



A Deep Study on Offensive Content Detection in Code-Mix Malayalam Language

Adolf P¹ ,* and Vijayakumar V²

¹ Research Scholar, Department of Computer Science, Sri Ramakrishna College of Arts and Science, Coimbatore, Tamil Nadu, India.

*Corresponding author. Email: adolfp37.pu@gmail.com

² Principal, Department of Computer Science, Nehru Arts and Science College, Coimbatore, Tamil Nadu, India.

veluvijay20@gmail.com

Abstract. Code-mixed languages are among the many user-generated communication techniques that have emerged as a result of social media's exponential expansion. In informal situations, the blending of two or more languages - code-mixing - has become an accepted practice in online communication. The use of abusive language in code-mixed text has increased in tandem with this trend, posing difficulties for maintaining inclusive and safe online spaces. Because regional and under-resourced languages frequently lack the resources and methods available for high-resource languages, it is especially important to identify objectionable idioms in these languages. One such instance is Malayalam, a Dravidian language, extensively spoken in India. Malayalam and English are often combined on social media sites to create code-mixed text that is widely utilized for communication across various user communities. However, because of dialect variations based on location, transliteration, spelling variation, trend-based variations, informal grammar, and the dynamic nature of online speech, identifying objectionable language in this context is more difficult than in monolingual literature. Robust techniques that can accurately predict the language and cultural subtleties of code-mixed information are needed to address these issues. In addition to analyzing current approaches, algorithms, strategies, and frameworks used in this developing field of study, this paper offers detailed study on Offensive language detection in Malayalam-English code-mixed social media. The paper acknowledges limits and emphasizes recent developments. In wider terms, the goal of this research is to explore and create useful applications in fields including cyber safety, content moderation, and the encouragement of positive online communication.

Keywords: Code-mixing, Malayalam-English, Offensive Language Detection, Social Media, Natural Language Processing (NLP).

1 Introduction

Social media platforms have become ubiquitous worldwide and are used by millions of people, leading to a tremendous amount of user-generated content generated every second on these platforms. Social media platforms provide a convenient and hassle-free way for people to express their opinions, views and talk about people, events, and happenings. Raw, unstructured data thus collected, formatted and analyzed, can serve as a rich resource for making decisions related to business strategies and policy making [25]. In this regard, research work in this domain was initially restricted to the English language as it belongs to a high resource language, with plenty of annotated corpora and natural language processing tools. The focus has switched to comprehending and identifying inappropriate information in regional and under-resourced languages, including their dialectal variations, as more users communicate in their mother tongues. This change has expanded the field of study, especially with the emergence of sentiment analysis, which is now strongly linked to the identification of objectionable words in native tongues. The problem of detecting objectionable language in low-resource languages is far more difficult than in high-resource languages like English. A large number of issues are related to the linguistic complexity of morphologically rich languages, the lack of extensive annotated data, and the shortage of feature extraction techniques. In this context, sentiment analysis is crucial since it makes it possible to extract valuable insights from user opinions that can further inform strategic and business decisions. However, most conversations and comments on social media are code-mixed, meaning that individuals blend their own native tongue with a more widely used language like English [24]. This blending is a reflection of the diverse and inclusive nature of online spaces, where people from many social and cultural backgrounds use distinctive language to express their opinions. Because it reflects genuine user sentiment, the study of such code-mixed content is very crucial. However, there is still a dearth of research in this area, particularly in languages with limited resources.

Millions of people in southern India speak Malayalam, a Dravidian language with a complicated morphology and a high degree of expressiveness. To express complex thoughts, it uses a multi-layered word generation mechanism [26]. The Malayalam script is an alpha-syllabary system that blends syllable-based and alphabetic writing principles. It was first formed from the Vattezhuth alphabet and later enlarged with symbols from the Grantha script [27], [28]. Malayalam is rarely used alone on social media, despite its linguistic diversity. Rather, it usually occurs in a code-mixed form, commonly intermingled with English, particularly in tweets, comments, and titles. Malayalam-English codemixed text, which frequently contains offensive terms, emerges as a result of this phenomena.

