



Multilingual deep learning framework for fake news detection using capsule neural network

Rami Mohawesh¹ · Sumbal Maqsood² · Qutaibah Althebyan^{1,3}

Received: 14 February 2023 / Revised: 30 March 2023 / Accepted: 31 March 2023 /
Published online: 9 May 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Fake news detection is an essential task; however, the complexity of several languages makes fake news detection challenging. It requires drawing many conclusions about the numerous people involved to comprehend the logic behind some fake stories. Existing works cannot collect more semantic and contextual characteristics from documents in a particular multilingual text corpus. To bridge these challenges and deal with multilingual fake news detection, we present a semantic approach to the identification of fake news based on relational variables like sentiment, entities, or facts that may be directly derived from the text. Our model outperformed the state-of-the-art methods by approximately 3.97% for English to English, 1.41% for English to Hindi, 5.47% for English to Indonesian, 2.18% for English to Swahili, and 2.88% for English to Vietnamese language reviews on TALLIP fake news dataset. To the best of our knowledge, our paper is the first study that uses a capsule neural network for multilingual fake news detection.

Keywords Fake news · Cross-language fake news · Machine learning · Deep learning

1 Introduction

Currently, fake news is more prevalent on social media channels as compared to traditional media (Nelson et al., 2020). As a result of this problem, many researchers are focusing on developing several fake news detection frameworks, which is a crucial and challenging task. A model to detect fake news aims to spot or identify misleading news by analysing

✉ Rami Mohawesh
rami.mohawesh@aau.ac.ae

Sumbal Maqsood
sumbal.maqsood@utas.edu.au

Qutaibah Althebyan
qutaibah.althebyan@aau.ac.ae; qaalthebyan@just.edu.jo

¹ Cybersecurity Department, College of Engineering, Al Ain University, Al Ain - Abu Dhabi, UAE

² School of Information Technology, University of Tasmania, Hobart, Tasmania, Australia

³ Software Engineering Department, Jordan University of Science and Technology, Irbid, Jordan

previously reviewed real and fake news. As a result, a high-quality and large-scale dataset is required to perform this task accurately and efficiently. Fake news and natural language processing (NLP) researchers face the difficult task of creating multilanguage models that can be used in any of the world's 7,000+ languages. Multimodal data from multiple languages can be challenging to collect and analyse simultaneously for a given task, making it essential to employ a framework that allows for the manageable generalisation of a visual language model (Nelson et al., 2020).

In late 2019, the biggest pandemic around the globe Coronavirus Disease, known as COVID-19, generated a massive amount of informative data about COVID-19 (Nelson et al., 2020). Platforms for spreading such information, like mass media and social media, make it possible for the information to reach a large number of the audience (Cinelli et al., 2020). Unfortunately, not all of the information is accurate or trustworthy. Some of the information spreading around those platforms can be categorised as misleading or even identified as false news. Notably, various countries may have different situations and strategies to control the spread of COVID-19, which also leads to a considerable amount of inappropriate news sharing (Apuke & Omar, 2021). For example, "The spread of COVID-19 is linked to 5G mobile networks", "Sunny weather protects you from COVID-19", and "Place a halved onion in the corner of your room to catch the COVID-19 germs". Social media facilitated the rapid dissemination of this and similar false news stories. During the early stages of the pandemic. The wave of misinformation was so massive that the authorities had coined a word for it: "infodemic" (Shapley, 2016). Meanwhile, a lot of fake news was produced in various languages to spread more easily to particular ethnic groups. Thus, it is very challenging for authoritative organisations to respond promptly to the spread of fake news. Failing to detect and stop multilingual COVID-19 misinformation can lead to mistrust, panic buying, social distancing, and refusal to get tested and vaccinated.

Researchers have been focusing on machine learning-based NLP (Natural Language Processing) strategies to prevent the spread of misinformation. To identify misinformation, one study (Ozbay & Alatas, 2020) employed twenty-three supervised machine learning models. Increasingly successful models based on deep learning approaches have recently been adopted to detect misinformation. For example, (Aggarwal et al., 2020) used BERT to identify misinformation for short text and had excellent results. Similarly, (Patwa et al., 2021) propose a model for detecting fake COVID-19 news. Research studies have demonstrated outstanding results in classifying COVID-19 news using machine learning methods.

Furthermore, NLP models trained in one language can be transferred to another without the need for extra annotated training data (Chu et al., 2021). In the process of adapting natural language processing (NLP) models to other languages, the zero-shot cross-lingual transfer is one method that shows promising results. Recent strategies for cross-lingual transfer have demonstrated that models can still generalise to a non-English language through the translation models for multiple languages or unsupervised language learning, followed by task-specific fine-tuning using only English annotation. This achievement can be traced back because several languages have significant common words in their vocabulary or structural in another way. For instance, several terms in the English and German vocabularies share a Latin root, such as the words "desk" and "tisch". All languages have a repetitive structure, and many of their morphologies or word orders are the same in many languages.

Visual information is definitely necessary for the cross-lingual transmission of vision-language models. Detecting fake news in news statements of varying lengths using multilingual models is a crucial but understudied step in this direction (Vo, 2022). Many visual concepts are universally recognised, despite the fact that human languages

may differ. Even though in English it is called “cat”, and in French, it is “chat”, both terms mean the same thing. In order to facilitate the transfer of vision-language models between languages, we learn to correlate phrases in various languages with perceptual notions by using this observation as a basis for our learning.

