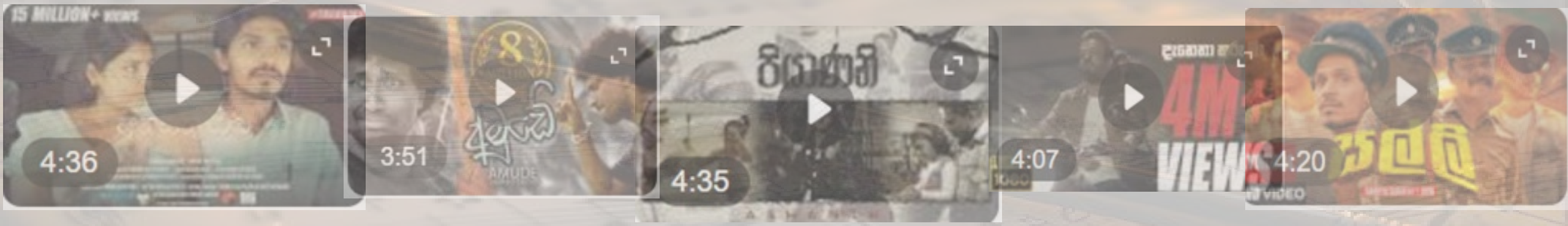
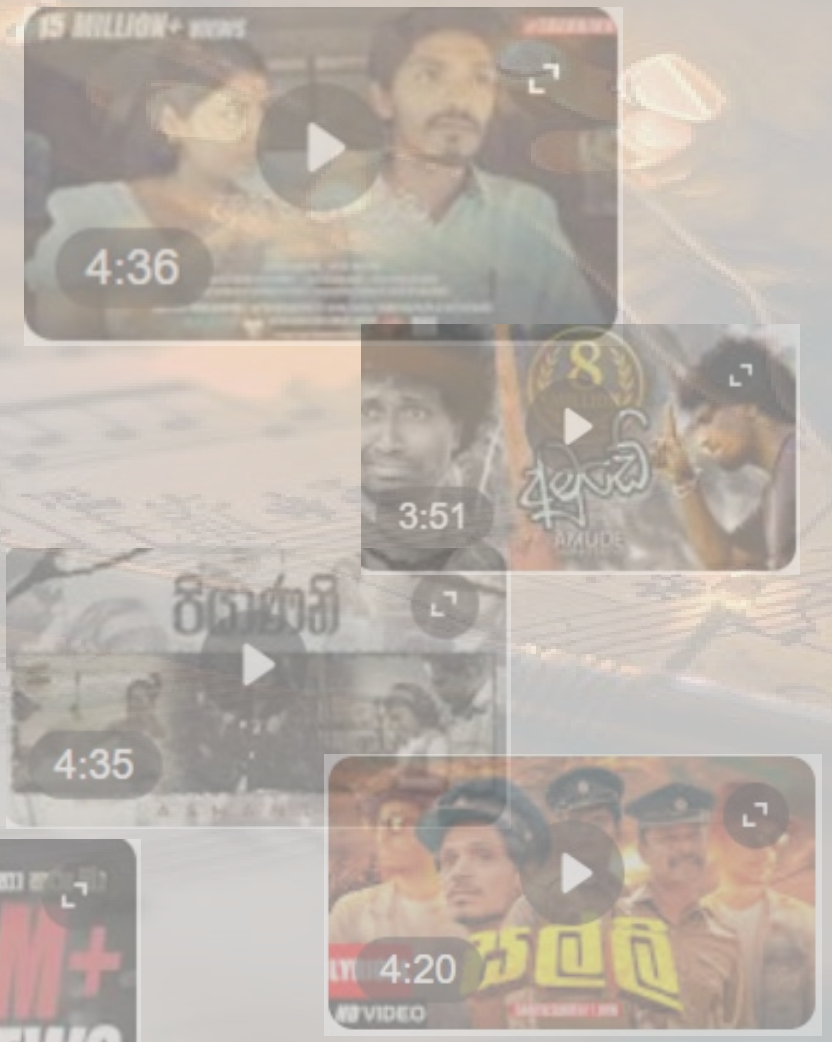
Valence–Arousal Modelling for Affective Analysis of Social Media Discourse on Sinhala Music Videos



W. M. Yomal De Mel Department of Computer Science & Engineering,
University of Moratuwa, Moratuwa,
Sri Lanka
Research Supervisor - Dr. **Nisansa de Silva**

