Findings of the Shared Task on Offensive Language Identification in Tamil, Malayalam, and Kannada

Bharathi Raja Chakravarthi[1], **Ruba Priyadharshini**[2], **Navya Jose**[3]
Anand Kumar M [4],**Thomas Mandl** [5], **Prasanna Kumar Kumaresan**[3],
Rahul Ponnusamy[3], **Hariharan R L** [4], **John Philip McCrae** [1] and **Elizabeth Sherly**[3]
[1]National University of Ireland Galway, [2]Madurai Kamaraj University,
[3]Indian Institute of Information Technology and Management-Kerala,
[4]National Institute of Technology Karnataka Surathkal,
[5]University of Hildesheim Germany
bharathi.raja@insight-centre.org

Abstract

Detecting offensive language in social media in local languages is critical for moderating user-generated content. Thus, the field of offensive language identification for under-resourced languages like Tamil, Malayalam and Kannada is of essential importance. As user-generated content is often code-mixed and not well studied for under-resourced languages, it is imperative to create resources and conduct benchmark studies to encourage research in under-resourced Dravidian languages. We created a shared task on offensive language detection in Dravidian languages. We summarize the dataset for this challenge which are openly available at https://competitions.codalab.org/competitions/27654, and present an overview of the methods and the results of the competing systems.

1 Introduction

The dawn of social media helped to bridge the gap between the borders and paved the way for the people to communicate or express their opinions more easily than any other time in human history (Edosomwan et al., 2011). But with this advent of social media, the platforms like YouTube, Facebook, and Twitter not only helped information sharing and networking but also became the place that target, defame and marginalize people merely on the basis of their physical appearance, religion or sexual orientation (Keipi et al., 2016; Benikova et al., 2018; Pamungkas et al., 2020). Social media has moulded itself into its specialised tool with which the people are verbally threatened and cornered not for what they have done but for who they are (Maitra and McGowan, 2012; Patton et al., 2013). Undoubtedly, the infiltration depth and width of this 'digital wonder' equipped the erstwhile 'invisible and socially paralysed' communities to play cards, if not extravagantly noticeable (Barnidge et al., 2019).

Research in hate speech detection (Kumar et al., 2018) or offensive language detection (Zampieri et al., 2020; Mandl et al., 2020) using Natural Language Processing (NLP) has significantly improved in recent years. However, the work on under-resourced languages is still limited (Chakravarthi, 2020). For example, under-resourced languages such as Tamil, Malayalam, and Kannada lack tools and datasets (Chakravarthi et al., 2020a,c; Thavareesan and Mahesan, 2019, 2020a,b). Recently, shared sentiment analysis for Tamil and Malayalam by Chakravarthi et al. (2020d) and offensive language identification in Tamil and Malayalam by Chakravarthi et al. (2020b) paved the wave for more research on Dravidian languages. Tamil, Malayalam and Kannada belong to the Dravidian language family and are spoken by some 220 million people in the Indian subcontinent, Singapore and Sri Lanka (Krishnamurti, 2003). It is essential to create NLP systems such as hate speech detection or offensive language detection in local languages since most user-generated content is in such languages.

The Dravidian language family is reported as being approximately 4500 years old by Bayesian phylogenetic analysis of cognate-coded lexical data (Kolipakam et al., 2018). The Dravidian languages scripts are first attested in the 580 BCE as Tamili[1] script inscribed on the pottery of Keezhadi, Sivagangai and Madurai district of Tamil Nadu, India (Sivanantham and Seran, 2019)[2] by Tamil Nadu State Department of Archaeology and Archaeological Survey of India. The modern Dravidian languages have its own scripts evolved from Tamili scripts. The scripts are alpha-syllabic, belonging to a family of the abugida writing systems that are partially alphabetic and partially syllable-based

---

[1]also called Damili or Dramili or Tamil-Brahmi
[2]Keeladi-Book-English-18-09-2019.pdf

(Albert et al., 1985; Rajendran, 2004). The writing system of Tamili was explained in the old grammar text Tolkappiyam which dates are various proposed between 9th century BCE to 6nd century BCE (Pillai, 1904; Swamy, 1975; Zvelebil, 1991; Takahashi, 1995) and in the Jaina work Samavayanga Sutta and Pannavana Sutta, these two Jain works date to 3rd-4th century BCE (Salomon, 1998).

Up until around the 15th-century CE, Malayalam was the west coast dialect of Tamil (Blackburn, 2006). Due to geographical isolation by the steep Western Ghats from the main speech community, the dialect eventually evolved into a distinct language by the 15th century (Sekhar, 1951). The first literary work in Malayalam was Ramacaritam ("Deeds of Rama") was written using Tamil Grantha script which was used to write Pali, Prakrit and foreign words in Tamil (Andronov, 1996). The first printed work in an Indian language and script was a Roman Catholic catechism translated by Henrique which is Thambiran Vanakkam (Doctrina Christam en Lingua Malauar Tamul in Portuguese) on 20 October 1,578 CE at Quilon, Venad (present day Kerala) (George, 1972). Similarly, Kannada also split from Tamil by 9th century CE; for example, Taayviru is a term of Kannada origin and is found in a Tamil inscription from the 4th-century CE. The Kappe Arabhatta record of 700 CE is the oldest extant form of Kannada poetry in the Tripadi metre, it is based in part on Kavyadarsha (Rawlinson, 1930; Hande et al., 2020).

