byteLLM@LT-EDI-2024: Homophobia/Transphobia Detection in Social Media Comments - Custom Subword Tokenization with Subword2Vec and BiLSTM

Durga Prasad Manukonda
ASRlytics
Hyderabad, India
mdp0999@gmail.com

Rohith Gowtham Kodali
ASRlytics
Hyderabad, India
rohitkodali@gmail.com

Abstract

This research focuses on Homophobia and Transphobia Detection in Dravidian languages, specifically Telugu, Kannada, Tamil, and Malayalam. Leveraging the Homophobia/Transphobia Detection dataset, we propose an innovative approach employing a custom-designed tokenizer with a Bidirectional Long Short-Term Memory (BiLSTM) architecture. Our distinctive contribution lies in a tokenizer that reduces model sizes to below 7MB, improving efficiency and addressing real-time deployment challenges. The BiLSTM implementation demonstrates significant enhancements in hate speech detection accuracy, effectively capturing linguistic nuances. Low-size models efficiently alleviate inference challenges, ensuring swift real-time detection and practical deployment. This work pioneers a framework for hate speech detection, providing insights into model size, inference speed, and real-time deployment challenges in combatting online hate speech within Dravidian languages.

1 Introduction

In light of the growing prevalence of online hate speech, this paper presents the findings of a workshop on detecting LGBTQ+ hate speech in social media comments. We focus on developing and evaluating models that can accurately identify and classify homophobic and transphobic slurs, offensive stereotypes, and other forms of hateful language within YouTube comment sections (Chakravarthi et al., 2024) (Chakravarthi et al., 2023) (Chakravarthi, 2023).

While a comment or post in the dataset may contain more than one sentence, the average sentence length in the corpus is one, and annotations are provided at the comment/post level. In the dynamic landscape of social media, concerns about hate speech targeting the LGBTQ+ community have gained prominence.

This research investigates the challenges of classifying individual social media comments or posts in low-resourced languages. It specifically focuses on Dravidian languages spoken in India—Telugu, Kannada, Tamil, and Malayalam—while acknowledging that the findings may not be universally applicable to other linguistic contexts. The goal is to develop systems that can effectively identify instances of homophobia or transphobia in these languages. Achieving this requires these systems to be adaptable and robust enough to handle the inherent diversity within Dravidian linguistics.

In the context of detecting Homophobia and Transphobia in social media, our research employs a unique approach that utilizes a custom-designed tokenizer and a Bidirectional Long Short-Term Memory (BiLSTM) architecture. A significant contribution of our work lies in the development of a tokenizer designed to streamline model sizes, enhance operational efficiency, and address the challenges associated with real-time deployment. This tokenizer not only minimizes the computational footprint but also optimizes the overall performance of the models. Its unique features empower effective handling of real-time scenarios, providing a versatile solution for deployment challenges.

The implementation of BiLSTM, coupled with our customized tokenizer, showcases significant improvements in the accuracy of hate speech detection, highlighting enhanced sensitivity to linguistic nuances. Compact-sized models effectively address inference challenges, ensuring rapid real-time detection and practical deployment. Our research establishes a pioneering framework for Homophobia and Transphobia Detection, providing insights into model size, inference speed, and the challenges associated with real-time deployment in combating online hate speech.

This paper outlines our methodology, technical advancements, and results, offering a compre-

Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion, pages 157–163
March 21, 2024

hensive examination of Homophobia/Transphobia Detection in social media comments in Dravidian languages. Our research aims to contribute not only to the specific challenges of Homophobia/Transphobia Detection in social media comments but also to a broader understanding of effective detection mechanisms applicable to diverse linguistic landscapes.

2 Related Work

Homophobia and transphobia detection in social media has garnered significant research attention due to its detrimental impact on the LGBTQ+ community. Various approaches have been explored, each addressing specific challenges and contributing to the development of robust detection systems.

Singh and Motlicek (2022) proposed a zero-shot learning framework for detecting homophobic and transphobic comments without labeled data, demonstrating its potential for resource-constrained scenarios.

Kumaresan et al. (2023) addressed the challenge of data scarcity in low-resource languages by presenting a fine-grained dataset and exploring cross-lingual transfer learning techniques. Their work highlights the effectiveness of transferring knowledge from resource-rich languages to improve detection accuracy in diverse linguistic settings.