Table of contents



Publications

01 Linguistic Analysis of Sinhala YouTube Comments on Sinhala Music Videos: A Dataset Study

24th ICTer International Conference

02 Sinhala Transliteration: A Comparative Analysis Between Rule-based and Seq2Seq Approaches

First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages (ACL)

03 GeeSanBhava: Sentiment Tagged Sinhala Music Video Comment Data Set

Communications in Computer and Information Science Advances in Computational Collective Intelligence (ICCCI 2025)

04 Team VYN at SemEval-2026 Task 3: Dimensional Aspect-Based Sentiment Analysis

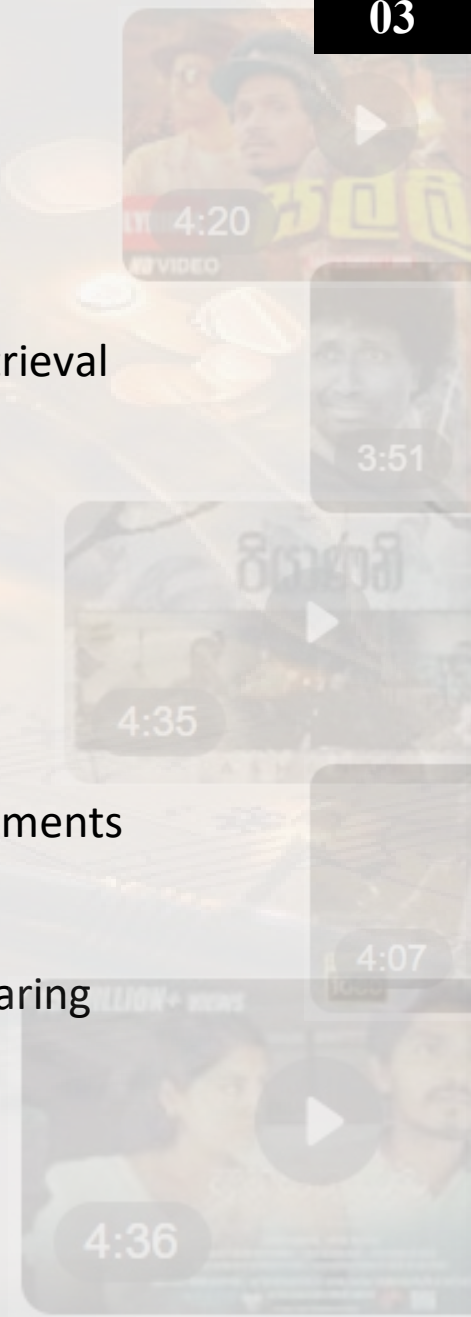
The 20th International Workshop on Semantic Evaluation (SemEval-2026)

