criteria.

The support values suggest that there are 1,679 instances of actual news texts and 1,846 instances of false news texts in the dataset, which creates a balanced foundation for evaluating the metrics. This investigation showcases the efficacy of the DistilBERT model in accurately discerning between authentic and fabricated news texts. It exhibits robust accuracy, recall, and F1-scores in both categories, establishing it as a dependable option for detecting false news based on text.

A.3.1.2 BERT

	Precision	Recall	F1-score	Support
True	1.00	1.00	1.00	8909
Fake	1.00	1.00	1.00	9744
accuracy			1.00	18653
macro avg	1.00	1.00	1.00	18653
weighted avg	1.00	1.00	1.00	18653

Table 11: Classifying report on using BERT

The performance metrics table (see Table 11) for the BERT model, when applied to text data from the Fakeddit dataset, exhibits remarkably high efficacy. The model attained a flawless score in all assessment criteria, including precision, recall, and F1-score. Each metric recorded a value of 1.00 for both real and false news classifications. BERT demonstrated perfect accuracy in its predictions, properly classifying all 8,909 instances of real news and all 9,744 cases of fake news.

The findings demonstrate the model's capacity to accurately differentiate between genuine and fabricated news articles, obtaining a flawless classification rate of 100%. The utilization of macro and weighted averages further highlights the flawless scores, emphasizing the uniformity of BERT's performance throughout the dataset. The results indicate that the BERT model is extremely dependable for identifying false news in text, offering precise and consistent categorization. This is essential for upholding the credibility of information sharing on social media platforms. This degree of effectiveness is a notable advancement in utilizing sophisticated transformer models for the automatic identification of bogus news.

A.3.1.3 KMeans

The graph (see Figure 15) depicts the distribution of the mean number of words per phrase in both fabricated and genuine news stories. The research was performed utilizing the KMeans clustering technique to distinguish between the two groups. The x-axis displays the mean number of words per phrase, while the y-axis reflects the probability distribution of these means within each group.

Based on the histogram, it is clear that both false and authentic news mostly contain sentences with an average length of 5 to 15 words. The highest likelihood for both groups is seen at roughly 10 words per phrase. Nevertheless, there exist nuanced variations in the distribution patterns. The actual news stories (shown by yellow bars) exhibit a little greater prevalence of shorter phrases, as evidenced by

