Fake News Detection and Classification with Multimodal Learning

Abstract

Fake news has become a major problem affecting many aspects of society, including politics, health, and the economy. Misinformation detection has become essential to reduce the impact of fake news. This study provides a comprehensive comparison of unimodal and multimodal approaches for fine-grained classification of fake news. I used the Fakeddit dataset and state-of-the-art deep learning techniques such as SVM, BiLSTM, BERT, and multimodal CNN. My results showed that the multimodal approach outperforms the unimodal approach that utilises text only. Furthermore, BERT achieved the highest accuracy in text classification, while GloVe dynamic word embeddings outperformed random initialization in CNN and BiLSTM architectures. Finally, my multimodal CNN performed best with 83% accuracy, demonstrating the effectiveness of combining text and image data for detecting fake news. These results demonstrate the potential for developing sophisticated tools to reliably and quickly detect fake news. Future research could consider including other modalities such as audio and video to improve the performance of the fake news detection model.

1. Introduction

As society becomes increasingly digitised, the proliferation of fake news, particularly on social media, has become a significant contributor to various problems within society. In recent years, the effects of this kind of misinformation[1] and disinformation[2] spreading can be seen in various parts of our society on climate change, global pandemic, vaccine distribution, racism, and politics, etc(Finneman & Thomas, 2018). Therefore, I aimed to build multimodal machine learning models to detect and categorise online fake news, which usually contains both images and texts.

According to the 2022 Pew Research Center survey, 95% of U.S. adults get their news from a digital device (e.g., smartphones, or computers) and 71% of U.S. adults get news from social media.

During recent years, fake news does not exist only in text form, but increasingly also includes both images and video with the original post. The majority of work has been focused on textual information only, such as unimodal approaches. Less research has been exploring multimodal approaches (for example Kumari & Ekbal, 2021), which use both texts and images to detect the fake news, acquiring better results than the unimodal approaches. Therefore, the goal of these studies address the problem of fake news detection as a binary classification task (that is, including classifying news as either true or fake). However, the most important

aim of this paper is to examine both unimodal and multimodal approaches to deal with a fine-grained classification of fake news.

To start, I use the Fakeddit dataset (Nakamura et al., 2020), made up of posts from Reddit consisting of about 800,000 samples from multiple categories of fake news which use both unimodal and multimodal approaches. If a post is detected as fake news, it is classified into one of six categories: true, misleading content, manipulated content, false connection, imposter content, and satire.

I examine different deep learning architectures for text classification such as Bidirectional long short-term memory (BiLSTM) (Hochreiter & Schmidhuber, 1997) and Bidirectional encoder representations from transformers (BERT) (Devlin et al., 2018). As a multimodal approach, I propose a CNN architecture that combines both texts and images to classify the fake news.
The main contributions of this paper are:
1. I propose a multimodal approach to detect and classify fake news using both texts and images.
2. I compare the performance of different deep learning architectures for text classification such as BiLSTM and BERT.

[1] means false or inaccurate information.
[2] means false information which is intended to mislead.

2. Related work

In the second decade of the current century, is really a revival of the concept of neural networks, a great number of various applications of deep learning techniques have emerged. Plenty of Natural Language Processing (NLP) advances are expected for the incorporation of deep neural network approaches[1,2].

Fake News Detection(Veronica Pérez-Rosas et al. (2018)) using binary fake news classification to achieve accuracies of 76% by using linear SVM classifiers on data annotated with linguistic features.

Recent work by Hansen et al. (2021) has focused on developing a model for assessing the accuracy of political claims based on the evidence provided. This model evaluates fake news by considering the reasoning behind it, specifically by examining the claims and the evidence presented to support them. Therefore, this research aims to understand how models can judge fake news based on the reasoning behind it, considering the claims and associated evidence.

Earlier research [3,4] has focused on studying the image features of data visualisations, such as the type of image and accompanying images. Jin et al. [5] fused text information with image data and used attention mechanisms with RNN on the image and LSTM on the text and social context to extract features and perform rumour detection on microblogs. This approach demonstrated improved detection results compared to using image or text information alone.

The key reason behind the multimodal approaches is that the majority of texts are accompanied by images, and the images may provide beneficial information to improve the results of the classification task (Baheti, 2020).

Currently, I am focusing on recent research on fake news detection, examining both unimodal and multimodal approaches.

2.1. Fake news detection-Unimodal approaches

Earlier research, a variety of machine learning algorithms were applied for fake news detection [6]. Therefore, comparing all algorithms was necessary. Dataset was tested using Support Vector Machines, Stochastic Gradient Descent, Gradient Boosting, Bounded Decision Trees, and Random Forests.