05 Continuous Valence–Arousal Modeling of Sinhala YouTube Comments with Hybrid Multi-Task Learning

In Submission

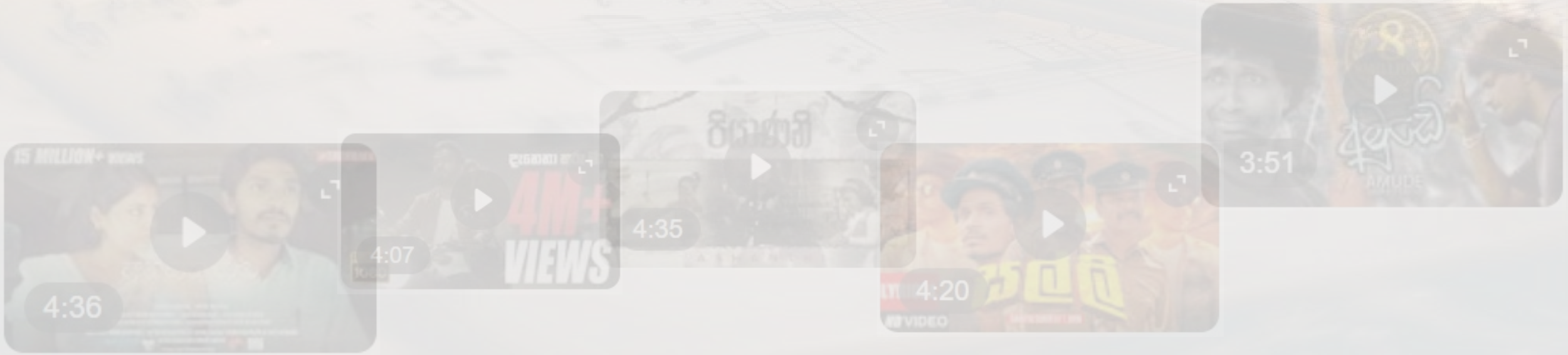
01. Introduction

- This research examine emotional dimensions in Sinhala music through Music Information Retrieval (MIR) and Music Emotion Recognition (MER).
- Limited exploration of Sri Lankan music within MIR; challenges due to scarcity of Sinhala NLP resources.
- Sinhala songs represent a rich cultural heritage and diverse musical genres in Sri Lanka.
- The objective of this research is to develop a comprehensive dataset of Sinhala YouTube comments to analyze linguistic patterns and emotional expressions.
- Integrate computational linguistics and NLP to identify frequent words and word pairs, comparing with general Sinhala corpora.



01. Introduction Cont.

- Utilization of Sinhala Wikipedia, newspaper articles, and previous NLP research to complement YouTube data.
- The research highlights the lexical diversity and unique linguistic traits in Sinhala music comments, enhancing understanding of listener emotions.
- This contribute to provide insights into how social media comments can serve as authentic sources for emotional analysis in underrepresented languages and music traditions.



02. Research Problem

Analyzing emotions in social media comments on songs, particularly in languages such as Sinhala, faces challenges due to limited NLP resources, with existing research mainly focusing on English, leaving a significant gap in understanding emotional responses to Sinhala music.

The research question explored in this thesis is:

"How to utilize emotion recognition on the comments for Sinhala songs on social media."

02. Related Work

2.1 Development of Sinhala Corpora

- This research examine emotional dimensions in Sinhala music through Music Information Retrieval (MIR) and Music Emotion Recognition (MER) [1].
- Early contributions by Upeksha et al. [2] and De Silva et al. [3] focused on compiling data from online sources such as news, academic content, and gazettes.
- A significant advancement came from Wijeratne et al. [4], who released a large dataset of Sinhala Facebook posts, making it publicly available.
- More recently, Hettiarachchi et al. [5] introduced a corpus of 506,932 news articles, significantly expanding the available data for Sinhala NLP.

[1] N. de Silva, “Survey on publicly available sinhala natural language processing tools and research,” arXivpreprint arXiv:1906.02358, 2019.

[2] D. Upeksha, C. Wijayarathna, M. Siriwardena, L. Lasandun, C. Wimalasuriya, N. De Silva, and G. Dias, “Implementing a corpus for sinhala language,” in Symposium on Language Technology for South Asia, vol. 2015, 2015, p. 3.

[3] N. de Silva, “Sinhala text classification: observations from the perspective of a resource poor language,” ResearchGate, 2015.

[4] Y. Wijeratne and N. de Silva, “Sinhala language corpora and stopwords from a decade of sri lankan facebook,” arXiv preprint arXiv:2007.07884, 2020.

[5] H. Hettiarachchi, D. Premasiri, L. Uyangodage, and T. Ranasinghe, “Nsina: A news corpus for sinhala,” arXiv preprint arXiv:2403.16571, 2024.

02. Related Work (Cnt.)

2.2 Music Information Retrieval (MIR)

- MIR strategies enable efficient access to music collections and benefit various groups, including industry professionals, researchers, and end-users [6].
- Techniques such as pitch histograms, Spiral Arrays, and graph structures are used to analyze and represent music data [7].

2.3 Music Emotion Recognition (MER)

- MER focuses on analyzing the emotional content in music using computational methods, with applications in personalized music recommendations and mood-based playlists [8,9].
- Despite challenges like subjective emotional perception, the field is growing and has potential implications for psychology and cognitive science [10,11].

[6] M. A. Casey, R. Veltkamp, M. Goto, M. Leman,C. Rhodes, and M. Slaney, “Content-based music information retrieval: Current directions and future challenges,” Proceedings of the IEEE, vol. 96, no. 4,pp. 668–696, 2008.

[7] F. Simonetta, S. Ntalampiras, and F. Avanzini, “Multimodal music information processing and retrieval:Survey and future challenges,” in 2019 international workshop on multilayer music representation and processing (MMRP). IEEE, 2019, pp. 10–18.

[8] S. Hizlisoy, S. Yildirim, and Z. Tufekci, “Music emotion recognition using convolutional long short termmemory deep neural networks,” Engineering Science and Technology, an International Journal, vol. 24,no. 3, pp. 760–767, 2021.

[9] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton,P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, “Music emotion recognition: A state of the art review,” in Proc. ismir, vol. 86, 2010, pp. 937–952.

