

SSNCSE_NLP@TamilNLP-ACL2022: Transformer based approach for Emotion analysis in Tamil language

Josephine Varsha & B. Bharathi

Department of CSE

Sri Siva Subramaniya Nadar College of Engineering

Kalavakkam - 603110

josephine2010350@ssn.edu.in

bharathib@ssn.edu.in

Abstract

Emotion analysis is the process of identifying and analyzing the underlying emotions expressed in textual data. Identifying emotions from a textual conversation is a challenging task due to the absence of gestures, vocal intonation, and facial expressions. Once the chatbots and messengers detect and report the emotions of the user, a comfortable conversation can be carried out with no misunderstandings. Our task is to categorize text into a predefined notion of emotion. In this thesis, it is required to classify text into several emotional labels depending on the task. We have adopted the transformer model approach to identify the emotions present in the text sequence. Our task is to identify whether a given comment contains emotion, and the emotion it stands for. The datasets were provided to us by the LT-EDI organizers [Sam-path et al. \(2022\)](#) for two tasks, in the Tamil language. We have evaluated the datasets using the pretrained transformer models and we have obtained the micro-averaged F1 scores as 0.19 and 0.12 for Task1 and Task 2 respectively.

1 Introduction

In today's world, the user has complete liberty to express their opinion on any topic in the form of comments, videos, reels, and reviews. Identifying emotions from a video or a graphic image is simple. By analyzing the body language, facial expressions, and speech modulation we can determine the emotion. However, the identification of emotion from a text is quite challenging due to the absence of discrete evidence. Emotions in the text are not only identified by their cue words such as happy, good, bore, hurt, hate, and fun, but also by the presence of interjections (e.g., "oopsie"), emoticons (e.g., "😄"), idiomatic expressions (e.g., "am on cloud nine"), metaphors (e.g., "sending clouds") and other descriptors mark the existence of emotions in the conversational text ([Thenmozhi et al., 2019](#); [Chakravarthi, 2020](#)). With the growth and

advancement of text messaging applications, it is possible to detect the emotion during conversation and proceed with the conversation with no miscommunications. In the last years, the recognition of emotions has become a multi-disciplinary research area ([Ghanghor et al., 2021a,b](#); [Yasaswini et al., 2021](#)). This plays an important role in HumanMachine interaction ([Ram and Ponnusamy, 2014](#)).

There are three main classification levels in Emotion Analysis: document-level, sentence-level, and aspect-level Emotion Analysis. Document-level Emotion analysis aims at classifying an opinion as a positive or a negative opinion or sentiment. Sentence-level emotion analysis strives to classify the emotion expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level EA will determine whether the sentence expresses positive or negative opinions. The authors [Wilson et al. \(2005\)](#) points out that emotional expressions are not necessarily subjective in nature. The authors [Liu \(2012\)](#) states that there is no fundamental difference between document and sentence level classifications because sentences are just short documents.

Tamil is one of the world's longest-surviving classical languages ([Anita and Subalalitha, 2019b,a](#); [Subalalitha and Poovammal, 2018](#); [Subalalitha, 2019](#)). According to A. K. Ramanujan, it is "the only language of modern India that is recognizably continuous with a classical history." Because of the range and quality of ancient Tamil literature, it has been referred to as "one of the world's major classical traditions and literatures." For about 2600 years, there has been a recorded Tamil literature. The earliest period of Tamil literature, known as Sangam literature, is said to have lasted from from 600 BC to AD 300. Among Dravidian languages, it possesses the oldest existing literature. The earliest epigraphic documents discovered on rock edicts and "hero

stones" date from the 6th century BC (Sakuntharaj and Mahesan, 2021, 2017,?, 2016; Thavareesan and Mahesan, 2019, 2020a,b, 2021a).