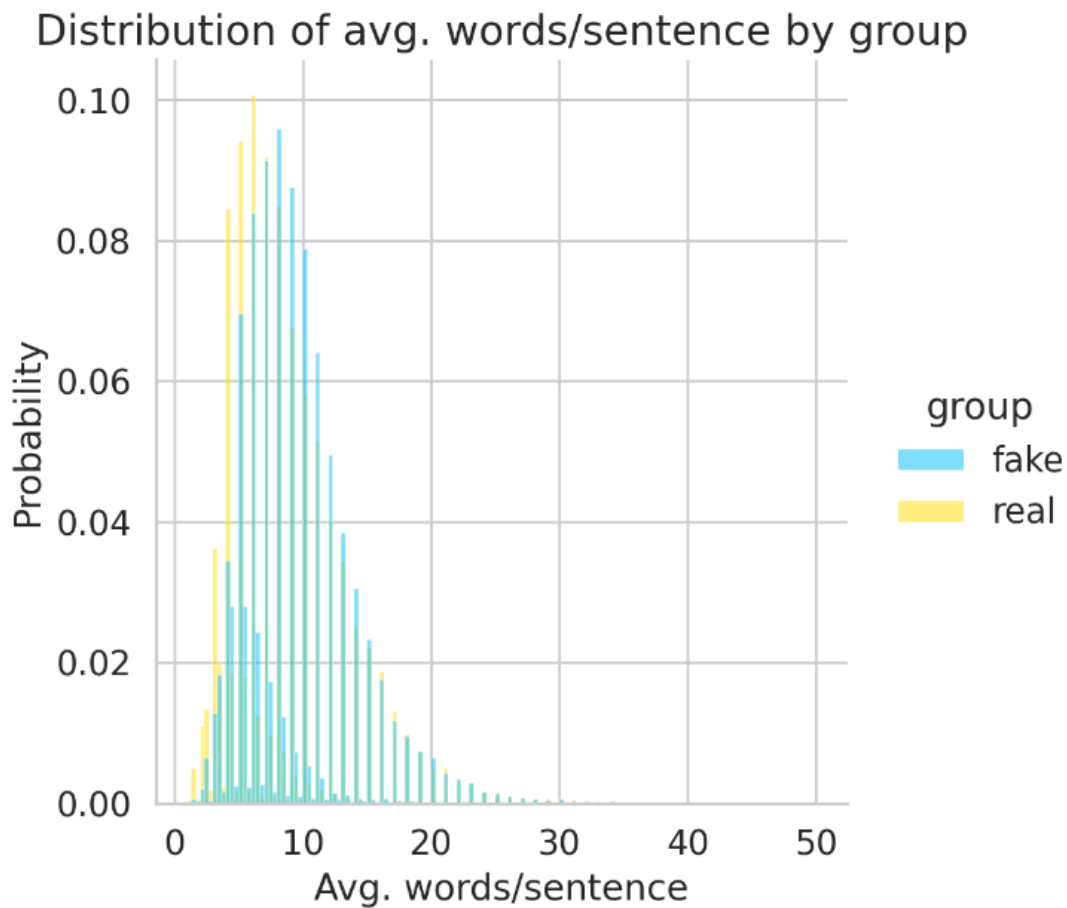


Figure 15: Avg. Words/Sentence

the increased probability density towards the lower values on the x-axis. Conversely, the counterfeit news stories (shown by blue bars) have a more dispersed pattern, characterized by a prominent elongation towards greater word counts per phrase.

The disparity indicates that authentic news articles often employ shorter, more succinct phrases, whereas fabricated news articles may have more diversity in sentence length, and occasionally include lengthier sentences. Comprehending these patterns is crucial for enhancing the accuracy of classification algorithms used in fake news detection models. This is because it brings attention to stylistic variations that may be utilized to improve the performance of the algorithms.

The provided box plot (see Figure 16) demonstrates the distribution of the average number of words per phrase for both false and actual news items. This analysis was conducted using the KMeans clustering method. The x-axis reflects the classification of groups as either false or real, while the y-axis indicates the average amount of words per phrase.

Both groups have comparable median values, around 10 words per phrase, as illustrated by the mid-

