X-CapsNet For Fake News Detection

Mohammad Hadi Goldani, Reza Safabakhsh, and Saeedeh Momtazi

Abstract—News consumption has significantly increased with the growing popularity and use of web-based forums and social media. This sets the stage for misinforming and confusing people. To help reduce the impact of misinformation on users' potential health-related decisions and other intents, it is desired to have machine learning models to detect and combat fake news automatically. This paper proposes a novel transformer-based model using Capsule neural Networks(CapsNet) called X-CapsNet. This model includes a CapsNet with dynamic routing algorithm paralyzed with a size-based classifier for detecting short and long fake news statements. We use two size-based classifiers, a Deep Convolutional Neural Network (DCNN) for detecting long fake news statements and a Multi-Layer Perceptron (MLP) for detecting short news statements. To resolve the problem of representing short news statements, we use indirect features of news created by concatenating the vector of news speaker profiles and a vector of polarity, sentiment, and counting words of news statements. For evaluating the proposed architecture, we use the Covid-19 and the Liar datasets. The results in terms of the F1-score for the Covid-19 dataset and accuracy for the Liar dataset show that models perform better than the state-of-the-art baselines.

Index Terms—Fake News Detection, COVID-19, BERT, Deep Convolutional Neural Networks, Multi-Layer Perceptron, Dynamic Routing Algorithm, Capsule Neural Networks.

Introduction

 \mathbf{I} N recent years, online social media have become a common platform for broadcasting news for political, commercial, and entertainment purposes. News is understood as any information intended to make the public aware of the events happening around them, which may affect them personally or socially [1].

People use social media to search for and consume news due to its ease, convenience, and rapid spread [2]. These platforms have brought both constructive and destructive impacts. Therefore, as an integral part of culture and society, social media is a double-edged sword [3]. People may manipulate and spread factual information for profit, or their entertainment in the form of fake news [4]. Fake news played a pivotal role in the 2016 United States presidential election campaign after the mass of false information leaked on Facebook over the last three months of the presidential election [5].

Misleading information can disrupt economies, reduce people's trust in their governments, or promote a specific product to make huge profits. For example, this has already happened with COVID-19. Misleading information about lockdowns, vaccinations, and death statistics have fueled panic over purchasing groceries, disinfectants, masks, and paper products. This has led to shortages that have disrupted the supply chain and exacerbated the gaps between supply and demand and food insecurity [6]. In addition, it caused a sharp decline in the international economy, severe losses in the value of crude oil, and the collapse of world stock markets [7]-[9]. Furthermore, due to the spread of COVID-19 and

E-mail: goldani@aut.ac.ir, Safa@aut.ac.ir, momtazi@aut.ac.ir

many people have lost faith in their governments, such as Italy and Iran [10], [11]. These all drive the world into an economic recession [9], [12], [13].

the shortage of medical protection products worldwide,

While a growing percentage of the population relies on social media platforms for news consumption, the reliability of the information shared remains an open issue. Fake news and many types of disinformation are rampant on social media, putting audiences around the world at risk. Therefore, detecting and mitigating the effects of disinformation is a crucial concern in studies where various approaches have been proposed, from linguistic indicators to deep learning models [14], [15].

Recently, the automatic detection of fake news is attracting a large number of researchers [16]–[18]. Early fake news detection methods often designed complete sets of handcrafted features based on news content, user profiles, and news propagation paths, then train classifiers to discriminate the truthfulness of the news [19]-[21]. However, it is challenging to design all-encompassing features, as fake news is usually created on different writing styles, types of topics, and social media platforms [22]. Therefore, many approaches based on deep neural networks [23]-[28] have been proposed to automatically learn patterns discriminated by propagation paths and news content [29].

The recent deep learning models improve the performance of fake news detection models, but the performance drops dramatically when the news content is short [30], [31]. As a solution, in this work, we propose a new model based on CapsNet and indirect features for detecting fake news. The DCNN and MLP models with different feature extraction sections can be parallelized with a CapsNet architecture that is enhanced by using margin loss as the loss function. We also compare varieties of word representation layers and finally use the Bidirectional Encoder Representations from Transformers (BERT) [32] and a robustly optimized BERT pretraining approach(RoBERTa) [33] in our proposed model.

M.H Goldani, R. Safabakhsh, and S. momtazi are with the Department of Computer Engineering, Amirkabir University of Technology, Tehran.

We show that the proposed models achieve better results than the state-of-the-art methods on the Covid-19 and Liar fake news datasets.

The rest of the paper is organized as follows: Section 2 reviews related work about fake news detection. Section 3 presents the model proposed in this paper. The datasets used for fake news detection and evaluation metrics are introduced in Section 4. Section 5 reports the experimental results, comparison with the baseline classification, and discussion. Finally, Section 6 summarizes the paper.

2 RELATED WORK

Social media have presently become the main source of information and news dissemination. This increases the challenges of spreading fake news. As a result, the identification of disinformation has been extensively studied in recent years, with the introduction of several tasks in the field. In this research, our focus is on the detection of fake news in two different domains, COVID-19 and politics, and is mainly based on the supervised method. A machine/deep learning model is trained based on the available data containing fake and real news. The model is then used to decide on the new news articles to find out if they are false or not. Most of the available studies investigating the fake news detection task have been conducted based on deep neural models, including CNN, Long Short Term Memory (LSTM), and BERT, which will be described in this section. The related works of the two domains are reviewed separately in this section.

2.1 COVID-19 domain

Since the appearance of the first case of COVID-19 on December 31, 2019, the World Health Organization (WHO) has declared Covid-19 as a pandemic emergency. As a source of COVID-19 information, tweets and social media news contain information or misinformation about COVID-19. For example, ordinary people become more eager to read more to know how to protect themselves [34]. Brennen et al. [35] examined the sources of misinformation about COVID-19. Their analysis revealed that most of the COVID-19 misinformation is fabricated from real information rather than invented. Therefore the detection of COVID-19 fake news has attracted data scientists.

Patwa et al. [36] provided a comprehensive dataset, the Covid-19 dataset, which includes fake and real news from Twitter. It includes 10,700 posts of the COVID-19 outbreak shared on social media. The real news was captured from 14 official Twitter accounts, and fake data were collected from social media and fact-checking websites. They performed a binary classification task (real vs. false) and evaluated four baselines of machine learning, namely Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gradient Boosting Decision Tree (GDBT). They achieved the best performance of 93.32 % F1-score with SVM on the test set.