However, user-generated content is often adopted to Roman script for typing due to historical reason and modern computer keyboard layouts. Hence, the majority of user-generated data for these Dravidian languages are code-mixed (Priyadharshini et al., 2020; Jose et al., 2020). Due to the growth of social media platforms across the world and the possibility of writing content without any moderation, users write content in multilingual code-switching without grammatical restrictions and using non-native scripts (Chakravarthi et al., 2019). In linguistics, code-switching is switching between two or more language in the same utterance. This explosion of code-mixed user-generated content in the social media platforms aroused interest in NLP. In this paper, we dicuss about the offensive language identification shared task for Tamil, Malayalam and Kannada languages and the participant's submissions.

2 Task Description

Offensive language identification for Dravidian languages at different levels of complexity were developed following the work of (Zampieri et al., 2019). It was customized to our annotation method from three-level hierarchical annotation schema. A new category **Not in intended language** was added to include comments written in a language other than the Dravidian languages. Annotations decision for offensive language categories were split into six labels to simplify the annotation process.

- **Not Offensive**: Comment/post does not have offence, obscenity, swearing, or profanity.

- **Offensive Untargeted**: Comment/post have offence, obscenity, swearing, or profanity not directed towards any target. These are the comments/posts which have inadmissible language without targeting anyone.

- **Offensive Targeted Individual**: Comment/post have offence, obscenity, swearing, or profanity which targets an individual.

- **Offensive Targeted Group**: Comment/post have offence, obscenity, swearing, or profanity which targets a group or a community.

- **Offensive Targeted Other**: Comment/post have offence, obscenity, swearing, or profanity which does not belong to any of the previous two classes.

- **Not in indented language**: If the comment is not in the intended language. For example, in the Malayalam task, if the sentence does not contain Malayalam written in Malayalam script or Latin script, then it is not Malayalam.

Sample of not-offensive comments in our dataset provided for the participants is given below with the corresponding English glosses..

- Tamil-English: *Paka thana poro movie la Enna irukunu baki ellam*

  English: "You will see what is the movie"

Sample offensive language text from the dataset

- Tamil-English: *Ghibran p—a vitta ungalukku vera aale kedaika laya da.*

  English: "Dont you get anyone else other than Ghibran V—a"

Language	Tamil	Malayalam	Kannada
Number of words	511,734	202,134	65,702
Vocabulary size	94,772	40,729	20,796
Number of comments	43,919	20,010	7,772
Number of sentences	52,617	23,652	8,586
Average number of words per sentence	11	10	8
Average number of sentences per comment	1	1	1

Table 1: Corpus statistics for Offensive Language Identification

Class	Tamil	Malayalam	Kannada
Not Offensive	31,808	17,697	4,336
Offensive Untargeted	3,630	240	278
Offensive Targeted Individual	2,965	290	628
Offensive Targeted Group	3,140	176	418
Offensive Targeted Others	590	-	153
Not in indented language	1,786	1,607	1,898
Total	43,919	20,010	7,772

Table 2: Offensive language Identification Dataset Distribution.

This is an example of Offensive Targeted Individual. Mr. Ghibran is a popular music composer in Tamil films, and the author tries to insult him.

- Malayalam-English: *Verupikkal manju ullathu ozhichal baki ellam super aanu ee padam.*

  English: "Except the presence of disgusting Manju, everything is super in this movie"

This is an example of Offensive Targeted Individual. Mrs. Manju Warrier is a popular actress in South Indian films, and the author tries to discredit her by clearly mentioning that her presence is disgusting.

- Malayalam-English: *Nalla trailer nu okke keri dislike adikunne ethelum thanthayillathavanmar aayirikum. poyi chavinedey...*

  English: "Those who dislike any trailers will probably be assholes. Go to hell..."

This is an example of Offensive Untargeted comment. This text have inadmissible language without targeting anyone.

2.1 Dataset

Data was compiled from different film trailers of Tamil, Kannada, and Malayalam languages from YouTube comments in the year 2019. The *YouTube*

Comment Scraper tool[3] was used to collect comments from YouTube social media. These comments were utilized to make the datasets for offensive language identification shared task. We collected comments that contain code-mixing at various levels of the text, with enough representation for each sentiment in all the three languages. *Langdetect library* [4] was used to detect different languages apart and eliminate the unintended languages. Our dataset contains different types of real-time code-mixing due to that fact that data was collected from social media. The dataset contains all forms of code-mixing ranging from purely monolingual texts in native languages to mixing of scripts, words, morphology, inter-sentential and intra-sentential switches.

