Metadata of the chapter that will be visualized in SpringerLink

Book Title	Intelligent Information and Database Systems	
Series Title		
Chapter Title	v3MFND: A Deep Multi-domain Multimodal Fake News Detection Model for Vietnamese	
Copyright Year	2023	
Copyright HolderName	The Author(s), under exclusive license to Springer Nature Switzerland AG	
Corresponding Author	Family Name	**Nguyen Thi**
	Particle	
	Given Name	**Cam-Van**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Vietnam National University (VNU), University of Engineering and Technology (UET)
	Address	Hanoi, Vietnam
	Email	vanntc@vnu.edu.vn
Author	Family Name	**Vuong**
	Particle	
	Given Name	**Thanh-Toan**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Vietnam National University (VNU), University of Engineering and Technology (UET)
	Address	Hanoi, Vietnam
	Email	18021279@vnu.edu.vn
Author	Family Name	**Le**
	Particle	
	Given Name	**Duc-Trong**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	Vietnam National University (VNU), University of Engineering and Technology (UET)
	Address	Hanoi, Vietnam
	Email	trongld@vnu.edu.vn
Author	Family Name	**Ha**
	Particle	
	Given Name	**Quang-Thuy**

Prefix	
Suffix	
Role	
Division	
Organization	Vietnam National University (VNU), University of Engineering and Technology (UET)
Address	Hanoi, Vietnam
Email	thuyhq@vnu.edu.vn

Abstract

Fake news become a critical problem on the Internet, especially social media. During the worldwide COVID-19 epidemic, social networking sites (SNSs) are primary sources to spread false news, which are incredibly difficult to detect and regulate them since they rapidly grow everyday. With multimedia technology advances, the content of social media news now is manifested via various modalities, such as text, photos, and videos. Approaches that learn the multimodal representation for detecting fake news have evolved in recent years. Additionally, there exist diverse content domains in news platforms. Exploiting data from these domains potentially solve the data sparsity problem as well as simultaneously boosting overall performance. In this paper, we propose an effective Deep Multi-domain Multimodal Fake News Detection model for Vietnamese, **v3MFND** for short. Extensive experiments on a real-life dataset reveal that **v3MFND** improves the performance of multi-domain multimodal fake news detection for Vietnamese considerably. An ablation study is also carried out to evaluate the role of each individual modality in the multimodal model.

Keywords (separated by '-')

Vietnamese fake news detection - Multimodal - Multi-domain

v3MFND: A Deep Multi-domain Multimodal Fake News Detection Model for Vietnamese

Cam-Van Nguyen Thi(✉), Thanh-Toan Vuong, Duc-Trong Le, and Quang-Thuy Ha

Vietnam National University (VNU), University of Engineering and Technology (UET), Hanoi, Vietnam
{vanntc,18021279,trongld,thuyhq}@vnu.edu.vn

Abstract. Fake news become a critical problem on the Internet, especially social media. During the worldwide COVID-19 epidemic, social networking sites (SNSs) are primary sources to spread false news, which are incredibly difficult to detect and regulate them since they rapidly grow everyday. With multimedia technology advances, the content of social media news now is manifested via various modalities, such as text, photos, and videos. Approaches that learn the multimodal representation for detecting fake news have evolved in recent years. Additionally, there exist diverse content domains in news platforms. Exploiting data from these domains potentially solve the data sparsity problem as well as simultaneously boosting overall performance. In this paper, we propose an effective Deep Multi-domain Multimodal Fake News Detection model for Vietnamese, **v3MFND** for short. Extensive experiments on a real-life dataset reveal that **v3MFND** improves the performance of multi-domain multimodal fake news detection for Vietnamese considerably. An ablation study is also carried out to evaluate the role of each individual modality in the multimodal model.

