Lidoma@LT-EDI 2024:Tamil Hate Speech Detection in Migration Discourse

M. Shahiki Tash, Z. Ahani, M. T. Zamir, O. Kolesnikova and **G. Sidorov**
Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC)
Corresponding: mshahikit2022@cic.ipn.mx

Abstract

The exponential rise in social media users has revolutionized information accessibility and exchange. While these platforms serve various purposes, they also harbor negative elements, including hate speech and offensive behavior. Detecting hate speech in diverse languages has garnered significant attention in Natural Language Processing (NLP). This paper delves into hate speech detection in Tamil, particularly related to migration and refuge, contributing to the Caste/migration hate speech detection shared task. Employing a Convolutional Neural Network (CNN), our model achieved an F1 score of 0.76 in identifying hate speech and signaling potential in the domain despite encountering complexities. We provide an overview of related research, methodology, and insights into the competition's diverse performances, showcasing the landscape of hate speech detection nuances in the Tamil language.

1 Introduction

The surge in Social Media platform users has led to a significant increase in information dissemination, granting immediate access to updated information with just a click. These platforms are used not only for social interaction but also for leisure and information retrieval (Sajjad et al., 2019; Ali et al., 2022a). There has been a notable surge in interest in social media analysis tasks within NLP (Bade, 2021), With a focus on emerging fields like identifying hopeful speech, there is a growing emphasis on advancing in this direction. (Yigezu et al., 2023a; Shahiki-Tash et al., 2023b) language identification (Tash et al., 2022; Balouchzahi et al., 2022a), fake news(Fazlourrahman et al., 2022), sentiment analysis (Tash et al., 2023; Yigezu et al., 2023b), and hate speech(Yigezu et al., 2023c) that researchers experimented with diverse models, including deep learning (Yigezu et al., 2022; Ahani et al., 2024), transformers (Tonja et al., 2022), and traditional machine learning techniques (Kanta and Sidorov, 2023).

However, along with its advantages, the widespread adoption of Social Media (Bade and Afaro, 2018) also brings negative aspects (Ali et al., 2022b). Users sometimes exhibit behavior that can be harmful, offensive, and even hateful toward various segments of society (Shahiki-Tash et al., 2023a).

Describing hate speech is complex, as Andrew Sellars argues against oversimplification of its definition and addressing methods (Sellars, 2016). There's disagreement regarding how hate speech refers to groups, with certain definitions associating it with minority groups or specific characteristics like race, religion, gender, or sexual orientation (Waltman and Mattheis, 2017).

This challenge has led to several shared tasks focused on detecting hate speech. In this context, our article centers on the analysis of Tamil user comments about migration and refuge using qualitative content analysis. As part of this effort, we participated in the Caste/migration hate speech detection shared task (Rajiakodi et al., 2024), which aims to develop models capable of identifying hate speech related to caste or migration.

The objective of this task is to create an automated classification system that predicts whether text, particularly on social media, contains caste/migration-related hate speech. We employed a CNN model for prediction, leveraging its successful track record in text classification within the literature. (Balouchzahi et al., 2023b,a). our proposed model obtained an F1 score of 0.76, yielding promising performance on the task of binary hate speech detection.

2 Related work

Motivated by the linguistic diversity across India, where languages like Tamil, Telugu, Kannada,

Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 184–189

Malayalam, Hindi, Punjabi, Bengali, Gujarati, Marathi, among others, are prevalent, researchers observed the limitations of models confined to English proficiency (Chakravarthi et al., 2020b). This prompted the development of a system capable of processing code-mixed languages for sentiment analysis. A significant hurdle in this endeavor has been the scarcity of labeled datasets. Notably, a few manually annotated datasets for offensive language and hate speech detection in Tamil (Chakravarthi et al., 2020b), Malayalam (Chakravarthi et al., 2020a), and Kannada (Hande et al., 2021) have been released, marking crucial contributions to the field. The study (Sánchez-Holgado et al., 2022) aimed to assess the relationship between online hate speech against migrants and refugees and social acceptance in Spain. Using Intergroup Contact and Mediated Intergroup Contact Theory, the research sought to validate hate speech as an indicator of social acceptance across Spanish provinces. Analyzing 97,710 tweets and secondary public data on migration, the study found no significant correlation between hate speech, foreign population proportions, and citizen attitudes toward immigrants. Despite fluctuations in hate speech presence from 2015 to 2020, no clear negative correlation emerged between foreign population proportions and hate speech on Twitter. Similarly, the anticipated negative correlation between attitudes toward migration and hate speech on Twitter could not be statistically confirmed.

The paper (Sanguinetti et al., 2018) outlines the development of a novel Twitter corpus comprising roughly 6,000 tweets annotated for hate speech targeting immigrants. This corpus aimed to serve as a reference dataset for monitoring hate speech through automated systems. The annotation scheme was meticulously crafted to encompass various factors influencing hate speech, resulting in a tagset beyond hate speech alone, including aggressiveness, offensiveness, irony, stereotype, and experimental intensity categories. While discussing the annotated data, the study focuses on hate speech intensity and its interrelation with stereotype, aggressiveness, and offensiveness. The findings indicate nuanced trends, showcasing implicit incitement in most hateful tweets. Stereotype prevalence is notably high in lower intensity degrees, indicating its role in implicit incitement.

The study (Anbukkarasi and Varadhaganapathy, 2022) achieved notable success in hate speech detection within code-mixed Tamil-English tweets using a synonym-based Bi-LSTM model. With an F1 score of 0.8169, the Bi-LSTM model outperformed other models evaluated, demonstrating its effectiveness in distinguishing hate and non-hate texts. Specifically, in classifying hate speech, the model attained an F1 score of 0.8110, while for non-hate texts, it achieved an F1 score of 0.8050

This study (Basava and Karri, 2021) tackles the pervasive issue of hate speech proliferation across social media platforms by introducing an ensemble system utilizing transformer models. Specifically, it aims to identify offensive language within code-mixed posts/comments in Dravidian Languages (Malayalam-English and Tamil-English). Situated within the framework of the Hate Speech and Offensive Content Identification in Dravidian-CodeMix (HASOC) (Chakravarthi et al., 2021) initiative, this research emphasizes the rising impact of hate speech online and the urgent need for robust detection methods. The ensemble method showcased promising performance during development, notably achieving scores of 0.93 for Tamil and 0.80 for Malayalam, utilizing the model HSU_TransEmb. However, when assessed on the test set, the performance declined, registering 0.66 for Tamil with the MuRIL model and 0.73 for Malayalam using HSU_TransEmb, indicating the necessity for more comprehensive datasets to enhance model robustness and efficacy in tackling hate speech in multilingual social media settings. (Jayanthi and Gupta, 2021) applied transformer-based models, utilizing a cased version of multilingual BERT and XLM-RoBERTa. Employing BERT at the sentence level, they transformed subword-level representations into word-level representations by averaging sub-token representations for improved classification. This innovative fusion architecture integrated a Bidirectional LSTM model to capture diverse word patterns, enhancing classification accuracy, and resulting in a 79.67% accuracy in classifying Tamil tweets.

3 Methodology

Convolutional Neural Networks (CNNs) excel in text classification tasks by utilizing convolutional and pooling layers to extract hierarchical features from sequential data, such as text (Balouchzahi et al., 2022b). These networks employ convolutional filters of varying sizes to detect n-gram features within the input text, followed by pooling layers that condense and aggregate the extracted

features. By learning local relationships between words and capturing essential patterns, CNNs effectively discern hate speech or offensive language within textual data. Their ability to model intricate relationships within text makes CNNs a potent tool in the realm of hate speech detection.