Galli et al. (2022) presented a detailed analysis of recent machine learning algorithms and their findings for fake news detection. This study provided the detail of datasets used for the experiments in terms of the efficiency and accuracy of the proposed models. Kanfoud and Bouramoul (2022) used standard language features for multilanguage sentiment analysis with high-accuracy results. The authors suggested that this is incredibly convenient for real-time applications. Zhang et al. (2022) proposed a technique for fake news detection based on the statement conflict to be classified as agree or disagree. The results showed the significance of the proposed technique and outperformed state-of-the-art methods using one dataset.

The fake news detection models based on textual contents (Yu et al., 2017, Shu, et al., 2019, Asghar et al., 2021) and visual features (Gupta et al., 2012, Qi et al., 2019) separately are ineffective with poor classification results. Thus, a multimodal approach has been introduced based on visual and textual features to improve the performance of fake news detection models. The convolutional neural network has been used in many studies to achieve the visual features for multimodal fake news detection (Wang et al., 2018, Yang et al., 2018, Singhal et al., 2019). However, CNN requires more training time and data to generalise the model and loses some crucial image features due to the pooling process that resultant in poor performance with combined with textual features. To deal with this issue, a capsule neural network has been proposed, which is faster than CNN and able to capture the most informative features from the image. In addition to this, we used LSTM to capture the sequence-to-sequence analysis to extract the textual features.

The goal of this study is to develop a multilingual framework based on capsule neural networks that can be used to improve the contextual representations of news text. However, with the zero-shot default, a significant efficiency gap for queries exists between English and languages other than English. Since large-scale language models and training have recently advanced, we propose a Multilingual technique to take advantage of the inadequate supervision provided by massive amounts of multilingual data in order to tackle this issue. Our technique has two significant advantages. Prior training on solely English data does not take advantage of the additional supervision provided by multilingual data and consequently does not improve the performance in any other language.

To detect fake news from multilingual COVID-19, we propose a new framework for the detection of multilingual fake news using transfer learning and capsule neural networks. Word embedding and n-gram feature extractors are used as feature extractors to develop models for detecting false news in various duration of press announcements.

1.1 Organisations

The paper is organised as follows: Section 2 illustrates the related works for detecting fake news. Section 3 describes the proposed framework in detail. Section 4 provides experimental settings, detailed descriptions of the dataset and dataset pre-processing. Section 5 presents the experimental results and discussion. Section 6 presents the conclusion and provides context for future work.

2 Related work

2.1 Cross-lingual transfer learning

A survey of fake news detection methods, including the primary categories and the end aims of these methods, has been published (Chu et al., 2021). These researchers employed both linguistic and social approaches in their research: linguistic methods, where they extracted relevant language patterns from the news content, and network approaches, where they used criteria such as message metadata to make decisions about fresh incoming news. Fake news detection is made more accessible with the help of Rubagotti and colleagues (Basile and Rubagotti (2018)) invention of an automated system known as CSI. The wording, response, and source of the incoming news are all factors taken into consideration by this detector when predicting the spread of false information. Three modules make up the model, the first of which extracts the articles' temporal representation. For classification purposes, the third and final module takes data from the first two modules, which are based on user behaviour. The results of their trials show that CSI improves accuracy. Known as ISOT, a novel dataset for detecting fake news, Ahmed et al. The data in this set was gathered only from publicly available records. N-gram models and six machine learning approaches were employed to identify bogus news in the ISOT dataset. This combination of TF-IDF extraction and linear support vector machine classification yielded the best results. New research by (Faruqui et al., 2014) attempts to determine whether or not a news article is bogus based on how many people have interacted with and/or liked it. They outlined two distinct approaches to categorisation. Using user interactions as characteristics, a logistic regression model is used in the first method. The subsequent is an innovative use of the crowdsourcing methods of the Boolean label. Both methods had impressive accuracy, as evidenced by the results. In addition, the trials demonstrated that taking into account the related informational data of people that engage along with the news is a significant feature in order to make a judgement regarding that news. Rather than short statements containing bogus news material, (Pérez-Rosas et al., 2017) offered two new datasets connected to seven distinct categories, and their datasets contain authentic news extracts. Linguistic variables, and language features like syntactic, semantic, and lexical, can help identify between fake and real news using a linear support vector machine classifier. Human performance in this area was compared to the performance of the produced system.

2.2 Transformer-based models

Later, multilingual mBERT was developed by the inventors of BERT by pre-training the BERT model using Wikipedia material in 104 languages (Wang et al., 2021). The multilingual mBERT to support several languages is mBERT. Later, many multilingual models were proposed following mBERT. In recent studies, large amounts of text were found to boost the model's performance. Because of this, XLMR model was proposed (Hu et al., 2021), which is pre-trained on CC-100, which is a much larger corpus of text than Wikipedia's for low-resource languages. Significantly larger models like XLM-RXL and XLM-RXXL (Goyal et al., 2021) are pre-trained on CC-100 and obtain much better results. mBERT, XLM-R, and its variations are trained on non-parallel data. As a result, it is considerably more manageable for a model to acquire cross-lingual representations if it is trained on parallel and non-parallel data.