Recently there are many research shared tasks on Tamil and other Dravidian languages conducted by researchers (Priyadharshini et al., 2021; Kumaresan et al., 2021; Chakravarthi and Muralidaran, 2021; Chakravarthi et al., 2020b; Sampath et al., 2022; Ravikiran et al., 2022; Chakravarthi et al., 2022; Bharathi et al., 2022; Priyadharshini et al., 2022). The goal of this task is to determine the emotional state of the user who writes comments. In this paper, we will look into the classification, and analyze the different emotions of the YouTube comments. Our focus lies in the study of emotion analysis in Tamil. The datasets were provided by the LT-EDI organizers in the Tamil language in two forms, namely, Task A and Task B, each consisting of a different number of comments or posts. The language constriction poses several challenges due to the limited resources available for the Tamil language. We have used multilingual models such as BERT, XLNet, and m-BERT transformer models to tackle this issue. In this paper, we investigate the efficacy of different learning models in identifying emotions. We then compare the F1-Score of the different transformer models for both datasets and conclude which is the better model.

The remainder of the paper is organized into 5 sections. Section 2 discusses the related works in the field of Artificial Intelligence, on emotion or sentiment analysis for both Tamil, and other languages. The methodology proposed for the model along with the models implemented are elaborately explained in the 3rd section of this paper. In section 4 the results and the observations are discussed. Section 5 concludes the paper.

2 Related Works

In this section, we will be reviewing the research work reported for emotion analysis from the text. The authors Abdelrahman et al. (2016) had proposed an architectural framework to identify the sentiments of both English and Tamil tweets. Tweets were gathered with the help of Twitter API. They had used the word sense disambiguation technique to determine the correct usage of the word sense and went about classifying the sentiments of tweets using a linear classifier like the Support Vector Machine.

K-means clustering and k-nearest neighbor clas-

sifier to predict the sentiments expressed in Tamil texts is used in Thavareesan and Mahesan (2021b). The data points are considered in two different ways for the clustering of the corpus; clustering by considering class-wise information and clustering without considering class-wise information. They extracted features using Tf, BoW, fastText, and word embeddings. The fastText and class-wise clustering method has yielded the best results of accuracy of 89.87

The authors of Jenarathanan et al. (2019) this paper had worked on ACTSEA: Annotated Corpus for Tamil & Sinhala Emotion Analysis, to develop emotion annotated twitter corpus in Sinhala and Tamil Languages. They had adopted the scalable semi-automatic approach and found it an effective process for creating a large-scale emotion corpus with acceptable quality. They've also concluded that it is useful for under-resourced languages.

A research work done by Vas aimed at creating a monolingual corpus for the Tamil language. They advanced the corpus solution and created the TamilEmo, a large dataset for fine-grained emotion detection that has been extensively annotated manually. They've further presented a detailed data analysis that illustrates that the accuracy of the annotations over the whole taxonomy with a high inter-annotator agreement in terms of Krippendorff's alpha

There are many research works on classifying the emotion of document sources into a single type of emotion. In Sharma et al. (2017), provides an insight on how to characterize a person's multiple emotions. The LEXicon Based Emotion Analyzer abbreviated as LEXER is employed to analyze the emotion underlying the text. The proposed method contains a dictionary that has different emotional values for words. Emotion values from the vocabulary are allotted to every expression that's being used in the text. A fuzzy set function is used to complement the emotional value of a negated word. This in comparison to polarity reversal is more realistic and reliable. The lexicon assigns an emotional value that is derived from a fuzzy set function. This is an efficient multi-emotion analyzer model which has still not been applied to the best of our understanding.

The authors Chakravarthi et al. (2020a) put forth a model that aimed at creating a gold standard code-mixed dataset for Malayalam-English that ensured providing enough data for research purposes. They

Language	Training	Development	Test
Task A	14208	3552	4440
Task B	30179	4269	4269

Table 1: Dataset description

used Krippendorff’s α method among the numerous approaches developed, to measure the degree of agreement between annotators. They used traditional machine learning algorithms such as Logistic regression (LR), Support vector machine (SVM), Decision tree (DT), Random Forest (RF), Multinomial Naive Bayes (MNB), K-nearest neighbors (KNN) on the newly annotated English-Malayalam dataset to show the insights about the dataset.