Dataset statistics (number of words, vocabulary size, number of comments, number of sentences, and average number of words per sentences) are given in Table 1 and class-wise distribution is given in Table 2. From the Table 1 we can see that vocabulary size is very big for Tamil and Malayalam this is due to the code mixing and complex morphology of these languages. Table 2 shows that all languages have the not-offensive class in the majority. In the case of Tamil, 71% of the total comments are not offensive, while Malayalam has

---

[3] https://github.com/philbot9/youtube-remarkscraper
[4] https://pypi.org/venture/langdetect/

85% non-offensive comments. But there is no consistent trends observable amongst offensive classes across the languages.

2.2 Training Phase

In the first phase, data is released for training and or development of offensive language detection models. Participants were given training data and validation dataset for preliminary evaluations or tuning of hyper-parameters. They were also given the option to perform cross-validation on the training data. In total, 119 participant register for the task and downloaded the data.

2.3 Evaluation Phase

In the second phase, test sets for evaluation are released for all three languages. Each participating team submitted their generated prediction for evaluation. Predictions are submitted via Google form to the organizing committer to evaluate the systems. CodaLab is an established platform to organize shared-tasks. However, we faced issues with running evaluation, so we choose to evaluate manually. The metrics used for evaluation is the weighted average F1 score.

3 Systems

3.1 System Descriptions

- Dowlagar and Mamidi (2021) used a pre-trained multilingual BERT transformer model with transliteration and class balancing loss for offensive content identification. They have ranked 2nd in Malayalam-English and 4th in Tamil-English languages.

- Li (2021) participated in all three of offensive language identification. They explored multilingual models based on XLM-RoBERTa and multilingual BERT trained on mixed data of three code-mixed languages. They solved the class-imbalance problem existed in the training data by class weights and class combination. Their model achieved a weighted average F1 scores of 0.75 (ranked 4th), 0.94 (ranked 4th) and 0.72 (ranked 3rd) on offensive language identification in code-mixed Tamil-English language, Malayalam-English language and Kannada-English language, respectively.

- Yasaswini et al. (2021) used various transfer learning-based models to classify a given post or comment in Tamil, Malayalam, and Kannada. They have fine-tuned transformer models to get better performance. The relatively high F1-scores of 0.9603, 0.7895 on Malayalam, Tamil were achieved by ULMFiT and 0.7277 on Kannada was achieved by Distil-BERT models.

- Vasantharajan and Thayasivam (2021) participated in Tamil, Kannada, and Malayalam language using the BERT fine-tuning strategies. Their model got a 0.96 F1- score for Malayalam, 0.73 F1-score for Tamil, and 0.70 F1-score for Kannada. Moreover, in the view of multilingual models, this modal ranked 3rd and achieved favorable results and confirmed the model as the best among all systems submitted to these shared tasks in these three languages. They got 2nd, 5th, 6th ranks in the leader-board for the Malayalam, Kannada, and Tamil test set respectively.

- Huang and Bai (2021) used the multilingual BERT model to complete in all three languages. For all three language data sets, they combined the Tf-Idf algorithm and the multilingual BERT model's output and introduced the CNN block as a shared layer.

- Zhao (2021) proposed a system based on the multilingual model XLM-Roberta and DPCNN. The test results on the official test data set confirm the effectiveness of their system. They achieved a weighted average F1-score of Kannada, Malayalam, and Tami language of 0.69, 0.92, and 0.76, respectively, ranked 6th, 6th, and 3rd.

- Ghanghor et al. (2021) used multilingual-BERT-base for offensive language identification, their system achieved a weighted F1 scores of 0.75 for Tamil, 0.95 for Malayalam, and 0.71 for Kannada. and ranked 3rd, 3rd and 4th for Tamil, Malayalam and Kannada respectively.

- Chen and Kong (2021) used multilingual BERT and TextCNN for semantic extraction and text classification. The results are not be ideal. They ranked 5th,5th, and 9th for Tamil, Malayalam and Kannada respectively.

- Renjit and Idicula (2021) used language-agnostic BERT (Bidirectional Encoder Rep-