[10] Y.-H. Yang and H. H. Chen, “Machine recognition of music emotion: A review,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 3,no. 3, pp. 1–30, 2012.

[11] S. Koelsch, “Investigating emotion with music: neuroscientific approaches,” Annals of the New York Academy of Sciences, vol. 1060, no. 1, pp. 412–418,2005.

02. Related Work (Cnt.)

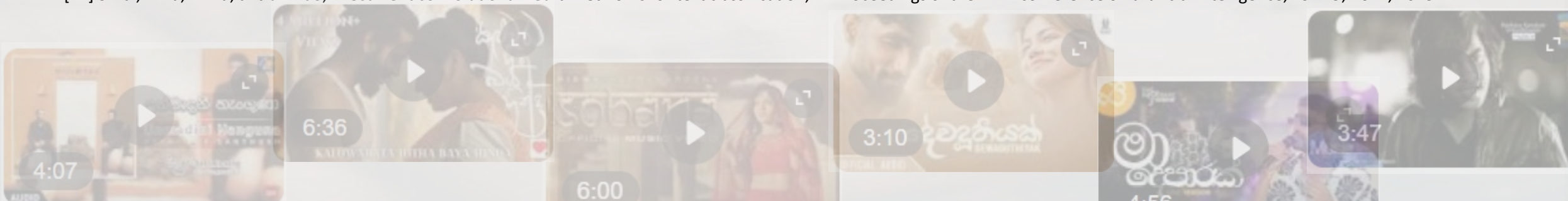
2.4 Word Embedding

- Word embeddings convert words into dense vectors that capture semantic and syntactic meanings, derived from large, unlabeled corpora. This helps in understanding word relationships effectively [12].
- Word2Vec, developed by Mikolov et al. [13], introduced two models: Continuous Bag of Words (CBOW) and Skip-Gram. These models predict word relationships by learning from the context, which made them influential in the NLP field.
- Word embeddings have significantly advanced modern NLP tasks, including sentiment analysis, machine translation, and information retrieval by enabling systems to understand and process text more accurately [14].

[12] L.-C. Yu, J. Wang, K. R. Lai, and X. Zhang, “Refining word embeddings using intensity scores for sentiment analysis,” IEEE/ACM Transactions on Audio,Speech, and Language Processing, vol. 26, no. 3, pp.671–681, 2017.

[13] T. Mikolov, K. Chen, G. Corrado, and J. Dean,“Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

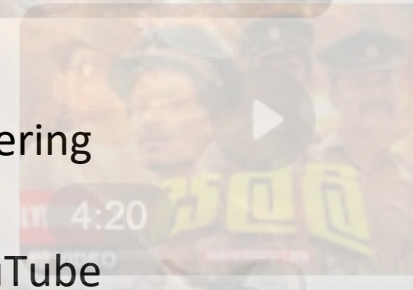
[14] S. Lai, L. Xu, K. Liu, and J. Zhao, “Recurrent convolutional neural networks for text classification,”in Proceedings of the AAAI conference on artificial intelligence, vol. 29, no. 1, 2015.

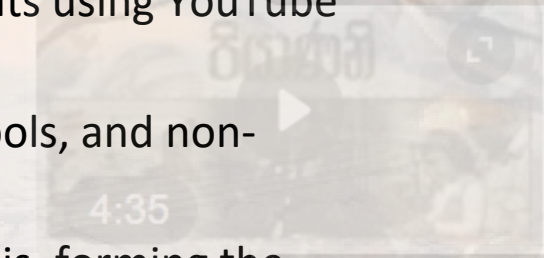


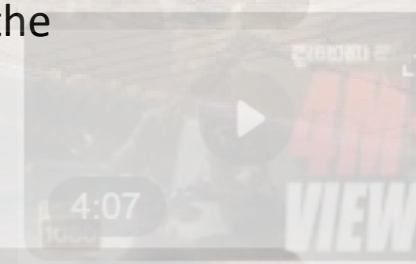
03. Data Preparation

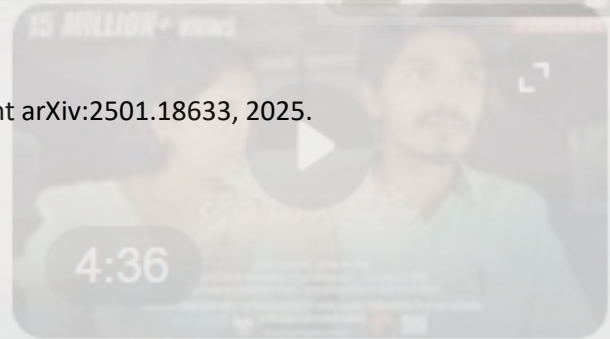
3.1 Data Collection and Initial Processing

- Data Source: Comments were collected from 27 YouTube videos featuring 20 Sinhala songs, covering various eras of Sinhala music.
- Extraction Method: Google App Script was used to extract a total of 93,116 comments using YouTube video IDs.
- Initial Data Cleaning: Preprocessing involved removing emojis, URLs, numbers, symbols, and non-Sinhala content to ensure data quality.
- Filtered Dataset: After cleaning, 75,656 comments were identified for further analysis, forming the basis for linguistic evaluation.