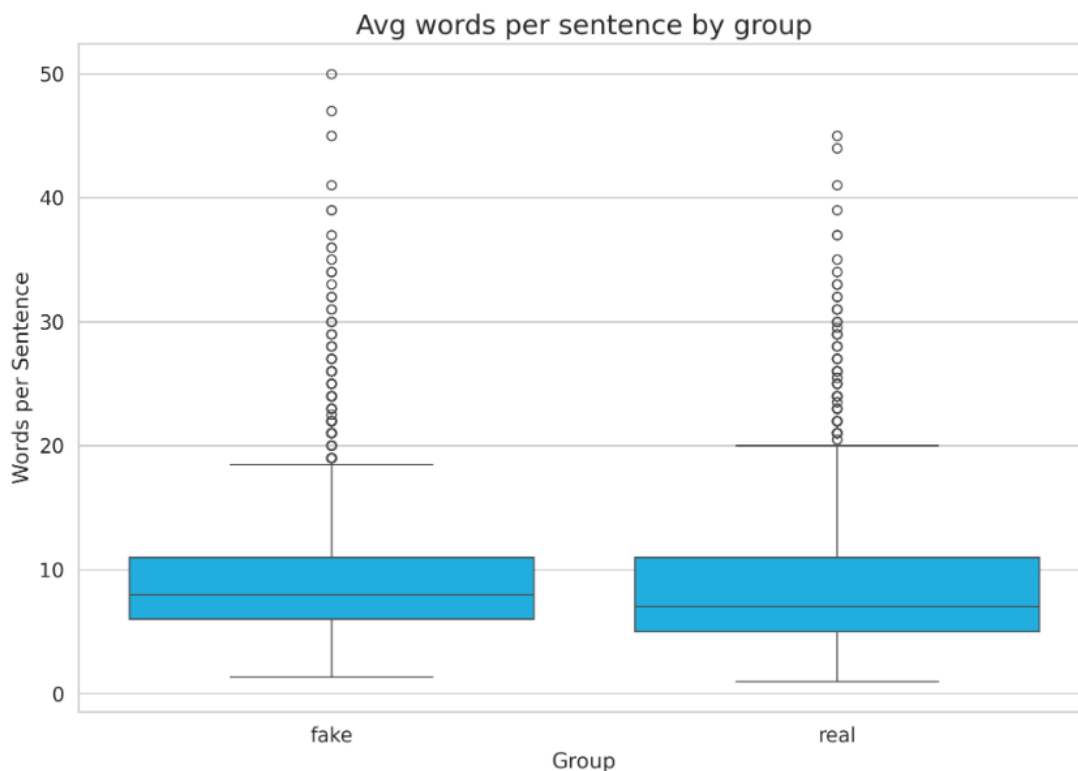


Figure 16: Avg. Words/Sentence By Group

dle line within each box. The boxes, which represent the interquartile range (IQR) between the 25th and 75th percentile, indicate that the majority of sentences in both false and actual news items are typically between 7 and 12 words in length.

Nevertheless, there are discernible disparities in the distribution and extreme values. The whiskers, emanating from the boxes, indicate that both groups possess a substantial amount of phrases surpassing the top quartile, with genuine news exhibiting a somewhat wider span. In addition, authentic news pieces have a higher number of outliers, with sentence lengths extending up to 50 words. This suggests a larger degree of variety and the occurrence of really lengthy phrases, distinguishing them from false news.

This research emphasizes that although both false and real news generally exhibit comparable average sentence lengths, real news exhibits greater variability in sentence length, occasionally incorporating significantly longer phrases. By examining the changes in sentence structure, we may gain valuable insights that can be used to improve false news detection algorithms. Specifically, we can focus on the stylistic variations between the two groups.

The scatter graphs (see Figure 17) above illustrate the outcomes of using the KMeans clustering algorithm on a slice of text data and comparing it to the true class labels. The x and y axes indicate the

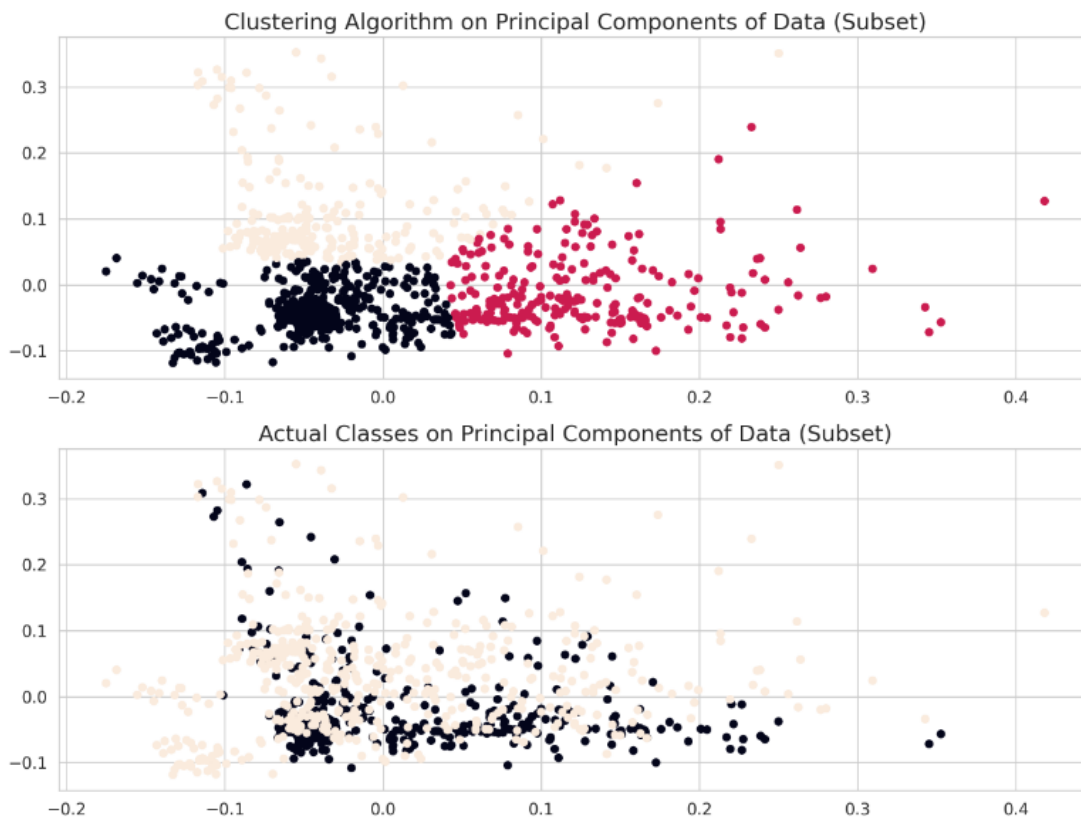


Figure 17: KMeans Clustering

principle components of the data, which are the main characteristics obtained by principle Component Analysis (PCA) to decrease the number of dimensions.