Team-Name	Precision	Recall	F1 Score	Rank
hate-alert (Saha et al., 2021)	0.78	0.78	0.78	1
indicnlp@kgp (Kedia and Nandy, 2021)	0.75	0.79	0.77	2
ZYJ123 (Zhao, 2021)	0.75	0.77	0.76	3
ALI-B2B-AI	0.75	0.78	0.76	3
SJ-AJ (Jayanthi and Gupta, 2021)	0.75	0.79	0.76	3
No offense (Awatramani, 2021)	0.75	0.78	0.76	3
NLP@CUET (Sharif et al., 2021)	0.75	0.78	0.76	3
Codewithzichao (Li, 2021)	0.74	0.77	0.75	4
hub (Huang and Bai, 2021)	0.73	0.78	0.75	4
MUCS (Balouchzahi et al., 2021)	0.74	0.77	0.75	4
bitions (Tula et al., 2021)	0.74	0.77	0.75	4
IIITK (Ghanghor et al., 2021)	0.74	0.77	0.75	4
OFFLangOne (Dowlagar and Mamidi, 2021)	0.74	0.75	0.75	4
spartans	0.74	0.78	0.75	4
cs (Chen and Kong, 2021)	0.74	0.75	0.74	5
Zeus	0.72	0.78	0.73	6
hypers (Vasantharajan and Thayasivam, 2021)	0.71	0.76	0.73	6
SSNCSE-NLP (B and Silvia A, 2021)	0.74	0.73	0.73	6
JUNLP (Garain et al., 2021)	0.71	0.74	0.72	7
snehan-coursera	0.7	0.73	0.71	8
IIITT (Yasaswini et al., 2021)	0.7	0.73	0.71	8
IRNLP-DAIICT (Dave et al., 2021)	0.72	0.77	0.71	8
CUSATNLP (Renjit and Idicula, 2021)	0.67	0.71	0.69	9
e8ijs	0.71	0.76	0.69	9
KU-NLP	0.66	0.75	0.67	10
cognizyr	0.68	0.74	0.66	11
Agilna	0.63	0.72	0.65	12
DLRG	0.57	0.74	0.64	13
Amrita-CEN-NLP (K et al., 2021)	0.64	0.62	0.62	14
KBCNMUJAL	0.75	0.56	0.62	14
JudithJeyafreeda (Andrew, 2021)	0.54	0.73	0.61	15

Table 3: Rank list based on F1-score along with other evaluation metrics (Precision and Recall) for Tamil language

Team-Name	Precision	Recall	F1 Score	Rank
hate-alert (Saha et al., 2021)	0.97	0.97	0.97	1
MUCS (Balouchzahi et al., 2021)	0.97	0.97	0.97	1
indicnlp@kgp (Kedia and Nandy, 2021)	0.97	0.97	0.97	1
bitions (Tula et al., 2021)	0.97	0.97	0.97	1
hypers (Vasantharajan and Thayasivam, 2021)	0.96	0.96	0.96	2
ALI-B2B-AI	0.96	0.97	0.96	2
OFFLangOne (Dowlagar and Mamidi, 2021)	0.96	0.96	0.96	2
SJ-AJ (Jayanthi and Gupta, 2021)	0.97	0.97	0.96	2
CUSATNLP (Renjit and Idicula, 2021)	0.95	0.95	0.95	3
IIITK (Ghanghor et al., 2021)	0.94	0.95	0.95	3
SSNCSE-NLP (B and Silvia A, 2021)	0.95	0.96	0.95	3
IRNLP-DAIICT (Dave et al., 2021)	0.96	0.96	0.95	3
Codewithzichao (Li, 2021)	0.93	0.94	0.94	4
JudithJeyafreeda (Andrew, 2021)	0.94	0.94	0.93	5
cs (Chen and Kong, 2021)	0.92	0.94	0.93	5
Zeus	0.92	0.95	0.93	5
Maoqin (Yang, 2021)	0.91	0.94	0.93	5
NLP@CUET (Sharif et al., 2021)	0.92	0.94	0.93	5
ZYJ123 (Zhao, 2021)	0.91	0.94	0.92	6
hub (Huang and Bai, 2021)	0.89	0.93	0.91	7
No offense (Awatramani, 2021)	0.9	0.93	0.91	7
e8ijs	0.89	0.92	0.9	8
snehan-coursera	0.9	0.89	0.89	9
KU-NLP	0.87	0.91	0.89	9
KBCNMUJAL	0.93	0.88	0.89	9
IIITT (Yasaswini et al., 2021)	0.84	0.87	0.86	10
Agilna	0.85	0.88	0.86	10
Amrita-CEN-NLP (K et al., 2021)	0.9	0.82	0.85	11
professionals (Nair and Fernandes, 2021)	0.89	0.84	0.85	11
JUNLP (Garain et al., 2021)	0.77	0.43	0.54	12

Table 4: Rank list based on F1-score along with other evaluation metrics (Precision and Recall) for Malayalam language