Shifath et al. [37] used eight different pre-trained transformer-based models with additional layers to build a stackable ensemble classifier and refined them for the proposed models. Their models were evaluated on the Covid-19

dataset and showed that the RoBERTa-CNN model achieves 96.49% F1 score on the test dataset.

Several supervised text classification algorithms were evaluated by Wani et al. [38] on the Covid-19 fake news detection dataset. Their models are based on CNN, LSTM, and BERT. They also assessed the importance of unsupervised learning in the form of a pre-trained language model and distributed word representations using an unlabeled corpus of COVID-19 tweets. They claimed that their model improved the fake news detection accuracy.

Samadi et al. [39] implemented three different neural classifiers with text representation models like BERT, RoBERTa, Funnel Transformer, and GPT2. They used Single Layer Perceptron (SLP), MLP, and CNN and tried to connect them to various contextualized text representation models. They compared the models' results and discussed their advantages and disadvantages. Finally, to corroborate the effectiveness of their approach, they selected the best model and compared their results with the most advanced models. Furthermore, they added a Gaussian noise layer to the combination of a contextualized text representation model with the CNN classifier. They claimed that the Gaussian noise layer could prevent overfitting in the learning process and, as a result, learns better about the Covid-19 and other data sets.

A two-stage automated pipeline model was developed by Vijjali et al. [40] for COVID-19 fake news detection. They used a state-of-the-art machine learning model for fake news detection. The first model used a novel fact-checking algorithm that searches for the most relevant facts regarding user claims for specific COVID-19 claims. The second model determines a claimant's level of truth by calculating the textual meaning between the claim and facts retrieved from a manually created Covid-19 dataset. They evaluated a set of models based on the classic text-based features as more contextualized Transformer-based models. They found that for both stages, the model pipelines based on BERT and ALBERT yield the best results.

Koloski et al. [41] leveraged several approaches and techniques for detecting COVID-19 fake news. They created several handcrafted features that capture the statistical distribution of characters and words in tweets. They showed that possible spatial representations were learned by capturing potentially relevant patterns from collections of n-grams of characters and features based on the words found in the tweets. For the assessment, they used various BERT-based representations to capture contextual information and the differences between fake and real COVID-19 news. Finally, they proved that the distilBERT tokenizer performs best with an F1 score of 97.05%.

2.2 Politic domain

Many approaches based on neural networks and deep learning models are used to detect fake news articles in the datasets provided for the political domain. Different approaches, including CNNs, RNNs, hybrid models as well as more recent models such as CapsNets were used for the task. Liar dataset was presented by Wang et al. [42]. Then they proposed a model that used statements and metadata together as inputs, a CNN for extracting features from

statements, and a BiLSTM (Bi-directional long short-term memory) network for extracting features from metadata. They demonstrated that their model significantly improved the accuracy.

Long et al. [43] proposed a model on the Liar dataset that incorporates speaker profiles as features, containing speaker position, party affiliation, title, and credit history, into an attention-based LSTM model. They used two ways to improve the model with speaker profiles; (1) considering them in the attention model; (2) incorporating them as additional input data. They demonstrated that this model improves the performance of the classifier on the Liar dataset.

The event adversarial neural network model was proposed by Wang et al. [44]. This model includes three main components: (1) the multimodal feature extractor, which uses CNN as its main module, (2) the fake news detector, which is a fully connected layer with softmax activation (3) the event discriminator that uses two completely connected layers and aims to classify the news in one of the K events based on the representations of the first components. A model based on CapsNet was also proposed for detecting fake news by Goldani et al. [30]. They applied different levels of n-grams and different embedding models to news items of various lengths. Four filters with 2,3,4, and 5 kernel sizes and convolutional n-gram layers with non-static embedding were used for long news statements. For short news, only a static embedding of two filters with kernel sizes of 3 and 5 was used. They showed that their model improves the accuracy of the state-of-the-art methods.

Choudhary et al. [45] used a language model to represent the news text to detect fake news. This linguistic model extracts information related to the news text's syntax, meaning, and readability. Due to the fact that this language model is time-consuming, they used a hierarchical model of neural networks to represent the features. In order to evaluate, the results obtained from sequential neural networks were compared with other machine learning methods and models based on LSTM. It was shown that the model based on hybrid neural networks could have better results than other methods.

Blackledge et al. [46] used transformer-based models to investigate the ability of these models to detect fake news and the generalizability of the models to detect fake news with different topics and models. They showed that the models could not naturally recognize news based on opinion and suspicion. Therefore, they proposed a new method to remove such news articles in the first step and then categorize them. In this article, it is shown that the generalizability using the proposed two-step method is able to improve the accuracy of the transformer-based models.

3 X-CapsNet for fake news detection

This section presents the models proposed for fake news detection in this paper. Depending on whether the news sentences are short or long, two different structures have been proposed to detect fake news. In these models, two parallel networks are concatenated: a CapsNet layer and a new size-based classification layer that uses a DCNN with pre-trained language models or an MLP layer with indirect features extracted from the input text. The concatenated

layer is added to a dense layer to be used for detecting fake news eventually. Figure 1 shows the architecture of the proposed model.

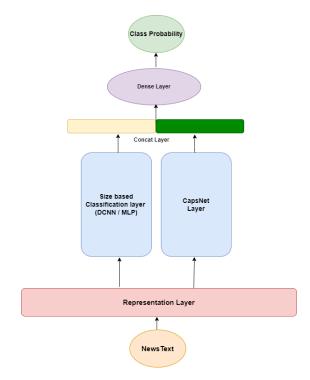


Fig. 1: Proposed model for fake news detection

A CapsNet with a representation layer based on a pretrained embedding model is used for all input text. In addition, when the news text is long, the DCNN model with three feature extractors with kernel sizes of 2, 3, and 4 is used. When the sentences are short, an MLP model that takes advantage of the indirect features of the news is used. In the following, we first review the pre-trained language models, including BERT with non-static embedding, which incrementally uptrains and updates the word embeddings in the training phase. Then we describe the classifiers used for the learning process.

