

# KU\_NLP@LT-EDI-EACL2021: A Multilingual Hope Speech Detection for Equality, Diversity, and Inclusion using Context Aware Embeddings

**Junaida M K**

Department of Information  
Technology  
Kannur University  
Kerala, India

junaidashukoor@gmail.com

**Ajees A P**

Department of Information  
Technology  
Kannur University  
Kerala, India

ajeesap87@gmail.com

## Abstract

Hope speech detection is a new task for finding and highlighting positive comments or supporting content from user-generated social media comments. For this task, we have used a Shared Task multilingual dataset on Hope Speech Detection for Equality, Diversity, and Inclusion (HopeEDI) for three languages English, code-switched Tamil and Malayalam. In this paper, we present deep learning techniques using context-aware string embeddings for word representations and Recurrent Neural Network (RNN) and pooled document embeddings for text representation. We have evaluated and compared the three models for each language with different approaches. Our proposed methodology works fine and achieved higher performance than baselines. The highest weighted average F-scores of 0.93, 0.58, and 0.84 are obtained on the task organisers' final evaluation test set. The proposed models are outperforming the baselines by 3%, 2% and 11% in absolute terms for English, Tamil and Malayalam respectively.

## 1 Introduction

In recent years, social media became an integral part of human life and people started spending more time on these platforms. But people are mindful of social media behaviour and putting less personal information in the public domain (Thavaresan and Mahesan, 2019, 2020a,b). Now social media behaviors changed quite dramatically and we are living not just in a pandemic, but also in an "infodemic", where fake news is becoming more common (Lima et al., 2020). Conversations on the internet are often a reflection of the conversations that one makes offline.

Several AI techniques are adopted to analyse the online comments in social media, which are intensified on the detection of negative com-

ments such as hate speech detection, offensive language identification and abusive language detection (Chakravarthi, 2020a). Hate speech is widely used in media, internet and public discourse which seen or read in the media expressing disapproval, hatred, or aggression towards minorities, could lead to violence and form negative impact on minorities. Hope speech detection is a new task related to Hate speech detection for finding and highlighting positive comments or supporting content, rather than just filtering hostile content (Chakravarthi et al., 2020a). For this task, YouTube comments/posts that offer support, reassurance, suggestions, inspiration and insight are recognized as hope speech. In bilingual and multilingual communities linguistic code-switching occurs in social groups (Chakravarthi et al., 2018, 2019; Chakravarthi, 2020b). It is incredibly important in many social groups, When an individual uses a group's dialect or accent, the audience is more receptive to the content (Jose et al., 2020; Priyadharshini et al., 2020). Recently, many researchers are focused on high resource languages using monolingual corpora but less attention is given to code-switching languages especially under resourced languages like Indian languages (Chakravarthi et al., 2020b; Mandl et al., 2020). In our work, we used Hope Speech dataset for Equality, Diversity and Inclusion (HopeEDI) not only in English but also code-switched Tamil and Malayalam (Chakravarthi, 2020a).

## 2 Related Work

The authors (Puranik et al., 2021; Ghanghor et al., 2021) introduced a novel task for detecting hostility diffusing content from comments in social media, dubbed hope-speech detection. The authors analysed and studied the importance of automatic identification of user-generated hope speech web content that diffuse tension and violence among

Language	Train	Dev	Test	Total
English	22,762	2,843	2,846	28,451
Tamil	16,160	2,018	2,020	20,198
Malayalam	8564	1070	1071	10,705

Table 1: Train-Development-Test Split

people in an international crisis. Finally, the obtained results are very promising and automatic recognition of hope speech may also find applications in many other contexts. But they restricted the definition of hope into diffuse tension and violence not considering the other perspectives of hope.