It is very difficult to identify offensive content in code-mixed Malayalam-English writing. This complexity is caused by a number of factors, including the absence of standardized and domain-specific datasets; the contextual diversity of data influenced by people, events, or incidents; the interaction of multiple languages within a single utterance; regional slang and dialectal variations; and the general lack of NLP research

resources devoted to code-mixed Malayalam. As a result, offensive language identification in this setting continues to be time-consuming and computationally demanding in addition to being challenging to understand.

This paper studies offensive content detection in Malayalam-English code-mixed text in order to address these issues, with an emphasis on current techniques, their shortcomings, and possible enhancements. The rest of this paper is organized as follows: A thorough review of relevant literature is presented in Section 2, the difficulties and constraints faced in this field of study are examined in Section 3, a generic model for offensive content detection in Malayalam-English code-mixed data is proposed in Section 4, performance evaluation methods are described in Section 5, and the work is concluded in Section 6 with a summary of findings and recommendations for future research.

2 Literature Review

The research works are categorized based on methods used such as machine learning, deep learning, natural language processing, data sets, pre-trained models. A review on different Deep Neural Network Models in Sentiment Analysis of Malayalam tweets consists of 1-D Convolutional Neural Networks (1-D CNN), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), Bi-directional LSTM (BiLSTM) and Gated Recurrent Unit (GRU) models [7]. Another survey is on Sentiment Analysis on Malayalam reviews which focus on key features of Natural Language Processing (NLP) such as Malayalam Part of Speech (PoS) tagging, stemming etc. [8].

A lexicon based sentiment analysis in Malayalam is done by Ashna M.P. et al. [17]. Domain- Independent sentiment analysis is done by V. Kasthoori et al. by using fuzzy logic algorithm and association tree rules [2]. A transformer-based pre-trained models with custom loss functions is proposed by Nikhil Kumar Ghanghor in hope speech detection [4]. Ying Chen et al. proposed a LSF - Lexical Syntactic Feature architecture to detect offensive content and to recognize the possible offensive users found in social media in English language [20]. Amir H et al. proposed multilevel detection of offensive content in English language [21]. Offensive language detection using artificial intelligence is proposed by Meredita Susanty et al. [22].

Support Vector Machine (SVM) is used for sentiment analysis of Malayalam film review by Deepu S. Nair [13]. A study based on Stochastic Gradient Descent (SGD) classifier and a Long Short-term Memory (LSTM) is proposed by B. S. R. Lakshmanan et al. on sentiment polarity analysis [5]. Social media sentiment analysis for Malayalam is proposed by M. Rahul et al. using Support Vector Machine [18]. Finite State Transducer based Morphology analysis for Malayalam Language is done by Santhosh Thottingal [12]. Sentiment analysis of Malayalam tweets using machine learning techniques such as Naive Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF) is proposed by S. Soumya et al. [15].

A research is proposed by Sai Muralidhar Jayanthi et al. by mBERT and XLM-RoBERTa models that leverage task adaptive pre-training of multilingual BERT models to identify offensive languages which includes Malayalam [3]. Sentence-level BERT and sentiment features are used in Malayalam-English code mixed sentiment

analysis is suggested by Y. P. Babu [6]. A study is done in offensive language identification in Dravidian languages by Kushal Kedia using BERT and RoBERTa models [10]. Sentiment analysis on multilingual code-mixing text using BERT including Malayalam-English is proposed by Huilin Sun et al. [14]. Sentiment Analysis and Offensive Language identification in Dravidian languages includes Malayalam is proposed by Adeep Hande et al. using mBERT and DistilBERT [16].