AQ1

AQ2

Keywords: Vietnamese fake news detection · Multimodal · Multi-domain

1 Introduction

Fake news has been confounded with phrases, e.g., rumor, false news, and misinformation, however there is no an universal definition [15]. The dissemination of fake news through the Internet has grown into a scourge of our time, especially through social networks in the news media. According to the Digital 2021 research[1], there are around 72 million Vietnamese users on board with the blazing-fast surge of SNSs, e.g., Facebook, Zalo, or Lotus, with a growth of 7 million accounts over the same period in 2020, equating to a penetration rate of 73.7%. SNSs are also an ideal platforms to spread fake news during the outbreak of the global Coronavirus (COVID-19) pandemic. It is extremely difficult to

[1] https://datareportal.com/reports/digital-2021-global-overview-report.

N.-T. Nguyen et al. (Eds.): ACIIDS 2022, LNAI 13757, pp. 1–13, 2023.
https://doi.org/10.1007/978-3-031-21743-2_49

detect and control these fake news when they grow exponentially everyday [11]. Once fake news can also cause extremely unpredictable consequences for society in general and people in particular, fake news detection becomes an important and urgent problem that needs to be addressed.

With the modern multimedia technology, one trend that cannot be overlooked is that more and more social media news now includes information in several modalities, such as text, images, and videos. Multiple information modalities provide greater proof of news events occurring and more attractive to readers, hence this kind of fake news is harder to detect and widespread easier. Methods for detecting fake news have steadily progressed from unimodal to multimodal techniques in recent years. The question of how to learn a joint representation that includes multimodal information has sparked a lot of interest in the scientific community. As an example, Shivangi et al. [14] introduce SpotFake, a multimodal framework for fake news detection using BERT [1]. As a variant improved from SpotFake, they propose another novel transfer learning-based fake news detection method named SpotFake+ [13], which is mainly based on XLNet [20]. In [4], Khattar et al. employs a bimodal auto-encoder in conjunction with a conditional classifier to build the MAVE model. In order to extract text and visual representation, they exploit bi-directional LSTMs and VGG-19 respectively. Probabilistic latent variable models are trained via maximizing the marginal probability of observed data. Song et al. [15] leverage an attention method to fuse a number of word embeddings and one image embedding to create fused features, and then extract essential features as a joint representation of the fused features. Generally, the modeling multiple modalities is efficient to improve the fake news detection rate.

In real-world scenarios, news platforms provide a variety of news in many disciplines, i.e., domains, on a daily basis. Potentially, levering data from these domains may help alleviate the data sparsity problem while also improving the performance of all domains. This raises the notion of multi-domain fake news detection (MFND).However, major domain shift and lack of labeled data in certain domains are two serious problems that make MFND becomes more challenging [19]. As an effort to solve MFND, Quiong Nan et al. [9] propose a model named MDFEND, whereby uses domain gate to combine different representations retrieved by mixture-of-experts in order to cope with multi-domain transfer and isolation. The limitation of MDFEND is just use the text feature while ignoring the connected images in each news.

Overall, these mentioned points motivate us to exploit the fake news detection problem in the direction of multi-domain multimodal methods. Because there is a dearth of multi-domain multimodal dataset for Vietnamese in fake news detection problems, it is critical to build an appropriate dataset. In this paper, we proposed a model named **v3MFND**: A Deep **M**ulti-domain **M**ulti**M**odal **F**ake **N**ews **D**ectection Model for **V**ietnamese. Our main contributions are as follows:

1. Construct a multi-domain multimodal dataset for Vietnamese fake news detection up on the ReINTEL dataset [6], a fake news dataset on Vietnamese SNSs.

2. Propose a multi-domain multimodal automatic fake news detection model for Vietnamese using advanced deep learning methods for text features and image features.
3. Conduct extensive experiments to assess the role of each individual modality in multimodal model.
4. Provide useful insight into several aspects of our approach for future research.

The rest of the paper is organized as follows. Section 2 delves into the details of the labeled Vietnamese multidomain multimodal fake news detection dataset. This is followed by a discussion of the proposed model in Sect. 3. In Sect. 4, we describe how we set up our experiment and go through the results in depth. Finally, we conclude the paper with Sect. 6.