Team-Name	Precision	Recall	F1 Score	Rank
SJ-AJ (Jayanthi and Gupta, 2021)	0.73	0.78	0.75	1
hate-alert (Saha et al., 2021)	0.76	0.76	0.74	2
indicnlp@kgp(Kedia and Nandy, 2021)	0.71	0.74	0.72	3
Codewithzichao (Li, 2021)	0.7	0.75	0.72	3
IIITK (Ghanghor et al., 2021)	0.7	0.75	0.72	3
e8ijs	0.7	0.74	0.71	4
NLP@CUET(Sharif et al., 2021)	0.7	0.74	0.71	4
ALI-B2B-AI	0.7	0.73	0.71	4
Zeus	0.66	0.75	0.7	5
hypers (Vasantharajan and Thayasivam, 2021)	0.69	0.72	0.7	5
bitions (Tula et al., 2021)	0.69	0.72	0.7	5
SSNCSE-NLP (B and Silvia A, 2021)	0.71	0.74	0.7	5
MUCS (Balouchzahi et al., 2021)	0.68	0.72	0.69	6
ZYJ123 (Zhao, 2021)	0.65	0.74	0.69	6
MUM	0.69	0.71	0.69	6
JUNLP (Garain et al., 2021)	0.62	0.71	0.66	7
OFFLangOne (Dowlagar and Mamidi, 2021)	0.66	0.65	0.65	8
cs (Chen and Kong, 2021)	0.64	0.67	0.64	9
hub (Huang and Bai, 2021)	0.65	0.69	0.64	9
No offense (Awatramani, 2021)	0.62	0.7	0.64	9
IRNLP-DAIICT (Dave et al., 2021)	0.69	0.69	0.64	9
JudithJeyafreeda (Andrew, 2021)	0.66	0.67	0.63	10
CUSATNLP (Renjit and Idicula, 2021)	0.62	0.63	0.62	11
Agilna	0.58	0.65	0.59	12
Amrita-CEN-NLP (K et al., 2021)	0.65	0.54	0.58	13
snehan-coursera	0.56	0.61	0.54	14
KBCNMUJAL	0.72	0.48	0.52	15
IIITT (Yasaswini et al., 2021)	0.46	0.48	0.47	16
Simon (Que et al., 2021)	0.6	0.3	0.33	17

Table 5: Rank list based on F1-score along with other evaluation metrics (Precision and Recall) for Kannada language

resentation from Transformers) for sentence embedding and a Softmax classifier. The language-agnostic representation based classification helped obtain good performance for all the three languages, out of which results for the Malayalam language are good enough to obtain a third position among the participating teams.

- K et al. (2021) implemented three deep neural network architectures such as a hybrid network with a Convolutional layer, a Bidirectional Long Short- Term Memory network (Bi-LSTM) layer and a hidden layer, a network containing a Bi- LSTM and another with a Bidirectional Recurrent Neural Network (Bi-RNN). In addition to that, they incorporated a cost-sensitive learning approach to deal with the problem of class imbalance in the training data. Among the three models, the hybrid network exhibited better training performance, and they submitted the predictions based on the same.

- Sharif et al. (2021) employed two machine learning techniques (LR, SVM), three deep learning (LSTM, LSTM+Attention) techniques and three transformers (m-BERT, Indic-BERT, XLM-R) based methods. They showed that XLM-R outperforms other techniques in Tamil and Malayalam languages while m-BERT achieves the highest score in the Kannada language. Their proposed models gained a weighted F1 score of 0.76 (for Tamil), 0.93 (for Malayalam), and 0.71 (for Kannada) with a rank of 3rd , 5th and 4th respectively.

- Dave et al. (2021) used TF-IDF character n-grams and pretrained MuRIL embeddings for text representation and Logistic Regression and Linear SVM for classification. Their best approach achieved ninth, third and eighth with weighted F1 score of 0.64, 0.95 and 0.71 in Kannada, Malayalam and Tamil on the test dataset respectively.

- Saha et al. (2021) presented an exhaustive exploration of different transformer models, also provided a genetic algorithm technique for ensembling different models. Their ensembled models trained separately for each language secured the first position in Tamil, the second position in Kannada, and the first position in Malayalam sub-tasks.

- Yang (2021) only participate in one of the language task- Malayalam. They used the transformer-based language model with BiGRU-Attention to complete this task. They ranked 5th in this task with a weighted average F1 score of 0.93 on the private leader board.

- Awatramani (2021) participated for Tamil language. They used mBERT-cased and XLM-RoBERTa. Their system was ranked 3 rd in the task leaderboard achieving an F1-score 0.76 for detecting offensive Tamil YouTube comments.

- Nair and Fernandes (2021) used Indic-BERT to generate word embeddings which is then fed into a 4-layer Multi Layer Perceptron (MLP) which does the multi-class classification task and achieve an F1 score of 0.85 for Malayalam language.

- Andrew (2021) did language-specific pre-processing before using several machine learning algorithms. For the Tamil language, they achieve a precision of 0.54, a recall of 0.73 and an F1-score of 0.61, a precision of 0.94, a recall of 0.94 and an F1-score of 0.93 for the Malayalam language and a precision of 0.66, a recall of 0.67 and an F1-score of 0.63 for the Kannada language.

- Que et al. (2021) used the XLM-Roberta model for pre-training for the Kannada language. They scored a very low score of 0.33 weighted

- Tula et al. (2021) proposed a multilingual ensemble based model from ULMFiT, DistilmBERT, and IndicBERT. Their model is able to handle both code-mixed data as well as instances where the script used is mixed (for instance, Tamil and Latin). They ranked number one for Malayalam dataset and ranked 4th and 5th for Tamil and Kannada, respectively.

- Jayanthi and Gupta (2021) system is an ensemble of mBERT and XLM-RoBERTa models which leverage task-adaptive pre-training of multilingual BERT models with a masked language modeling objective. They ranked 1st

for Kannada, 2nd for Malayalam and 3rd for Tamil.