3.1 Representation layer

Word embedding is one of the most widely used techniques in Natural Language Processing (NLP), and the goal is to learn a low-dimensional vector representation for a word in a text [47]. The power of word embedding algorithms such as Word2Vec [48], FastText [49], and GloVe [50] in capturing semantic and syntactic word relationships has been proven. This capability has facilitated various NLP tasks, such as aspect extraction, part-of-speech tagging, and sentiment analysis [51]. The idea of distributional semantics states that words occurring in the same context tend to have similar meanings. In fact, word embeddings reveal hidden relationships between words that can be used during a training process. However, the static embedding techniques mentioned above only capture limited semantic information due to providing a unique static embedding vector for a word in different contexts. Therefore, in recent years, researchers have implemented various deep transformerbased word representation techniques that can take contextual information into account to generate embedding vectors for the same word in different contexts.

BERT [32] is an unsupervised deep model that uses the transformer architecture [52] and has been trained on a huge text corpus using two different scenarios: (1) the masked language model that learns the relationships between words by using their adjacency in a sentence, and (2) the next sentence prediction that learns relations between sentences.

RoBERTa [33] is similar to BERT with some hyperparameters tuning and modifications to its learning process. Liu et al. [33] claimed that BERT was undertrained. Therefore, in addition to a larger dataset with longer sequences for training the model, they trained the model with longer batches. They also compared various alternative training approaches together and, as a result, claimed that for enhancing the performance of the learning process, the next sentence prediction loss can be removed.

GPT2 is a large transformer-based language model trained on various texts taken from web pages on the Internet [53]. The GPT2 architecture is similar to the OpenAI GPT model [54], and is fine-tuned using the four tasks: text classification, text similarity, text consequence, and question answering.

Funnel Transformer is an efficient encoder-decoder architecture that reduces the input features' resolution (length) using a pool operation and embeds them into a lower-dimensional vector [55]. In this model, the decoder is an additional part of the architecture. It is used to simulate a masked language or, in the ELECTRA pre-learning task [56], a new method of learning the self-supervised representation of a language.

In fake news detection, Goldani et al. [30] showed that when the training data size is large enough, the model's performance can improve by using non-static embedding. Therefore, we also use a non-static setting in our model to update the text representation model during the training phase. The recent deep learning models improve the performance of fake news detection models, but the performance drops dramatically when the news content is short. To resolve this problem, we extract more features from the news in addition to the features extracted from the word embedding of sentences. The new features include the signs and information in the sentences [57] along with the history of the speaker profile. More specifically, we use the following indirect features:

- Count of words (length of a news article)
- Count of unique words
- Count of letters
- Count of stop words
- Polarity score
- Subjectivity score
- History of the speaker profile

3.2 DCNN layer for long news article

In recent years, different variations of CNNs have been used in the task of fake news detection [31], [38], [58]–[60]. A CNN architecture with convolutional and pooling layers can accurately extract features from local rendering to global

rendering that indicate the powerful representational capabilities of CNNs. In order to extract more robust features for the learning process, the CNN needs to be enhanced with more identifying information. This requires that the intra-cluster similarity and inter-cluster dissimilarity of the learned features be maximized. For this goal, one of the most commonly used loss functions that are used with softmax in CNNs for fake news tasks is the margin loss [31]. Using this loss function avoids overlapping problems and helps the model mitigate overfitting problems [31]. In Figure 2, the computational flow of the DCNN classifier is demonstrated. Zhong et al. [61] showed that fake news detection could be investigated by adopting a standard text classification model consisting of an embedding layer, a onedimensional convolutional layer, a max-pooling layer, and finally, a prediction-based output layer [61]. Our proposed model is motivated by the concept of multiple parallel channels-variable-size-based neural networks considering three different filter sizes 2, 3, and 4 as n-gram convolutional layers for feature extraction.

In this section, we present our fake news detection model for long news. The proposed model includes a pre-trained embedding model and two parallel classifiers. It reaps the benefits of both DCNN and CapsNet as two different neural network architectures that are used as classifiers.

Figure 2 shows the proposed model. In this architecture, four parallel neural networks have been used. These parallel networks include three different n-grams convolutional layers for feature extraction and a CapsNet layer that includes the primary capsule layer, a convolutional capsule layer, and a feed-forward capsule layer that was previously introduced by Yong et al. [62]. Moreover, in the next layer, the outputs of CNNs and CapsNet go through a global max-pooling and a leaky-ReLU (Rectified Linear Unit) and concatenate. Then after using two dense layers, the final output predicts the label of the input news article. With this architecture, the models can learn more meaningful and extensive text representations on different n-gram levels according to the length of the text.

3.3 MLP layer for short news article

Due to the MLP's ability to learn, handle noisy or incomplete data and solve complex problems in real-time, this method is applied for the proposed classifier for the detection of short fake news [63]. Figure 3 shows the proposed model for detecting short news instances.

In this model, indirect features are added in the training phase because of the short size of the news article and the need for more features for detecting fake news in addition to the transformer-based representation layer.

The model is designed, implemented, and evaluated using the open-source Python software, and the Keras library. It consists of a fully connected 4-layer MLP Neural Network(NN) architecture with an input layer that takes data from the indirect characteristics of the input news text, two hidden layers to process the inputs, and an output layer that indicates the output.

Each layer is a computational abstraction consisting of a fixed number of computational data entities called neurons. Neurons are the building block of the NN that allow the

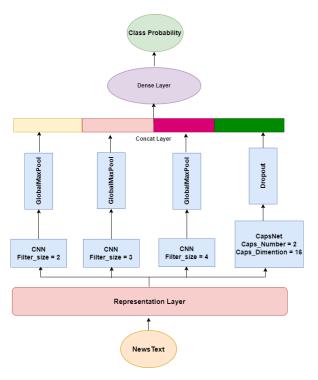


Fig. 2: Proposed model for fake news detection in long news statements

NN to learn from the data and adjust its weights. Neurons in different layers are interconnected through weights that define the weight matrix between the layers.

The model is designed sequentially by adding layers one after the other, starting with the input layer, the first hidden layer, the second hidden layer, and finally, the output layer. The number of input-outputs in the feature dataset determines the neurons in the NN input and output layer. The inner layers contain an arbitrary/specific number of neurons calculated empirically based on standard rules. Each level is defined by the number of nodes/neurons and the triggering function. The first layer has 12 input nodes corresponding to 12 feature attributes of the news indirect feature vector.