2.3 Fake news detection using capsule network

Sabour et al. (2017) introduced a capsule neural network, and it performed better than CNNs using the MINIST dataset. In the field of computer science, neural networks include capsule networks that make an effort to perform converse graphics processing. This method attempts, in order to figure out how the objective process that creates an image works backwards. The capsule neural network consists of numerous capsules to perform functions and attempt to forecast parameters for instantiation and a specific object located in a specific position (Sabour et al., 2017).

Goldani et al. (2021) proposed a fake news detection model using a capsule network using two datasets, ISOT and LIAR. They used different embedding models to analyse the distance length of new items. The results demonstrate a 7.8% improvement in ISOT, a 3.1% improvement in the validation set, and a 1% improvement in the test set of the LIAR dataset when compared to the state-of-the-art approaches. The bidirectional encoder representations from transformers (BERT) model is used by Palani et al. (2022) to extract textual information, preserving word semantic links. Unlike CNN, the CapsNet model collects an image's most informative visual elements. Combining these variables creates a richer data representation that helps identify fake news. The proposed model outperformed the SpotFake + model using Politifact and Gossipcop datasets.

Sun et al. (2022) proposed a fake news detection model using bi-GRU and capsule networks. This approach considered the contextual relationship in its entirety without sacrificing any of the text's spatial structural data. The suggested model employs a two-way GRU (BiGRU) network to enhance the capsule network, which can enhance the accuracy of rumour detection. The proposed model outperforms benchmark rumour detection algorithms on the Twitter dataset by 6.1% under static routing and 6.7% under dynamic routing. Braşoveanu and Andonie (2019) proposed a hybrid approach incorporating machine learning, semantics, and NLP. They provided a semantic false news detection system based on text-extracted sentiment, entities, and facts. The findings of this study revealed that semantic characteristics increase false news categorisation accuracy.

Word embedding models and their variants are discussed in detail in this section. For the second part of our study, we suggested two capsule neural network models for fake news identification based on the duration of the news statements.

2.3.1 Word embedding methods variations

Syntactic and semantic information can be extracted from words using dense word representation. Word embedding refers to the mapping of word representations to a low-dimensional space. Closely related words are clustered together in the vector space in these representations. (Mikolov et al., 2017) proposed word2vec in 2013, a set of mathematical methods for generating word embeddings from raw text that has shown to be highly efficient. In order to build these models, enormous amounts of text are used to train neural networks with two layers of input. Using these methods, each word can be represented as a vector in a vector space with hundreds of dimensions. Similar-sounding words are clustered together in this area.

Word2vec vectors like “Google News”, which is trained on 100 billion Google news words, are now available. Using these pre-trained vectors for initialising word vectors is a popular way to increase text processing performance, especially in the absence of an extensive,

supervised training dataset. Deep neural networks can employ these distributed vectors for text classification (Mobiny & Van Nguyen, 2018). However, further refinement of these pre-trained embeddings is possible.

The authors (Mobiny & Van Nguyen, 2018) were the first who used multiple learning settings for vector representation of words using word2vec and demonstrated their superiority over standard pre-trained embeddings in a CNN model. These are the options: pre-trained vectors are utilised as inputs for the framework of the neural network in the static word embedding model. There are no changes made to these vectors while they are being trained.

This model uses the pre-trained vectors at the beginning of the learning process. Although the vectors are adjusted for optimal performance for each job during training, they are based on training data for that specific target task. It uses two sets of word2vec vectors, one non-static and the other during training fine-tuned (Table 1).

3 The proposed framework – multilingual-fake model

We proposed a multilingual capsule network learning-based model with multilingual embeddings integrated with semantics infusion. There are several layers in the proposed architecture for effective multilanguage fake news detection. The semantic infusion is used for extra lexical semantics (Fig. 1).

Let there be two input layers w_1, w_2, \dots, w_n for O_s input, which is the source language and w_1, w_2, \dots, w_n for T_s inputs for translated target language provided to the embedding layer. The embedding layer receives the input sequence, which contains a sequence of words as text. This input sequence is a matrix $X \in \mathbb{R}^{e \times V}$ where V is vocabulary size, and e represents a vector of words. This input also includes extra semantic infusion for fake news detection. The feature extraction layer contains the forward hidden units and backward hidden units aligned with the bidirectional Long Short-Term Memory (Bi-LSTM) network. The concept of using LSTM is to extract the contextual feature relationship with the local features. The forward feature set is denoted as h_t^f , and backward denoted as h_t^b with total n units from the output of bidirectional LSTM as:

$$h_t = \left[h_t^f, h_t^b \right] \in \mathbb{R}^{2 \times n} \quad (1)$$

This feature set is provided to the capsule network. There are two layers for the capsule network, the first one is primary, and the second is a convolutional capsule layer. In order to describe spatial location correlations between features, such as the local order of words and their semantics, the first primary capsule layer extracts instantiation parameters (Zhao et al., 2018).