Ashraf et al. (2022) explored an SVM-based model, achieving notable F1-scores and emphasizing the importance of automatic detection for timely intervention. Their work demonstrates the effectiveness of traditional machine learning algorithms in hate speech detection tasks.

Sharma et al. (2023) investigated deep learning techniques for Dravidian languages, highlighting the superiority of IndicBERT (Kakwani et al., 2020) in addressing low-resource language challenges. This study demonstrates the potential of deep learning models for capturing language-specific features and improving detection accuracy.

(Swaminathan et al., 2022) employed a hybrid approach combining word embeddings, SVM classifiers, and BERT-based transformers (Devlin et al., 2019), achieving promising results. Their work showcases the potential of combining diverse techniques to leverage their strengths and enhance detection performance.

Chakravarthi et al. (2022) investigated the use of pseudolabeling for automated homophobia/transphobia detection, demonstrating significant improvements in model performance. Their work emphasizes the importance of robust evaluation and highlights the potential of pseudolabeling for improving model accuracy.

The study presents ConBERT-RL(Raj et al., 2024), a novel framework using Reinforcement Learning and a concatenated CM-BERT representation, excelling in offensive comment classification for transliterated Tamil in English with a 90% and 93% micro-average accuracy improvement. It effectively captures language-specific features and nuances, demonstrated through t-SNE visualization and graph network comparisons.

The broader literature underscores the global prevalence of offensive language, emphasizing the need for protection and proactive measures to mitigate its impact on vulnerable communities ((Gkotsis et al., 2016); (Oswal, 2021); (Díaz-Torres et al., 2020); (Wang et al., 2019)).

This review highlights significant advancements in homophobic and transphobic comment detection. Despite notable progress, various challenges persist, necessitating further exploration and development. These challenges include the expansion to encompass more low-resource languages, the creation of robust models tailored for code-mixed content, the integration of contextual information, and the exploration of Explainable AI techniques. Addressing these challenges will contribute to a more comprehensive and effective approach to combating online hate speech.

3 Dataset

3.1 Embedding Datasets

Our research draws upon a substantial corpus sourced from the AI4Bharath[1] datasets for Telugu, Tamil, Malayalam, and Kannada. Specifically, we harnessed the initial 5,000,000 lines from the Telugu corpus (1.3GB), 9,492,782 lines from the Tamil corpus (980MB), 11,512,628 lines from the Malayalam corpus (1.2GB), and 15,000,000 lines from the Kannada corpus (1.5GB). These datasets serve as a rich and diverse source of linguistic content, covering an array of topics relevant to our research. This linguistic variety is instrumental in fostering the development of embeddings that are not only robust but also generalizable, crucial for the success of our research endeavors.

[1] https://github.com/AI4Bharat/indicnlp_corpus

3.2 Homophobia/Transphobia Datasets

Our research, undertaken as part of LT-EDI@EACL 2024 [2], is focused on Dravidian languages—Tamil, Telugu, Kannada, and Malayalam—selected for their shared linguistic roots and the intricate process of developing high-quality embeddings for each. This cohesive group forms the basis for our study, aiming to understand and address online hate speech in these languages.

The tasks involved in our study include identifying discriminatory comments based on sexual orientation or gender identity, utilizing datasets covering Dravidian languages to ensure comprehensive representation. Additionally, the focus on categorizing YouTube comments aids in recognizing instances of homophobia and transphobia, facilitating a deeper analysis of their manifestation in online discourse (Chakravarthi et al., 2022) (Kumaresan et al., 2023).

Language	Train	Dev	Test
Telugu	9,050	1,940	1,939
Malayalam	3,114	1,213	866
Kannada	10,066	2,157	2,156
Tamil	2,662	666	833

Table 1: Homophobia/Transphobia Detection Dataset Statistics

Table 1 details the dataset distribution across languages, providing training, development, and test sets for system development and evaluation.

4 Methodology

This section unveils the details of our innovative architecture, integrating two crucial components: a dynamic Subword Embeddings module and a robust BiLSTM Classification module. We explore data preprocessing, subword tokenization, embedding training, and orchestration of our advanced classifier.

4.1 Preprocessing and Tokenization

This section delineates the procedures for data preprocessing and tokenization applied in the Shared Task on Homophobia/Transphobia Detection in social media comments.

