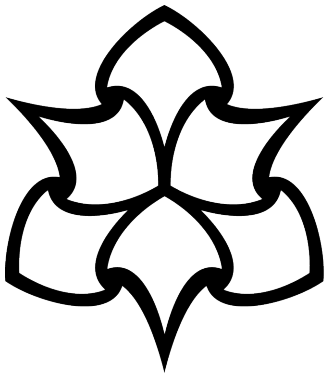

Introduction to Special Issue on Misinformation, Fake News and Rumor Detection in Low-Resource Languages

The online information chaos is undoubtedly a non-trivial combination of misinformation (honest mistakes), dis-information (rumors and fake news), and mal-information (leaks & cyberhate). The special issue presents the use of the latest state-of-the-art techniques and novel solutions to false information resolution on varying platforms, for low-resource monolingual content or multilingual code-mix and code-switch content. An open call-for-papers was issued for this special issue and the response from the research community was overwhelming. After undergoing exhaustive peer-review the guest editors have carefully selected twelve papers.

The papers fall under four key cyber issue categories that lead to social media toxicity, namely: offensive language detection, fake news detection, rumour detection, and spam reviews. A brief summary of the research papers included in each category is listed as follows:

Offensive Language Detection:
Cyberspace is an environment for the deployment of a variety of hostile, direct, and indirect behavioral strategies to attack people or groups. Existing studies on identification of these hostile behaviors and attacks are primarily using English texts and datasets.

Ranasinghe and Zampieri [1] apply this cross-lingual contextual word embeddings and transfer learning to the available English datasets containing offensive language to make predictions in low-resource languages. They project predictions on comparable data in Arabic, Bengali, Danish, Greek, Hindi, Spanish, and Turkish. They use the benchmark datasets of SemEval-2019 Task 5 and OffensEval 2020. Their model achieved competitive performance and confirms the robustness of cross-lingual contextual embeddings and transfer learning for this task. The work by Sangwan and Bhatia [2] puts forward a text based denigration detection model based on attention residual networks having ResNet blocks. They named their mode **Hindi Denigrate Comment-Attention Residual Network (HDC-ARN)** which is validated on Hindi comments on trending events in India, including Tablighi Jamaat spiked Covid-19 and Sushant Singh Rajput Death. Their model achieved an F1 score of 0.642. Another work by Bhowmick et al. [3] on derogatory social media post detection adds pictures with text to propose a multimodal framework. They use Memes pictures along with textual remarks and comments for derogatory post identification. By training an encapsulated transformer network. Their proposed framework combine multilingual text analysis using **Bidirectional Encoder Representations from Transformers (BERT)** along with Meme picture comprehension using face recognition OCR model. Also, a new Facebook meme-post dataset is created by the authors and baseline results are provided.

Fake News Detection:
Fake news has become a major topic of study among researchers and the general public. Because of the huge stakes, automatic detection of fake news has arisen as one of the major issues. Most of the

work for detecting fake news focuses on the English language. However, automated detection of fake news is important irrespective of the language used for spreading false information.

Das et al. [4] annotate and detect legitimacy in the Bengali e-papers. Both shallow and deep classifiers are used with the semantic properties of Bengali sentences to annotate legitimacy in Bengali. The annotation model is validated using supervised machine learning algorithms with the lexical features, syntactic features, and domain-specific features. Jain et al. [5] put forward a study employing various deep learning and machine learning techniques for fake news detection in Bengali. They use a publicly available "ban fake news dataset" to validate the models. The deep learning techniques used by them include l**ong short-term Memory (LSTM),** and **Bi-directional long short-term Memory (BiLSTM),** and it is concluded that BiLSTM gives superlative accuracy of 55.92%. Various machine learning techniques used by them are Naive Bayes, **Passive Aggressive Classifier (PAC),** and Random Forest. Out of the machine learning techniques, Random Forest outperforms other methods with an accuracy of 62.37%. Similarly, fake news detection model in low resource Urdu language has also been proposed based on CNN based feature extraction and ensemble learning classification. Saeed et al. [6] propose a model based on deep contextual semantics learned from CNN. The features extracted from CNN combined with n-gram features and are used to train ensemble learning **Majority Voting (MV)** classifier. It has been concluded that deep learning extracted contextual semantics combined with standard features help in improving results of conventional ensemble learning. The proposed framework outperforms the state-of-the-art Urdu fake news detection models.