Seq2Seq deep neural network for detecting the emotions from textual conversations which include a sequence of phrases are adopted in [Thenmozhi et al. \(2019\)](#). The Seq2Seq model is adopted and the sequence of n words is mapped with a target label(n:1 mapping). The sequence was vectorized and sent to the bidirectional LSTM for encoding and decoding.

A study on how sentiment communicates in Dravidian social media language in a code-mixed setting was taken up as a shared task by [Chakravarthi et al. \(2021\)](#). The results of the sentiment analysis shared task on Tamil, Malayalam, and Kannada are presented. The top-performing systems involved the application of attention layers on the contextualized word embeddings.

3 Methodology and Data Pre-processing

In this section, we have illustrated our implementation of the pre-trained machine learning transformer models in detail. Further, we investigate the performance of the various transformer models in the coming sections. The architecture of the proposed model and the steps are given in the Fig. 1.

The dataset provided by the LT-EDI organizers [Sampath et al. \(2022\)](#) for the Tamil Tasks A and B, consisted of 22,200 and 38,717 posts/comments respectively. The details are given in Table 1.

In task A we were provided with data annotated for 8-10 emotions for social media comments in Tamil. In task B we were provided with data annotated for fine-grained 30 emotions for social media comments in Tamil [Sampath et al. \(2022\)](#).

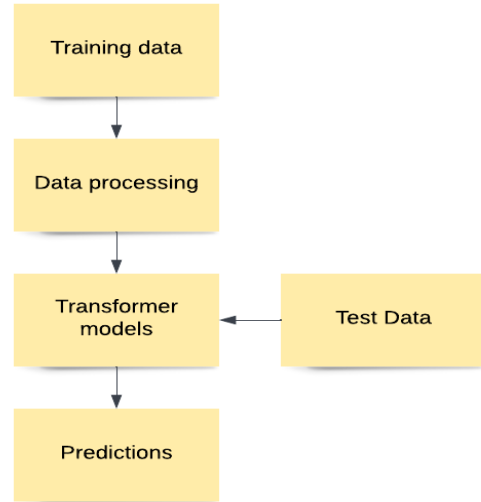


Figure 1: The architecture of the proposed system

3.1 Data-set Analysis

The goal of this task is to identify whether a given comment contains emotions, and which emotion it represents. A comment or post within the corpus may contain more than one sentence but the average sentence length of the corpora is 1. The annotations in the corpus are made at a comment or post level [Sampath et al. \(2022\)](#).

The dataset provided by LT-EDI 2021 organizers, consisted of the training set, development set, and test set of 14208, 3552, and 4440 instances respectively for Task A text, and 30179, 4269, and 4269 instances for the Task B text. The dataset contained text sequences that include user utterances along with the context, followed by the offensive class label. The task is to identify the emotion underlying the text and label them accordingly.

3.2 Data Pre-processing

Data pre-processing is essential for any machine learning problem. The given dataset of YouTube comments shows signs of irregularities in spelling and words. Firstly, the dataset is cleaned and processed before classifying.

- Hashtags, HTML tags, mentions, and URLs are removed
- Annotate emojis, emoticons, and replace them with the text they represent
- Convert uppercase characters to lowercase
- To expand abbreviations

- Remove special characters
- Remove accented characters
- Reduce lengthened words
- Remove extra white spaces

We've implemented data processing with the use of the nltk package, abbreviated as the Natural Language Toolkit, built to work with the NLP (Natural Language Processing). It provides various text processing libraries for classification, tokenization, parsing, semantic reasoning, etc. For our model, we've only used the regular expression (re) module. The re. sub() function was used to clean and scrape the text, remove URLs, remove numbers, and remove tags.