The data points in the top plot are assigned different colors according to the clusters determined by the KMeans algorithm. The clustering algorithm successfully partitioned the data into several groups, demonstrating its capability to detect underlying patterns within the text data. The distinctiveness of the clusters indicates that KMeans algorithm is capable of discerning several categories of data points, albeit with some instances of overlapping.

The lower plot displays the factual classifications of the data points, once again shown by distinct colors. This figure functions as a reference point for assessing the effectiveness of the KMeans clustering algorithm. By juxtaposing the clustering outcomes with the real categories, one may evaluate the precision and efficiency of the clustering process.

Based on the comparison, it is clear that although the KMeans clustering method has successfully identified separate clusters, it does not precisely match the actual distribution of classes. There are inconsistencies, namely in the areas where the clusters intersect. This suggests that although KMeans is capable of detecting some patterns in the data, there is still potential for enhancing the accuracy

of text data classification.

In summary, this visual comparison effectively showcases the strengths and weaknesses of KMeans clustering in detecting patterns in text data. This serves as a basis for enhancing classification algorithms to enhance precision.

A.3.2 Image Unimodal

A.3.2.1 ResNet-34

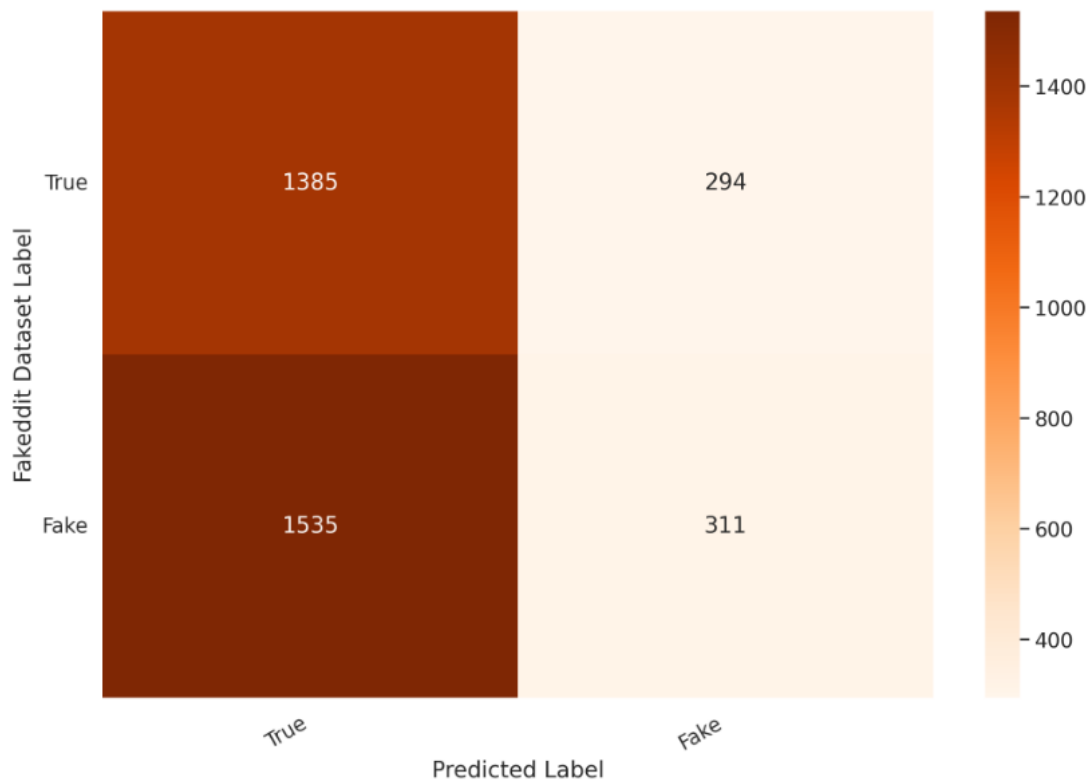


Figure 18: Confusion matrix with ResNet34

The confusion matrix (see Figure 18) provided demonstrates the efficacy of the ResNet34 model in detecting false news when applied to picture data from the Fakeddit dataset. The matrix displays the quantities of true positives, true negatives, false positives, and false negatives. More precisely, the algorithm accurately classified 1,385 photos as actual news (true positives) and 311 images as fake news (true negatives). Nevertheless, it erroneously categorized 294 authentic news photographs as counterfeit (false positives) and 1,535 counterfeit news images as authentic (false negatives). This highlights a notable difficulty in effectively differentiating between authentic and counterfeit news using only picture data, emphasizing the need to strengthen the model's accuracy through improved feature extraction or integration with other data modalities. The significant prevalence of false neg-

atives, wherein misinformation is erroneously categorized as accurate, is particularly worrisome and highlights the necessity of enhancing the model to mitigate these inaccuracies.

	Precision	Recall	F1-score	Support
True	0.50	0.94	0.94	1679
Fake	0.53	0.88	0.66	1846
accuracy			0.52	3525
macro avg	0.52	0.51	0.43	3525
weighted avg	0.52	0.52	0.44	3525

Table 12: Classifying report on using Resnet-34