Sentiment Analysis of tweets in Malayalam using Long Short-Term Memory units and Convolutional Neural nets is done by S. Sachin kumar et al. [19]. Offensive language detection in Tamil, Malayalam and Kannada code-mixed with English Language is proposed by Omar Sharif by using multiple methods including machine learning, deep learning and transformers [11]. A code-mixed dataset is created for English-Malayalam by manual annotation is done by Bharathi Raja et al. [1]. Gowtham Ramesh et al. proposed a collection for 11 Indic languages named Samanantar including Malayalam and creates neural machine translator models [9]. Vijayakumar V, Hari Prasad D presented a review on details about machine and deep learning approach in order to detect and prevent multimodal and multilingual cyberbullying [23]. The detailed review of literature related to offensive content detection in code-mixed Malayalam Language is presented in Table 1.

Table 1. Offensive content detection in Code-Mix language - Literature review

Authors	Year	Method / Technology / Algorithm	Outcome
Sai Muralidhar Jayanthi, Akshat Gupta	2021	mBERT and XLM- RoBERTa models	Offensive Language Identification in Dravidian languages
Nikhil Kumar Ghanghor, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Ruba Priyadarshini, Sajeetha Thavareesan, Bharathi Raja Chakravarthi	2021	transformer-based pre-trained models	Hope Speech Detection
Omar Sharif, Eftekhar Hossain and Mohammed Moshiul Hoque	2021	Machine Learning, Deep Learning, Transformers	Offensive Language Detection
Bharathi Raja Chakravarthi, Navya Jose, Shardul Suryawanshi, Elizabeth Sherly John P. McCrae	2020	By manual annotation	Code-Mixed Malayalam-English dataset
BalaSundaraRaman Lakshmanan, Sanjeeth Kumar Ravindranath	2020	Stochastic Gradient Descent (SGD), Long Short-term Memory (LSTM)	Sentiment Classifier Polarity
Yandrapati Prakash Babu, Rajagopal Eswari and K Nimmi	2020	Sentence BERT	Sentiment Analysis in Malayalam-English mixed
Kushal Kedia, Abhilash Nandy	2020	BERT and	Offensive Language

		RoBERTa	Identification
Huilin Sun, Jiaming Gao and Fang Sun	2020	BERT	Sentiment Analysis
S. Soumya and Pramod K.V.	2020	Machine Learning	Sentiment Analysis
V. Kasthoori, B. Soniya and V. Jayan	2019	fuzzy logic and association tree rules	Domain-Independent Sentiment Analysis
Santhosh Thottingal	2019	Stuttgart Finite State Transducer(SFST)	morphology analyzing
Deepu S. Nair, Jisha P. Jayan, Rajeev R.R, Elizabeth Sherly	2015	Support Vector Machine	Sentiment Analysis

3 Challenges and Limitations

The detection of offensive content in Malayalam-English code-mixed social media text presents unique and multifaceted challenges. These difficulties stem from linguistic intricacies, computational constraints, and limited resource availability, collectively making it a highly complex research problem mentioned in Table 2.

Table 2. Offensive content detection - Challenges

	Romanization and Spelling Variability
Linguistic Intricacies	Dialect and Slang Variations
	Contextual Ambiguity
	Pre-trained Models Limited to Monolingual Malayalam
Computational Constrains	Imbalanced Class Distribution
	Feature Extraction Limitations
	Evaluation and Benchmarking
	Computational Complexity and Resource Constraints
	Scarcity of Annotated Datasets
Limited Resources Availability	Complexity of Code-Mixing
	Domain-Specific Variability
	Multimodal Offensive Content

3.1 Linguistic Intricacies

Romanization and Spelling Variability. Challenge emerges from the inconsistent Romanization of Malayalam words, which is highly prevalent across social media platforms is discussed by Nikhil Kumar Ghanghor et al. [4]. Malayalam does not have a

recognized Roman script convention, in contrast to languages with established transliteration requirements. Users consequently adopt improvised spellings that are frequently excessively shortened, phonetic, or influenced by casual digital communication methods. This high degree of unpredictability adds noise to the text, making basic tasks like feature extraction, vocabulary creation, and tokenization much more challenging. Furthermore, because semantically identical words might be represented as different tokens, the inconsistency interferes with embedding learning. In addition to making preprocessing more difficult, this irregularity drastically lowers the system effectiveness when handling data with mixed Malayalam and English codes.