Detecting fake news classification uses a capsule network model (Goldani et al. 2021) built on CNN and pre-trained word embeddings over the ISOT (Ahmed et al., 2017) and LIAR (Wang, 2017) datasets. The ISOT dataset consists of a collection of fake and true news samples from Reuters and Kaggle, but the LIAR dataset is made up of short articles which they are classified into six classes: pants-fire, false, barely-true, half-true, mostly-true and true. Therefore, they used both binary and multi-class fake news classification. The best accuracy accomplished with the proposed model is 99.8 % for the ISOT dataset, binary classification and 39.5 % for the LIAR dataset, multi-class classification.

The LIAR dataset (Girgis et al. 2018) carried out fake news classification using three different models: vanilla Recurrent Neural Network, Gated Recurrent Unit (GRU) and LSTM. The GRU model achieves an accuracy of 21.7%, quite a better performance than the LSTM (21.66 %) and the vanilla RNN (21.5 %) models.

In my review, I found that deep learning architectures are effective for binary classification of fake news, but their performance for fine-grained classification is lower. BERT has shown strong results in various text classification tasks and has specifically been used for multi-classification of fake news. However, my research suggests that using multimodal approaches, combining both text and image data, can significantly improve the performance of fake news detection.

2.2. Fake news detection-Multimodal approaches

Suggest a model by Giachanou et al. (2020) to implement multimodal classification of news samples as either true or fake. In order to the BERT model (Devlin et al., 2018) is applied to acquire textual representations. The authors implement the VGG(Visual Geometry Group) network and use 16 layers followed by a LSTM layer and a mean pooling layer for the visual features. The authors use the dataset which is retrieved from the FakeNewsNet collection (Shu et al., 2020). Conclusively, the authors use 2,745 fake news and 2,714 real samples collected from the GossipCop posts. Therefore, the model achieves an F1-score of 79.55 %.

The other authors (Kirchknopf et al. 2021) use binary classification of fake news over the Fakeddit dataset to perform four dissimilar modalities of data. More specifically, the authors utilise the textual content of the news which is related to comments, the images and the remaining metadata belonging to other modalities. The good accuracy achieved is 95.5%.
Author Kaliyar et al. (2020) use a binary classification of fake news for the DeepNet model. This model is built of one embedding layer, three convolutional layers, one LSTM layer, seven dense layers, ReLU for activation and finally the softmax function for the binary classification. The model is evaluated on the Fakeddit and BuzzFeed (Kaggle, a) datasets. The BuzzFeed dataset include news articles collected during the U.S. election and they are classified as either true or fake. The model achieves an accuracy of 86.4% on the Fakeddit dataset (binary classification) and 95.2% over the BuzzFeed dataset.

Research by Bahad et al. [7] proposed on fake news detection using GloVe pre-trained word embedding. It combines it with several deep learning architectures such as CNN, Recurrent Neural Network (RNN), Unidirectional Long Short-Term Memory (LSTM), and Bidirectional LSTM. The results obtained from these studies were varied. Research results with a value of 98.75% accuracy using the Bidirectional LSTM. The study also tested it using the Fake or Real News Dataset[8] and obtained 91.48% accuracy using Unidirectional LSTM.

My conclusion from this review, most multimodal approaches that were assessed only for the binary classification of fake news on the Fakkeddit dataset. Only a few studies (such as Kang et al., 2021) have explored the multi-class classification of fake news using a reduced version of this dataset. In contrast, the approach proposed in this research utilises a deep convolutional network, while some other multimodal approaches have achieved similar performance using a simpler CNN.

3. Methods

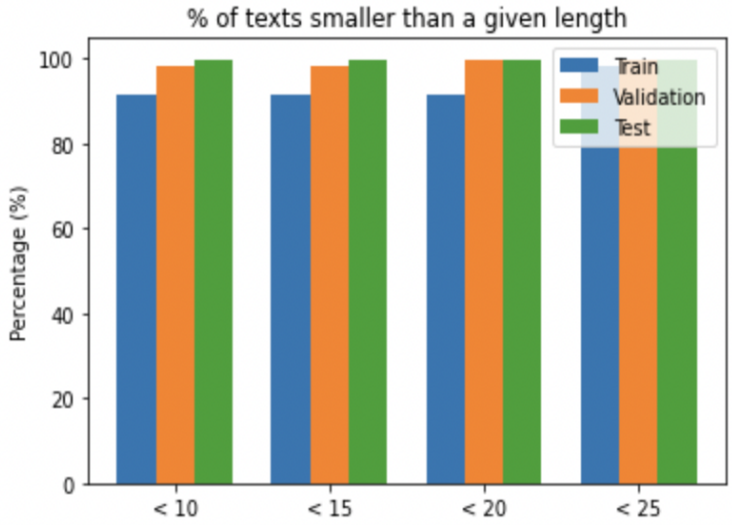
3.1. Unimodal approaches
Three unimodal models only using the texts are proposed: BiLSTM
and BERT.