2 M2-ReINTEL: Vietnamese Multi-domain Multimodal Dataset

ReINTEL dataset[2] is a fake news dataset collected from Vietnamese SNSs [6]. The most straightforward information, i.e., news content, is gathered for each item of news, as well as different modalities, i.e., photos, sequential signals, i.e., timestamp, and social context, i.e., number of likes, number of shares. Table 1 summarises the characteristics of this dataset. It is worth noting that, in addition to the text of the news gathered, the urls associated with the photos used in the news is also provided; however, not all news feature photos.

Table 1. Example of common features in ReINTEL dataset

Feature	Example	Explaination
id	1952_public	unique id for a news post on SNSs
uid	7d31f701ce1abfc1fc 0b7b311debc99d	the anonymized id of the owner
text	The State Bank of Vietnam reduced a series of operating interest rates from March 17. (Ngân hàng Nhà nước giảm hàng loạt lãi suất điều hành từ 17/3)	the text content of the news
timestamp	1584336550	the time when the news is posted
image_links	_	image urls associated with the news
nb_likes	4	the number of likes
nb_comments	0	the number of comment
nb_shares	0	the number of shares
label	0	unreliable news are labeled 1, reliable news are labeled 0

[2] https://vlsp.org.vn/vlsp2020.

The dataset includes 4825 news, which only 3583 news have visual features. All of news are manually labeled with ten domains namely *science, health, politics, education, economics, disaster, military, sports, entertainment, society.*

We have two experts who independently label each record, then do a match between these two results. For labels that do not agree, we review the record together and come up with a final label. However, because the content of the news is quite long, the domains covered in those news come from many different domains, it is quite difficult to assign a label. We selected the more popular domain among the domains considered. In the future, multi-label label and simultaneous identification may be considered for implementation.

Because this data was gathered during the COVID-19 epidemic in Vietnam, there are a lot of linked articles, which is why the disaster domain is so popular. A simple statistic is shown in Table 2. To make identification and future referencing easier, we gave this multi-domain multimodal dataset for fake news detection the name **M2-ReINTEL**.

Table 2. The statistics of dataset M2-ReINTEL

Domain		Science	Health	Politics	Education	Economic
Real	no_image	66	62	92	92	223
	has_image	6	30	15	54	33
Fake	no_image	2	14	31	7	18
	has_image	0	4	10	1	5
Total		74	110	148	154	279
Domain		Disaster	Military	Sport	Entertainment	Society
Real	no_image	902	19	62	35	1443
	has_image	433	3	35	30	371
Fake	no_image	274	3	0	1	237
	has_image	107	1	1	6	97
All		1716	26	98	72	2148

We can quickly identify a severe imbalance in this dataset by looking at the data statistics table. The difference in fake/real labels, the quantity of articles between domains, and the heterogeneity in the number of photos all contribute to this imbalance. Only popular domains have a large number of related articles such as: *society, disaster, economics, politics.* This is also true in practice, when a large amount of fake news that confuses public opinion originates from ordinary social concerns, breaking news related to the economic and political situation, and especially news about epidemics and disasters.

3 v3MFND: A Deep Multi-domain Multimodal Fake News Dectection Model for Vietnamese

In this section, we present our novel model named **v3MFND**: A Deep **M**ulti-domain **M**ulti**M**odal **F**ake **N**ews **D**ectection Model for **V**ietnamese. Figure 1

depicts the overall architecture of **v3MFND**. The model mainly consists of four components: pre-processing phase, modalities representation learning, fusion layer, and a classification layer.