[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

03. Data Preparation (Cnt.)

3.2 Linguistic Segmentation and Transliteration

- Language Segregation: The dataset was compared against the NLTK English corpus to separate purely English comments.
- Focus on Sinhala Content: 35,428 comments were identified as containing Sinhala characters and incorporated into the final dataset.
- Transliteration Process: Google Transliterator API and Ruled based Transliteration was used to convert 30,633 English-scripted comments to Sinhala script, successfully transliterating 28,043.

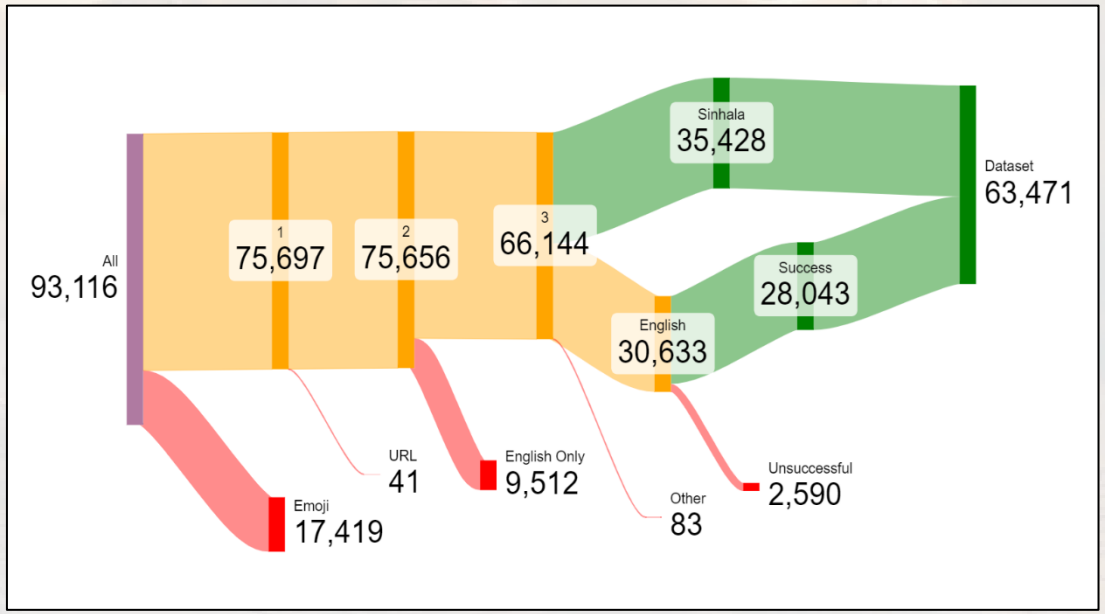


Fig. 1: Stages of the Data Preparation Process

[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

03. Data Preparation (Cnt.)

3.2 Linguistic Segmentation and Transliteration (Cont.)

Transliteration Process:

- Ruled based transliteration and Google Transliterator API used to convert 30,633 English-scripted comments to Sinhala script, successfully transliterating 28,043.
- Final Dataset: Post-transliteration, the dataset consisted of 63,471 Sinhala comments, representing 68% of the initial data.

Latin Sequence	Sinhala Character	Latin Sequence	Sinhala Character
a	අ	aa	ආ
A	ඇ	Aa	ඈ
i	ඉ	ie	ඊ
u	උ	uu	ඌ
e	එ	ea	ඒ
I	ඓ	o	ඔ
ka	ක	ga	ග
ma	ම	ya	ය
ra	ර	ba	බ
ca	ච	ja	ජ
ta	ට	la	ල
Da	ඩ	wa	ව
tha	ත	sa	ස
da	ද	ha	හ
na	න	pa	ප
Na	ණ	La	ළ
mI	ඹ	thI	ඵ
Ka	ඛ	Ga	ඝ
cha	ඡ	Tha	ඨ
Dha	ඪ	dha	ධ
Pa	ඵ	bha	භ
fa	ෆ	Ba	ඹ
GNa	ඦ	KNa	ඤ
jha	ඣ	Lu	ළු
Luu	ළූ	Sa	ෂ
sha	ශ	GNa	ඦ
ki	කි	ku	කු
ke	කෙ	ko	කො
kaa	කා	kAa	කෑ
kIe	කී	kei	කේ
gi	ගි	gu	ගු
ge	ගෙ	go	ගො
gaa	කා	gAa	ගෑ
gIe	ගී	gei	ගේ
goe	ගෝ	guu	ගූ
gau	ගෞ	\n	ං