Chakravarthi (2020a) constructed a multilingual Hope Speech dataset for Equality, Diversity and Inclusion (HopeEDI) containing user-generated comments from YouTube for English and two low-resource languages, Malayalam and Tamil. The authors considered much more perspectives support, reassurance, suggestions, inspiration and insight of the hope and EDI. To facilitate future research on encouraging positivity, the authors make this dataset publicly available and created several baselines to benchmark the proposed dataset.

### 3 Materials and Methods

This section describes the dataset used for our experiments and technical description of the proposed methodology.

#### 3.1 Dataset

For our work, we used the Shared Task dataset on Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI 2021- EACL 2021 (Chakravarthi and Muralidaran, 2021). The dataset contains YouTube comments from English, code-switched Tamil and Malayalam. This is considered a multilingual resource to allow cross-lingual studies and approaches. The corpus consists of a total of 59,354 comments from YouTube videos, where 28,451 comments are in English, 20,198 comments are in Tamil, and the remaining 10,705 comments are in Malayalam. The dataset was manually annotated with three different labels: Hope Speech, Not-Hope Speech and Other languages, where Other languages refer to comments that were not in the intended language. The Figure 1 shows, distribution of three classes in shared task dataset. For our work, we have used the training, validation and test set of the shared task as depicted in Table 1.

#### 3.2 Methods

Hope speech detection is a form of text classification, which classify sentences or documents into specified categories. Most current state of art approaches to text classification rely on a technique called text embedding. The embeddings of words in a sentence is used to make a vector representation of the sentence. The sentence embeddings can be achieved in many ways. It could be done by convolutional neural networks (CNN) (Kim, 2014), by averaging word vectors (Iyyer et al., 2015) or by using Recurrent Neural Networks(RNNs) (Dai and Le, 2015).

In our proposed method we have used Deep contextual and fixed (non-contextual) embeddings to derive word representations. Then, puts them into an RNN or does a pooling operation on overall word embeddings to obtain a text representation. Finally, a softmax layer accepts the text representations to get the actual class label (Akbik et al., 2019b). We have implemented three models for each language using contextualized string embeddings (*flair*) in a **FLAIR** framework (Akbik et al., 2019a).

**Contextualized String Embedding (*Flair*) :** (Akbik et al., 2018) proposed a novel word embedding from internal states of a trained character language model and termed as *contextual string embeddings*. The two primary factors powering contextual string embeddings (*Flair*) are, words are trained as characters and embeddings are contextualised by their surrounding text. Hence, it is treated as character language model(charLM) and it perhaps the biggest benefit of using over a word-based language model when a word has not been seen in the training data. The Flair embeddings are obtained by training two LSTM -based language models, forward and backward. The final word embedding is the concatenation of two specific hidden states from two language models. Let,  $t_1, t_2, \dots, t_n$  be the character indices. Furthermore, let  $h_t^f, h_t^b$  be the hidden states at character position  $t$  for forward and backward LM respectively. The *Flair* embedding  $w_k$  for the word  $w_k$  is defined as given in Equation 1.

$$w_k = \left[ h_{t_{k+1}-1}^f; h_{t_k-1}^b \right] \quad (1)$$

**Stacked Embedding :** In this method we combine different embeddings, such as *word2vec*, *Glove*, *FastText* along with embeddings generated from *Flair* language models(Akbik et al., 2019a).