Dialect and Slang Variations. Malayalam is a linguistically rich language with many dialects and a dynamic collection of slang terms. When it comes to offensive language identification, this variability becomes very difficult. Slang and culturally ingrained terms, which can have very diverse meanings in different cultures, are frequently used to convey obscene or abusive material [26]. These idioms are challenging for neural network models to identify and interpret since they are rarely included in formal dictionaries and are typically omitted from pre-trained embedding. Additionally, a term's offensiveness is frequently context-dependent; depending on usage and audience, the same word may be interpreted as amusing, neutral, or disparaging.

Contextual Ambiguity. The fact that offensive intent is not often communicated directly adds an additional layer of complication to the detection of offensive language. Alternatively, it might be quietly expressed through irony, humor, sarcasm, or culturally particular allusions that are difficult to understand without context. In many conversational contexts, words or phrases that seem insulting in isolation may have neutral or even positive meanings [27]. On the other hand, depending on tone and context, seemingly innocuous phrases can be used as weapons to convey disparaging overtones. Computational models run the danger of either overestimating or underestimating the presence of offensive intent in Malayalam-English code-mixed settings if they are unable to catch these subtle indicators.

3.2 Computational Constraints

Pre-trained Models Limited to Monolingual Malayalam. Pre-trained language models that have been created for multilingual applications, such as mBERT, MuRIL, and IndicBERT, are mostly designed for either broad multilingual environments or monolingual activities. The complexities present in code-mixed data are not directly addressed by the models. Because of this, they usually perform poorly when used to Malayalam-English texts, where linguistic boundaries are quite flexible and traditional grammatical structures are regularly broken. It is challenging for models trained on standard language corpora to capture both the syntactic inconsistencies and semantic complexities present because code-switching in this environment does not follow predictable principles.

Imbalanced Class Distribution. The difficulty of class imbalance found in the majority of datasets presents another challenge. Most samples are neutral or non-offensive, with objectionable content making up only a small portion of the total data. Machine learning models become biased in favor of the dominant class as a result of this disparity, which creates extremely unbalanced datasets. Because of this, these models frequently attain high overall accuracy by accurately detecting neutral content, but they have difficulty successfully identifying offending instances. Careful data augmentation, resampling techniques, or cost-sensitive learning methods are needed to address this issue in order to improve the resilience of offensive language detection systems and guarantee equitable representation of minority classes.

Feature Extraction Limitations. The hybrid nature of code-switching creates highly irregular patterns, and frequent spelling variations and non-standard Romanization further fragment the text; consequently, the extracted features tend to be sparse, inconsistent, and poorly representative of the underlying semantics. While deep learning methods have advanced the field, significant challenges were encountered when dealing with out-of-vocabulary (OOV) terms, creative spellings, and informal digital expressions that dominate social media discourse.

Evaluation and Benchmarking. The lack of a common benchmark dataset or evaluation methodology for Malayalam-English offensive content detection is a significant obstacle. Furthermore, evaluation methods vary greatly, with researchers using a variety of performance criteria like accuracy, F1-score, or precision-recall trade-offs, which makes it challenging to compare models directly. This discrepancy hinders the creation of trustworthy baselines and slows down the field's overall advancement because advancements noted in one study might not be transferable to another. To enable fair comparisons and promote more systematic developments in Malayalam-English code-mixed offensive language identification, it will be necessary to develop standardized benchmark corpora and evaluation methodologies.

Computational Complexity and Resource Constraints. The primary obstacle to experimentation is the computational requirements for training and fine-tuning transformer-based models, such as XLM-RoBERTa and MuRIL, which are envisioned to have billions of parameters and, as a result, typically require substantial computational infrastructure in the form of high-end GPUs or TPUs, large-scale memories, and extensive storage capacity. This is the main reason why state-of-the-art approaches for Malayalam-English code-mixed offensive language detection have not yet become widely used.