Tab. 1: Transliteration rules.

[16] Y. De Mel, K. Wickramasinghe, N. de Silva, and S. Ranathunga, “Sinhala transliteration: A comparative analysis between rule-based and Seq2Seq approaches,” in Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages, pp. 166–173, 2025.

03. Data Preparation (Cnt.)

3.3 Data Composition and Methodological Significance

- Dataset Composition: The dataset comprises 63,471 Sinhala comments, with varying engagement across songs.
- Notable Insights: High engagement was observed for songs like "අන් සතු ඔබ" (14,100 comments), while others like "පියාණනී" had fewer interactions (130 comments).
- Importance of Rigorous Processing: Systematic extraction and thorough data cleaning were crucial in maintaining dataset integrity.
- Multilingual Data Handling: The use of the Google Transliterator API enhanced the dataset's linguistic diversity, vital for comprehensive analysis.

Artist	Song	Comment Count
Thisara Weerasinghe	අන් සතු ඔබ	14100
Sarith and Surith	සල්ලි සල්ලි	12254
Dinesh Gamage	දැනෙනා තුරු මා	5378
Methun SK	කාරි නෑ සඳ	3935
Saman Lenin	අමුඬේ	3757
BnS	උන්මාද ප්‍රේම ගීය	3292
Danith Sri	එහෙම දේවල් නෑ හිතේ	2825
Sanuka Wick	පෙරවදනක්	2634
Raini Charuka	කළුවරට හිත බය	2596
Ridma	සොබනා	2058
Prageeth Perera	කෝමලියා	1871
Abisheka and Mihdu	දන්නවාද ආදරේ නීතිය	1722
Sanuka Wick	සරාගයේ	1506
Sandeep Jayalath	නුරාවී	1372
Sanuka	මෝහිනී	1319
Saman Lenin	අම්බරුවෝ	809
Victor Rathnayake	තනිවෙන්නට මගේ ලොවේ	791
Nanda Malini	මා සඳට කැමති බව	735
Victor Rathnayake	අපෙ හැඟුම් වලට	387
Ashanthi	පියාණනී	130

Tab. 2: Comment Counts by Artist and Song

[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

03. Data Preparation (Cnt.)

3.4 Data Annotation

- Developed GeeSanBhava, a large-scale human-annotated Sinhala music emotion dataset.
- Utilized Russell's Valence-Arousal model for fine-grained emotion representation.
- Achieved 84.96% inter-annotator agreement, demonstrating high annotation reliability.
- Weighted average Comment based emotion vector and song based emotion vector was compared.
- Comment based emotion vector was standardized to eliminate biases.

$$c_{z,i} = \frac{c_i - \frac{\sum_{j=0}^{n} c_j}{n}}{\max_{k=0 \to n} \left(c_k - \frac{\sum_{j=0}^{n} c_j}{n} \right)}$$

[17] Y. De Mel and N. de Silva, “GeeSanBhava: Sentiment tagged Sinhala music video comment data set,” in International Conference on Computational Collective Intelligence, pp. 157–171, Springer, 2025.

03. Data Preparation (Cnt.)

3.5 Evaluation of Comment Classification Using Existing Model

- Initial objective was to finetune existing pretrained model.Closest reference was conducted by Ranathunga and Liyanage on Sinhala News paper comments[17].
- This research only classify the comments into Negative and Positive sub classes.
- The pretrained models were not available thus the model was trained by the procedure given by the researchers
- Four pre-trained embedding models were used: word2vec [18], fastText [19], SinBERT [20],and sentence embeddings (SBERT) [21]
- Multiple traditional machine learning classifiers were employed to train and evaluate the models - Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost.
- The performance of these classifiers was tested using various combinations of embedding techniques

[17] Y. De Mel and N. de Silva, “GeeSanBhava: Sentiment tagged Sinhala music video comment data set,” in International Conference on Computational Collective Intelligence, pp. 157–171, Springer, 2025.
[18] S. Ranathunga and I. U. Liyanage, “Sentiment analysis of Sinhala news comments,” Transactions on Asian and Low-Resource Language Information Processing, vol. 20, no. 4, pp. 1–23, 2021.
[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.
[20] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.
[21] V. Dhananjaya, P. Demotte, S. Ranathunga, and S. Jayasena, “Bertifying Sinhala: A comprehensive analysis of pre-trained language models for Sinhala text classification,” in Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 7377–7385, 2022.
[22] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3982–3992, 2019.