3.1.1. Preprocessing for deep learning models

Preprocess corpus texts by removing stop words, punctuation, numbers, and extra spaces. Each text is then tokenized, part-of-speech tagged, lemmatized using the WordNetLemmatizer, and then converted to a series of integers. This conversion is accomplished by first learning the vocabulary of the corpus and building a dictionary that maps each word to a unique integer. The resulting string of integers preserves the original order of the words in the text.

Pads and truncates a sequence of integers so that all inputs to a deep learning model have the same length. This causes some information to be lost when the sequence breaks. To determine the optimal length, I calculated the percentage of text under 10, 15, 20, and 25 characters and found that 98% of the text was under 15 characters. For this reason, I chose to use a length of 15 tokens after padding and truncation(Figure 1).

Figure 1: % of texts smaller than a given length



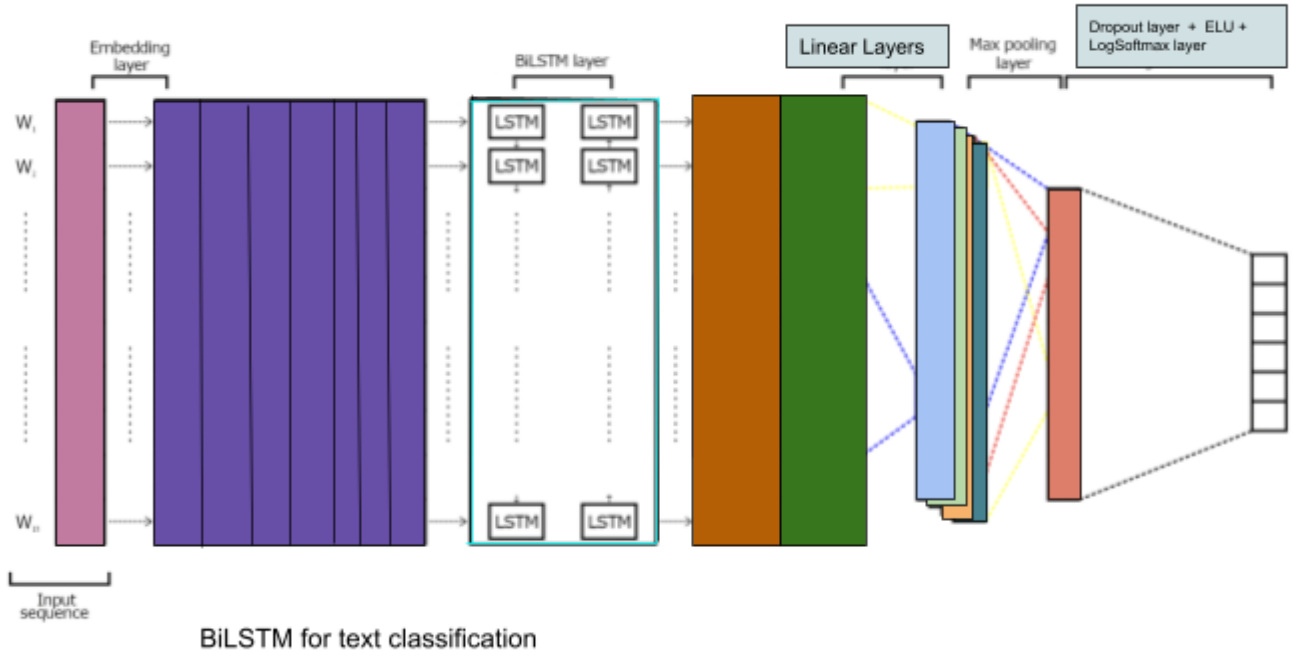
A deep learning model takes a sequence of word embeddings as input and an embedding layer transforms each integer value in the sequence into a vector of word embeddings. This produces a 15-by-300 matrix (where 300 is the dimension of the word embeddings) for each vectorized document. I use both randomly initialised Grove word embeddings and pre-trained Grove word embeddings(Pennington et al., 2014)., with a dynamic approach (which allows the model to continue training word embeddings) and a static approach (which allows the model to be prohibited from training).

3.1.2. BiLSTM

The BiLSTM architecture for fake news text classification uses an embedding layer as the first layer. The embedding matrix is initialised using both random initialization and pretrained 300-dimensional GloVe word embeddings. The choice of 300-dimensional embedding over

small dimensions (such as 50, 100, or 200) is based on research showing that larger dimensions lead to better performance (Patel & Bhattacharyya, 2017).