3.3 Limited Resources Availability

Scarcity of Annotated Datasets. The lack of extensive, standardized, and methodically annotated datasets is one of the main obstacles in the study of Malayalam-English

code-mixed text. The few datasets that are now available are usually small, restricted to specific domains, and frequently gathered from websites that are prohibited, like YouTube or Twitter. Training models with enough robustness and generalizability is severely hampered by this limited coverage. Charavarthi et al. conducted the first research on the Malayalam language dataset [1]. The lack of good quality corpora still limits meaningful progress in natural language processing for Malayalam-English code-mixed data.

Complexity of Code-Mixing. Code-mixing is a linguistic trend that poses unique and challenging issues for natural language processing. In talks, speakers frequently switch between Malayalam and English, sometimes many times in a single sentence. Since these changes often occur without regular patterns, it is difficult to formulate accurate structural principles. The result is literature with inconsistent sentence patterns, hybridized grammar, and overlapping linguistic traditions. An additional degree of ambiguity is introduced by the highly context-dependent semantic meanings of code-mixed phrases. Because of this, existing models struggle to parse, understand, and generate reliable results for content that is code-mixed between Malayalam and English.

Domain-Specific Variability. The diversity of unpleasant information on various social media sites presents another difficulty. Character restrictions, audience demographics, and social standards all have an impact on how each platform promotes communication. For example, YouTube comments frequently have conversational, informal tones that are significantly influenced by regional phrases and slang. Twitter posts, on the other hand, are usually brief and frequently employ emoji, hashtags, code-mixing shortcuts, and abbreviations to express meaning. Because of these stylistic variations, offensive language appears on various platforms in a variety of ways. The domain-specific nature of offensive language is highlighted by the fact that models trained on data from one platform typically perform poorly when applied to another.

Multimodal Offensive Content

It's critical to understand that objectionable social media content extends beyond linguistic expressions. Memes, altered photos, brief videos, and even emoji sequences with sarcastic or implicit meanings are examples of multimodal formats that are used in a large amount of abusive or disparaging communication. These non-textual or text-plus-visual combinations are just as, if not more, damaging than plain text since they frequently increase the effect of offensive content [23]. Addressing this gap requires incorporating multimodal learning approaches that can jointly process text, imagery, and symbolic cues to achieve more comprehensive offensive content detection.

4 Generic Model of Offensive Content Detection in Code-Mix Malayalam Language

The code-mixed Malayalam language is taken as input for the detection system. The inputs are preprocessed before taking to the system. A NLP model is trained and used for detection purposes. The output is categorized as content detected and not detected. The model is shown in figure 1.

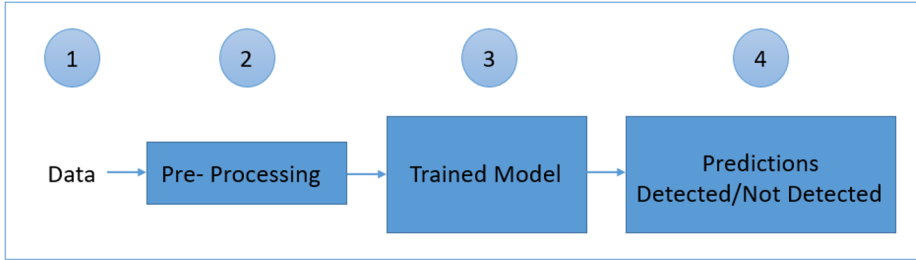


Fig. 1. Offensive Content Detection - Diagram

4.1 Step by step explanation:

STEP 1. The code mixed Malayalam language dataset is selected from different data sources such as website, social media and language forums etc.