[2] https://codalab.lisn.upsaclay.fr/competitions/16056

4.1.1 Preprocessing Pipeline

Our comprehensive preprocessing involved normalization, cleaning (removing noise like URLs and hashtags), and transliteration using the indic_transliteration library[3] for uniform processing.

4.1.2 Subword Tokenization

Post-preprocessing, we implemented a custom subword tokenizer, "VowelToken," for each language. This approach aimed to enhance granularity, capturing morphemic and grammatical information crucial for detecting linguistic nuances related to homophobia/transphobia. Leveraging subword tokens enables the embedding model to learn more precise and informative representations, potentially improving detection performance.

The proposed VowelToken subword tokenizer exhibits universality, utilizing linguistic principles based on vowel boundaries for accurate segmentation across diverse languages, including Dravidian languages. Its rule-based design focuses on identifying and segmenting words based on consistent vowel boundary patterns, enhancing precision and reliability in the tokenization process. Refer to Table 2 for the preprocessing and tokenization statistics of each language corpus.

4.2 Subword Embeddings Module

The Subword2Vec module obtains subword embeddings using the Word2Vec method by Mikolov et al. (2013). The module's initialization involves specifying critical parameters: vocabulary size (V), minimum frequency (f_{min}), and embedding dimension ($d_{subword}$). Subword counts are collected to construct a subword vocabulary (S), and embeddings are trained using Stochastic Gradient Descent (SGD).

The module's initialization involves specifying critical parameters, starting with the vocabulary size (V) that sets the upper limit for subword consideration. Additionally, the minimum frequency parameter (f_{min}) serves as the threshold for subword inclusion based on frequency. The embedding dimension ($d_{subword}$), characterizing the dimensionality of subword embeddings, is also defined. These parameters collectively configure the module during the initialization process, a pivotal aspect of our research.

Subword counts are collected from the corpus to construct a subword vocabulary (S). The sub-

[3] https://github.com/indic-transliteration/indic_transliteration_py

Language	Total Words	Total Subtokens	Subtokens (Count >= 2)	Emb. Size(MB)
Telugu	179,732,317	22,596	13,405	6.6
Tamil	174,349,374	15,065	9,406	4.5
Kannada	399,312,707	17,889	12,173	5.8
Malayalam	117,054,028	19,155	14,190	5.92

Table 2: Preprocessing and Tokenization Statistics with Embedding(Emb.) Sizes of 100-dimensional Model

word splitting process is executed based on vowels, excluding subwords with counts below f_{min}. This process is mathematically expressed as:

S = {s | s is a $subword$, $count(s) \geq f_{min}$, $|S| \leq V$}

$$S = \{s \in \mathcal{W} \mid count(s) \geq f_{min}, |S| \leq V\} \quad (1)$$

The subword splitting process involves dividing the input word into subwords based on vowel boundaries. Consonant prefixes and suffixes are included in the subwords when applicable, and special tokens "_" (start of subword) are added to the first letter. Subword embeddings ($E_{subword}$) are initialized as a random matrix with dimensions ($|S|$, $d_{subword}$).

The training phase employs Stochastic Gradient Descent (SGD) (Tian et al., 2023) to train subword embeddings. The objective is to minimize the Mean Squared Error (MSE) loss (L) between subword pairs. The SGD update is expressed as:

$$E_{subword}^{(t+1)} = E_{subword}^{(t)} - \eta \nabla L(E_{subword}^{(t)}) \quad (2)$$

Here, t represents the training iteration, η is the learning rate, and ∇L is the gradient of the loss function. Training subword embeddings is a crucial step in refining the model's representation of subword relationships.

4.3 BiLSTM Classifier

The BiLSTM architecture, inspired by Ghosh et al. (2020), plays a crucial role in the fake news classification task. It consists of two essential components: a subword embedding layer and a bi-directional LSTM layer.