03. Data Preparation (Cnt.)

3.6 Continuous Valence–Arousal

- Reformulated emotion recognition from **discrete classification** to **continuous Valence–Arousal (V–A) regression** for finer-grained affect prediction.
- Initial classification experiments showed **low F1 scores**, motivating the transition to continuous emotion modeling.
- Validated the proposed **multi-task framework** on an **English Facebook V–A benchmark** before applying it to Sinhala.
- Adopted **multi-task learning**, jointly optimizing **V–A regression** and **auxiliary V–A classification** to improve robustness and generalization.

03. Data Preparation (Cnt.)

3.6 Continuous Valence–Arousal (Cont.)

- Extended the methodology to **Sinhala social media comments**, providing one of the first large-scale studies on **continuous emotion modeling**.
- Applied **joint Valence–Arousal oversampling** to address class imbalance while preserving untouched validation and test sets.
- Conducted a comprehensive **embedding ablation study** using **FastText, SinBERT, and SinLLaMA** representations.
- Proposed a **gated fusion architecture** integrating **SinBERT-large, FastText, SinBERT, and SinLLaMA**, achieving the best continuous emotion prediction performance.

03. Data Preparation (Cnt.)

3.6 Continuous Valence–Arousal (Cont.)

The final architecture integrates SinBERT-large contextual representations with FastText, sentence-level SinBERT, and SinLLaMA embeddings through a gated fusion mechanism, achieving the strongest predictive performance for continuous emotion modeling.

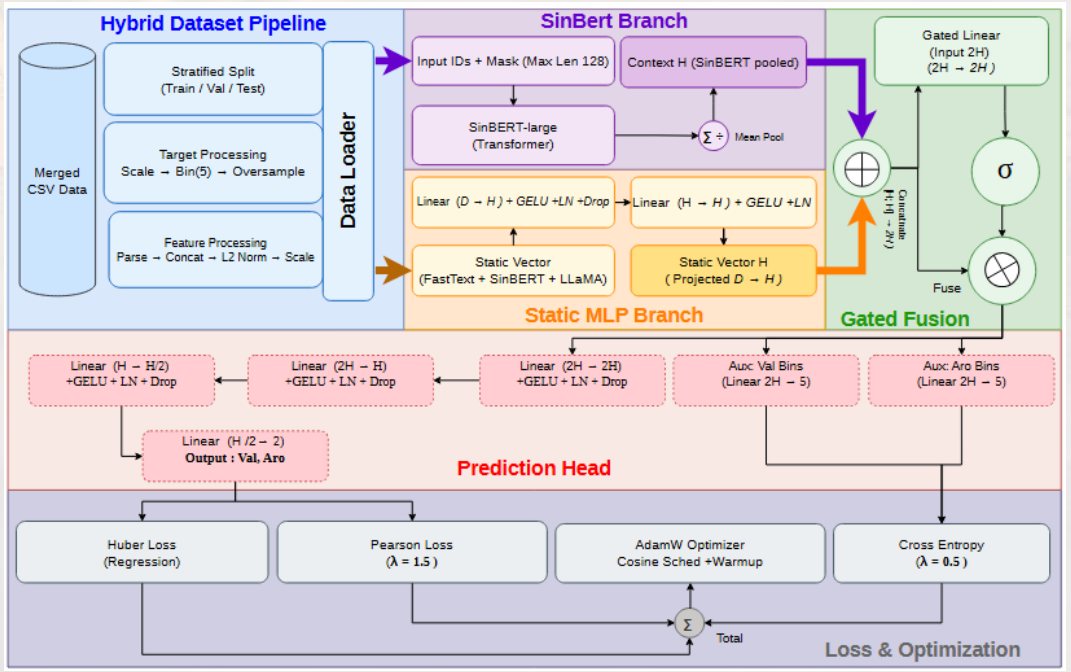


Fig. 2: Hybrid multi-task architecture for valence--arousal prediction in Sinhala social media comments. The model integrates contextual representations from SinBERT with static external embeddings (FastText, SinBERT sentence vectors, and SinLLaMA embeddings) using gated fusion, and jointly optimizes regression and auxiliary classification objectives.