The performance metrics table (see Table 12) for the ResNet34 model, applied just to picture data from the Fakeddit dataset, provides valuable insights into the model's efficacy. The accuracy for recognizing real news photographs is 0.50, meaning that only 50% of the images categorized as true are accurately detected. In contrast, the accuracy for counterfeit news photos is somewhat greater at 0.53. The recall, which quantifies the model's capacity to accurately recognize all pertinent instances, is particularly elevated for real news at 0.94, indicating that the model effectively detects the majority of true news photos. Nevertheless, the fake news recall rate is at 0.88, which, although commendable, suggests the presence of some incorrect rejections. The F1-score, which is calculated as the harmonic mean of precision and recall, is 0.66 for false news. This figure accurately represents the balance between precision and recall. The model's overall accuracy is 0.52, indicating that it properly identifies just over 50% of the photos. The accuracy, recall, and F1-score macro and weighted averages consistently indicate a modest performance across classes, with values around 0.52. The support values reveal that there are 1,679 authentic news photographs and 1,846 fabricated news images in the dataset, which serves as a significant foundation for these measures. This investigation indicates that the ResNet34 model has satisfactory recall, especially for factual news. However, its overall precision and accuracy reveal areas that require enhancement, maybe through improved feature extraction or the incorporation of multi-modal data.

A.3.3 Analysis and Comparison

The table (see Table 13) displays the unimodal classification outcomes obtained from the Fakeddit dataset using various models and setups. Three models, including KMeans, DistilBERT, and BERT, were utilized for text data. Each model was optimized using AdamW and employed Softmax as the loss function. The KMeans model demonstrated its efficacy in clustering text data for false news identification by achieving an accuracy of 93%. DistilBERT, a more efficient and quicker iteration of BERT,

Dataset	Get Data	Function	Model	My Output
Fakeddit	Text Data	Optimizer: AdamW, Loss_function: Softmax	KMeans	93%
Fakeddit	Text Data	Optimizer: AdamW, Loss_function: Softmax	DistilBERT	83%
Fakeddit	Text Data	Optimizer: AdamW, Loss_function: Softmax	BERT	100%
Fakeddit	Image	Optimizer: AdamW, Loss_function: Softmax	ResNet34	52%

Table 13: Unimodal Classification Results

attained an accuracy of 83%, demonstrating a robust performance while somewhat less successful than KMeans. BERT, renowned for its exceptional text processing skills, attained a flawless accuracy of 100%, demonstrating its resilience in effectively managing textual material for the categorization of bogus news.

On the other hand, when the ResNet34 model was used with picture data, the resulting accuracy was just 52%. This implies that although ResNet34 is a robust model for classifying images, it may not be as efficient at detecting false news solely based on picture data. The significant disparity in performance between the text and picture models underscores the intricacy and difficulties linked to unimodal false news identification. When complex models such as BERT are used to evaluate text data, the results demonstrate impressive accuracy. However, relying just on picture data, even with sophisticated models like ResNet34, may not offer enough discriminating capability to reliably detect bogus news. The results highlight the possible necessity for multimodal techniques that integrate both text and picture data in order to enhance the overall classification performance.