STEP 2. Pre-processing functions are applied to the collected dataset to improve the detection accuracy like tokenization - the process of converting text into tokens before transforming it into vectors, stop words removal – removing commonly used words (i.e. “the”, “a”, “an”), normalization - a processes analyze the words such as highly unstructured text like 2morrow is converted as tomorrow which ensures that words treated equally, and lemmatization - removes the grammar tense and transforms each word into its original form.

STEP 3. The detection models are developed using machine learning and deep learning algorithms (CNN, LSTM, transformers) or pre-trained models like BERT, ROBERTa, XLM-ROBERTa etc .

STEP 4. The predictions of the NLP model are classified as whether the content is of-fensive, non-offensive and neutral.

5 Performance Measures for Offensive Content Detection

In offensive content detection tasks, evaluating the model’s performance requires robust and reliable metrics that account for imbalanced datasets and multi-class scenarios often found in social media text. It is observed that performance evaluations use precision, recall, F1-score, Accuracy.

Accuracy measures the proportion of correct predictions (both offensive and non-offensive) out of the total predictions. Though accuracy provides a general performance

overview, it can be misleading in imbalanced datasets where one class (e.g., non-offensive content) dominates. For example, a model predicting every comment as "non-offensive" may achieve high accuracy but fail completely at detecting offensive content. Recall measures the ability of the model to identify all the offensive content present in the dataset. High recall means the system is catching most offensive comments, but it may also bring in more false positives if not balanced with precision. Precision measures the proportion of correctly predicted offensive content out of all the instances that the model labeled as offensive. A higher precision indicates that when the model predicts an instance as offensive, it is likely to be correct. Precision is especially important in scenarios where false positives (incorrectly flagging non-offensive content) have serious implications, such as unnecessary censorship or wrongful flagging of user content. The F1-score which is the harmonic mean of precision and recall, balance both the metrics into a single score. The F1-score is useful when there is an imbalance between offensive and non-offensive samples, as it provides a better measure of the model's ability to handle both false positives and false negatives.

Mathematically, these can be defined as:

$$Precision = \frac{TP}{(TP + FP)} \quad (SEQ \text{ equation } \textcolor{red}{1} \textcolor{red}{1} \text{ MERGEFORMAT } 1)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (SEQ \text{ equation } \textcolor{red}{1} \textcolor{red}{1} \text{ MERGEFORMAT } 2)$$

$$F1 = \frac{(Precision * Recall)}{(Precision + Recall)} \quad (SEQ \text{ equation } \textcolor{red}{1} \textcolor{red}{1} \text{ MERGEFORMAT } 3)$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (SEQ \text{ equation } \textcolor{red}{1} \textcolor{red}{1} \text{ MERGEFORMAT } 4)$$

where TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative

Additional to the above researchers considered Macro-Averaging which calculates the metrics independently for each class also takes average, treating all classes equally and Weighted-Averaging which takes frequency of each class into account, giving more weight to dominant classes.

6 Findings and Conclusion

Major contributions in the field of offensive content detection in code-mixed Malayalam language is discussed in this paper. Many research works are analyzed based on database, algorithms etc. Based on the literature study a lack of study area is identified. A lack of standard data set, pre-trained models and effective offensive content detection was observed in code-mixed Malayalam language. This article will be useful for the researchers to get an overview of the major contributions done in offensive content detection in code-mixed Malayalam.

This research undertook an extensive review of the existing body of work in domain of offensive content detection in Malayalam-English code-mixed text, a rapidly evolving area within natural language processing and computational linguistics. This study examined a variety of topics by critically evaluating a large number of research, including the development and curation of datasets, preprocessing methods, machine learning and deep learning approaches, and the use of sophisticated transformer-based models for offensive language detection. From rule-based and lexicon-driven techniques to complex deep learning and pre-trained neural language models like mBERT, XLM-R, and IndicBERT, the analysis showed a consistent evolution in research methodologies. These developments show how contextual representation learning is becoming increasingly important for managing the complicated, hybrid character of code-mixed text data.