Embedding layer is initialised with a pretrained embedding matrix of 300 as input size. Embedding weights are set as trainable if train_parameters is set to True, otherwise frozen.
The BiLSTM layer takes the embedded input and processes it using a two-layer bidirectional LSTM network. The hidden size of each LSTM unit is fixed at 100.
Applies a max pooling layer to the output of the BiLSTM layer to reduce the length of the output sequence and summarise information from the entire sequence.
The output of the max pooling layer passes through two linear layers (self.linear1 and self.linear2) between which activation functions (ELU and log-softmax) are applied, the resulting vector is fed into the logsoftmax function to generate the predicted classes. The output size of the first linear level is 64 and the output size of the second linear level is equal to the number of labels (labels = 6). A dropout layer is added between two linear layers to prevent overfitting. Early stopping is employed to determine the optimal number of epochs.



BiLSTM for text classification

3.1.3. BERT

Instead of randomly initialised grove embeddings, vectors provided by BERT are now used to represent input tokens. BERT, unlike GloVe, takes into account the context of each word (Pennington et al., 2014). Text preprocessing follows the same steps as above, but uses the BertTokenizer class from the Transformers library (Face) for tokenization. This is because it has its own vocabulary and eliminates the need to train a custom tokenizer. [CLS] and [SEP] tokens are added to the beginning and end of each tokenized sequence(Isabel Segura., 2021).

Attention masking is a technique used in Natural Language Processing (NLP) to prevent the model from attending to padded tokens during the training process. The padded tokens are added to sequences so that all sequences have the same length, which is a requirement for many NLP models.
This is used to train the BERT model for text classification. This model uses the BertForSequenceClassification class from the Transformers library. This model is loaded from a pre-trained 'bert-base-uncased' checkpoint (12 layers, 768 hidden size, 12 heads and 110 million parameters). This checkpoint has been fine-tuned on a large corpus of English text. The number of output labels is set to 6, attention weights and hidden state outputs are turned off.
The optimizer is defined using the AdamW optimizer with a learning rate of 2e-5, a weight drop of 0.01, and a small epsilon of 1e-8.
Training is run for 3 epochs(Devlin et al., 2018) and compute the accuracy of the model's predictions. Gradually decrease the learning rate during training using the get_linear_schedule_with_warmup method. Training is done with random seeds for reproducibility.
At the end of each epoch, the script calculates the accuracy of the model on the validation set and stores it in a list of validation accuracy. After each epoch, the script prints out the average training loss, the time taken for the epoch, and the validation accuracy. The model is set to evaluation mode before evaluating the validation data to ensure that batch normalisation and dropout layers do not affect the evaluation.
Random seeds are set for reproducibility. For each epoch, the training process is performed. The progress of the training process is reported every 40 batches and the training loss is accumulated. The gradients are clipped to prevent exploding gradients and the model's parameters are updated using the optimizer and learning rate scheduler. After the training process, the model is saved.

3.2. Multimodal approach(CNN)

My multimodal approach combines text and image information to classify news articles. A CNN is used to process both the text and corresponding image, producing a 6-dimensional output vector from which the predicted class is derived. Before feeding the data into the network, I preprocess the text by following the same steps outlined in 3.1.1. For the images, standardise their shape to 560 x 560. In the network, separate operations are applied to the text and image inputs.

An embedding layer is created using the nn.Embedding method to represent the textual data (titles) in the network. This layer contains 110,688 unique word embeddings of 300 dimensions each. The embedding weights are initialised with the pretrained embedding matrix and no gradients are computed during training to ensure that the embedding matrix remains fixed.

Images are processed using two convolution layers with a ReLU activation function and a max pooling layer in between.
Titles are processed in four different convolution layers. The title tensor first passes through the embedding layer, then is processed by each of the four convolution layers. A ReLU activation function is applied to the output of the convolutional layer, followed by a max pooling operation with a filter of size (2 x 2).
The outputs of the image and title CNNs are concatenated and passed through the ReLU activation function to two linear layers 3 feature maps of shape (137 x 137). The final layer is a linear layer with 6 output nodes representing the predicted class probabilities. A softmax activation is applied to the output and the result is used to compute the negative log-likelihood loss using the nn.NLLLoss criterion.

Models are trained using the AdamW optimizer, an optimization algorithm that adjusts the learning rate of model parameters based on gradients.
The training loop processes images and titles in batches of 60, each iteration of the loop processing batches of 60 samples. Images are read, padded and truncated to a uniform size of 560x560 and stored in the images tensor. The titles are input tokenized, stored in the titles tensor and passed through the embed layer. Labels are stored in the labels tensor.
After completing a batch of 60 samples, the gradients are reset to zero, the images and titles are passed through the network, and the predicted class probabilities are computed. Then the negative log-likelihood loss is calculated between the predicted probability and the actual label and the gradient is calculated with respect to the loss. Gradients are used by the optimizer to update model parameters and record training accuracy.