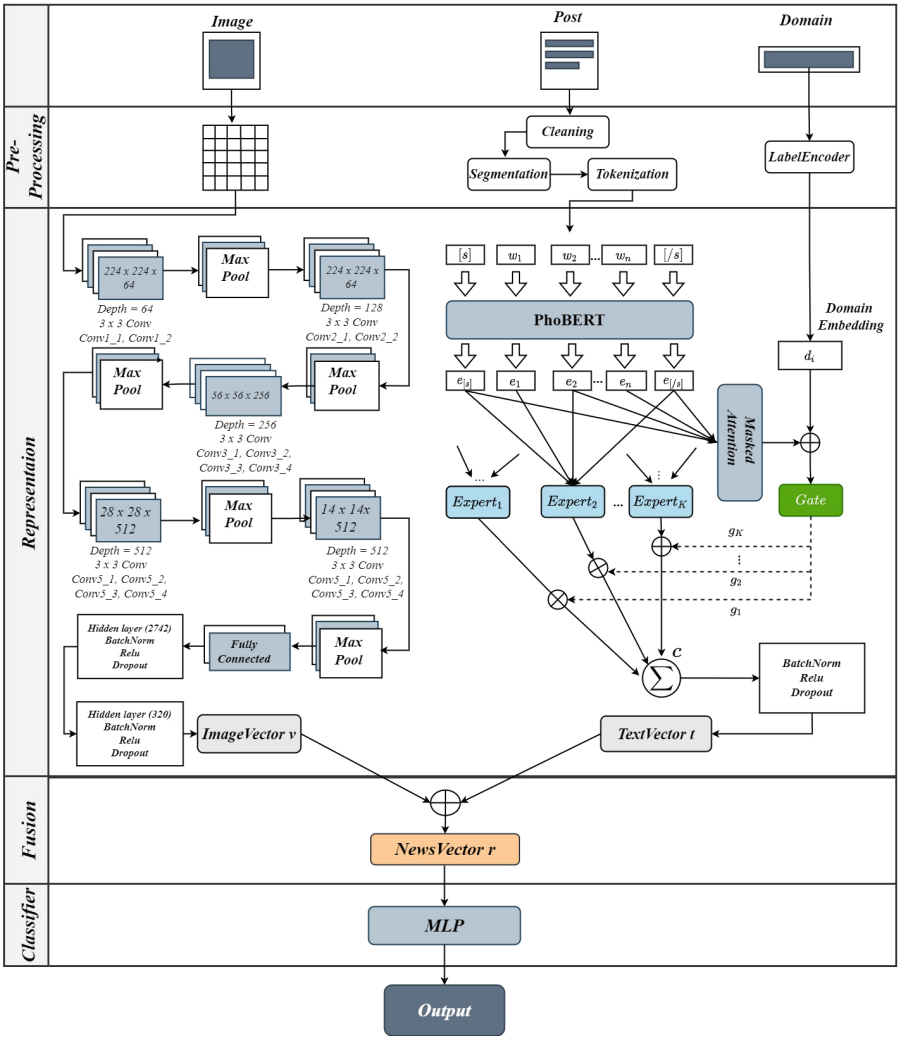

Fig. 1. An overview of v3MFND architecture.

Given a dataset $\mathcal{D}$, with input consisting of two modalities: $\mathcal{C}$ containing the contents of n news items C_i, $\mathcal{V}$ consists of l news' photos V_{ij}, $1 \leq i \leq n$, $= 1 \leq j \leq l$. The input dataset is fed through a separate pre-processing phase for each modality. Text data will be cleaned, normalized, then segmented

using VNCoreNLP [18] and tokenized using PhoBERT Tokenizer [10]. Because one news may have zero or many images, a consistent procedure is applied for synchronization: (1) randomly select an image to be attached as the featured image in the post, (2) for posts without images, use a black image (value 0) as the attacked image, and (3) the images are then resized to $224 \times 224 \times 3$ size. We are also performing the label encoder with the domain data.

3.1 Multimodal Representation Learning

Textual Feature Extractor: For each news content C_i, the text content after going through pre-processing phase is tokenized into a sequence of tokens denoted as $W = [w_1, w_2, \ldots, w_n]$, where n is the total number of tokens. To correctly interpret input, we additionally add special tokens [< s>] and [</s>] into W to obtain new sequence $W_s = [[<s>], w_1, w_2, \ldots, w_n, [</s>]]$.

Pre-trained language models, particularly BERT [1] based on the architecture of Transformer [17], have lately gained a lot of traction and have achieved state-of-the-art improvements on different NLP tasks. PhoBERT [10] is the best pre-trained BERT model for Vietnamese based on RoBERTa [7], which optimized the BERT pre-training strategy for more robust results. Thus, instead of a vector representation of the text, we leveraged PhoBERT to extract word embeddings contain textual information, which is denoted as $E = [e_{[<s>]}, e_1, e_2, \ldots, e_n, e_{[</s>]}]$. Each word embedding e_i is a textual feature since it provides information about the entire text content.