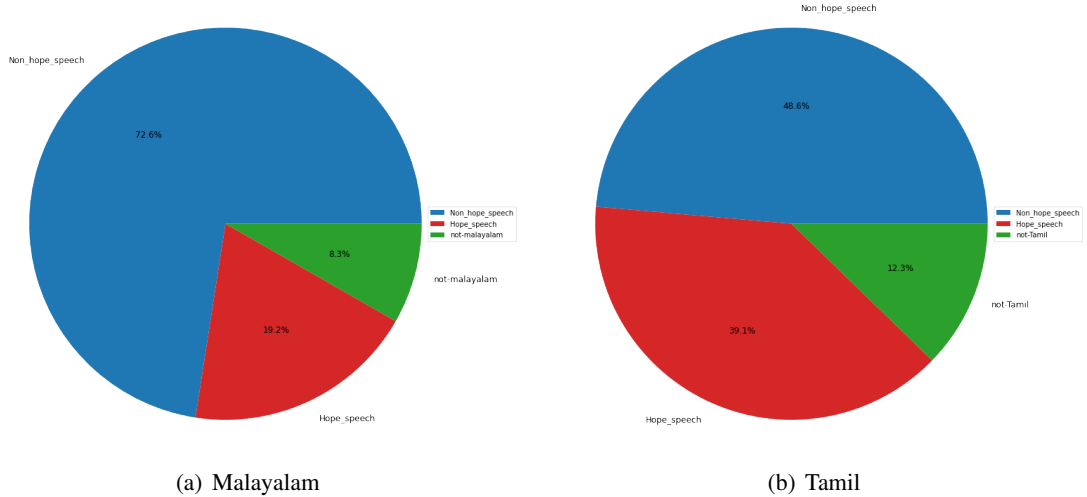


Figure 1: Distribution of three classes over the Hope Speech Detection for Equality, Diversity, and Inclusion at LT-EDI 2021-EACL 2021 Shared Task dataset for English, Malayalam and Tamil languages

**Document Pool Embeddings :** It is a simple document embedding, calculate a pooling operation (mean or max or min) over a list of token embeddings in a document. The default operation is mean which gives the mean of all words in the sentence (Akbik et al., 2019a).

## 4 Experiments

### 4.1 Experimental Setup

We have utilized the FLAIR framework (Akbik et al., 2019a) for all our experiments with GPU (12 GB) provided by Google Colab. We have trained three models for each language. The first model employed a pretrained *flair* LM for word representation and RNN for text representation (*FLAIR+RNN*). The second model combines pretrained *flair* LM and pretrained word embeddings (PWE) for word representation and RNN is used for text representation (*FLAIR+PWE+RNN*). The third model apply the same word representation and pooled document embedding for text representation (*FLAIR+PWE+Pooled*). The Gated Recurrent Unit (GRU) is used for document RNN embedding and max pool operation is used for pooled document embedding. The Table 2 shows the hyperparameters settings for all our experiments.

For all three languages, We have adopted pretrained *flair* embeddings from FLAIR framework (*en-forward*, *en-backward*, *ml-forward*, *ml-backward*, *ta-forward* and *ta-backward*) (Akbik et al., 2019a). For PWE in English we have also taken the twitter embeddings from same framework. The Malay-

Parameter	Value
Learning rate	0.1
Patience	5
Batch size	32
Anneal factor	0.5
Word dropout	0.05
Loss function	Cross Entropy
Epochs	20

Table 2: Hyper Parameters

alam and Tamil language used the PWE from IndicFT, FastText-based word embeddings (11 languages), which is a 300-dimensional word embeddings for each language on IndicCorp, recently published large monolingual sentence level corpora for 11 Indian languages (Kakwani et al., 2020).

### 4.2 Evaluation

The performance of a text classification model is usually noted as F1-score (harmonic mean of precision and recall) since accuracy can often be misleading in an imbalanced class distribution (Akosa, 2017). For this task, the dataset having imbalanced class distribution as in Figure 1. Due to the imbalance problem, we measured our system performance in terms of weighted averaged Precision, weighted averaged Recall and weighted averaged F-Score across all the three classes. Weighted averaged calculations use average of the support-weighted mean per label.

Language	Classifier	Weighted avg		
		P	R	F
English	<i>DT</i>	0.90	0.90	0.90
	<i>MNB</i>	0.60	0.58	0.56
Tamil	<i>LR</i>	0.58	0.57	0.56
Malayalam	<i>DT</i>	0.73	0.76	0.73

Table 3: Baseline results from (Chakravarthi, 2020a).