4. Evaluation

4.1. Dataset

The Fakeddit dataset (Nakamura et al., 2020) is a collection of over one million instances of Reddit, including text, images, comments, as well as user-submitted posts and comments made in response to those postings. The dataset is multimodal, allowing both binary and fine-grained classification. Classification categories include true news and five different types of fake news. Data is collected from various topics on Reddit and divided into training, validation, and test partitions. Data imbalance can complicate the task of classifying poorly represented categories.

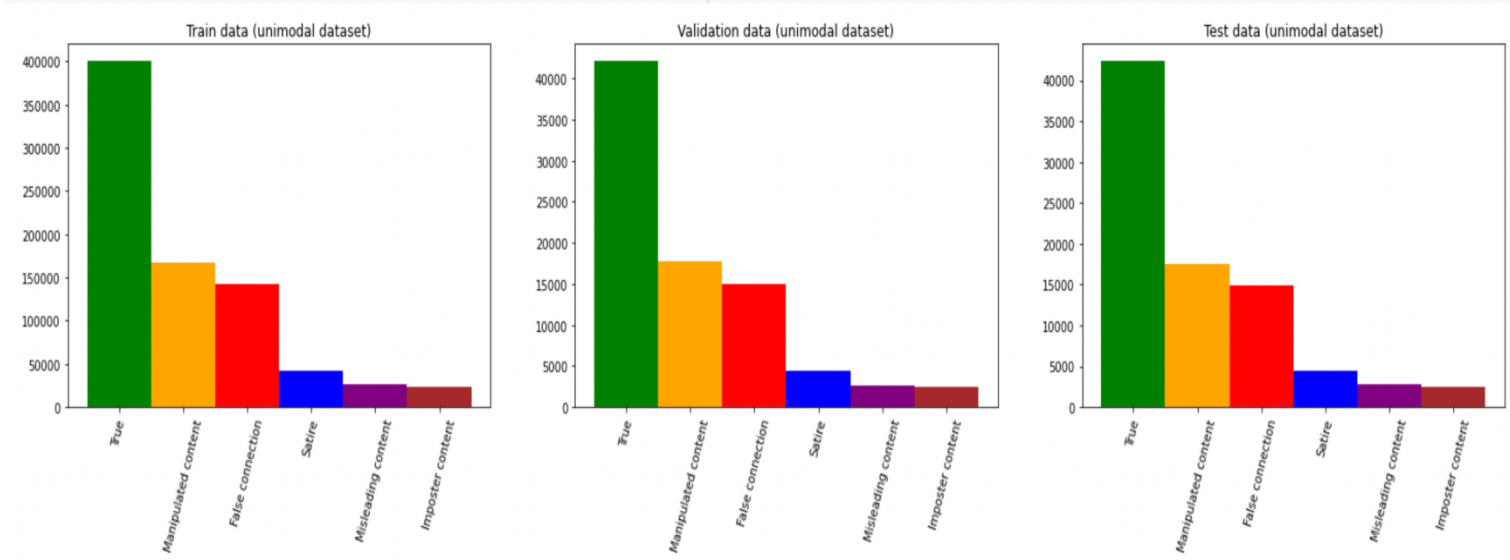
The categories are:

Class	Description	Label
True	Real and accurate information	0
Manipulated content	Altered or manipulated information	1
False connection	Incorrectly connecting two events or facts	2
Satire	Humorous or exaggerated information	3
Misleading content	Inaccurate or incomplete information	4
Imposter content	False information posing as real	5