The size of the input layer is 12×1. The hidden layer1 is the second layer. It is the first inner layer with 64 neurons in the optimized design and uses the ReLU activation function to operate on the inputs. ReLU is the most used activation function as it overcomes the problem of escape gradients during backpropagation and is suitable for large NNs. Each neuron in this layer is connected to all inputs of the weighted input layer. The size of the merge weight matrix of the input layer and the hidden layer is 12×64. The second hidden layer is the third layer. It has 32 neurons in the optimized form and uses the ReLU activation function. Each neuron in this layer is connected to the outputs of all neurons of the hidden layer1 with associated weights. The dimension of the matrix of the joining weights of hidden layer1 and hidden layer 2 is 32×32. The output layer is the fourth layer with 32 output nodes concatenated with the CapsNet output, and finally, a vector with 64 dimensions is fed to the dense layer.

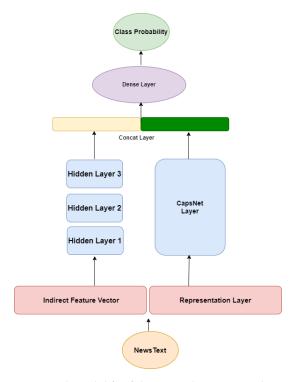


Fig. 3: Proposed model for fake news detection in short news statements

3.4 CapsNet layer

After the success of CapsNets in various NLP tasks [64], different models based on CapsNets have been used for fake news detection in recent years [30], [65], [66].

As mentioned in subsection 3.1, there are different ways for encoding a text. In this work, we use internal word embedding encoding, which means that in CapsNet, the input text is encoded. In this case, we use a 100-dimensional vector to represent a word with a batch size of 50. After evaluations, such dimensions prove to be sufficient to train CapsNet effectively.

Considering the fake news detection with real and fake classes, low-level capsules should catch the most important words in the text. These are the words that significantly affect the classification. After that, two high-level capsules(real and fake classes) through dynamic routing detect the dependencies between significant words and recursively combine them to get the correct prediction. Algorithm 1 shows the dynamic routing procedure, in which r in STEP 2 in line 3 shows a hyperparameter that can be used for the training phase. In the "for loop", the scaler product $a_{ij} = v_j.u_{i|i}$ is defined as the agreement. The agreement is a log likelihood and is added to the initial logit, b_{ij} . The initial coefficients are refined iteratively. This operation measures the agreement between the output v_j of each capsule j in the above layer and the prediction $u_{i|i}$ that was made by capsule i.

Consequently, the assumption is that low-level capsules will detect the words that significantly affect the classification of the text, and high-level capsules will detect low-level capsules and maximize the prediction value.

In the original architecture of CapsNet [67], 2D convolution was used for image processing because the input is

an image. It is important to consider the pixels because the surrounding pixels bring additional information. However, for fake news detection and in the case of text classification, it is not necessary to consider the surrounding pixels since it is not an image. A 1D convolution operation at the first and primary layers is used in this case.

In the original architecture, the decoder framework reconstructs the input image and uses the decoder as a regularization method [67]. However, in the text classification task, there is no reason to use the decoder because the task is only to classify the input into predefined categories. Therefore, the decoder structure is removed from the proposed architecture. Instead of using the decoder as the regularization method, the proposed models use a dropout layer against overfitting [51].

```
Algorithm 1: Dynamic Routing algorithm
```

```
Input: u_{i|j}, r, l
Output: v_j

1 STEP 1: for all capsule i in layer l and capsule j in layer (l+1): b_{ij} \leftarrow 0

2 STEP 2: iterative routing:
3 for i in range (r) do
4 | for all capsule i in layer l: c_i \leftarrow softmax(b_{ij})
5 | for all capsule j in layer l+1: s_j \leftarrow \Sigma_i c_{ij} u_{j|i}
6 | for all capsule j in layer l+1: v_j \leftarrow squash(s_j)
7 | for all capsule i in layer i and capsule i and capsule i in layer i and capsule i and capsul
```

The CapsNet architecture includes a standard convolutional layer called an n-gram convolutional layer that acts as a feature extractor. The second layer maps the scalar features into a capsule representation called the primary capsule layer. The outputs of this capsule are fed to the new layer called the convolutional capsule layer. In this layer, each capsule is only connected to the local area of the layer below. In the final step, the previous layer's output is flattened and fed through the feed-forward capsule layer. For this layer, all capsules in the output are considered to be of a specific class. This architecture uses the maximum margin loss to train the model as presented in Figure 4 [62]. A CapsNet with two capsules of dimension 16 followed by a leaky-ReLU has been chosen as a parallelized neural network in the proposed model.

3.5 Fully connected layer

The functionality of a dense layer is considered a linear operation in which all inputs are connected to all outputs by some weights. We use two dense layers to make the proposed model inherently dense. In the proposed model, the first dense layer takes the output of the concat layer, and then the second dense layer predicts the final output.

4 EVALUATION

The proposed model is evaluated in this section using different datasets for fake news detection.

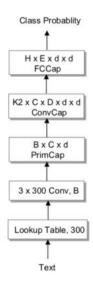


Fig. 4: The architecture of capsule network proposed for text classification by [62]

4.1 Dataset

We use two datasets from different domains for the evaluation of the model. These datasets include the Covid-19 dataset and the Liar dataset.

4.1.1 The Covid-19 dataset

When the COVID-19 pandemic began, social media users shared more and more misinformation and unconfirmed news about the Coronavirus. This motivated researchers to collect datasets from social media and propose machine learning models to evaluate their methods. The Covid-19 dataset is one of the recent datasets proposed by [36] that is a comprehensive dataset including fake and real news from Twitter. This dataset includes 10,700 posts about the COVID-19 outbreaks that were shared on social media. COVID-19 fake articles were collected from fact-checking websites and social media. Moreover, the real news was obtained from 14 official Twitter accounts. Table 1 shows the statistics of the dataset.

TABLE 1: Covid-19 dataset statistics by [36]

Dataset	La	Total	
Dataset	Real	Fake	10141
Training	3360	3360	4420
Validation	1120	1020	2140
Test	1120	1020	2140

4.1.2 The Liar dataset

Another fake news dataset used for evaluating different models is the Liar dataset. This dataset contains 12,800 short political news texts from the United States in 6 different categories and is accessible from the POLITIFACT.COM website. Every news text has been validated on this site by a human agent. Therefore, the dataset is divided into 6 categories: true, false, mostly-True, half-true, barely-True ¹

and pants-fire². The distribution of labels is 1050 pants-fire labels, and the number of the other categories is between 2063 and 2638 [42]. For each news article, the metadata, such as speaker profiles, are taken into account in addition to news statements. This metadata includes valuable information about the new speaker's name, topic, job, state, party, and overall credit history. The total number of credit histories includes false counts, mostly-true counts, barely-true counts, half-true counts, and pants-fire counts. Table 2 demonstrates the statistics of the Liar dataset.