The assessment identifies important gaps and enduring difficulties in the area despite these developments. One of these is the lack of substantial, standardized, high-quality annotated datasets for code-mixed text between Malayalam and English, which still makes it difficult to train and test reliable models. The linguistic complexity of code-mixing, which is typified by inconsistent grammar, numerous dialectal variations, inconsistent Romanization, and region-specific slang, is another noteworthy challenge that presents substantial obstacles for both feature engineering and contextual embedding techniques.

It is challenging for models trained in one environment to properly generalize to another due to the significant differences in content and communication styles across platforms such as Facebook, Twitter, and YouTube. Performance constraints are made worse by problems like class imbalance, in which non-offensive information predominates in datasets that are available, especially when it comes to attaining high recall for offensive classes.

This research has important implications for practice. The amount of user-generated content using derogatory, damaging, or abusive language is increasing along with social media usage. Online moderation, user safety, and the development of healthier digital spaces all depend on the development of dependable, scalable, and context-aware algorithms for offensive material detection.

This review also provides a valuable input to the future research. Addressing current gaps requires collaborative efforts to:

- Develop standardized, large-scale annotated datasets that capture the diversity of Malayalam-English code-mixing.
- Fine-tune or build context-aware pre-trained models that effectively represent hybrid language structures.
- Explore multi-task and multimodal approaches that combine offensive language detection with related tasks such as sentiment analysis, stance detection, and emotion recognition.
- Establish benchmarking frameworks to enable consistent evaluation and comparison of methodologies.

In summary, this study offers a comprehensive synthesis of the state of research in offensive content detection for Malayalam-English code-mixed language. It highlights the urgent issues that must be resolved in addition to identifying current contributions and new trends. For researchers, practitioners, and developers working to develop reliable, context-sensitive systems for offensive content detection in under-resourced, code-mixed language settings, this review is an essential resource because it offers both a thorough understanding of current limitations and practical insights.

Acknowledgments. The authors thank the Sri Ramakrishna College of Arts and Science who provided the support for doing the academic paper.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chakravarthi, B. R., Jose, N., Suryawanshi, S., Sherly, E., & McCrae, J. P. (2020, May). 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)* (pp. 177-184).
2. Kasthoori, V., Soniya, B., & Jayan, V. (2018). Domain-independent sentiment analysis in malayalam. In *Computational Intelligence: Theories, Applications and Future Directions-Volume II: ICCI-2017* (pp. 151-160). Singapore: Springer Singapore.
3. Jayanthi, S. M., & Gupta, A. (2021, April). Sj_aj@ 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* (pp. 307-312).
4. Ghanghor, N., Ponnusamy, R., Kumaresan, P. K., Priyadharshini, R., Thavareesan, S., & Chakravarthi, B. R. (2021, April). IIITK@ 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* (pp. 197-203).
5. Lakshmanan, B., & Ravindranath, S. K. (2020). Theedhum Nandrum@ Dravidian-CodeMix-FIRE2020: a sentiment polarity classifier for YouTube comments with code-switching between Tamil, Malayalam and English. *arXiv preprint arXiv:2010.03189*.