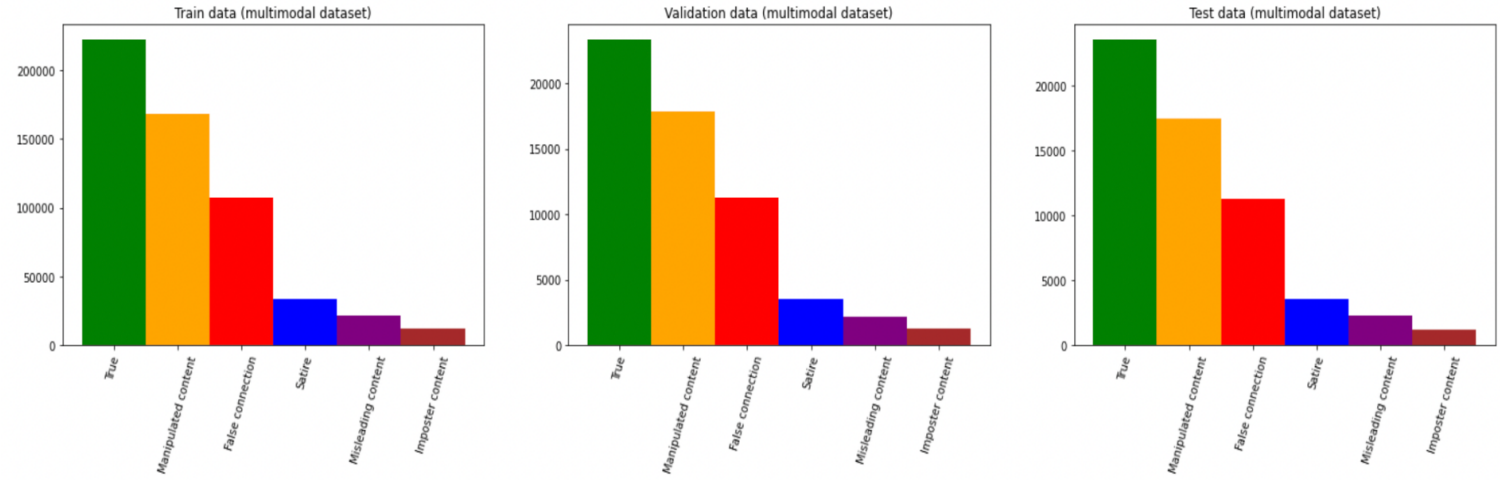
The Fakeddit dataset is divided into training, validation, and test partitions and is available in two versions: unimodal (text only) and multimodal (text and images). The text of all instances in the multimodal dataset are also included in the unimodal dataset.

Figures show the class distributions for the unimodal and multimodal data sets, respectively. The class distributions are similar on both the dataset and the training, validation, and test splits. However, the record is disproportionate, with more instances of the True, Manipulated Content, and False Connection class than parody, misleading content, and Imposter content classes. This imbalance can complicate the task of classifying underestimated classes.

Class distribution (unimodal dataset)



Class distribution (multimodal dataset)



4.2. Results

This section presents the model results by showing the recall, precision, and F1 score for each class. To compare the performance of the models, I calculated not only the overall accuracy, but also the recall, precision, and F1 metrics computed only for the five classes of fake news, excluding the True class categories. Calculate micro and macro averages. To determine which model is best at detecting False content, I use F1 micro-score and accuracy as comparison metrics.

4.2.1. Unimodal approaches

BiLSTM results

In initial experiments with bidirectional LSTM (BiLSTM), the weights in the embedding layer were initialised with random values and updated during the training process and using random initialization and pre-trained Glove vectors.

Results of BiLSTM with random initialization(Table 1)

	precision	recall	f1-score
0	0.77	0.83	0.80
1	0.49	0.37	0.42
2	0.69	0.54	0.61
3	0.29	0.07	0.11
4	0.74	0.85	0.79
5	0.71	0.59	0.64
accuracy			0.73
micro avg	0.70	0.66	0.68
macro avg	0.58	0.48	0.51

This model achieved an accuracy of 73%, a micro F1-score of 68%, and a macro F1-score of 51% (Table 1). The classes of True and Misleading content had the highest F1-scores of 79%, which may be due to the fact that they are the majority classes in the dataset. Conversely, the model achieved the lowest F1-score of 11% for the minority class of Satire (as seen in Table 1). This suggests that the results for the different classes are related to the number of instances per class. However, the model achieved an F1-score of 42% for the second minority class of Manipulated content.

Interestingly, despite only exploiting the textual content of the news, the model achieved an F1-score of 61% for classifying instances of False connection, where the text and image are not in accordance.

Additionally, I also explored BiLSTM with static (as seen in Table 1) and dynamic (as seen in Table 2) GloVe embeddings. In both models, the embedding layer was initialised with pretrained GloVe vectors. When dynamic training was chosen, these vectors were updated during the training process, while they remained fixed during the training process if static training was chosen. The model with dynamic vectors outperformed the one with static vectors, with a slightly improved accuracy (roughly one percentage point). However, in terms of micro F1-score, the static model was better than the dynamic one. Both models had the similar macro F1-score around 51%. The results for the classes showed no significant differences, except for the class of Satire. Updating the pre-trained GloVe vectors resulted in a decrease of 11% points in F1-score for this class.

Results of BiLSTM with static Glove vectors(Table 2)

	precision	recall	f1-score
0	0.77	0.83	0.80
1	0.49	0.36	0.41
2	0.68	0.56	0.61
3	0.14	0.00	0.00
4	0.74	0.85	0.79
5	0.68	0.60	0.63
accuracy			0.73
micro avg	0.70	0.66	0.68
macro avg	0.54	0.47	0.49

In conclusion, the BiLSTM model has achieved a good performance in terms of accuracy and F1-score for some of the classes, while the performance is relatively low for some other classes.