Let there be $W \in \mathbb{R}^{(2 \times n) \times m}$ where m presents the capsule dimensionality. For the features extracted from features layer h_t , each capsule is $p_i \in \mathbb{R}^d$ given in the equation below, where g shows the linear function to shorten the vector length between 0 to 1 points given in the equation.

$$p_i = g(W^T h_t + b) \quad (2)$$

$$g(p) = \frac{\|p\|^2}{1 + \|p\|^2} \frac{p}{\|p\|} \quad (3)$$

Table 1 Summary of the existing research

Reference	Machine Learning Model	Major findings
(Basile and Rubagotti (2018))	SVM	Works only with a single language
(Faruqui et al., 2014)	logistic regression	Works only with a single language
(Pérez-Rosas et al., 2017)	SVM	Works only with a single language
(Wang et al., 2021)	mBERT	The mBERT model allows for a standardised representation of information across languages. However, cross-language performance is not affected by the lexical overlap between languages
(Sabour et al., 2017)	capsule neural network	Works only with a single language
(Chen et al., 2018)	deep adversarial network (ADAN)	However, the well-known deep learning-based approaches and word embedding methods still have significant drawbacks due to poor performance in retraining semantic information between documents over cross languages
(Hu et al., 2021)	XLMR model	As a result, it is considerably more accessible for a model to acquire cross-lingual representations if it is trained on parallel and non-parallel data. Further, poor performance in retraining semantic information between documents over cross languages

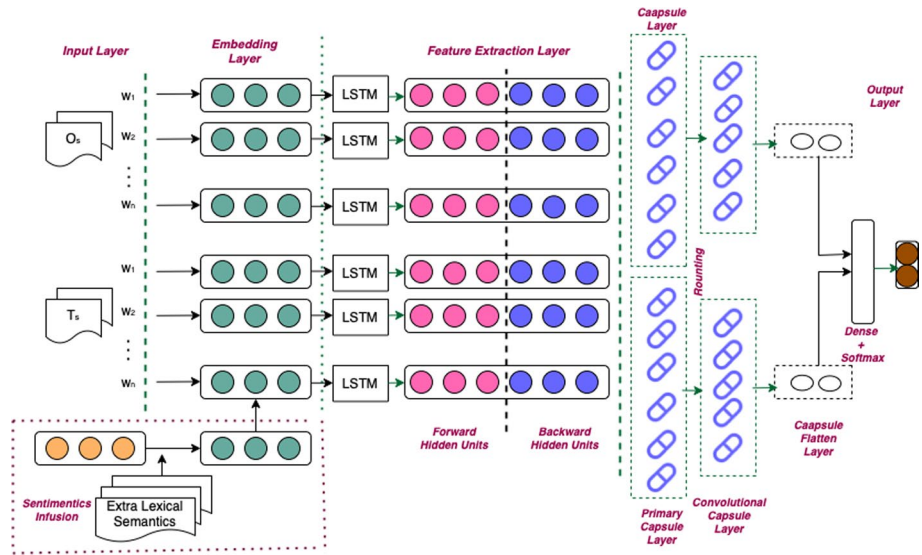


Fig. 1 Proposed multilingual deep learning architecture for Fake News Detection using capsule neural network

There is a connection that exists between the capsules in the primary and convolutional capsule layer. The max pooling in the convolutional neural network might neglect some important spatial relationship features by just focusing on specific positions if it exists (Zhao et al., 2018). To make the capsule network more effective, dynamic routing is used to learn the connection weights and ignore irrelevant information (Srivastava et al., 2018). This dynamic connection between these two capsule layers is given below, where p_i shows the primary and v_k presents the convolutional capsule layer.

$$c_{k|i} = softmax(b_{k|i}) \tag{4}$$

$$v_k = g\left(\sum_k c_{k|i} p_{k|i}\right) \tag{5}$$

$$b_{k|i} = b_{k|i} + p_{k|i} \cdot v_k \tag{6}$$

Where $b_{k|i}$ is the connection between primary and convolutional capsule layers. The output capsules are flattened after the routing. Using the SoftMax function, the parallel architecture is concatenated to obtain the final probability of fake news detection.

4 Experiments

We propose a multilingual framework for fake news detection using a deep learning model based on a capsule neural network. The experiments were conducted using one dataset. This section provides the detail of implementation setup, evaluation metrics, dataset, and pre-processing.

Input: Set of input news articles, w_1, w_2, \dots, w_n for O_s for the source language.
 w_1, w_2, \dots, w_n for T_s for translated target language; $X \in \mathbb{R}^{e \times V}$
Semantic Infusion; Extra Lexical Semantic analysis

Output: Label output {Fake, Genuine}

Where “0” indicate the fake news; “1” presents real news.