4.3.1 Sub-Word Embedding Layer

The Sub-Word Embedding Layer operates on an input word sequence $x = [w_1, w_2, ..., w_n]$ utilizing a subword embedding function. Each word w_i is mapped to its corresponding subword embeddings, denoted as $w_{i1}, w_{i2}, ..., w_{in}$, where n represents the number of subwords for the i-th word. The final word embedding for w_i, denoted as $\mathbf{e}_i$, is obtained by summing the embeddings of its constituent subwords:

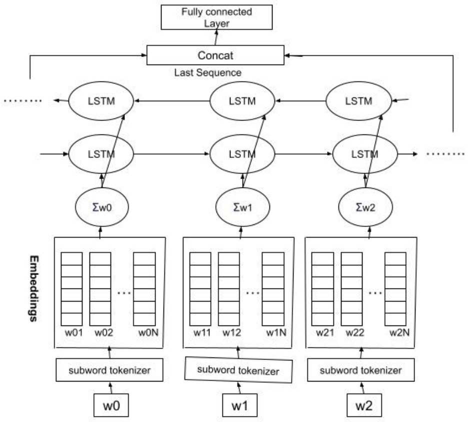

Figure 1: The unfolded architecture of BiLSTM classifier with three 3 word example sample.

$$\mathbf{e}_i = w_{i1} + w_{i2} + \ldots + w_{in} \quad (3)$$

The output of this layer is a tensor X_{embed} of dimensions $1 \times n \times d_{embed}$, where d_{embed} signifies the size of each word embedding.

$$X_{embed} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n] \quad (4)$$

Here, $\mathbf{e}_i$ represents the word embedding for the i-th word in the sequence, and n is the length of the sequence.

4.3.2 Bi-directional LSTM Layer

The Bi-directional LSTM Layer engages with the embedded sequence X_{embed} to adeptly capture contextual information. Configured with an input size of d_{embed} (matching the embedding size) and a hidden size of d_{hidden}, the bidirectional LSTM ensures the seamless flow of information both in forward and backward directions. The resulting output, denoted as $blstm\_out$, takes the form of a tensor with dimensions $1 \times n \times (2 \times d_{hidden})$, as it concatenates the hidden states from both directions.

$$blstm\_out = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_n] \quad (5)$$

In essence, the BiLSTM layer processes the input sequence and produces hidden states $\mathbf{h}_i$ for each word in the sequence.

The forward pass of the model is mathematically expressed as follows:

$$\mathbf{h}_i^{(f)}, \mathbf{h}_i^{(b)} = BiLSTM(\mathbf{e}_{1:i}, \mathbf{e}_{i:n}), \quad \forall i \in \{1, ..., n\} \quad (6)$$

Here, $\mathbf{h}_i^{(f)}$ and $\mathbf{h}_i^{(b)}$ symbolize the forward and backward hidden states at position i, respectively. The BiLSTM function operates on subword embeddings $\mathbf{e}_{1:i}$ and $\mathbf{e}_{i:n}$ for each i in the sequence.

4.3.3 Classifier Output

The final prediction, denoted as y, is derived by applying a linear transformation to the last hidden state in the forward direction ($\mathbf{h}_n^{(f)}$) using weights matrix W and bias b.

$$y = W\mathbf{h}_n^{(f)} + b \quad (7)$$

This linear transformation allows the model to make predictions based on the learned representations from the BiLSTM layer.

Figure 1 illustrates the unfolded architecture of the BiLSTM Classifier, providing a visual representation of the sequence processing and contextual information capture. This design adeptly integrates subword embeddings with a BiLSTMbased approach, showcasing adaptability and potential across various natural language processing applications.

5 Experimental Setup

Our experimental setup is designed to demonstrate the effectiveness of our proposed approach in the context of the Shared Task on Homophobia/Transphobia Detection in social media comments. We conducted experiments using a 100-dimensional embedding model tailored to each language. The embedding sizes were determined based on linguistic characteristics and dataset scale, as outlined in Table 2. Specifically, for Telugu, Tamil, Kannada, and Malayalam, the embedding sizes were 6.6 MB, 4.5 MB, 5.8 MB, and 5.92 MB, respectively.

Subword tokenization was facilitated by the custom VowelToken subword tokenizer, designed for universality and based on linguistic principles using vowel boundaries. This tokenizer ensured accurate segmentation across diverse languages, including Dravidian languages. Its rule-based design focused on identifying and segmenting words based on consistent vowel boundary patterns, enhancing precision and reliability in the tokenization process.