Language	Model	Weighted avg		
		P	R	F
English	<i>FLAIR+RNN</i>	0.91	0.92	0.91
	<i>FLAIR+PWE+RNN</i>	0.92	0.93	0.91
	<i>FLAIR+PWE+Pooled</i>	0.92	0.93	<b>0.93</b>
Tamil	<i>FLAIR+RNN</i>	0.58	0.58	0.56
	<i>FLAIR+PWE+RNN</i>	0.60	0.59	<b>0.58</b>
	<i>FLAIR+PWE+Pooled</i>	0.62	0.60	0.56
Malayalam	<i>FLAIR+RNN</i>	0.78	0.82	0.79
	<i>FLAIR+PWE+RNN</i>	0.82	0.86	0.83
	<i>FLAIR+PWE+Pooled</i>	0.84	0.85	<b>0.84</b>

Table 4: Our models results from proposed methodology.

## 5 Results and Analysis

This section presents the results and analysis of our experiments, which we have explained in the previous sections. As well as, we present the baseline results obtained by the authors (Chakravarthi, 2020a), to compare the performance with our proposed models. The final evaluation results obtained for participants’ submissions by task organisers are also presented in overview paper (Chakravarthi and Muralidaran, 2021). The best-performing classifiers on HopeEDI datasets are considered as baseline models for this task (See Table 3). The Table 4 depict the precision, recall and F-score results of trained models for three languages described in section 4.1.

As shown, all our proposed models for Hope EDI detection task in three languages work well and show significant improvement than the best-performed baseline models. The results indicate that a Hope EDI detection classifier with good precision and recall can be constructed using deep learning approaches. It may be due to the representational power of pre-trained word embeddings or language models to capture semantic and lexical structure. It has virtually replaced the feature engineering part of supervised machine learning classifiers.

For English and Malayalam, our model with

stacked word embeddings and pooled document embedding achieved the highest performance on the HopeEDI dataset with a weighted average F-Score of 0.93 and 0.84 respectively. But in Tamil, stacked PWE with RNN document embedding performed better than other models with weighted average F-Score of 0.58. Furthermore, it can be seen that all Malayalam language models (0.79,0.83,0.84) achieved greater improvement than baseline models(0.73). Other two languages exhibit slight improvement in proposed models than baseline models.

In most cases, it can be observed from the Table 4 that the *flair* embedding with PWE gives higher performance than the standalone embeddings. This indicates that combining contextual embedding with non-contextual embeddings achieves noticeably better outcomes. This finding is in line with the findings of (Akbik et al., 2018). The Figures 2 and 3 represents the class level precision, recall and f-scores of best-performed baseline models and best performed proposed models for all three languages. It can be noticed from these figures that the class "not-english" have no significant performance in both models due to the imbalance in shared task dataset (Chakravarthi, 2020a).

Also, we can observe that the "Non-hope-speech" class for English shows similar perfor-

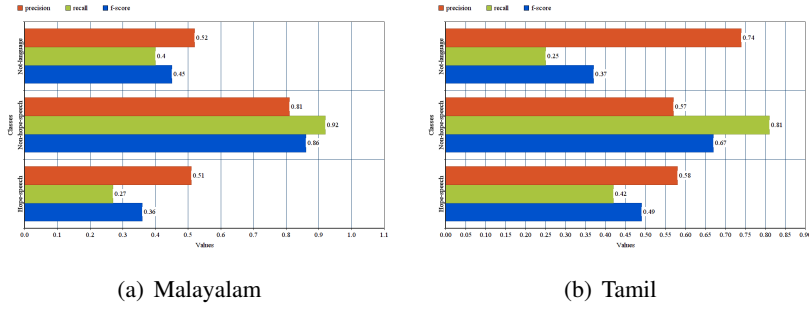


Figure 2: Class level precision, recall and f-scores of baseline models (Chakravarthi, 2020a) for three languages.