Regexp() module we were able to extract the tokens from the string by using the regular expression with the RegexpTokenizer() method. Tokenizing is a crucial step when it comes to cleaning the text. It is used to split the text into words or sentences, splitting it into smaller pieces that still hold its meaning outside the context of the rest of the text. When it comes to analyzing the text, we need to tokenize by word and tokenize by sentence. This is how unstructured data is turned into structured data, which is easier to analyze.

3.3 Model Description

The dataset text was classified using 3 transformer models, namely BERT, XLNet, and m-BERT

- BERT: BERT stands for Bidirectional Encoder Representations from Transformers. BERT is a pre-trained model for the top 104 languages of the world on Wikipedia (2.5B words) with 110 thousand shared word piece vocabulary, using masked language modeling (MLM) objective, which was first introduced in [Devlin et al. \(2018\)](#). BERT uses bi-directional learning to gain context of words from left to right context simultaneously. This is optimized by the Masked Language Modelling. The MLM is different from the traditional recurrent neural networks (RNNs), which generally see the word one after the other. This model randomly masks 15% of the words in the input and predicts the masked words when the entire masked sentence is run through the model.

- XLNet: The XLNet transformer model was proposed in 'XLNet: Generalized Autoregressive Pre-training for Language Understanding' [Yang et al. \(2019\)](#). It is pre-trained using an autoregressive model (a model that predicts future behavior based on past behavior) which enables learning bidirectional contexts by maximizing the expected likelihood over all permutations of the factorization order and overcomes the limitations of BERT thanks to its autoregressive formulation [Yang et al. \(2019\)](#). It integrates the Transformer-XL mechanism with a slight improvement in the language modeling approach.
- m-BERT: m-BERT is a pre-trained model on a large corpus of multilingual data. It is trained on the top 104 languages with the largest Wikipedia using a masked language modeling (MLM) objective. It was first introduced in [Devlin et al. \(2018\)](#)

4 Results and Analysis

The BERT (Bidirectional Encoder Representations from Transformers) models and XLNET were used for the Task A dataset. The BERT model operates on the principle of an attention mechanism to learn contextual relations between words. The transformer encoder used is bidirectional, unlike the other directional methods which read input sequentially. BERT reads the entire sequence of text at once. This bidirectional property of the encoder has made it very useful for classification tasks. The BERT models BERT and m-BERT were trained for 5 epochs. XLNet does not suffer from pre-train fine-tune discrepancy since it does not depend on data corruption. We have trained the XLNet model for 5 epochs. The bert-base-uncased model showed the best F1-Score of 0.19, 0.08 for Task A and Task B respectively.

4.1 Task A

The accuracy obtained by the BERT model was found to be 0.35, XLNet, and m-BERT showed an accuracy of 0.34, and 0.34 respectively. The bert-base-uncased model (BERT) showed the best performance with a weighted F1 score of 0.19. The weighted precision, weighted recall, weighted F1 score, and accuracy are given in the Table 2.

Pre-trained model	Precision	Recall	F1-score	Accuracy
bert-base-uncased	0.18	0.35	0.19	0.35
xlnet-base-cased	0.12	0.34	0.18	0.34
hline bert-base-multilingual-uncased hline	0.12	0.34	0.18	0.34

Table 2: Performance analysis of the proposed system using development data for Task A

Pre-trained model	Precision	Recall	F1-score	Accuracy
bert-base-uncased	0.05	0.19	0.08	0.19
xlnet-base-cased	0.02	0.16	0.04	0.16
hline bert-base-multilingual-uncased hline	0.13	0.20	0.12	0.20

Table 3: Performance analysis of the proposed system using development data for Task B

4.2 Task B

For Task B, the training data was run for the 3 transformer models. The training data with the best F1 score is run with the test data. The bert-base-multilingual-uncased model yielded the best results, with an F1 score of 0.12. The weighted precision, weighted recall, weighted F1 score, and accuracy is given in the Table 3.