Other than conventional machine learning and deep learning models, newer models based on transformers have also been explored by researchers for fake news detection in low resource languages. Samadi et al. [7] propose a BERT model for fake news detection in the Persian language. The BERT is used with a pool-based representation for providing a representation for the whole document and a sequence representation for providing a representation for every token of the document. Single layer perceptron and convolutional neural networks are used with the BERT model to detect fake news as well as to extract features. The proposed model is evaluated on three datasets including two Persian rumor datasets and a newly proposed TAJ dataset made up of news from different news agencies. The results show that the proposed BERT model achieves superior performance to the previous baseline results.

In another work, De et al. [8] utilize the BERT algorithm for domain-agnostic multilingual fake-news classification. Effectiveness of language-agnostic feature transfer across different languages was evaluated. Language independent feature transfer of model was evaluated by performing cross-domain transfer experiments. The authors also introduce a multilingual multi-domain fake news detection dataset in five languages and seven different domains.

Apart from classification of fake news, identification of the source of fake news has also been studied. Dhall et al. [9] propose a blockchain and keyed watermarking-based framework for checking the integrity of posts on social media and messaging platforms. The model ensures accountability of the originator of the post. Backtracking and forward tracking of the fake news are used to curb the spread. The proposed framework offers a proactive as well as reactive solution for curtailment of fake and vicious news on social media.

Rumor Detection:
With social media's beneficial applications, such as socialization and news broadcasting, it also serves as a platform for disseminating rumors. Rumors can endanger the security of society in normal or critical situations. Therefore, it is important to detect rumors in all languages.

Gumaei et al. [10] propose a rumor detection model based on **eXtreme gradient boosting (XGBoost)** classifier and validate their model on public dataset comprising Arabic tweets. Their model

uses content-based, user-based, and topic-based features and it is concluded that the model outperforms existing methodologies identifying rumors in Arabic text. In a work by Nagadeh et al. [11], the role of content in spreading rumors is explored. They propose a content-based model to verify the Persian rumors on Telegram and Twitter. The model uses fusion of semantic, pragmatic, and syntactic information. ParsBERT and parallel CapsNets are used with contextual word embeddings of the source rumor which are then concatenated with the extracted pragmatic and syntactic features of the rumor. The results validate the model's effectiveness in the early stage of rumor when social only limited content information is available lacking rumors' social and structural features.

Spam Reviews:
Review spam is a persistent and detrimental problem that can cause loss in consumer trust and it is imperative to develop methods to help businesses separate truthful reviews from fake ones.

Najadat et al. [12] propose a keyword-based method for detecting Arabic spam reviews in Facebook comments. Term's weight, TF-IDF are used with filter feature section methods including information gain, chi-squared, deviation, correlation, and uncertainty to extract keywords from text. Baseline machine learning algorithms including C4.5, kNN, SVM, and Naive Bayes are used to detect Arabic spam reviews. It is concluded that decision tree outperforms other classifiers in the prediction of spam reviews.

REFERENCES

[1] T. Ranasinghe and M. Zampieri. 2021. Multilingual offensive language identification for low-resource languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[2] S. R. Sangwan and M. P. S. Bhatia. 2021. Denigrate comment detection in low-resource Hindi language using attention-based residual networks. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[3] R. S. Bhowmick, I. Ganguli, J. Paul, and J. Sil. 2021. A multimodal deep framework for derogatory social media post identification of recognized person. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[4] S. Das, P. Rai, and S. Chatterji. 2021. Deep level analysis of legitimacy in Bengali news sentences. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[5] R. Jain, D. K. Jain, Dharna, and N. Sharma. 2021. Fake news classification: A quantitative research description. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[6] R. Saeed, H. Afzal, H. Abbas, and M. Fatima. 2021. Enriching conventional ensemble learner with deep contextual semantics to detect fake news in Urdu. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[7] M. Samadi, M. Mousavian, and S. Momtazi. 2021. Persian fake news detection: Neural representation and classification at word and text levels. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[8] A. De, D. Bandyopadhyay, B. Gain, and A. Ekbal. 2021. A transformer based approach to multilingual fake news detection in low-resource languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[9] S. Dhall, A. D. Dwivedi, S. K. Pal, and G. Srivastawa. 2021. Blockchain based framework for reducing fake or vicious news spread on social media/messaging platforms. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[10] A. Gumaei, M. S. Al-Rakhami, M. H. Hassan, V. H. C. De Albuquerque, and D. Camacho. 2021. An effective approach for rumor detection of Arabic tweets using extreme gradient boosting method. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[11] Z. J. Nagadeh, M. R. F. Derakhshi, and A. Sharifi. 2021. A deep content-based model for Persian rumor verification. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

[12] H. Najadat, M. Alzubaidi, and I. Qarqaz. 2021. Detecting Arabic spam reviews in social networks based on classification algorithms. *ACM Transactions on Asian and Low-Resource Language Information Processing*.