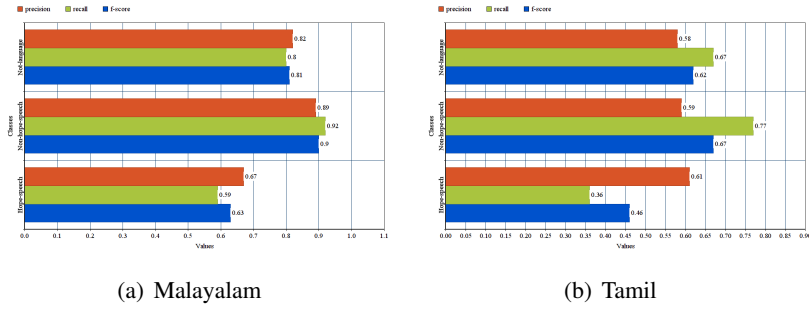


Figure 3: Class level precision, recall and f-scores of best-performed models from our experiments for three languages.

mance in both models and "Hope-speech" class hold a remarkable improvement over baseline models. In Tamil and Malayalam the majority class "Non-hope-speech" possesses approximately similar F-scores for baseline and proposed the best model. Another notable and interesting observation from Figure 2 and 3 is that the two minority classes in Malayalam dataset, "Hope-speech" and "not-malayalam" shows comparable performance approximately 30% above F-score than the baseline model. Also, "not-Tamil" class shows approximately 25% above F-score than the baseline model.

These higher scores indicate that deep learning techniques can classify the correct class regardless of label distribution better than traditional machine learning technique. This may be because our deep learning techniques rely on pre-trained embeddings and language model, which is a contextualized word representation allowing a word to be associated with multiple word vectors, whereas the classical techniques rely on merely manually selected features.

## 6 Conclusion

We have presented a deep learning technique with contextual aware embeddings for Hope speech de-

tection task in three languages: English, Malayalam and Tamil (code-switched). We have used contextual string embedding (flair) and pre-trained word embeddings (PWE) for word representations and Recurrent Neural Network (RNN) and pooled document embedding with max pool operation for text representations. All the three models were evaluated for each language and compared the baseline models using the dataset given by Hope Speech Detection for Equality, Diversity, and Inclusion (HopeEDI) Shared Task organizers. All model performances are measured using weighted average F-score due to the imbalanced class distribution in the dataset.

We have obtained the highest F-scores 0.93, 0.58 and 0.84 for three languages English, Tamil and Malayalam respectively, which significantly improved performance over baselines (0.90, 0.56, and 0.73). For the minority class 'not-language', the proposed best model improved 35% and 25% performance than baselines for Malayalam and Tamil, respectively. Based on these observations, we conclude that the deep learning models with contextual string embeddings are well suited for HopeEDI detection task with an imbalanced dataset. We could also achieve good performance results with moder-

ate resources (one GPU and a small corpus), even without optimizing hyperparameters.

The performance can be improved further by fine-tuning hyperparameters and pre-trained contextual embeddings, incorporating different attention mechanisms and increasing the size of the training dataset.