- B and Silvia A (2021) described an automatic offensive language identification from Dravidian languages with various machine learning algorithms. They achieved F1 scores of 0.95 for Malayalam, 0.7 for Kannada and 0.73 for task2-Tamil on the test-set.

- Garain et al. (2021) used IndicBERT and BERT architectures, to facilitate identification of offensive languages for Kannada-English, Malayalam-English, and Tamil-English code-mixed language pairs extracted from social media. F1 score for language pair Kannada-English as 0.62, 0.71, and 0.66, respectively, for language pair Malayalam-English as 0.77, 0.43, and 0.53, respectively, and for Tamil-English as 0.71, 0.74, and 0.72, respectively.

- Balouchzahi et al. (2021) Two models, namely, COOLI-ensemble and COOLI-Keras were trained with the char sequences extracted from the sentences combined with words as features. Out of the two proposed models, the COOLI-Ensemble model (best among our models) obtained the first rank for the Ma-En language pair with 0.97 weighted F1- and fourth and sixth rank with a 0.75 and a 0.69 weighted F1-score for Ta-En and Kn-En language pairs respectively.

- Kedia and Nandy (2021) leveraged existing state of the art approaches in text classification by incorporating additional data and transfer learning on pre-trained models. Their final submission is an ensemble of an AWD-LSTM based model along with 2 different transformer model architectures based on BERT and RoBERTa. They achieved weighted-average F1 scores of 0.97, 0.77, and 0.72 in the Malayalam-English, Tamil-English, and Kannada-English datasets ranking 1st, 2nd, and 3rd on the respective tasks.

4 Results and Discussion

The offensive language identification shared task was organized for three languages Tamil, Malayalam, and Kannada. Many participants connecting to all three languages had submitted their solutions as described in the previous section. Here in this section, we'll be highlighting the results of all three

Language	Team-Name	Rank
Tamil	hate-alert	1
	indicnlp@kgp	2
	ZYJ123,ALI-B2B-AI, SJ-AJ,No offense, NLP@CUET	3
Malayalam	hate-alert, MUCS, indicnlp@kgp,bitions	1
	hypers, ALI-B2B-AI, OFFLangOne,SJ-AJ	2
	CUSATNLP, IIITK, SSNCSE-NLP, IRNLP-DAIICT	3
Kannada	SJ-AJ	1
	hate-alert	2
	indicnlp@kgp, Codewithzichao, IIITK	3

Table 6: Overall Results with Top Three Ranks

languages, which have ranked top positions on the leaderboard. Table 6 shows the overall results and teams which are placed in the top three positions.

The team hate-alert (Saha et al. (2021)) came at the first position in Tamil and Malayalam languages and second in Kannada as well, and the team indicnlp@kgp (Kedia and Nandy (2021)) came second in Tamil, first in Malayalam, and third in Kannada. The team SJ-AJ (Jayanthi and Gupta (2021)) came at the first position in Kannada and third and second in Tamil and Malayalam respectively.As per the F1-scores apart from Malayalam, the other two languages got scores around 0.75, whereas 0.97 was the former's score. As the vocabulary size of Malayalam is more, it has produced better results compared to Tamil (which is having the maximum), and Kannada, where vocabulary is less but has got similar results to that of Tamil.

5 Conclusion

We presented the results of the first shared task on Offensive Language Identification in Tamil, Malayalam, and Kannada relying on a thoroughly annotated data set based on human judgements. The task setup provided an opportunity to test model in multilingual scenarios along with code-mixing phenomenon. However, several teams reach high

performance in all three languages. We found out that the type of embedding strongly influence the results. We hope this task and dataset will intrigue and facilitate further research on under-resourced Dravidian languages.

Acknowledgments

References

Devasagayam Albert et al. 1985. *Tolkāppiyam Phonology and Morphology: An English Translation*, volume 115. International Institute of Tamil Studies.

JudithJeyafreeda Andrew. 2021. JudithJeyafreedaAndrew@DravidianLangTech-EACL2021:Offensive language detection for Dravidian Code-mixed YouTube comments. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Mikhail Sergeevich Andronov. 1996. *A grammar of the Malayalam language in historical treatment*, volume 1. Otto Harrassowitz Verlag.

Vasudev Awatramani. 2021. No Offense@DravidianLangTech-EACL2021: Offensive Tamil Identification and beyond the performance . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Bharathi B and Agnusimmaculate Silvia A. 2021. SSNCSE NLP@DravidianLangTech-EACL2021: Offensive Language Identification on Multilingual Code Mixing Text. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Fazlourrahman Balouchzahi, Aparna B K, and H L Shashirekha. 2021. MUCS@DravidianLangTech-EACL2021:COOLI-Code-Mixing Offensive Language Identification. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Matthew Barnidge, Bumsoo Kim, Lindsey A. Sherrill, Žiga Luknar, and Jiehua Zhang. 2019. Perceived exposure to and avoidance of hate speech in various communication settings. *Telematics and Informatics*, 44:101263.