3.1 Dataset

The dataset (Chakravarthi, 2020, 2022) is formatted in CSV (Comma-Separated Value), featuring columns labeled "Text" and "Tag". The "Text" column contains the textual content, while the "Tag" column signifies whether a comment is categorized as caste/migration hate speech, indicated by values: 1 for caste/migration hate speech and 0 for non-caste/migration hate speech (Chakravarthi et al., 2022).

The exemplification of the dataset structure is illustrated in Table 1.

Table 1: Tamil comments and their labels

Text	Tag
Ippadiye solli tamilanai izhivu paduthuvathey indha sangi kumbal, dhaanda, tamilians are getting educated, they want better life, mostly looking for decent job.	0
Freedom app eh. Bunda Advertisement Vera ya	1
Like this one day all these North Indias are going to chase every Tamilians from Tamilian Nadu. This is very dangerous. Need to probe into this and I request that all the Tamil people not to give these North Indians any accommodation. We need to save our Rights and control North Indians heavy migration. These people are hooligans.	1
it's nothing wrong people travel to earn money but in same time native people also need work hard for better life...lucky Brother you know hindi to communicate to Vadakans...Nice review	0

3.2 Classification algorithm

The classification algorithm we've designed encompasses several sequential steps, each contributing to the overall process. Below, we'll elaborate on these stages to provide a comprehensive understanding of our classification methodology.

3.3 Cleaning Data

The initial part of the code involves data cleaning functions like "remove_emoji", "remove_url", and "clean_text". These functions are applied to both the training and test datasets to eliminate emojis, URLs, special characters, and punctuation from the text. It ensures that the text is sanitized for further processing and analysis.

3.4 Padding

Tokenizer and padding functions from Keras are employed to convert text data into sequences of integers and ensure uniform sequence length. The "Tokenizer" converts text to numerical sequences, and "pad_sequences" ensures uniform length for modeling purposes, enhancing compatibility with neural network layers.

3.5 Label Encoding

Label encoding is performed using "LabelEncoder" from Scikit-learn to convert categorical labels into numerical format, preparing them for model training. Additionally, one-hot encoding ("tf.keras.utils.to_categorical") is applied to represent categorical labels as binary vectors.

3.6 Model Architecture

The neural network architecture comprises several layers: an embedding layer, a 1D convolutional layer ("Conv1D"), global max pooling, dropout, and a dense layer. Regularization techniques like L2 regularization are employed to prevent overfitting. The model summary provides a detailed overview of the architecture, including layer types, output shapes, and parameters.

3.7 Model Compilation and Training

The model is compiled using a categorical cross-entropy loss function and the Nadam optimizer. The code then trains the model using the training dataset ("train_ds") for 50 epochs, with validation performed on the validation dataset ("valid_ds"). Training history is recorded to monitor model performance and convergence.

3.8 Model Evaluation and Prediction

After training, the model is utilized to generate predictions on the test data ("x_test"), providing insights into the model's performance on unseen data. Additionally, metrics like classification reports or confusion matrices were derived to evaluate model performance comprehensively.

4 Results

The competition observed diverse performances in the detection of hate speech in the Tamil language. Prominent teams securing positions 1-3 demonstrated commendable M_F1 scores ranging from 0.82 to 0.80, indicating the effectiveness of their strategies. In contrast, the bottom-ranking teams (15-16) encountered challenges, attaining lower scores of 0.49 and 0.38, respectively. The 6th position achieved by our team, with an M_F1 score of 0.76, underscores the complexities involved in addressing nuances of hate speech in Tamil. Although our approach exhibited competence, the competitive environment and intricate nature of the task underscore the necessity for further refinement in areas such as data handling, feature engineering, and model fine-tuning. A detailed presentation of the results is available in Table 2.