Fake news detection model process:

1: Word Embedding

The following converts the textual input news article data into word embeddings:

2: for *textual-news-articles* in O_s and T_s

3: for *textual-news-articles* do

4: Sentence tokenise

5: Map vocabulary IDs

6: Set the sequence length

7: Create segment IDs

8: Provides the embedded input to LSTM for sequential analysis with backward and forward hidden layers:

9: Obtain the parameters with initial weights

10: Calculate the weights; Calculate the forward and backward hidden states

11: Combine the forward and backward states; $h_t = [h_t^f, h_t^b] \in \mathbb{R}^{2 \times n}$

12: Output the Hidden states $h = [h_1, \dots, h_n], \in \mathbb{R}$

13: Obtain the content and semantic features from LSTM

14: Obtain the visual features from Capsule Neural Network

15: For Each h_t Features;

16: Calculate capsule features using Linear Function; $g(W^t h_t + b)$

17: Dynamic routing to learn the connection weights; $c_{k|i} = \text{softmax}(b_{k|i})$

18: Extracted the visual and content features

19: The output label class using Dense with SoftMax activation layer

20: Return fake/ genuine class output

Algorithm 1: The proposed Multilingual-Fake algorithm

4.1 Evaluation metric

To evaluate our model’s effectiveness, we used the TALLIP dataset created by De et al. (2021), which contains six different languages: English, Hindi, Swahili, Vietnamese, and Indonesian, where each dataset has six different domains. The dataset contains text news only.

To carry out the experiments, datasets were divided into three categories: training, validation, and testing. For the experiments of our proposed model, we used classification that can be evaluated using classification accuracy. The accuracy presents the ratio of accurate predictions of fake news to the total number of samples that can be calculated by the equation given below. For the evaluation of the proposed model, accuracy and F1- measure metrics have been used, which can be defined as follow:

Table 2 Summary statistics information of the dataset

Dataset Domain	Language dataset		Multilingual dataset	
	Training	Testing	Training	Testing
Domain-Independent Data	600	380	3000	1900
Celebrity	300	200	1500	1000
Sports	50	30	250	150
Politics	50	30	250	150
Entertain	50	30	250	150
Technology	50	30	250	150
Education	50	30	250	150
Business	50	30	250	150

Accuracy: Full estimate of correctly classified instances and can be calculated by

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{False Negative} + \text{True Negative}}$$

F1- measure shows an average of precision and recall and is computed by

$$\text{F1 - measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.2 Implementation details

We used the Python language¹ to implement our model. We set the maximum sentence length to 200. We trained the model multiple times to get the best results in terms of performance metrics. All experiments have been done using the Colab pro tool.²

For our proposed model implementation, the training batch size for all datasets was set to 64. The dropout was 0.5, and the regularisation coefficient was set to 0.001. The learning rate was set to 0.001 using the Adam algorithm (Kingma & Ba, 2015) to train the proposed model. The number of layers of LSTM was set to 2, and the hidden layer dimension was set to 512. The number of gated attention layers was set to 6, and the number of training epochs was set to 30. It is worth mentioning that we trained the model several times to get the best parameters for our models. Our model is developed with a batch size of 256 and a learning rate of 2–105 using the AdamW optimiser (Loshchilov and Hutter, 2019).

4.3 Dataset

TALLIP-fake news datasets This dataset was created by De et al. (2021) and contained six different languages English, Hindi, Swahili, Vietnamese, and Indonesian, where each dataset has six different domains (Technology, Business, Education, Politics, Celebrity News, Entertainment). The dataset consists of three columns: news text Label (Fake/Genuine),

¹ <http://www.python.org>

² <https://colab.research.google.com/>

topic, and ID (sentence identification number). Table 2 shows a summary statistics information of the dataset.

4.3.1 Data pre-processing

Throughout this stage, datasets were pre-processed to remove noise like URLs, stop words, emojis, POS tagging, tokenisation, etc. The NLTK³ toolkit was used for pre-processing (Aghakhani et al., 2018); it is an Open-Source library widely used. Further, we used the CoreNLP tool⁴ (Manning et al., 2014) For POS tagging, which is used for different languages.

5 Results and discussion

5.1 Comparison with the state-of-the-art methods

In this sub-section, we have selected a range of state-of-the-art approaches for comparison to demonstrate our proposed framework's efficiency. According to the current work, we used accuracy as an evaluation metric (Li et al., 2020).

Baseline approaches:

- **mBERT** (Devlin et al., 2019): The traditional BERT model's multilingual pre-trained language variant is mBERT (mBERT). The BERT model to build multilingual representations of texts employs a self-supervised learning strategy supported by shared common word tokens, like recent empirical investigations (Li et al., 2020), (Srivastava & Sutton, 2017). Then, these multilingual language representations are employed to assist the multilingual fake news classification task.
- **XLM** (Conneau et al., 2020): Considered a significant effort for multilingual text classification. In this model, Conneau et al. (2020) introduced common sub-word lexicon mapping frameworks to construct the integrated embedding structures for words in different languages for the multilingual text classification task.
- **XLM-RoBERTa** (Conneau et al., 2020): is the most recent learning strategy for multilingual textual representations based on transformers. This model is an extended version of the BERT model trained on the large-scale text that consists of one hundred languages. Massive experimental experiments conducted by (Conneau et al., 2020) showed that the XLM-RoBERTa model outperformed the conventional mBERT in a range of multilingual text classification tasks.
- **MGL** (Li, et al., 2020): is a preliminary GNN-based approach for addressing several cross-language text classification tasks. In this model, Li et al. (2020) developed a graph model learning strategy for the multilingual knowledge transfer between the source and the target languages. In summary, the MGL works as dynamic meta-graphs to capture the semantic links between words and documents in various languages.