To evaluate the impact of these subword embeddings, we seamlessly integrated them into our BiLSTM-based model architecture. The ClassificationModel includes a Sub-Word Embedding Layer, Bi-directional LSTM Layer, and Linear Classification Layer, utilizing subword embeddings from VowelToken. The BiLSTM layer features an input size of 100 and a hidden size of 128. Additionally, the model utilizes the Adam optimizer with a learning rate of 0.001 during training.

Datasets were partitioned into training, development, and test sets based on the distribution outlined in Table 1. Model training utilized the Adam optimizer with a learning rate of 0.001 and a batch size of 64. Early stopping was implemented, with a patience setting of 10 epochs based on development set performance. Evaluation metrics, including recall, precision, F1 score, and accuracy, were used to measure the model's effectiveness.

6 Experimental Results and Discussions

The effectiveness of our subword tokenization is evident in the remarkably low perplexity (less than 1.2) achieved after just one training epoch for embeddings. Despite the constraints of limited training time and data, this result underscores the efficacy of our subword tokenization approach.

The table 3 reveals the macro average F1-Scores (M_F1-scores) for the Homophobia/Transphobia Detection Task on the test sets across various languages. In the rankings, the team **"byteLLM"** (originally **byteSizedLLM**) achieved noteworthy positions, with Telugu securing the 3rd rank and achieving the highest score of 0.959, closely trailing the top score of 0.971. Malayalam also performed well, obtaining a commendable M_F1-Score of 0.891 and ranking 3rd, with a top score of 0.942. Similarly, Tamil secured the 3rd rank with a score of 0.801. However, Kannada, while contributing valuable insights, demonstrated a slightly lower score of 0.922 and secured the 6th rank. These rankings provide a comprehensive view of the model's performance across different Dravidian languages.

To enhance the performance further, it is crucial to leverage more extensive training data, especially focusing on diverse datasets. Given the

Language	M_F1-Score	Rank	Top Score
Telugu	0.959	3	0.971
Malayalam	0.891	3	0.942
Kannada	0.922	6	0.948
Tamil	0.801	3	0.880

Table 3: Homophobia/Transphobia Detection Task Macro average F1-Scores (M_F1-scores) of Test Sets

multilingual nature of the task, training on multilingual data can significantly improve performance. While we trained on 4.3GB of L3Cube-HingCorpus data, the model produced 6.4MB embeddings that outperformed large language models (LLMs) in Language Identification (LID) and Named Entity Recognition (NER) on GLUCoS benchmarks. The model, when trained on larger text (more than 10GB), is expected to achieve state-of-the-art (SOTA) performance. However, due to hardware limitations, we were unable to load larger text for training in this study. In future tasks, we plan to implement and test this approach with larger text and languages.

7 Conclusion and Future Work

Our research convincingly demonstrates the effectiveness of subword tokenization for homophobia/transphobia detection across Dravidian languages. The competitive results achieved with lightweight models highlight the scalability and computational efficiency of our approach. Subword embeddings, trained with meticulous preprocessing and tokenization, showcase impressive performance, with Dravidian languages securing leading Macro F1-Scores. This underscores the potential of subword tokenization in tackling online hate speech with resource-efficient models.

Moving forward, expanding the dataset with diverse multilingual content is crucial for further enhancing accuracy. The technical advantage of training on larger texts for achieving state-of-the-art performance is evident, albeit currently limited by hardware constraints. Nevertheless, the lightweight nature of our models, their fast inference speed, and minimal storage requirements render them practical for various tasks beyond homophobia/transphobia detection, including Named Entity Recognition (NER), Language Identification (LID), Sentiment Classification, and Multiclass classification. We plan to explore their applicability in generative AI for future research, potentially opening doors to even more impactful applications.

References

Nsrin Ashraf, Mohamed Taha, Ahmed Abd Elfattah, and Hamada Nayel. 2022. NAYEL @LT-EDI-ACL2022: Homophobia/transphobia detection for equality, diversity, and inclusion using SVM. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 287–290, Dublin, Ireland. Association for Computational Linguistics.

Bharathi Raja Chakravarthi. 2023. Detection of homophobia and transphobia in YouTube comments. *International Journal of Data Science and Analytics*, pages 1–20.

Bharathi Raja Chakravarthi, Adeep Hande, Rahul Ponnusamy, Prasanna Kumar Kumaresan, and Ruba Priyadharshini. 2022. How can we detect Homophobia and Transphobia? Experiments in a multilingual code-mixed setting for social media governance. *International Journal of Information Management Data Insights*, 2(2):100119.