Table 2: Performance Rankings of Hate Speech Detection Models in Tamil Language

Team name	M_F1	Rank
Transformers - Kriti Singhal	0.82	1
kubapok - Jakub Pokrywka	0.81	2
CUET_NLP_Manning	0.80	3
BITS_Graph4NLP	0.77	4
Algorithmalliance	0.76	5
lidoma - Moein Tash	0.76	6
CUET_NLP_GoodFellows	0.75	7
quartet - shaun Allan	0.73	8
KEC_AI_DSNLP_	0.65	9
selam - Selam Abitte	0.62	10
byteSizedllm	0.61	11
SSN-nova - Ankitha Reddy	0.59	12
WordWizards_tamil	0.54	13
KEC_DL_KSK - Kalaivani K.S.	0.49	14
Habesha - mesay gemeda	0.38	15

5 limitations

1. The study encounters a limitation stemming from the absence of hyperparameter tuning in the experimental setup. Optimal hyperparameter configurations are crucial in fine-tuning the performance of machine learning models, and their absence in our experiments could impact the overall effectiveness of our approach.

2. Another constraint in our methodology lies in the omission of experiments specifically designed to address the challenge of imbalanced datasets. Hate speech detection tasks often contend with imbalances between the number of instances belonging to different classes. Strategies such as oversampling, undersampling, or utilizing specialized algorithms for imbalanced datasets could be explored to enhance the model's ability to handle such data distribution challenges.

3. Our study is also constrained by the lack of incorporation of any feature selection techniques. Feature selection plays a vital role in enhancing model interpretability, reducing computational complexity, and potentially improving predictive performance. Future iterations of our methodology could benefit from the integration of feature selection methods to identify and retain the most informative features.

4. An additional limitation is the absence of any ensemble model in our experimental framework. Ensemble models, which combine predictions from multiple models, often contribute to improved generalization and robustness. Integrating ensemble techniques, such as bagging or boosting, could offer a more comprehensive and resilient hate speech detection system. This represents an avenue for future research to explore and enhance the overall performance of our approach.

6 Conclusion

This research delves into the realm of hate speech detection in Tamil, with a particular emphasis on themes related to migration and refuge within the framework of the Caste/migration hate speech detection shared task. Leveraging a Convolutional Neural Network (CNN), our model exhibited a commendable F1 score of 0.76, demonstrating its efficacy in identifying hate speech amidst inherent complexities. The analysis sheds light on the competitive landscape, uncovering diverse performances across teams with scores ranging from 0.38 to 0.82. These variations underscore the challenges inherent in addressing hate speech nuances in the Tamil language. As part of our future endeavors, we intend to enhance our approach by expanding our dataset and incorporating transformer models, aiming to further improve the accuracy of hate speech detection in this linguistic context.

Ethics Statement

We affirm our commitment to ethical research practices and compliance with ACL guidelines in conducting and presenting our study. No ethical concerns or conflicts of interest arose during the course of this research.

Acknowledgments

The work was done with partial support from the Mexican Government through the grant A1-S-47854 of CONACYT, Mexico, grants 20241816, 20241819, and 20240951 of the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional, Mexico. The authors thank the CONACYT for the computing resources brought to them through the Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje of the Laboratorio de Supercómputo of the INAOE, Mexico and acknowledge the support of Microsoft through the Microsoft Latin America PhD Award.

References

Zahra Ahani, Moein Shahiki Tash, Yoel Ledo Mezquita, and Jason Angel. 2024. Utilizing deep learning models for the identification of enhancers and super-enhancers based on genomic and epigenomic features. *arXiv preprint arXiv:2401.07470*.

Raza Ali, Umar Farooq, Umair Arshad, Waseem Shahzad, and Mirza Omer Beg. 2022a. Hate speech detection on twitter using transfer learning. *Computer Speech Language*, 74:101365.

Raza Ali, Umar Farooq, Umair Arshad, Waseem Shahzad, and Mirza Omer Beg. 2022b. Hate speech detection on twitter using transfer learning. *Computer Speech & Language*, 74:101365.