BERT results.(Table 3)

	precision	recall	f1-score
0	0.84	0.88	0.86
1	0.67	0.55	0.60
2	0.69	0.67	0.68
3	0.62	0.39	0.48
4	0.76	0.79	0.78
5	0.76	0.68	0.72
accuracy			0.79
micro avg	0.72	0.69	0.71
macro avg	0.70	0.62	0.65

Table 7 shows the evaluation results of the BERT model. Outperforms all previous unimodal deep learning approaches with 79% accuracy and 71% micro-F1, BERT pre-trained contextual text representations with context-free GloVe vectors and random initialization. It shows its superiority compared to neural networks.

Further analysis of the results shows that the BERT performs better in all classes compared to previous deep learning models. BERT's performance seems to be directly proportional to the number of training instances for a given class. True and Misleading content receive the highest F1-scores of 86% and 78%, while Satire performs the worst with an F1-score of 48%.

The overall accuracy of the model is 79%, which means the model correctly predicts 79% of the samples in the test set. The micro average of precision, recall, and F1-score is 72%, 69%, and 71%, respectively, which indicates a relatively good performance of the model.

4.2.2. Multimodal approach(Table 4)

	precision	recall	f1-score
0	0.76	0.93	0.83
1	0.70	0.75	0.72
2	0.80	0.53	0.64
3	0.51	0.10	0.16
4	1.00	0.99	1.00
5	0.87	0.60	0.71
accuracy			0.83
macro avg	0.77	0.65	0.68
weighted avg	0.83	0.83	0.82

The evaluation of the multimodal approach, combining both text and image information, is shown in table 4. The approach achieved an accuracy of 83% and a weighted average F1-score of 82%.

Compared to the other models, the multimodal approach performs well in terms of precision and recall, particularly in True with a precision of 76% and recall of 93%, and Misleading content with a precision and recall of 100%. On the other hand, Satire performs the worst with a precision of 51% and recall of only 10%.

The macro average F1-score of the multimodal approach is 68%, which is lower than the micro average F1-score. The results suggest that the multimodal approach effectively leverages both text and image information to improve the performance compared to unimodal models.

4.2.3. Comparison of the best models(table 5)

Model	Precision	Recall	F1-score	Accuracy
SVM	0.71	0.72	0.71	0.72
BiLSTM (Dynamic + GloVe)	0.70	0.66	0.68	0.73
BERT	0.70	0.62	0.76	0.79
Multimodal CNN	0.83	0.83	0.82	0.83

I used the GridsearchCV technique to get the best hyper-parameter tuning for SVM modals as a baseline. The table 5 compares the performance of four various models: Support Vector Machine (SVM), Bidirectional Long Short Term Memory (BiLSTM) with Dynamic Vectors and GloVe, BERT(Bidirectional Encoder Representations Transformers) and Multimodal Convolutional Neural Network (CNN).
Based on the results shown in table 5 which I discovered that the best model among the three is the multimodal CNN with an accuracy of 83% and an F1-score of 82% The SVM model has an accuracy of 72% and an F1-score of 71% which is lower compared to the multimodal CNN. The BiLSTM (Dynamic + GloVe) model also has lower performance with an accuracy of 73% and an F1-score of 68% BERT has an accuracy of 79% and an F1-score of 76% which is a little better than the BiLSTM model.
In conclusion, the results show that the Multimodal CNN is the best model from the three others for this particular problem. The combination of both text and image features may be providing complementary information, leading to a better performance compared to the single modality models.

The comparison shows that the multimodal CNN model outperforms other models in this study, including the BERT model. In conclusion, the results of this study provide intuition into the performance of various models for this task and could serve as a reference for future research in this area.

5. Conclusions

Fake news can have devastating effects on various sectors of society, including politics, healthcare and business. Therefore, it is crucial to develop effective and accurate tools to detect misinformation[1]. This study presents a comprehensive comparison of unimodal and multimodal deep learning methods for detailed classification of fake news. To the best of my knowledge, this is one of the first studies of its kind. The results of the study show that multimodal approaches are better than single-modal text-based procedures. The BERT model proved to be the best model for text classification, and the use of GloVe dynamic word embeddings improved the performance of the BiLSTM models compared to random

initialization. These results provide valuable information for future research in the field of fake news detection.

It is suggested that future research looks into other cutting-edge multimodal models that can be used for this task. Consider experimenting with models like VGG-19, and ImageNet, which have demonstrated strong performance on computer vision tasks. I can also explore the effectiveness of various pre-trained language models, such as LSTM, GRU or BERT, to see if they can enhance the performance of the multimodal CNN model.
Examining different methods for feature extraction from texts and images is another area that requires more study. For instance, I might try optimising pre-trained models like ResNet or Inception using the task-specific dataset. To enhance the model's performance on this task, I could also experiment with transfer learning techniques.