Bharathi Raja Chakravarthi, Prasanna Kumar Kumaresan, Ruba Priyadharshini, Paul Buitelaar, Asha Hegde, Hosahalli Lakshmaiah Shashirekha, Saranya Rajiakodi, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, José Antonio García-Díaz, Rafael Valencia-García, Kishore Kumar Ponnusamy, Poorvi Shetty, and Daniel García-Baena. 2024. Overview of Third Shared Task on Homophobia and Transphobia Detection in Social Media Comments. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity and Inclusion*, Malta. European Chapter of the Association for Computational Linguistics.

Bharathi Raja Chakravarthi, Rahul Ponnusamy, Malliga Subramanian, Paul Buitelaar, Miguel Ángel García-Cumbreras, Salud María Jiménez-Zafra, José Antonio García-Díaz, Rafael Valencia-García, and Nitesh Jindal. 2023. Overview of Second Shared Task on Homophobia and Transphobia Detection in English, Spanish, Hindi, Tamil, and Malayalam. In *Proceedings of the Third Workshop on Language Technology for Equality, Diversity and Inclusion*, Varna, Bulgaria. Recent Advances in Natural Language Processing.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

María José Díaz-Torres, Paulina Alejandra Morán-Méndez, Luis Villasenor-Pineda, Manuel Montes-y Gómez, Juan Aguilera, and Luis Meneses-Lerín. 2020. Automatic detection of offensive language in social media: Defining linguistic criteria to build a Mexican Spanish dataset. In *Proceedings of the Second Workshop on Trolling, Aggression and Cyberbullying*, pages 132–136, Marseille, France. European Language Resources Association (ELRA).

Koyel Ghosh, Dr. Apurbalal Senapati, and Dr. Ranjan Maity. 2020. Technical domain identification using word2vec and BiLSTM. In *Proceedings of the 17th International Conference on Natural Language Processing (ICON): TechDOfication 2020 Shared Task*, pages 21–26, Patna, India. NLP Association of India (NLPAI).

George Gkotsis, Anika Oellrich, Tim Hubbard, Richard Dobson, Maria Liakata, Sumithra Velupillai, and Rina Dutta. 2016. The language of mental health problems in social media. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 63–73, San Diego, CA, USA. Association for Computational Linguistics.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In *Findings of EMNLP*.

Prasanna Kumar Kumaresan, Rahul Ponnusamy, Ruba Priyadharshini, Paul Buitelaar, and Bharathi Raja Chakravarthi. 2023. Homophobia and transphobia detection for low-resourced languages in social media comments. *Natural Language Processing Journal*, page 100041.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space.

Nikhil Oswal. 2021. Identifying and categorizing offensive language in social media.

Vivek Suresh Raj, Chinnaudayar Navaneethakrishnan Subalalitha, Lavanya Sambath, Frank Glavin, and Bharathi Raja Chakravarthi. 2024. ConBERT-RL: A policy-driven deep reinforcement learning based approach for detecting homophobia and transphobia in low-resource languages. *Natural Language Processing Journal*, 6:100040.

Deepawali Sharma, Vedika Gupta, and Vivek Kumar Singh. 2023. *Detection of Homophobia amp; Transphobia in Malayalam and Tamil: Exploring Deep Learning Methods*, page 217–226. Springer Nature Switzerland.

Muskaan Singh and Petr Motlicek. 2022. IDIAP submission@LT-EDI-ACL2022: Homophobia/transphobia detection in social media comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 356–361, Dublin, Ireland. Association for Computational Linguistics.

Krithika Swaminathan, Bharathi B, Gayathri G L, and Hrishik Sampath. 2022. SSNCSE_NLP@LT-EDI-ACL2022: Homophobia/transphobia detection in multiple languages using SVM classifiers and BERT-based transformers. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 239–244, Dublin, Ireland. Association for Computational Linguistics.

Yingjie Tian, Yuqi Zhang, and Haibin Zhang. 2023. Recent advances in stochastic gradient descent in deep learning. *Mathematics*, 11(3).

Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. 2019. Systematic literature review on the spread of health-related misinformation on social media. *Social Science Medicine*, 240:112552.