³ https://www.nltk.org/_modules/nltk/tag.html

⁴ <https://nlp.stanford.edu/software/tagger.shtml>

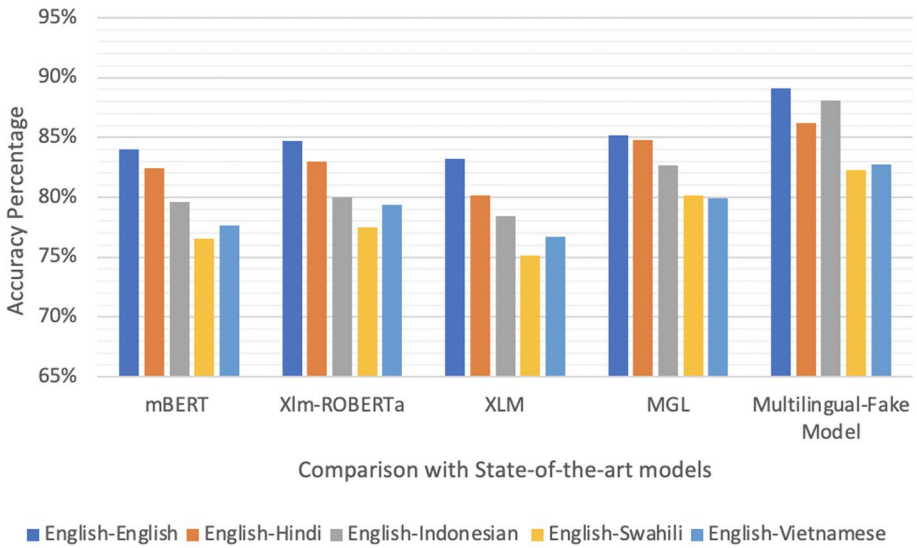


Fig. 2 Comparison with the state-of-the-art models using Accuracy

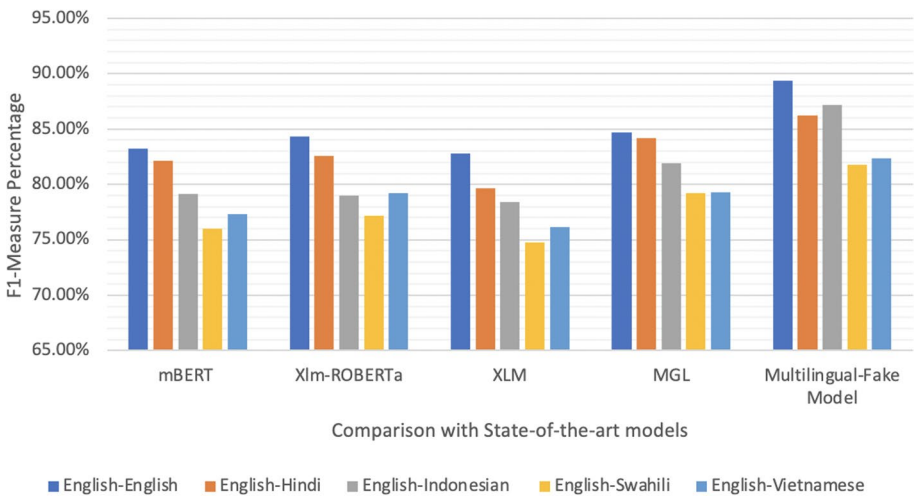


Fig. 3 Comparison with the state-of-the-art models using F-measure

As shown in Figs. 2, 3 and Tables 3, 4, the state-of-the-art methods of fake news detection are either based on advanced deep learning or graph-based model. Our proposed framework achieves better classification performance than the state-of-the-art methods on all the datasets languages. Compared to the MGL, the accuracy increased by 5%. The Table results show the effectiveness of the proposed framework in detecting multilingual fake news. This shows that semantic and contextual information is

Table 3 Comparison with the state-of-the-art models using Accuracy

Models	Dataset language				
	English-English	English-Hindi	English-Indonesian	English-Swahili	English-Vietnamese
Accuracy					
<i>mBERT</i> (Wang et al., 2021)	84%	82.43%	79.63%	76.58%	77.67%
<i>Xlm-ROBERTa</i> (Conneau et al., 2020)	84.72%	83.01%	79.98%	77.45%	79.34%
<i>XLNet</i> (Conneau et al., 2020)	83.20%	80.15%	78.46%	75.13%	76.70%
<i>MGL</i> (Li, et al., 2020)	85.15%	84.80%	82.63%	80.12%	79.90%
Multilingual-Fake Model (Ours)	89.12%	86.21%	88.10%	82.30%	82.78%

Table 4 Comparison with the state-of-the-art models using F-measure

Models	Dataset language				
	English-English	English-Hindi	English-Indonesian	English-Swahili	English-Vietnamese
<i>F-measure</i>					
<i>mBERT</i> (Wang et al., 2021)	83.24%	82.14%	79.11%	75.99%	77.34%
<i>Xlm-ROBERTa</i> (Conneau et al., 2020)	84.34%	82.56%	79.01%	77.13%	79.19%
<i>XLNet</i> (Conneau et al., 2020)	82.80%	79.68%	78.37%	74.77%	76.12%
<i>MGL</i> (Li, et al., 2020)	84.68%	84.16%	81.89%	79.19%	79.28%
Multilingual-Fake Model (Ours)	89.34%	86.23%	87.17%	81.78%	82.36%

beneficial for multilingual fake news detection. This can explain the improvement of our proposed framework in terms of accuracy. It is worth mentioning that, in addition to the efficiency and strength of our proposed framework, other essential factors affect its performance, such as feature extraction, which is very important when dealing with textual data.