In MDFEND [9], the authors leverage advantage of Mixture-of-Expert [3,8] to extract the news' representations for multiple domains. Because each expert specializes in a specific domain, the news representation retrieved by a single expert may only contain incomplete information, and hence may not fully describe the features of news content. Thus, it motivates us to use multiple experts networks, e.g., TextCNN [5] in **v3MFND**. A representation extracted by an expert network denoted as follows:

$$p_i = \Phi_i(E, \gamma_i) \tag{1}$$

where Φ_i is an expert network, $1 \leq i \leq K$, γ_i represents the parameters to be learned and K is a hyper-parameter that indicates the number of TextCNN model.

MDFEND [9] also employs a domain gate with the domain embedding as well as sentence embedding as input to guide the selection process. In order to assist with domain-specific representation extraction. We construct a domain embedding $D = [d_1, d_2, \ldots, d_m]$ where m is the total number of domains. This selection method produces a vector v representing each expert's weight ratio as:

$$g = SoftMax(Gate(D \oplus S; \rho)) \tag{2}$$

where $Gate(D \oplus S; \rho)$ is a domain gate feed-forward network, S is sentence-level embedding obtained from mask-attention network and ρ is the parameters in

the domain gate. Softmax funtion is used to nomalize the output of $Gate(.;\rho)$. The news' final textual representation vector is obtained via:

$$t = \sum_{i=1}^{K} g_i p_i \tag{3}$$

Visual Feature Extractor: VGG19 (a.k.a., VGGNet-19), a pre-trained model on the ImageNet dataset [12], is a variant of the VGG model which in short consists of 19 layers including 16 convolution layers, 3 Fully connected layers, 5 MaxPool layers and 1 SoftMax layer. In **v3MFND**, the VGG-19 model is used to learn different visual features. We extract the output of the VGG-19 convolutional network's second last layer, denoted as o. The final visual representation v is acquired by passing o via a fully connected layer to reduce down to a final dimension of length 14 as follows:

$$v = \sigma(W.o) \tag{4}$$

where W is the weight matrix of the fully connected layer in the visual feature extractor.

3.2 Multimodal Fusion

To obtain the desired news representation r, the two feature vectors received from separate modalities are fused together using a simple concatenation approach. It is not only combines different visual features, but also reflects the dependencies between textual features, visual features and meta-data in the same news.

$$r = t \oplus v \tag{5}$$

Additionally, we also conducts experiments with some other pseudo-combination methods besides concat, which are addition and average.

- **Sum:** the vectors are added together to produce a representative vector representing the entire news by the following formula:

$$r = t + r \tag{6}$$

- **Average:** the vectors are averaged to give a representative vector representing the entire news using the following formula:

$$r = \frac{t + r}{2} \tag{7}$$

3.3 Learning

The final feature vector of the news is fed into the classifier, which is a Multilayer Perception (MLP) network with a SoftMax function to make prediction as follows:

$$\hat{y} = SoftMax(r) \tag{8}$$

The fake news detector's purpose is to determine whether or not the news is fake, the loss function is set to Binary Cross-Entropy as follows.

$$L(\theta) = -ylog(\hat{y} - (1-y)log(1-\hat{y}) \tag{9}$$

where θ indicates all of the proposed model's learnable parameters, and $y \in \{0, 1\}$ signifies the ground-truth label.

4 Experimental Setup

4.1 Dataset

The ReINTEL dataset [6], which was utilized in the evaluation campaign for the 7th International Workshop on Vietnamese Language and Speech Processing (VLSP 2020) on detecting fake news on Vietnamese SNSs, was used to evaluate our model with the new version named M2-ReINTEL. Details of M2-ReINTEL dataset and how we proceed to label to obtain the final experimental dataset are described in Sect. 2 above.