TABLE 2: The Liar dataset statistics provided by [42]

Liar Dataset Statistics	
Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length (tokens)	17.9
Top-3 Speaker Affiliations	
Democrats	4,150
Republicans	5,687
None (e.g., FB posts)	2,185

4.2 Evaluation metrics

In our experiments, the classification accuracy, precision, recall, and F1-score are used as evaluation metrics. The accuracy is the ratio of correct predictions of the news label to the total number of news samples and is computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{1}$$

Precision shows the percentages of the reported fake news that are correctly detected:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall estimates the ratio of the correctly detected fake news:

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-score is the harmonic mean of precision and recall:

$$F1 - score = 2 \times \frac{Precision + Recall}{Precision2 \times Recall}$$
 (4)

In these equations, TP represents the number of True Positive results, TN represents the number of True Negative results, FP represents the number of False Positive results, and FN represents the number of False Negative results.

5 RESULTS

This section evaluates the proposed models on the Covid-19 and Liar datasets on different representation layers. The results are compared to other baseline methods, and the performance of parallel layers is evaluated separately. In the end, in the discussion subsection, a series of experiments on the dataset are discussed.

2. very false

5.1 Classification results on the Covid-19 dataset

5.1.1 Classification results on different representation on the Covid-19 dataset

Table 3 shows the evaluation of the proposed model on the Covid-19 dataset using different representation layers and routing iterations for the dynamic routing algorithm of the CapsNet. As it can be seen, the best result belongs to the model with RoBERTa as representation layer.

TABLE 3: Classification results on different representations on the Covid-19 dataset

Model	RL	Acc	Prec	Rec	F1
	BERT	96.77	96.55	97.48	97.00
DCNN	Funnel	97.05	96.85	97.20	97.02
CapsNet	GPT2	97.00	96.63	97.48	97.05
	RoBERTa	97.34	97.21	97.38	97.29

5.1.2 Classification results on different routing iterations

Figure 5 shows evaluations on the Covid-19 dataset in terms of the F1-score for different routing iterations for the dynamic routing algorithm of the CapsNet model. As a result, for long news statements with COVID-19 subjectivity, one repetition is sufficient to achieve the best result. This shows that combining higher hierarchies for data with more sentences is unnecessary to achieve better results, and the best results are obtained by repeating dynamic routing once.

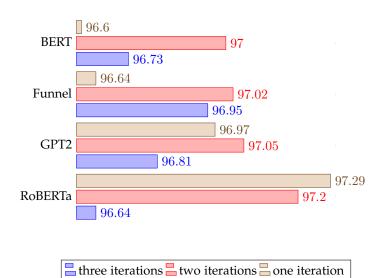


Fig. 5: Classification results on different routing iterations on the Covid-19 dataset

5.1.3 Classification results on the Covid-19 dataset

After the presentation of the Covid-19 dataset by [36], different machine learning and deep learning models were evaluated for fake news detection on this dataset. [36] used conventional machine learning models, including the DT, LR, SVM, and GDBT. [37] proposed an MLP connected to the RoBERTa's pooled output by utilizing additional data for training. [38] evaluated many models, including a softmax layer connected to the BERT for prediction. [39] proposed a CNN connected to the RoBERTa's pooled

output and showed that the performance of this model is better than the previous models. Table 4 compares our proposed model's results with the state-of-the-art models on the Covid-19 test set. The results in terms of the F1-score show that the DCNN-CapsNet model can perform better than the state-of-the-art baselines. Moreover, it should be mentioned that recall is an important factor in fake news detection since missing any fake news article has its own negative side effects. As can be seen in the tabulated results, we achieved the best recall among competitors.

TABLE 4: Comparison of proposed model result with the result of other models on the Covid-19 test set.

Model	Acc	Prec	Rec	F1
DT [36]	85.37	85.47	85.37	85.39
LR [36]	91.96	92.01	91.96	91.96
SVM [36]	93.32	93.33	93.32	93.32
GDBT [36]	86.96	87.24	86.96	86.96
RoBERTa-MLP [37]	96.68	97.12	95.880	96.49
BERT-MLP [38]	95.79	98.94	92.15	95.43
RoBERTa-CNN [39]	97.43	98.30	96.27	97.27
Proposed Model	97.34	97.21	97.38	97.29

5.1.4 Performance of parallel layers

Table 5 shows the proposed model's performance compared to the two parallel models. The results show that using different feature extractors for CNN and adding CapsNet, which aims at keeping detailed information about the location of the object and its pose throughout the network, can improve the performance of the baseline models. The best result is achieved when both models are used together.

TABLE 5: Comparison of proposed model result with the result of parallel layers on the Covid-19 test set.

Model	Acc	Prec	Rec	F1
RoBERTa-CapsNet	93.22	93.92	91.91	96.31
RoBERTa-CNN	97.43	98.30	96.27	97.27
Proposed Model	97.34	97.21	97.38	97.29

5.2 Classification results on the Liar dataset

5.2.1 Classification results on different representations on the Liar dataset

Table 6 shows the evaluation of the proposed model on the Liar dataset using different representation layers and routing iterations for the dynamic routing algorithm of the CapsNet. As it can be seen, the best result belongs to the model with RoBERTa as the representation layer.

TABLE 6: Classification results on different representations on the Liar dataset

Model	Representation	Validation	Test
	BERT	38.55	35.70
MLP	Funnel	37.69	35.46
CapsNet	GPT2	42.91	39.67
	RoBERTa	41.19	41.77

5.2.2 Classification results on different routing iterations

Figure 6 shows evaluations on the Liar dataset in terms of the accuracy for different routing iterations for the dynamic routing algorithm of the CapsNet model. As a result, more repetition is needed for short news statements with political subjectivity to achieve the best result. This shows that combining higher hierarchies for data with more sentences is necessary to achieve better results, and the best results are obtained by repeating dynamic routing twice.