6 Conclusion

This research article proposes a multilingual capsule network learning-based model with multilingual embeddings integrated with semantics infusion. Our model is made up of two parallel architectures for source and target languages. BiLSTM is used to extract contextual characteristics, and a Capsule Network is used to capture hierarchically positioned relationships. Experiments conducted on TALLIP datasets demonstrated that the classification performance had been greatly improved and outperformed state-of-the-art algorithms by approximately 3.97% for English to English, 1.41% for English to Hindi, 5.47% for English to Indonesian, 2.18% for English to Swahili, and 2.88% for English to Vietnamese languages for the multilingual fake news detection task, demonstrating the efficacy of our proposed model. The experimental results on these two well-known datasets showed improvement in terms of accuracy by 7.8% on the ISOT dataset and 3.1% on the validation set, and 1% on the test set of the LIAR dataset.

Author contributions Rami initialised the project and managed the study. Rami collected the data and performed a formal analysis. Rami analysed the data. Rami wrote the initial manuscript. All authors contributed to the editing of the paper.

Data availability All datasets are open-source, and the sources are cited.

Declarations

Ethics approval and consent to participate No ethical issue is involved.

Competing interests No competing interest.

References

- Aggarwal, A., et al. (2020). Classification of fake news by fine-tuning deep bidirectional transformers based language model. *EAI Endorsed Transactions on Scalable Information Systems*, 7(27), e10–e10. <https://doi.org/10.4108/eai.13-7-2018.163973>
- Aghakhani, H., et al. (2018). Detecting deceptive reviews using generative adversarial networks. In 2018 IEEE Security and Privacy Workshops (SPW), IEEE. <https://doi.org/10.48550/arXiv.1805.10364>
- Apuke, O.D., & Omar, B. (2021). Fake news and COVID-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, 56: 101475. <https://doi.org/10.1016/j.tele.2020.101475>
- Asghar, M. Z., et al. (2021). Exploring deep neural networks for rumor detection. *Journal of Ambient Intelligence and Humanized Computing*, 12, 4315–4333. <https://doi.org/10.1007/s12652-019-01527-4>
- Basile and Rubagotti (2018). Basile, A., & Rubagotti, C. (2018). CrotoneMilano for AMI at Evalita2018. A performant, cross-lingual misogyny detection system. EVALITA Evaluation of NLP and Speech Tools for Italian 12: 206. <https://doi.org/10.4000/BOOKS.ACCADEMIA.4734>.