S Anbukkarasi and S Varadhaganapathy. 2022. Deep learning-based hate speech detection in code-mixed tamil text. *IETE Journal of Research*, pages 1–6.

Girma Yohannis Bade. 2021. Natural language processing and its challenges on omotic language group of ethiopia. *Journal of Computer Science Research*, 3(4):26–30.

Girma Yohannis Bade and Akalu Assefa Afaro. 2018. Object oriented software development for artificial intelligence. *American Journal of Software Engineering and Applications*, 7(2):22–24.

Fazlourrahman Balouchzahi, Sabur Butt, A Hegde, Noman Ashraf, HL Shashirekha, Grigori Sidorov, and Alexander Gelbukh. 2022a. Overview of colikanglish: Word level language identification in code-mixed kannada-english texts at icon 2022. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, pages 38–45.

Fazlourrahman Balouchzahi, Sabur Butt, Grigori Sidorov, and Alexander Gelbukh. 2023a. Reddit: Regret detection and domain identification from text. *Expert Systems with Applications*, 225:120099.

Fazlourrahman Balouchzahi, Anusha Gowda, Hosahalli Shashirekha, and Grigori Sidorov. 2022b. Mucic@ tamilnlp-acl2022: Abusive comment detection in tamil language using 1d conv-lstm. In *Proceedings of the Second Workshop on Speech and Language Technologies for Dravidian Languages*, pages 64–69.

Fazlourrahman Balouchzahi, Grigori Sidorov, and Alexander Gelbukh. 2023b. Polyhope: Two-level hope speech detection from tweets. *Expert Systems with Applications*, 225:120078.

Sai Naga Viswa Chaitanya Basava and Anjali Poornima Karri. 2021. Transformer ensemble system for detection of offensive content in dravidian languages. In *Working Notes of FIRE 2021-Forum for Information Retrieval Evaluation (Online). CEUR*.

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. 2022. Hope speech detection in youtube comments. *Social Network Analysis and Mining*, 12(1):75.

Bharathi Raja Chakravarthi, Anand Kumar M, John P McCrae, B Premjith, KP Soman, and Thomas Mandl. 2020a. Overview of the track on hasoc-offensive language identification-dravidiancodemix. In *FIRE (Working notes)*, pages 112–120.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, Subalalitha Cn, John McCrae, Miguel Ángel García, Salud María Jiménez-Zafra, Rafael Valencia-García, Prasanna Kumaresan, Rahul Ponnusamy, Daniel García-Baena, and José García-Díaz. 2022. Overview of the shared task on hope speech detection for equality, diversity, and inclusion. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 378–388, Dublin, Ireland. Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Vigneshwaran Muralidaran, Ruba Priyadharshini, and John P McCrae. 2020b. Corpus creation for sentiment analysis in code-mixed tamil-english text. *arXiv preprint arXiv:2006.00206*.

BR Chakravarthi, PK Kumaresan, R Sakuntharaj, AK Madasamy, S Thavareesan, S Chinnaudayar Navaneethakrishnan, and T Mandl. 2021. Overview of the hasoc-dravidiancodemix shared task on offensive language detection in tamil and malayalam. In *Working Notes of FIRE 2021-Forum for Information Retrieval Evaluation*. CEUR.

B Fazlourrahman, BK Aparna, and HL Shashirekha. 2022. Coffitt-covid-19 fake news detection using fine-tuned transfer learning approaches. In *Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2*, pages 879–890. Springer.

Adeep Hande, Siddhanth U Hegde, Ruba Priyadharshini, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Sajeetha Thavareesan, and Bharathi Raja Chakravarthi. 2021. Benchmarking multi-task learning for sentiment analysis and offensive language identification in under-resourced dravidian languages. *arXiv preprint arXiv:2108.03867*.