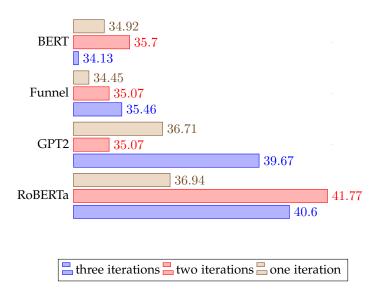


Fig. 6: Classification results on different routing iterations on the Liar dataset

5.2.3 Classification results on the Liar dataset

Table 7 compares our proposed model's results with the state-of-the-art models on the Liar validation and test set. The results in terms of accuracy show that the MLP-CapsNet model can perform better than the state-of-the-art baselines.

TABLE 7: Comparison of proposed model result with the result of other models on the Liar validation and test set.

Model	Validation(%)	Test(%)
Hybrid CNN [42]	24.60	24.10
LSTM attention [43]	37.80	38.50
Two stage BERT model [68]	-	40.58
CapsNet [30]	40.90	39.50
CNN with margin loss [31]	44.40	41.60
Proposed model	41.19	41.77

5.2.4 Performance of parallel layers

Table 8 shows the proposed model's performance compared to the two parallel models. The results show that using indirect features and adding CapsNet, which aims at keeping detailed information about the location of the object and its pose throughout the network, can improve the performance of the baseline models. The best result is achieved when both models are used together.

TABLE 8: Comparison of proposed model result with the result of parallel layers on the Liar validation and test set.

Model	Validation(%)	Test(%)
CapsNet	40.90	39.50
CNN with margin loss	44.40	41.60
Proposed model	41.19	41.77

5.3 Discussion

This section further analyzes the training set of the Covid-19 dataset for real and fake news labels. Figures 7 and 8 show the word clouds for the real and fake news of the training set after omitting the stopwords, respectively. From the word clouds and most frequent words, we see an overlap of the important words across fake and real news. Therefore for more analysis, we list the ten most frequent words in real and fake news after removing the stopwords:

- Fake news: covid, Coronavirus, people, claim, trump, virus, say, vaccine, new, and case
- Real news: case, covid, new, state, test, number, death, India, total, day

If we ignore the common words of the two groups, we find that among the fake news, words about sensitive quotes and reports such as the vaccine, Coronavirus, and the names of politicians are more frequent. Also, statistics about the number of infected cases and related words in this domain, such as number, death, day, and names of countries, have been repeated more frequently in the real news.

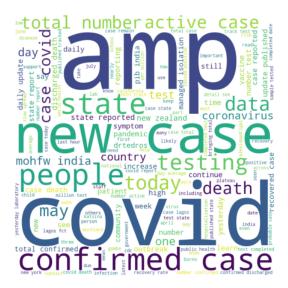


Fig. 7: Word cloud of real news

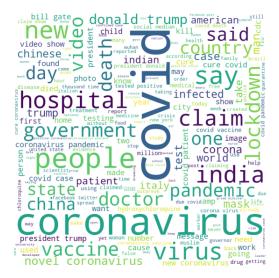


Fig. 8: Word cloud of fake news

We also analyze the polarity and subjectivity of real and fake news of the Covid-19 training set. Polarity is a float that lies in the range of [-1,1], where -1 means negative statement and 1 means a positive statement. Objective refers to factual information, whereas subjective sentences generally refer to emotion or judgment and personal opinion. Subjectivity is also a float that lies in the range of [0,1] [69].

Figure 9 and 10 show the polarity and subjectivity based on the frequency of real and fake news, respectively. We can see that "zero" is the most frequent type of polarity, and in both real and fake news, positive polarity is more than negative polarity. However, for both classes, negative polarity for fake news is more frequent than for real news, while positive polarity is more pronounced in real news.

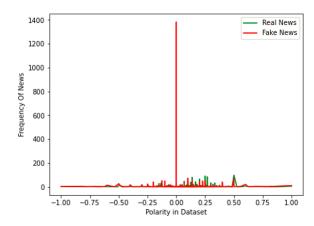


Fig. 9: Polarity of real and fake news

In figure 10, we can see in both fake and real news that zero subjectivity is more frequent; but for fake news, it is obviously more. It is also observed that the subjectivity distribution for fake news is higher than that for real news.

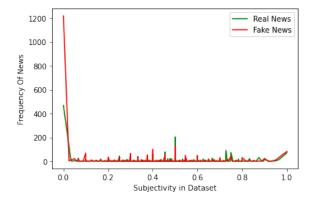


Fig. 10: Subjectivity of real and fake news

Figure 11 shows the different methods of fake news sharing with COVID-19 subject on social media proposed by [70]. This study's model was developed with the U&G theory and previous studies and includes six sharing methods: entertainment, socialization, pass time, altruism, information seeking, and information sharing.

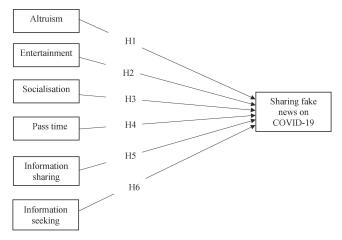


Fig. 11: methods of fake news sharing on COVID-19 proposed by [70].

In order to further experiments on the subjectivity of fake news, the fake news with the subjectivity of one in Figure 10 is extracted and classified. In Table 9, we can see most of the fake news with high subjectivity shared for information sharing purposes on social media. Also, socialization is another frequent method that is used in fake news sharing. As a result, one of the methods of spreading fake news is sharing interesting information and then using members to re-share that news on social networks.

TABLE 9: Distribution of different sharing methods of fake news on COVID-19

Method of	Number of news	
Fake new sharing		
Socialization	16	
Alturism	3	
Information seeking	3	
Information sharing	49	

6 CONCLUSION

This paper proposes X-CapsNet for detecting long and short fake news statements. DCNN-CapsNet with margin loss has been proposed for detecting long fake news statements, and MLP-CapsNet with indirect features for short fake news statements. DCNN-CapsNet uses four parallel neural networks. These parallel networks include three different n-grams convolutional layers for feature extraction and a CapsNet layer. MLP-CaCapsNet has been proposed to solve the problem of short fake news statements that, in addition to the representation layer, we use an indirect features vector created by concatenating news speaker profile information and sentiment, polarity, and sentence information of a fake news article. Different pre-trained representation models and different iterations of the dynamic routing algorithm for the CapsNet have been used to evaluate the proposed models. Finally, models have been tested on two recent wellknown datasets in the field, namely the Covid-19 with long fake news statements and the Liar datasets as a dataset with short news statements. Our result shows that using these models can improve the performance of state-of-theart baselines.