- Braşoveanu, A.M., & Andonie, R. (2019). Semantic fake news detection: a machine learning perspective. In *Advances in Computational Intelligence: 15th International Work-Conference on Artificial Neural Networks, IWANN 2019, Gran Canaria, Spain, June 12–14, 2019, Proceedings, Part I 15*, Springer. https://doi.org/10.1007/978-3-030-20521-8_54
- Chen, X., et al. (2018). Adversarial deep averaging networks for cross-lingual sentiment classification. *Transactions of the Association for Computational Linguistics*, 6, 557–570. https://doi.org/10.1162/tacl_a_00039
- Chu, S. K. W., et al. (2021). Cross-Language fake news detection. *Data and Information Management*, 5(1), 100–109. <https://doi.org/10.2478/dim-2020-0025>
- Cinelli, M., et al. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10(1), 1–10. <https://doi.org/10.1038/s41598-020-73510-5>
- Conneau, A., et al. (2020). Unsupervised cross-lingual representation learning at scale. *ACL*. <https://doi.org/10.18653/v1/2020.acl-main.747>
- De, A., et al. (2021). A transformer-based approach to multilingual fake news detection in low-resource languages. *Transactions on Asian and Low-Resource Language Information Processing*, 21(1), 1–20. <https://doi.org/10.1145/3472619>
- Devlin, J., et al. (2019). Pre-training of deep bidirectional transformers for language understanding In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, MN: Association for Computational Linguistics: 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Faruqui, M., et al. (2014). Retrofitting word vectors to semantic lexicons. arXiv preprint arXiv:1411.4166. <https://doi.org/10.48550/arXiv.1411.4166>
- Galli, A., et al. (2022). A comprehensive Benchmark for fake news detection. *Journal of Intelligent Information Systems*, 59(1), 237–261. <https://doi.org/10.1007/s10844-021-00646-9>
- Goldani, M.H., et al. (2021). Detecting fake news with capsule neural networks. *Applied Soft Computing*, 101: 106991. <https://doi.org/10.48550/arXiv.2002.01030>
- Goyal, N., et al. (2021). Larger-scale transformers for multilingual masked language modeling. REPL4NLP. <https://doi.org/10.18653/v1/2021.repl4nlp-1.4>
- Gupta, M., et al. (2012). Evaluating event credibility on twitter. *Proceedings of the 2012 SIAM international conference on data mining, SIAM*. <https://doi.org/10.1137/1.9781611972825.14>
- Hu, J., et al. (2021). Explicit alignment objectives for multilingual bidirectional encoders. NAACL. <https://doi.org/10.48550/arXiv.2010.07972>
- Kanfoud, M. R., & Bouramoul, A. (2022). SentiCode: A new paradigm for one-time training and global prediction in multilingual sentiment analysis. *Journal of Intelligent Information Systems*, 59(2), 501–522. <https://doi.org/10.1007/s10844-022-00714-8>
- Kingma, D.P., & Ba, J. (2015). Adam: A Method for Stochastic Optimisation. CoRR abs/1412.6980. <https://doi.org/10.48550/arXiv.1412.6980>
- Li, Z., et al. (2020). Learn to cross-lingual transfer with meta graph learning across heterogeneous languages. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. <https://doi.org/10.18653/v1/2020.emnlp-main.179>
- Manning, C.D., et al. (2014). The Stanford CoreNLP natural language processing toolkit. *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*. <https://doi.org/10.3115/v1/P14-5010>
- Mikolov, T., et al. (2017). Advances in pre-training distributed word representations. arXiv preprint arXiv:1712.09405.
- Mobiny, A., & Van Nguyen, H. (2018). Fast capsnet for lung cancer screening. *Medical Image Computing and Computer Assisted Intervention—MICCAI 2018: 21st International Conference, Granada, Spain, September 16–20, 2018, Proceedings, Part II 11*, Springer. <https://doi.org/10.48550/arXiv.1806.07416>
- Nelson, L. M., et al. (2020). US public concerns about the COVID-19 pandemic from results of a survey given via social media. *JAMA Internal Medicine*, 180(7), 1020–1022. <https://doi.org/10.1001/jamainternmed.2020.1369>
- Ozbay, F.A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: Statistical Mechanics and its Applications*, 540: 123174. <https://doi.org/10.1016/j.physa.2019.123174>
- Palani, B., et al. (2022). CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT. *Multimedia Tools and Applications*, 81(4), 5587–5620. <https://doi.org/10.1007/s11042-021-11782-3>
- Patwa, P., et al. (2021). Overview of constraint 2021 shared tasks: Detecting english covid-19 fake news and hindi hostile posts. *International Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation*, Springer.

- Pérez-Rosas, V., et al. (2017). Automatic detection of fake news. arXiv preprint arXiv:1708.07104. <https://doi.org/10.48550/arXiv.1708.07104>
- Qi, P., et al. (2019). Exploiting multi-domain visual information for fake news detection. 2019 IEEE international conference on data mining (ICDM), IEEE. <https://doi.org/10.48550/arXiv.1908.04472>
- Sabour, S., et al. (2017). Dynamic routing between capsules. *Advances in neural information processing systems* 30.
- Shapley, L. S. (2016). 17. A value for n-person games. *Contributions to the Theory of Games (AM-28), Volume II*, Princeton University Press: 307–318.
- Shu, K., et al. (2019). defend: Explainable fake news detection. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. <https://doi.org/10.1145/3292500.3330935>
- Singhal, S., et al. (2019). Spofake: A multi-modal framework for fake news detection. 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), IEEE. <https://doi.org/10.1109/BigMM.2019.00-44>
- Srivastava, A., & Sutton, C. (2017). Autoencoding Variational Inference For Topic Models. ICLR. <https://doi.org/10.48550/arXiv.1703.01488>
- Srivastava, S., et al. (2018). Identifying aggression and toxicity in comments using capsule network. Proceedings of the first workshop on trolling, aggression and cyberbullying (TRAC-2018).
- Sun, X., et al. (2022). Rumour detection technology based on the BiGRU_capsule network. *Applied Intelligence* 1–17. <https://doi.org/10.1007/s10489-022-04138-3>
- Vo, T. (2022). An integrated topic modelling and graph neural network for improving cross-lingual text classification. *Transactions on Asian and Low-Resource Language Information Processing*. <https://doi.org/10.1145/3530260>
- Wang, Y., et al. (2018). Eann: Event adversarial neural networks for multi-modal fake news detection. Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining. <https://doi.org/10.1145/3219819.3219903>
- Wang, Z., et al. (2021). Cross-lingual text classification with heterogeneous graph neural network. *ACL*. <https://doi.org/10.18653/v1/2021.acl-short.78>
- Yang, Y., et al. (2018). TI-CNN: Convolutional neural networks for fake news detection. arXiv preprint arXiv:1806.00749. <https://doi.org/10.48550/arXiv.1806.00749>
- Yu, F., et al. (2017). A convolutional approach for misinformation identification. *IJCAI*. <https://doi.org/10.24963/ijcai.2017/545>
- Zhang, D., et al. (2022). Fake news detection based on statement conflict. *Journal of Intelligent Information Systems*, 59(1), 173–192. <https://doi.org/10.1007/s10844-021-00678-1>
- Zhao, W., et al. (2018). Investigating capsule networks with dynamic routing for text classification. arXiv preprint arXiv:1804.00538. <https://doi.org/10.18653/v1/D18-1350>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.