4.2 Baselines

To our best knowledge, there is no multimodal multi-domain model has been proposed up to now. In order to evaluate the effectiveness of our proposed model v3MFND, we seek to compare against two baseline models namely MDFEND [9] (multi-domain model) and SpotFake [14] (multimodal model).

The dataset also provides metadata, which is also mined by many competing teams [16], so we conduct more experiments to evaluate the role of metadata in the model. We also examine different fusion operations, such as average and sum, in addition to the concat given in the model.

4.3 Experimental Settings

After tuning the proposed model on various settings to find the optimal value, the final settings is shown in Table 3.

4.4 Evaluation Metrics

For each experiment, we report Area Under the ROC Curve (AUC-ROC) as the performance measure. The fundamental reason is that due to data imbalance, traditional binary classification model assessment metrics including accuracy, precision, recall, and F-1 score [2] do not appropriately assess the model's efficacy. The performance is rounded to four digits.

Table 3. Parameters configuration

Parameters	10 domains	5 domains
num "expert"	6	3
Optimizer	Adam	
Learning rate	0.01	0.001
Batch size	32	
MLP	1 dense layer (384 unit) + batch norm 1d + relu + drop_out = 0.4	
Resize image	dense layer (2742 unit) + batch norm 1d + relu + drop_out = 0.4 dense layer (320 unit) + batch norm 1d + relu + drop_out = 0.4	
Resize metadata	dense layer (320 unit) + batch norm 1d + relu + drop_out = 0.4	
Norm text data (fusion)	batch norm 1d + relu + drop_out = 0.4	
Max len of sentence	170	
Word embedding vector dimension	768	

5 Results

In this section, we will report and discuss the experimental results of the multimodal multi-domain deep learning model applied to Vietnamese (v3MFND) and baseline models on M2-ReINTEL.

We conducted experiments using dataset on all 10 domains and the 5 domains with the largest amount of labels due to the imbalance of data regarding domains and labels (disaster, social, economics, health and politics). Notice that, in the model/fusion type column, for example v3MFND - concat, it means the v3MFND model with the fusion type is concat. With a value of "–", it means that the result cannot be calculated. The result values in bold are the highest values by domain, underlined the highest of the fusion types.

With the results presented in Table 4 and Table 5, we can see that the proposed model v3MFND, with concat in the fusion phase, performs well across the board. The MDFEND model scored 0.9753 and 0.9548 in the Disaster and Social domains, respectively, whereas v3MFN scored 0.9294, which was lower

Table 4. Results on 10 data domains with measurement AUC-ROC

Model/Fusion type		Disasters	Education	Entertainment	Economic	Health
MDFEND		**0.9753**	*1.0*	–	**0.9710**	0.875
SpotFake		0.9216	*1.0*	–	1.0	1.0
$v3MDFN_{meta}$	concat	0.9597	*1.0*	–	**0.9710**	0.875
	mean	0.9534	*1.0*	–	0.9565	1.0
	sum	0.9516	*1.0*	–	0.8406	1.0
$v3MDFN_{img+meta}$	concat	0.8324	*1.0*	–	0.3623	1.0
	mean	0.7903	*1.0*	–	0.6377	0.5
	sum	0.7126	*1.0*	–	0.5362	1.0
$v3MDFN_{ProposedModel}$	concat	0.7781	*1.0*	–	0.7826	1.0
	mean	0.8988	*1.0*	–	0.942	1.0
	sum	0.8313	*1.0*	–	0.7681	1.0
Model/Fusion type		Military	Politics	Science	Society	Sport
MDFEND		–	0.6429	–	**0.9548**	–
SpotFake		–	0.7302	–	0.8359	–
$v3MDFN_{meta}$	concat	–	0.8036	–	0.9277	–
	mean	–	0.6786	–	0.9243	–
	sum	–	**0.8750**	–	0.9294	–
$v3MDFN_{img+meta}$	concat	–	0.7679	–	0.7049	–
	mean	–	0.7857	–	0.6794	–
	sum	–	0.7143	–	0.6085	–
$v3MDFN_{ProposedModel}$	concat	–	0.6250	–	0.6696	–
	mean	–	0.6250	–	0.7356	–
	sum	–	0.7679	–	0.8286	–