References:

https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/

Veronica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic de- tection of fake news. In *Proceedings of the 27th International Conference on Computational Linguis- tics*, pages 3391–3401, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Finneman, T., & Thomas, R. J. (2018). A family of falsehoods: Deception, media hoaxes and fake news. Newspaper Research Journal, 39 , 350–361. doi:https://doi.org/10.1177/0739532918796228.

Casper Hansen, Christian Hansen, and Lucas Chaves Lima. 2021. Automatic fake news detection: Are models learning to reason?

1. O’Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision. In Proceedings of the Science and Information Conference, Leipzig, Germany, 1–4 September 2019; pp. 128–144.

2. Deng, L.; Liu, Y. Deep Learning in Natural Language Processing, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2018.

Wu K, Yang S, Zhu KQ (2015) False rumours detection on Sina Weibo by propagation structures. In: 2015 IEEE 31st international conference on data engineering, pp 651–662

Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2016) Novel visual and statistical image features for microblogs news verification. IEEE Transactions on Multimedia 19(3):598–608

Jin Z, Cao J, Guo H, Zhang Y, Luo J (2017) Multi-modal fusion with recurrent neural networks for rumour detection on microblogs. In: Proceedings of the 25th ACM international conference on multimedia, pp 795–816

Shlok Gilda,”Evaluating machine learning algorithms for fake news detection”, 2017 IEEE 15th Student conference on Research and Development, INSPEC Accession Number: 17613664.

Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. Applied Soft Computing, 101 , 106991.

Ahmed, H., Traore, I., & Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In International conference on intelligent, secure, and dependable systems in distributed and cloud environments (pp. 127–138). Springer.

Wang, W. Y. (2017). "Liar, Liar Pants on Fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648.

Girgis, S., Amer, E., & Gadallah, M. (2018). Deep Learning Algorithms for Detecting Fake News in Online Text. In 2018 13th International Conference on Computer Engineering and Systems (ICCES) (pp. 93–97). doi:10.1109/ ICCES.2018.8639198.

Baheti, P. (2020). Introduction to Multimodal Deep Learning. Retrieved from https://heartbeat.comet.ml/
introduction-to-multimodal-deep-learning-630b259f9291..

Giachanou, A., Zhang, G., & Rosso, P. (2020). Multimodal multi-image fake news detection. In 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 647–654). IEEE.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pretraining of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805,.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2020). FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. Big data, 8 , 171–188.

Kirchknopf, A., Slijepcevic, D., & Zeppelzauer, M. (2021). Multimodal Detection of Information Disorder from Social Media. arXiv preprint arXiv:2105.15165,.

Kaliyar, R. K., Kumar, P., Kumar, M., Narkhede, M., Namboodiri, S., & Mishra, S. (2020). DeepNet: An Efficient Neural Network for Fake News Detection using News-User Engagements. In 2020 5th International Conference on Computing, Communication and Security (ICCCS) (pp. 1–6).

Bahad P., Saxena P. and Kamal R.Procedia Comput. Sci., 165 (2019), pp. 74-82,.

https://github.com/joolsa/fake_real_news_dataset

Kang, Z., Cao, Y., Shang, Y., Liang, T., Tang, H., & Tong, L. (2021). Fake news detection with heterogeneous deep graph convolutional network. In K. Karlapalem, H. Cheng, N. Ramakrishnan, R. K. Agrawal, P. K. Reddy, J. Srivastava, & T. Chakraborty (Eds.), Advances in Knowledge Discovery and Data Mining (pp. 408–420). Cham: Springer International Publishing.

Kumari, R., & Ekbal, A. (2021). AMFB: Attention based multimodal Factorized Bilinear Pooling for multimodal Fake News Detection. Expert Systems with Applications, 184 , 115412.

Nakamura, K., Levy, S., & Wang, W. Y. (2020). Fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection. In Proceedings of the 12th Language Resources and Evaluation Conference (pp. 6149–6157). Marseille, France: European Language Resources Association.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9 , 1735–1780.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Convolutional networks. In Deep learning (pp. 321–362). Cambridge, MA, USA: MIT press.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pretraining of deep bidirectional transformers for language understanding.

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021b). Deep Learning–based Text Classification: A Comprehensive Review.

Patel, K., & Bhattacharyya, P. (2017). Towards lower bounds on number of dimensions for word embeddings. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532–1543).

Patel, K., & Bhattacharyya, P. (2017). Towards lower bounds on number of
dimensions for word embeddings. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short
Papers) (pp. 31–36).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pretraining of deep bidirectional transformers for language understanding.
arXiv preprint arXiv:1810.04805,.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532–1543).

Kai Nakamura, Sharon Levy, and William Yang Wang. 2020. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection.

Isabel Segura-Bedmar (2021). Computer Science Department, Universidad Carlos III de Madrid, Avenida de la.