4.3 Performance Metrics

In this task we have evaluated the models based on the macro average of Precision, Recall, and F1 Score. They provide us with an evaluation of the performance of the ML algorithm. We’ve used classification metrics for our research. Classification Metrics evaluate a model’s performance and tell you how good or bad the classification is, but each of them evaluates it in a different way.

Precision: Precision is the ratio of true positives and total positives predicted. As the name goes, Precision refers to the accuracy of the classification algorithm.

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

Recall: Recall may be defined as the number of positives returned by our ML model. Recall is the measure of the model correctly identifying the True Positives (TP).

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

F1 Score: The F1-score metric uses a combination of precision and recall. The F1 score is the

harmonic mean of the two. Since it takes both, Precision, and Recall into account, it is more useful than accuracy.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5 Conclusions

In this paper, we have investigated the baseline accuracy of different models as well as their variants on the datasets, and also proposed an approach for identifying emotions from the text. We have achieved F1 scores of 0.19 and 0.12 for Task A and Task B respectively. Due to the time constraint and not promising results, we had not submitted our results to the organizers. Identifying emotions based on text is quite a challenge and the performance of this model can further be enhanced by adopting favorable features.

References

- Samir E. Abdelrahman, Umar Farooq, and Tej Prasad Dhamala. 2016. Sentiment analysis of english and tamil tweets using path length similarity based word sense disambiguation.
- R Anita and CN Subalalitha. 2019a. An approach to cluster tamil literatures using discourse connectives. In *2019 IEEE 1st International Conference on Energy, Systems and Information Processing (ICESIP)*, pages 1–4. IEEE.
- R Anita and CN Subalalitha. 2019b. Building discourse parser for thirukkural. In *Proceedings of the 16th International Conference on Natural Language Processing*, pages 18–25.
- B Bharathi, Bharathi Raja Chakravarthi, Subalalitha Chinnudayar Navaneethakrishnan, N Sripriya,

- Arunaggiri Pandian, and Swetha Valli. 2022. Findings of the shared task on Speech Recognition for Vulnerable Individuals in Tamil. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*. Association for Computational Linguistics.
- Bharathi Raja Chakravarthi. 2020. 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, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John P McCrae. 2020a. A sentiment analysis dataset for code-mixed malayalam-english. *arXiv preprint arXiv:2006.00210*.
- 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*, pages 61–72, Kyiv. Association for Computational Linguistics.
- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Thenmozhi Durairaj, John Phillip McCrae, Paul Buitaleer, Prasanna Kumar Kumaresan, and Rahul Ponnusamy. 2022. Findings of the shared task on Homophobia Transphobia Detection in Social Media Comments. In *Proceedings of the Second 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*, pages 21–24.
- Bharathi Raja Chakravarthi, Ruba Priyadharshini, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, Elizabeth Sherly, John P McCrae, Adeep Hande, Rahul Ponnusamy, Shubhanker Banerjee, et al. 2021. Findings of the sentiment analysis of dravidian languages in code-mixed text. *arXiv preprint arXiv:2111.09811*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.
- Nikhil Ghanghor, Parameswari Krishnamurthy, Sajeetha Thavareesan, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2021a. IITK@DravidianLangTech-EACL2021: Offensive language identification and meme classification in Tamil, Malayalam and Kannada. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 222–229, Kyiv. Association for Computational Linguistics.
- Nikhil Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021b. 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*, pages 197–203, Kyiv. Association for Computational Linguistics.
- Rajenthiran Jenarathanan, Yasas Senarath, and Uthayasanker Thayasivam. 2019. Actsea: annotated corpus for tamil & sinhala emotion analysis. In *2019 Moratuwa Engineering Research Conference (MERCon)*, pages 49–53. IEEE.
- Prasanna Kumar Kumaresan, Ratnasingam Sakuntharaj, Sajeetha Thavareesan, Subalalitha Navaneethakrishnan, Anand Kumar Madasamy, Bharathi Raja Chakravarthi, and John P McCrae. 2021. Findings of shared task on offensive language identification in Tamil and Malayalam. In *Forum for Information Retrieval Evaluation*, pages 16–18.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- Ruba Priyadharshini, Bharathi Raja Chakravarthi, Subalalitha Chinnaudayar Navaneethakrishnan, Thenmozhi Durairaj, Malliga Subramanian, Kogilavani Shanmugavadivel, Siddhanth U Hegde, and Prasanna Kumar Kumaresan. 2022. Findings of the shared task on Abusive Comment Detection in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Ruba Priyadharshini, Bharathi Raja Chakravarthi, Sajeetha Thavareesan, Dhivya Chinnappa, Durairaj Thenmozhi, and Rahul Ponnusamy. 2021. Overview of the DravidianCodeMix 2021 shared task on sentiment detection in Tamil, Malayalam, and Kannada. In *Forum for Information Retrieval Evaluation*, pages 4–6.
- C.Sunitha Ram and R. Ponnusamy. 2014. An effective automatic speech emotion recognition for tamil language using support vector machine. In *2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, pages 19–23.
- Manikandan Ravikiran, Bharathi Raja Chakravarthi, Anand Kumar Madasamy, Sangeetha Sivanesan, Ratnavel Rajalakshmi, Sajeetha Thavareesan, Rahul Ponnusamy, and Shankar Mahadevan. 2022. Findings of the shared task on Offensive Span Identification in code-mixed Tamil-English comments. In *Proceedings of the Second Workshop on Speech and*

- Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2016. [A novel hybrid approach to detect and correct spelling in tamil text](#). In *2016 IEEE International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 1–6.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2017. [Use of a novel hash-table for speeding-up suggestions for misspelt tamil words](#). In *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, pages 1–5.
- Ratnasingam Sakuntharaj and Sinnathamby Mahesan. 2021. [Missing word detection and correction based on context of tamil sentences using n-grams](#). In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 42–47.
- Anbukkarasi Sampath, Thenmozhi Durairaj, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Subalalitha Chinnudayar Navaneethakrishnan, Kogilavani Shanmugavadivel, Sajeetha Thavareesan, Sathiyaraj Thangasamy, Parameswari Krishnamurthy, Adeep Hande, Sean Benhur, and Santhiya Ponusamy, Kishor Kumar Pandiyan. 2022. Findings of the shared task on Emotion Analysis in Tamil. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.
- Shikhar Sharma, Piyush Kumar, and Krishan Kumar. 2017. [Lexer: Lexicon based emotion analyzer](#). In *International Conference on Pattern Recognition and Machine Intelligence*, pages 373–379. Springer.
- C. N. Subalalitha. 2019. [Information extraction framework for kurunthogai](#). *Sādhanā*, 44(7):156.
- CN Subalalitha and E Poovammal. 2018. Automatic bilingual dictionary construction for tirukural. *Applied Artificial Intelligence*, 32(6):558–567.
- 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 (MERCCon)*, 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.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2021a. [Sentiment analysis in tamil texts using k-means and k-nearest neighbour](#). In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 48–53.
- Sajeetha Thavareesan and Sinnathamby Mahesan. 2021b. [Sentiment analysis in tamil texts using k-means and k-nearest neighbour](#). In *2021 10th International Conference on Information and Automation for Sustainability (ICIAfS)*, pages 48–53.
- D Thenmozhi, Aravindan Chandrabose, Srinethe Sharavanan, et al. 2019. [Ssn_nlp at semeval-2019 task 3: Contextual emotion identification from textual conversation using seq2seq deep neural network](#). In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 318–323.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 347–354.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. [Xlnet: Generalized autoregressive pretraining for language understanding](#). *CoRR*, abs/1906.08237.
- Konthala Yasaswini, Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. [IIIT@DravidianLangTech-EACL2021: Transfer learning for offensive language detection in Dravidian languages](#). In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*, pages 187–194, Kyiv. Association for Computational Linguistics.