REFERENCES

- P. K. Verma, P. Agrawal, V. Madaan, and R. Prodan, "Mcred: multimodal message credibility for fake news detection using bert and cnn," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2022.
- [2] X. Zhang and A. A. Ghorbani, "An overview of online fake news: Characterization, detection, and discussion," *Information Processing & Management*, vol. 57, p. 102025, 2019.
- [3] K. Ma, C. Tang, W. Zhang, B. Cui, K. Ji, Z. Chen, and A. Abraham, "Dc-cnn: Dual-channel convolutional neural networks with attention-pooling for fake news detection," *Applied Intelligence*, pp. 1–16, 2022.
- [4] A. Bondielli and F. Marcelloni, "A survey on fake news and rumour detection techniques," *Information Sciences*, vol. 497, pp. 38–55, 2019.
- [5] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *Journal of economic perspectives*, vol. 31, no. 2, pp. 211–36, 2017.
- [6] M. K. Elhadad, K. F. Li, and F. Gebali, "Detecting misleading information on covid-19," *Ieee Access*, vol. 8, pp. 165 201–165 215, 2020.
- [7] M. Mhalla, "The impact of novel coronavirus (covid-19) on the global oil and aviation markets," *Journal of Asian Scientific Research*, vol. 10, no. 2, pp. 96–104, 2020.
- [8] C. Albulescu, "Coronavirus and oil price crash," arXiv preprint arXiv:2003.06184, 2020.
- [9] N. J. Gormsen and R. S. Koijen, "Coronavirus: Impact on stock prices and growth expectations," The Review of Asset Pricing Studies, vol. 10, no. 4, pp. 574–597, 2020.
- [10] T. Ling, G. Hoh, C. Ho, and C. Mee, "Effects of the coronavirus (covid-19) pandemic on social behaviours: From a social dilemma perspective," *Technium Soc. Sci. J.*, vol. 7, p. 312, 2020.
- [11] S.-Y. Ren, R.-D. Gao, and Y.-L. Chen, "Fear can be more harmful than the severe acute respiratory syndrome coronavirus 2 in controlling the corona virus disease 2019 epidemic," World journal of clinical cases, vol. 8, no. 4, p. 652, 2020.
- [12] R. Baldwin and B. W. Di Mauro, "Economics in the time of covid-19: A new ebook," VOX CEPR Policy Portal, pp. 2–3, 2020.
- [13] L. Sułkowski, "Covid-19 pandemic; recession, virtual revolution leading to de-globalization?" *Journal of Intercultural Management*, vol. 12, no. 1, 2020.
- [14] R. Bal, S. Sinha, S. Dutta, R. Joshi, S. Ghosh, and R. Dutt, "Analysing the extent of misinformation in cancer related tweets," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, 2020, pp. 924–928.

- [15] S. A. Memon and K. M. Carley, "Characterizing COVID-19 misinformation communities using a novel twitter dataset," CEUR Workshop Proceedings, vol. 2699, 2020.
- [16] K. Shu, D. Mahudeswaran, and H. Liu, "Fakenewstracker: a tool for fake news collection, detection, and visualization," *Computational and Mathematical Organization Theory*, vol. 25, no. 1, pp. 60–71, 2019.
- [17] S. Ghosh and C. Shah, "Towards automatic fake news classification," Proceedings of the Association for Information Science and Technology, vol. 55, no. 1, pp. 805–807, 2018.
- [18] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic Detection of Fake News," in *Proceedings of the International Conference on Computational Linguistics*,, 2018, p. 3391–3401.
- [19] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675–684.
- [20] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2012, pp. 171–175.
- [21] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on sina weibo," in *Proceedings of the ACM SIGKDD workshop on mining data semantics*, 2012, pp. 1–7.
- [22] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake News Detection on Social Media: A Data Mining Perspective," in ACM SIGKDD Explorations Newsletter, vol. 19, no. 1, 2017, pp. 22–36.
- [23] S. R. Sahoo and B. B. Gupta, "Multiple features based approach for automatic fake news detection on social networks using deep learning," Applied Soft Computing, vol. 100, p. 106983, 2021.
- [24] A. Sedik, A. A. Abohany, K. M. Sallam, K. Munasinghe, and T. Medhat, "Deep fake news detection system based on concatenated and recurrent modalities," *Expert Systems with Applications*, p. 117953, 2022.
- [25] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multiclass fake news detection," in *Proceedings of the 27th international* conference on computational linguistics, 2018, pp. 1546–1557.
- [26] M. Davoudi, M. R. Moosavi, and M. H. Sadreddini, "Dss: A hybrid deep model for fake news detection using propagation tree and stance network," *Expert Systems with Applications*, vol. 198, p. 116635, 2022.
- [27] M.-Y. Chen, Y.-W. Lai, and J.-W. Lian, "Using deep learning models to detect fake news about covid-19," ACM Transactions on Internet Technology, 2022.
- [28] J. Zhang, L. Cui, Y. Fu, and F. B. Gouza, "Fake news detection with deep diffusive network model," arXiv preprint arXiv:1805.08751, 2018.
- [29] Q. Liao, H. Chai, H. Han, X. Zhang, X. Wang, W. Xia, and Y. Ding, "An integrated multi-task model for fake news detection," *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [30] M. H. Goldani, S. Momtazi, and R. Safabakhsh, "Detecting fake news with capsule neural networks," *Applied Soft Computing*, vol. 101, p. 106991, 2021.
- [31] M. H. Goldani, R. Safabakhsh, and S. Momtazi, "Convolutional neural network with margin loss for fake news detection," *Information Processing & Management*, vol. 58, no. 1, p. 102418, 2021.
- [32] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pretraining of deep bidirectional transformers for language understanding," in *Proceedings of NAACL-HLT*, 2019, pp. 4171–4186.
- [33] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," arXiv preprint arXiv:1907.11692, 2019.
- [34] D. S. Abdelminaam, F. H. Ismail, M. Taha, A. Taha, E. H. Houssein, and A. Nabil, "Coaid-deep: An optimized intelligent framework for automated detecting covid-19 misleading information on twitter," *Ieee Access*, vol. 9, pp. 27840–27867, 2021.
- [35] J. S. Brennen, F. M. Simon, P. N. Howard, and R. K. Nielsen, "Types, sources, and claims of covid-19 misinformation," 2020.
- [36] P. Patwa, S. Sharma, S. Pykl, V. Guptha, G. Kumari, M. S. Akhtar, A. Ekbal, A. Das, and T. Chakraborty, "Fighting an infodemic: Covid-19 fake news dataset," in *International Workshop on Com-bating Online Hostile Posts in Regional Languages during Emergency Situation*. Springer, 2021, pp. 21–29.
- [37] S. Shifath, M. F. Khan, M. Islam et al., "A transformer based approach for fighting covid-19 fake news," arXiv preprint arXiv:2101.12027, 2021.