## References

- Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019a. [FLAIR: An easy-to-use framework for state-of-the-art NLP](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pages 54–59, Minneapolis, Minnesota. Association for Computational Linguistics.
- Alan Akbik, Tanja Bergmann, and Roland Vollgraf. 2019b. [Pooled contextualized embeddings for named entity recognition](#). 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)*, pages 724–728, Minneapolis, Minnesota. Association for Computational Linguistics.
- Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. [Contextual string embeddings for sequence labeling](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- J. Akosa. 2017. Predictive accuracy : A misleading performance measure for highly imbalanced data.
- Bharathi Raja Chakravarthi. 2020a. [HopeEDI: A multilingual hope speech detection dataset for equality, diversity, and inclusion](#). In *Proceedings of the Third Workshop on Computational Modeling of People’s Opinions, Personality, and Emotion’s in Social Media*, pages 41–53, Barcelona, Spain (Online). Association for Computational Linguistics.
- Bharathi Raja Chakravarthi. 2020b. *Leveraging orthographic information to improve machine translation of under-resourced languages*. Ph.D. thesis, NUI Galway.
- Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae. 2018. [Improving wordnets for under-resourced languages using machine translation](#). In *Proceedings of the 9th Global Wordnet Conference*, pages 77–86, Nanyang Technological University (NTU), Singapore. Global Wordnet Association.
- Bharathi Raja Chakravarthi, Mihael Arcan, and John P. McCrae. 2019. [WordNet gloss translation for under-resourced languages using multilingual neural machine translation](#). In *Proceedings of the Second Workshop on Multilingualism at the Intersection of Knowledge Bases and Machine Translation*, pages 1–7, Dublin, Ireland. European Association for Machine Translation.
- Bharathi Raja Chakravarthi, M Anand Kumar, John Philip McCrae, Premjith B, Soman KP, and Thomas Mandl. 2020a. Overview of the track on HASOC-Offensive Language Identification-DravidianCodeMix. In *Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2020)*. CEUR Workshop Proceedings. In: CEUR-WS. org, Hyderabad, India.
- Bharathi Raja Chakravarthi and Vigneshwaran Muralidaran. 2021. Findings of the shared task on Hope Speech Detection for Equality, Diversity, and Inclusion. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.
- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P. McCrae. 2020b. [Overview of the Track on Sentiment Analysis for Dravidian Languages in Code-Mixed Text](#). In *Forum for Information Retrieval Evaluation, FIRE 2020*, page 21–24, New York, NY, USA. Association for Computing Machinery.
- Andrew M. Dai and Quoc V. Le. 2015. [Semi-supervised sequence learning](#).
- Nikhil Kumar Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IITK@LT-EDI-EACL2021: Hope Speech Detection for Equality, Diversity, and Inclusion in Tamil, Malayalam and English. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, Online.
- Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. [Deep unordered composition rivals syntactic methods for text classification](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1681–1691, Beijing, China. Association for Computational Linguistics.
- Navya Jose, Bharathi Raja Chakravarthi, Shardul Suryawanshi, Elizabeth Sherly, and John P. McCrae. 2020. [A Survey of Current Datasets for Code-Switching Research](#). In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 136–141.
- Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. [IndicNLPsuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian](#)

- languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.
- Yoon Kim. 2014. [Convolutional neural networks for sentence classification](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
- Diego Laurentino Lima, Maria Antonieta Albanez A. de Medeiros Lopes, and Ana Maria Brito. 2020. [Social media: friend or foe in the COVID-19 pandemic?](#) *Clinics*, 75.
- Thomas Mandl, Sandip Modha, Anand Kumar M, and Bharathi Raja Chakravarthi. 2020. [Overview of the HASOC Track at FIRE 2020: Hate Speech and Offensive Language Identification in Tamil, Malayalam, Hindi, English and German](#). In *Forum for Information Retrieval Evaluation, FIRE 2020*, page 29–32, New York, NY, USA. Association for Computing Machinery.
- Ruba Priyadharshini, Bharathi Raja Chakravarthi, Mani Vegupatti, and John P. McCrae. 2020. [Named Entity Recognition for Code-Mixed Indian Corpus using Meta Embedding](#). In *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 68–72.
- Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. [IIIT@LT-EDI-EACL2021-Hope Speech Detection: There is always hope in Transformers](#). In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2019. [Sentiment Analysis in Tamil Texts: A Study on Machine Learning Techniques and Feature Representation](#). In *2019 14th Conference on Industrial and Information Systems (ICIIS)*, pages 320–325.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020a. [Sentiment Lexicon Expansion using Word2vec and fastText for Sentiment Prediction in Tamil texts](#). In *2020 Moratuwa Engineering Research Conference (MERCOn)*, pages 272–276.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2020b. [Word embedding-based Part of Speech tagging in Tamil texts](#). In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, pages 478–482.