6. Babu, Y. P., Rajagopal, E., & Nimmi, K. (2020). CIA_NITT@ Dravidian-CodeMix-FIRE2020: Malayalam-English Code Mixed Sentiment Analysis Using Sentence BERT And Sentiment Features. In *FIRE (Working Notes)* (pp. 566-573).
7. Soumya, S., & Pramod, K. V. (2019, November). Sentiment analysis of Malayalam tweets using different deep neural network models-case study. In *2019 9th International Conference on Advances in Computing and Communication (ICACC)* (pp. 163-168). IEEE.
8. Mathews, D. M., & Abraham, S. (2018). Sentiment analysis on Malayalam language: a survey. *Int. J. Manag. Technol. Eng*, 8, 1046-1054.
9. Ramesh, G., Doddapaneni, S., Bheemaraj, A., Jobanputra, M., Ak, R., Sharma, A., ... & Khapra, M. S. (2022). Samanantar: The largest publicly available parallel corpora collection for 11 indic languages. *Transactions of the Association for Computational Linguistics*, 10, 145-162.
10. Kedia, K., & Nandy, A. (2021, April). indicnlp@ kgp at DravidianLangTech-EACL2021: Offensive language identification in Dravidian languages. In *Proceedings of the first workshop on speech and language technologies for dravidian languages* (pp. 330-335).
11. Sharif, O., Hossain, E., & Hoque, M. M. (2021, April). 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* (pp. 255-261).
12. Thottingal, S. (2019, August). Finite state transducer based morphology analysis for Malayalam language. In *Proceedings of the 2nd Workshop on Technologies for MT of Low Resource Languages* (pp. 1-5).
13. Soumya, S., & Pramod, K. V. (2020). Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*, 6(4), 300-305.
14. Sun, H., Gao, J., & Sun, F. (2020). HIT_SUN@ Dravidian-CodeMix-FIRE2020: Sentiment Analysis on Multilingual Code-Mixing Text Base on BERT. In *FIRE (Working Notes)* (pp. 517-521).
15. Soumya, S., & Pramod, K. V. (2020). Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*, 6(4), 300-305.
16. Hande, A., Hegde, S. U., Priyadharshini, R., Ponnusamy, R., Kumaresan, P. K., Thavaresan, S., & Chakravarthi, B. R. (2021). Benchmarking multi-task learning for sentiment analysis and offensive language identification in under-resourced dravidian languages. *arXiv preprint arXiv:2108.03867*.
17. Ashna, M. P., & Sunny, A. K. (2017, July). Lexicon based sentiment analysis system for malayalam language. In *2017 International conference on computing methodologies and communication (ICCMC)* (pp. 777-783). IEEE.
18. Rahul, M., Rajeev, R. R., & Shine, S. (2018). Social media sentiment analysis for Malayalam. *Int. J. Comput. Sci. Eng*, 6(06), 48-53.
19. Kumar, S. S., Kumar, M. A., & Soman, K. P. (2017, November). Sentiment analysis of tweets in malayalam using long short-term memory units and convolutional neural nets. In *International Conference on Mining Intelligence and Knowledge Exploration* (pp. 320-334). Cham: Springer International Publishing.
20. Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012, September). Detecting offensive language in social media to protect adolescent online safety. In *2012 international conference on privacy, security, risk and trust and 2012 international confernece on social computing* (pp. 71-80). Ieee.
21. Razavi, A. H., Inkpen, D., Uritsky, S., & Matwin, S. (2010, May). Offensive language detection using multi-level classification. In *Canadian Conference on Artificial Intelligence* (pp. 16-27). Berlin, Heidelberg: Springer Berlin Heidelberg.

22. Susanty, M., Rahman, A. F., Normansyah, M. D., & Irawan, A. (2019, March). Offensive language detection using artificial neural network. In *2019 International Conference of Artificial Intelligence and Information Technology (ICAIT)* (pp. 350-353). IEEE.
23. Vijayakumar, V., & Hari Prasad, D. (2021). A Study on deep learning algorithms for multi-modal and multilingual cyberbullying detection. *Indian Journal of Applied Research*, 11(7), 50-54.
24. Kester, E., & Kester, E. (2024, September 12). Code-Switching and Code-Mixing – What you need to know. Bilinguistics. <https://bilinguistics.com/code-switching-and-code-mixing/>
25. Chaffey, Dave. “Global Social Media Statistics Research Summary.” Smart Insights, 14 Feb. 2025, www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/
26. Wikipedia contributors. (2026, March 29). Malayalam script. Wikipedia. https://en.wikipedia.org/wiki/Malayalam_script
27. Grandha script & malayalam language. <http://www.keralaculture.org/malayalam-script/243> Accessed: 2026/03/12.
28. “Vatteluttu - Wikiwand.” Wikiwand, 2023, www.wikiwand.com/en/Vatteluttu. Accessed 2026/03/29.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