- [38] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating deep learning approaches for covid19 fake news detection," in International Workshop on Combating Online Ho stille Posts in Regional Languages during Emergency Situation. Springer, 2021, pp. 153–163.
- [39] M. Samadi, M. Mousavian, and S. Momtazi, "Deep contextualized text representation and learning for fake news detection," *Informa*tion Processing and Management, vol. 58, no. 6, p. 102723, 2021.
- [40] R. Vijjali, P. Potluri, S. Kumar, and S. Teki, "Two stage transformer model for covid-19 fake news detection and fact checking," in Proceedings of the 3rd NLP4IF Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda, 2020, pp. 1–10.
- [41] B. Koloski, T. Stepišnik-Perdih, S. Pollak, and B. Škrlj, "Identification of covid-19 related fake news via neural stacking," in International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation. Springer, 2021, pp. 177–188.
- [42] W. Y. Wang, "" Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection," in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 2017, p. 422–426.
- [43] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake news detection through multi-perspective speaker profiles," in *Proceed*ings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2017, pp. 252–256.
- [44] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, " Eann: Event Adversarial Neural Networks for Multi-Modal Fake News Detection," in Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining, ACM, 2018, pp. 849–857.
- [45] A. Choudhary and A. Arora, "Linguistic feature based learning model for fake news detection and classification," *Expert Systems with Applications*, vol. 169, p. 114171, 2021.
- [46] C. Blackledge and A. Atapour-Abarghouei, "Transforming fake news: Robust generalisable news classification using transformers," arXiv preprint arXiv:2109.09796, 2021.
- [47] Y. Li, Q. Pan, T. Yang, S. Wang, J. Tang, and E. Cambria, "Learning word representations for sentiment analysis," *Cognitive Computa*tion, vol. 9, no. 6, pp. 843–851, 2017.
- [48] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
- [49] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Asso*ciation for Computational Linguistics, vol. 5, pp. 135–146, 2017.
- [50] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
- [51] A. Khikmatullaev, J. Lehmann, and K. Singh, "Capsule neural networks for text classification," Ph.D. dissertation, Ph. D. dissertation, 04 2019, 2019.
- [52] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in neural information processing systems, 2017, pp. 5998–6008.
- [53] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever et al., "Language models are unsupervised multitask learners," OpenAI blog, vol. 1, no. 8, p. 9, 2019.
- [54] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever et al., "Improving language understanding by generative pre-training," 2018.
- [55] Z. Dai, G. Lai, Y. Yang, and Q. Le, "Funnel-transformer: Filtering out sequential redundancy for efficient language processing," Advances in neural information processing systems, vol. 33, pp. 4271– 4282, 2020.
- [56] K. L. Clark, M. Le, Q. Manning, and C. ELECTRA, "Pre-training text encoders as discriminators rather than generators," in *Interna*tional conference on learning representations, 2020.
- [57] G. K. Shahi, A. Dirkson, and T. A. Majchrzak, "An exploratory study of covid-19 misinformation on twitter," Online social networks and media, vol. 22, p. 100104, 2021.
- [58] J. A. Nasir, O. S. Khan, and I. Varlamis, "Fake news detection: A hybrid cnn-rnn based deep learning approach," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100007, 2021.
- [59] K. L. Tan, C. P. Lee, and K. M. Lim, "Fn-net: A deep convolutional neural network for fake news detection," in 2021 9th International

- Conference on Information and Communication Technology (ICoICT). IEEE, 2021, pp. 331–336.
- [60] R. K. Kaliyar, A. Goswami, P. Narang, and S. Sinha, "Fndneta deep convolutional neural network for fake news detection," *Cognitive Systems Research*, vol. 61, pp. 32–44, 2020.
- [61] B. Zhong, X. Xing, P. Love, X. Wang, and H. Luo, "Convolutional neural network: Deep learning-based classification of building quality problems," *Advanced Engineering Informatics*, vol. 40, pp. 46–57, 2019.
- [62] M. Yang, W. Zhao, J. Ye, Z. Lei, Z. Zhao, and S. Zhang, "Investigating Capsule Networks with Dynamic Routing for Text Classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 2018, pp. 3110–3119.
- [63] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using mlp classifier," Applied Acoustics, vol. 180, p. 108077, 2021.
- [64] Q. Jiang, L. Chen, R. Xu, X. Ao, and M. Yang, "A challenge dataset and effective models for aspect-based sentiment analysis," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 6280–6285.
- [65] B. Palani, S. Elango, V. Viswanathan K et al., "Cb-fake: A multi-modal deep learning framework for automatic fake news detection using capsule neural network and bert," Multimedia Tools and Applications, pp. 1–34, 2021.
- [66] S. Sridhar and S. Sanagavarapu, "Fake news detection and analysis using multitask learning with bilstm capsnet model," in 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence). IEEE, 2021, pp. 905–911.
 [67] S. Sara, F. Nicholas, and H. Geoffrey E, "Dynamic routing between
- [67] S. Sara, F. Nicholas, and H. Geoffrey E, "Dynamic routing between capsules," in *Proceedings of the International Conference on Neural Information Processing Systems (NIPS)*, 2017, p. 3859–3869.
- [68] C. Liu, X. Wu, M. Yu, G. Li, J. Jiang, W. Huang, and X. Lu, "A two-stage model based on bert for short fake news detection," in International conference on knowledge science, engineering and management. Springer, 2019, pp. 172–183.
- [69] Parthvi Shah, "Sentiment Analysis using TextBlob," https://towardsdatascience.com/my-absolute-go-to-forsentiment-analysis-textblob-3ac3a11d524 (Accessed: 2020-06-27), 2020.
- [70] O. D. Apuke and B. Omar, "Fake news and covid-19: modelling the predictors of fake news sharing among social media users," *Telematics and Informatics*, vol. 56, p. 101475, 2021.