Sai Muralidhar Jayanthi and Akshat Gupta. 2021. Sj_aj@ dravidianlangtech-eacl2021: Task-adaptive pre-training of multilingual bert models for offensive language identification. *arXiv preprint arXiv:2102.01051*.

Selam Kanta and Grigori Sidorov. 2023. Selam@ dravidianlangtech: Sentiment analysis of code-mixed dravidian texts using svm classification. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 176–179.

Saranya Rajiakodi, Bharathi Raja Chakravarthi, Rahul Ponnusamy, Prasanna Kumar Kumaresan, Sathiyaraj Thangasamy, Bhuvaneswari Sivagnanam, and Charmathi Rajkumar. 2024. Overview of Shared Task on Caste and Migration Hate Speech Detection. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Muhammad Sajjad, Fatima Zulifqar, Muhammad Usman Ghani Khan, and Muhammad Azeem. 2019. Hate speech detection using fusion approach. In *2019 International Conference on Applied and Engineering Mathematics (ICAEM)*, pages 251–255. IEEE.

Patricia Sánchez-Holgado, Javier J Amores, and David Blanco-Herrero. 2022. Online hate speech and immigration acceptance: A study of spanish provinces. *Social sciences*, 11(11):515.

Manuela Sanguinetti, Fabio Poletto, Cristina Bosco, Viviana Patti, and Marco Stranisci. 2018. An italian twitter corpus of hate speech against immigrants. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*.

AF Sellars. 2016. Defining hate speech (research publication no. 2016–20). *Cambridge, MA: Berkman Klein Center*.

Moein Shahiki-Tash, Jesús Armenta-Segura, Zahra Ahani, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2023a. Lidoma at homomex2023@ iberlef: Hate speech detection towards the mexican spanish-speaking lgbt+ population. the importance of preprocessing before using bert-based models. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023)*.

Moein Shahiki-Tash, Jesús Armenta-Segura, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2023b. Lidoma at hope2023iberlef: Hope speech detection using lexical features and convolutional neural networks. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), co-located with the 39th Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), CEUR-WS. org*.

M Shahiki Tash, Z Ahani, Al Tonja, M Gemeda, N Hussain, and O Kolesnikova. 2022. Word level language identification in code-mixed kannada-english texts using traditional machine learning algorithms. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, pages 25–28.

Moein Tash, Jesus Armenta-Segura, Zahra Ahani, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2023. Lidoma@ dravidianlangtech: Convolutional neural networks for studying correlation between lexical features and sentiment polarity in tamil and tulu languages. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 180–185.

Atnafu Lambebo Tonja, Mesay Gemeda Yigezu, Olga Kolesnikova, Moein Shahiki Tash, Grigori Sidorov, and Alexander Gelbuk. 2022. Transformer-based model for word level language identification in code-mixed kannada-english texts. *arXiv preprint arXiv:2211.14459*.

Michael S Waltman and Ashely A Mattheis. 2017. Understanding hate speech. In *Oxford research encyclopedia of communication*.

Mesay Gemeda Yigezu, Girma Yohannis Bade, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2023a. Multilingual hope speech detection using machine learning.

Mesay Gemeda Yigezu, Tadesse Kebede, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2023b. Habesha@ dravidianlangtech: Utilizing deep and transfer learning approaches for sentiment analysis. In *Proceedings of the Third Workshop on Speech and Language Technologies for Dravidian Languages*, pages 239–243.

Mesay Gemeda Yigezu, Olga Kolesnikova, Grigori Sidorov, and Alexander Gelbukh. 2023c. Transformer-based hate speech detection for multi-class and multi-label classification.

Mesay Gemeda Yigezu, Atnafu Lambebo Tonja, Olga Kolesnikova, Moein Shahiki Tash, Grigori Sidorov, and Alexander Gelbukh. 2022. Word level language identification in code-mixed kannada-english texts using deep learning approach. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON): Shared Task on Word Level Language Identification in Code-mixed Kannada-English Texts*, pages 29–33.