Darina Benikova, Michael Wojatzki, and Torsten Zesch. 2018. What Does This Imply? Examining the Impact of Implicitness on the Perception of Hate Speech. In *Language Technologies for the Challenges of the Digital Age*, pages 171–179, Cham. Springer International Publishing.

Stuart H Blackburn. 2006. *Print, folklore, and nationalism in colonial South India*. Orient Blackswan.

Bharathi Raja Chakravarthi. 2020. *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. 2019. Comparison of Different Orthographies for Machine Translation of Under-Resourced Dravidian Languages. In *2nd Conference on Language, Data and Knowledge (LDK 2019)*, volume 70 of *OpenAccess Series in Informatics (OASIcs)*, pages 6:1–6:14, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Bharathi Raja Chakravarthi, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly, and John Philip McCrae. 2020a. A sentiment analysis dataset for code-mixed Malayalam-English. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 177–184, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, M Anand Kumar, John Philip McCrae, Premjith B, Soman KP, and Thomas Mandl. 2020b. Overview of the track on "HASOC-Offensive Language Identification-DravidianCodeMix". In *Proceedings of the 12th Forum for Information Retrieval Evaluation*, FIRE '20.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John Philip McCrae. 2020c. Corpus creation for sentiment analysis in code-mixed Tamil-English text. In *Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-resourced languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL)*, pages 202–210, Marseille, France. European Language Resources association.

Bharathi Raja Chakravarthi, Ruba Priyadharshini, Vigneshwaran Muralidaran, Shardul Suryawanshi, Navya Jose, Elizabeth Sherly, and John P. McCrae. 2020d. 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.

Shi Chen and Bing Kong. 2021. cs@DravidianLangTech-EACL2021: Offensive Language Identification Based On Multilingual BERT Model. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Bhargav Dave, Shripad Bhat, and Prasenjit Majumder. 2021. IRNLPDAIICT@DravidianLangTech-EACL2021:Offensive Language identification in Dravidian Languages using TF-IDF Char N-grams and MuRIL . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Suman Dowlagar and Radhika Mamidi. 2021. OFFLangOne@DravidianLangTech-EACL2021: Transformers with the Class Balanced Loss for Offensive Language Identification in Dravidian Code-Mixed text . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Simeon Edosomwan, Sitalaskshmi Kalangot Prakasan, Doriane Kouame, Jonelle Watson, and Tom Seymour. 2011. The history of social media and its impact on business. *Journal of Applied Management and entrepreneurship*, 16(3):79–91.

Avishek Garain, Atanu Mandal, and Sudip Kumar Naskar. 2021. JUNLP@DravidianLangTech-EACL2021: Offensive Language Identification in Dravidian Langauges . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Karimpumannil Mathai George. 1972. *Western influence on Malayalam language and literature*. Sahitya Akademi.

Nikhil Ghanghor, Bharathi Raja Chakravarthi, Ruba Priyadharshini, Sajeetha Thavareesan, and Parameswari Krishnamurthy. 2021. IIITK@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*. Association for Computational Linguistics.

Adeep Hande, Ruba Priyadharshini, and Bharathi Raja Chakravarthi. 2020. KanCMD: Kannada CodeMixed dataset for sentiment analysis and offensive language detection. In *Proceedings of the Third Workshop on Computational Modeling of People's Opinions, Personality, and Emotion's in Social Media*, pages 54–63, Barcelona, Spain (Online). Association for Computational Linguistics.

Bo Huang and Yang Bai. 2021. HUB@DravidianLangTech-EACL2021: Identify and Classify Offensive Text in Multilingual Code Mixing in Social Media. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Sai Muralidhar Jayanthi and Akshat Gupta. 2021. SJAJ@DravidianLangTech-EACL2021: Task-Adaptive Pre-Training of Multilingual BERT models for Offensive Language Identification. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. 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)*.

Sreelakshmi K, Premjith B, and Soman KP. 2021. AmritaCENNLP@DravidianLangTech-EACL2021: Deep Learning-based Offensive Language Identification in Malayalam, Tamil and Kannada . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Kushal Kedia and Abhilash Nandy. 2021. indicnlp@kgp@DravidianLangTech-EACL2021: Offensive Language Identification in Dravidian Languages . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Teo Keipi, Matti Näsi, Atte Oksanen, and Pekka Räsänen. 2016. *Online hate and harmful content: Cross-national perspectives*. Taylor & Francis.

Vishnupriya Kolipakam, Fiona M Jordan, Michael Dunn, Simon J Greenhill, Remco Bouckaert, Russell D Gray, and Annemarie Verkerk. 2018. A bayesian phylogenetic study of the dravidian language family. *Royal Society open science*, 5(3):171504.

Bhadriraju Krishnamurti. 2003. *The Dravidian languages*. Cambridge University Press.

Ritesh Kumar, Atul Kr. Ojha, Shervin Malmasi, and Marcos Zampieri. 2018. Benchmarking aggression identification in social media. In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018)*, pages 1–11, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Zichao Li. 2021. Codewithzichao@DravidianLangTech-EACL2021: Exploring Multilingual Transformers for Offensive Language Identification on Code Mixing Text . In *Proceedings of the First Workshop*

on Speech and Language Technologies for Dravidian Languages. Association for Computational Linguistics.