Table 5. Average results across 10 data domains with measurement AUC-ROC

Model/Fusion type		All
MDFEND		**0.9576**
SpotFake		0.8872
$v3MDFN_{meta}$	concat	0.7178
	mean	0.7046
	sum	0.6669
$v3MDFN_{img+meta}$	concat	0.7195
	mean	0.7641
	sum	0.8198
$v3MDFN_{ProposedModel}$	concat	0.9404
	mean	0.9364
	sum	0.9399

than MDFEND's 0.0254. (2.6 %). All models projected properly in the Health and Education domains, which is owing to a severe data scarcity that caused the model to overfit. The SpotFake model received the highest score of 1.0 in the Ecocomics domain. The rationale is similar to that of Education and Health.

The v3MFND model with fusion as 'concat' produced the best results of 1.0 in the Political domain. The AUC - ROC measure cannot be used to calculate the results in the domains of entertainment, military, science, and sports because the number of news news items in these domains is very small, and the majority of these news items are real news. As a result, when dividing the data set, these domains only have real news.

The models that deal with the multi-domain problem, MDFEND and v3MFND, give better overall results than the SpotFake model. In addition, with the use of more metadata in the v3MFND model, the model not only ineffective, but also reduced the performance of the model. Besides, in the v3MFND model, in most domains, the 'concat' fusion gives higher results than other fusion methods.

Table 6. Results on 5 data domains with measurement AUC-ROC

Model/Fusion type		Disasters	Finance	Health	Politics	Society	All
$MDFEND$		0.9662	**1.0**	**1.0**	**0.8333**	0.8724	0.9263
$SpotFake$		0.944	0.9697	0.9394	0.7407	0.805	0.8876
$v3MDFN_{meta}$	concat	0.8006	0.6875	0.6	0.875	0.7546	0.7748
	mean	0.6935	0.625	0.65	0.8333	0.7544	0.6673
	sum	0.7504	0.9062	0.45	0.8333	0.7357	0.7309
$v3MDFN_{img\_meta}$	concat	0.7134	0.8125	0.5	**0.9583**	0.6413	0.6812
	mean	0.8728	0.9062	0.9	0.75	0.7904	0.8266
	sum	0.7169	0.6875	1.0	0.6667	0.706	0.72435
$v3MDFN_{ProposedModel}$	concat	**0.9754**	**1.0**	**1.0**	0.625	**0.8858**	**0.9375**
	mean	0.9594	**1.0**	**1.0**	0.7917	0.8332	0.9080
	sum	0.9657	**1.0**	**1.0**	0.625	0.8493	0.9157

In Table 6, we can see that our proposed model v3MFND as concat fusion gives the best results on most domains and on the entire evaluation set. Specifically, on the Disaster, Social, Economics and Health domain, the v3MFND model achieved 0.9754, 0.8858, 1.0 and 1.0, respectively, with the AUC-ROC measure. As for the entire evaluation set, the v3MFND model reached 0.9375.

From the above results table, we can also see that MDFEND and v3MFND models with multi-domain problem handling give better results in most domains and overall better results than SpotFake model. In addition, with the use of more metadata on the v3MFND model, the overall performance of the model was reduced, but on some domains such as Politics, the $v3MFND_{metadata}$ model gave the highest result of 0.9583 with a precision of 0.9583.

6 Conclusion

In this study, we investigate the problem of detecting multi-domain, multi-modal fake news for Vietnamese. As the first contribution, we build a multi-domain, multi-method fake news dataset for Vietnamese up on the ReINTEL dataset, whereby assigns domain names to news items. Subsequently, we propose a multi-domain, multi-method fake news detection framework for Vietnamese named v3MFND using advanced deep learning models to extract features for multi-modal data (text, images) and multi-domain data problem solving. Furthermore, extensive experiments on the newly constructed dataset demonstrate the effectiveness of v3MFND compared against baselines for the fake new detection task.