Ishani Maitra and Mary Kate McGowan. 2012. *Speech and harm: Controversies over free speech*. Oxford University Press on Demand.

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.

Srinath Nair and Dolton Fernandes. 2021. professionals@DravidianLangTech-EACL2021. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Endang Wahyu Pamungkas, Valerio Basile, and Viviana Patti. 2020. Do You Really Want to Hurt Me? Predicting Abusive Swearing in Social Media. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 6237–6246, Marseille, France. European Language Resources Association.

Desmond Upton Patton, Robert D. Eschmann, and Dirk A. Butler. 2013. Internet banging: New trends in social media, gang violence, masculinity and hip hop. *Computers in Human Behavior*, 29(5):A54 – A59.

MS Purnalingam Pillai. 1904. *A Primer of Tamil Literature*. Ananda Press.

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)*.

Qinyu Que, Gang Wang, and Shuangjun Jia. 2021. Simon @ DravidianLangTech-EACL2021: Detecting Offensive Content in Kannada Language . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

S Rajendran. 2004. Strategies in the formation of compound nouns in tamil. *Languages Of India*, 4.

HG Rawlinson. 1930. The kadamba kula—a history of ancient and mediæval karnataka.(studies in indian history of the indian historical research institute, st. xavier's college, bombay, no. 5).

Sara Renjit and Sumam Mary Idicula. 2021. CUSATNLP@DravidianLangTech-EACL2021:Language Agnostic Classification of Offensive Content in Tweets. In *Proceedings*

of the First Workshop on Speech and Language Technologies for Dravidian Languages. Association for Computational Linguistics.

Debjoy Saha, Naman Paharia, Debajit Chakraborty, Punyajoy Saha, and Animesh Mukherjee. 2021. Hate-Alert@DravidianLangTech-EACL2021: Ensembling strategies for Transformer-based Offensive language Detection . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Richard Salomon. 1998. *Indian epigraphy: a guide to the study of inscriptions in Sanskrit, Prakrit, and the other Indo-Aryan languages*. Oxford University Press on Demand.

A. C. Sekhar. 1951. [Evolution of Malayalam]. *Bulletin of the Deccan College Research Institute*, 12(1/2):1–216.

Omar Sharif, Eftekhar Hossain, and Mohammed Moshiul Hoque. 2021. NLP-CUET@DravidianLangTech-EACL2021: Offensive Language Detection from Multilingual Code-Mixed Text using Transformers. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

R Sivanantham and M Seran. 2019. Keeladi: An Urban Settlement of Sangam Age on the Banks of River Vaigai. *India: Department of Archaeology, Government of Tamil Nadu, Chennai*.

BGL Swamy. 1975. The Date of Tolkappiyam: A Retrospect. *Annals of Oriental Research (Madras), Silver Jubilee*, 292:317.

Takanobu Takahashi. 1995. *Tamil love poetry and poetics*, volume 9. Brill.

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.

Debapriya Tula, Prathyush Potluri, Shreyas MS, Sumanth Doddapaneni, Pranjal Sahu, Rohan Sukumaran, and Parth Patwa. 2021. Bitions@DravidianLangTech-EACL2021: Ensemble of Multilingual Language Models with

Pseudo Labeling for Offense Detection in Dravidian Languages. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Charangan Vasantharajan and Uthayasanker Thayasivam. 2021. Hypers@DravidianLangTech-EACL2021: Offensive language identification in Dravidian code-mixed YouTube Comments and Posts. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Maoqin Yang. 2021. Maoqin @ DravidianLangTech-EACL2021: The Application of Transformer-Based Model . In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Konthala Yasaswini, Karthik Puranik, Adeep Hande, Ruba Priyadharshini, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. IIITT@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*. Association for Computational Linguistics.

Marcos Zampieri, Shervin Malmasi, Preslav Nakov, Sara Rosenthal, Noura Farra, and Ritesh Kumar. 2019. Predicting the type and target of offensive posts in social media. 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 1415–1420, Minneapolis, Minnesota. Association for Computational Linguistics.

Marcos Zampieri, Preslav Nakov, Sara Rosenthal, Pepa Atanasova, Georgi Karadzhov, Hamdy Mubarak, Leon Derczynski, Zeses Pitenis, and Çağrı Çöltekin. 2020. SemEval-2020 task 12: Multilingual offensive language identification in social media (OffensEval 2020). In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1425–1447, Barcelona (online). International Committee for Computational Linguistics.

Yingjia Zhao. 2021. ZYJ123@DravidianLangTech-EACL2021: Offensive Language Identification based on XLM-RoBERTa with DPCNN. In *Proceedings of the First Workshop on Speech and Language Technologies for Dravidian Languages*. Association for Computational Linguistics.

Kamil V Zvelebil. 1991. Comments on the Tolkappiyam Theory of Literature. *Archív Orientální*, 59:345–359.