As future works, the challenge with learning data is that the existing data set employed by the model is still not diverse and general, thus additional data (in particular, fake news items) on some domains, such as sports, entertainment, military, and science, is required. The imbalance of data between domains and the imbalance of labels (fake news, real news) are issues we haven't solved yet. These imbalances occur in tandem rather than their own. Furthermore, when labeling, we discovered that certain items belong not just to one news domain, but to a group of domains. Therefore, our next research will focus on resolving multi-domain data imbalance issues, including not only data disequilibrium between domains but also label imbalance.

AQ3

Acknowledgment. .

References

1. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
2. Hossin, M., Sulaiman, M.N.: A review on evaluation metrics for data classification evaluations. Int. J. Data Mining Knowl. Manage. Process **5**(2), 1 (2015)
3. Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E.: Adaptive mixtures of local experts. Neural Comput. **3**(1), 79–87 (1991)
4. Khattar, D., Goud, J.S., Gupta, M., Varma, V.: MVAE: multimodal variational autoencoder for fake news detection. In: The World Wide Web Conference, pp. 2915–2921 (2019)
5. Kim, Y.: Convolutional neural networks for sentence classification. CoRR abs/1408.5882 (2014). https://arxiv.org/abs/1408.5882
6. Le, D.T., et al.: ReINTEL: a multimodal data challenge for responsible information identification on social network sites. In: Proceedings of the 7th International Workshop on Vietnamese Language and Speech Processing, pp. 84–91. Association for Computational Lingustics, Hanoi, Vietnam (2020). https://aclanthology.org/2020.vlsp-1.16
7. Liu, Y., et al.: Roberta: a robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)

8. Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., Chi, E.H.: Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1930–1939 (2018)
9. Nan, Q., Cao, J., Zhu, Y., Wang, Y., Li, J.: Mdfend: multi-domain fake news detection. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pp. 3343–3347 (2021)
10. Nguyen, D.Q., Nguyen, A.T.: Phobert: pre-trained language models for vietnamese. arXiv preprint arXiv:2003.00744 (2020)
11. Shu, K., Wang, S., Liu, H.: Beyond news contents: the role of social context for fake news detection. In: Proceedings of the twelfth ACM International Conference on Web Search and Data Mining, pp. 312–320 (2019)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
13. Singhal, S., Kabra, A., Sharma, M., Shah, R.R., Chakraborty, T., Kumaraguru, P.: Spotfake+: a multimodal framework for fake news detection via transfer learning (student abstract). In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13915–13916 (2020)
14. Singhal, S., Shah, R.R., Chakraborty, T., Kumaraguru, P., Satoh, S.: Spotfake: a multi-modal framework for fake news detection. In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), pp. 39–47. IEEE (2019)
15. Song, C., Ning, N., Zhang, Y., Wu, B.: A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks. Inf. Process. Manage. **58**(1) (2021)
16. Tuan, N.M.D., Minh, P.Q.N.: Reintel challenge 2020: a multimodal ensemble model for detecting unreliable information on vietnamese sns. arXiv preprint arXiv:2012.10267 (2020)
17. Vaswani, A., et al.: Attention is all you need. Adv. Neural Inf. Process. Syst. **30** (2017)
18. Vu, T., Nguyen, D.Q., Nguyen, D.Q., Dras, M., Johnson, M.: Vncorenlp: a vietnamese natural language processing toolkit. arXiv preprint arXiv:1801.01331 (2018)
19. Weiss, K., Khoshgoftaar, T.M., Wang, D.: A survey of transfer learning. J. Big Data **3**(1), 1–40 (2016)
20. Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., Le, Q.V.: Xlnet: generalized autoregressive pretraining for language understanding (2019). https://doi.org/10.48550/ARXIV.1906.08237, https://arxiv.org/abs/1906.08237

Author Queries

Chapter 49

Query Refs.	Details Required	Author's response
AQ1	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	
AQ2	As per Springer style, both city and country names must be present in the affiliations. Accordingly, we have inserted the city and country names in the affiliation. Please check and confirm if the inserted city and country names are correct. If not, please provide us with the correct city and country names.	
AQ3	Kindly note that the "Acknowledgment" has no content. Please check and confirm.	