04. Results and Discussion

4.1 Overview of Data Analysis Process

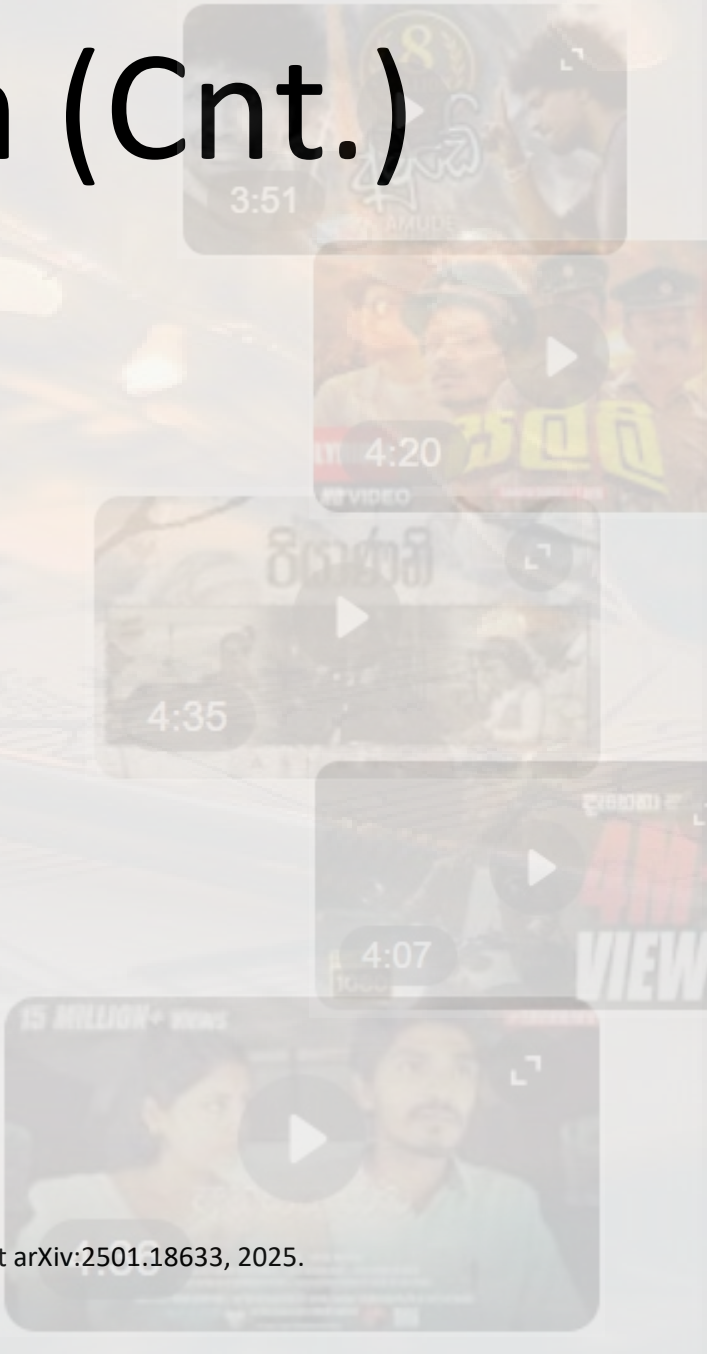
- Comprehensive analysis of 67,959 unique words in Sinhala YouTube comments.
- Rigorous filtration refined dataset to 54,834 unique Sinhala words.
- Examined word frequencies and word pairs to identify key linguistic patterns.
- Highlighted the importance of understanding domain-specific language use.

[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

04. Results and Discussion (Cnt.)

4.2 Insights from Word Frequencies and Word Pairs

- Analysis of 328,922 distinct word pairs reveals linguistic diversity.
- High-frequency words indicate prevalent themes, e.g., "ලස්සනයි," "සුපිරි."
- Top word pairs suggest emotional and aesthetic expressions.
- Analysis aids in understanding cultural and emotional nuances in comments.



[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

04. Results and Discussion (Cnt.)

4.3 Derivation of Stop Words

- Identified 964 frequent words using standardized Z-scores.
- Words with z-scores > 3.0 marked as potential stop words.
- Compared against existing Sinhala and English stop word lists.
- 182 words matched English stop words; 53 matched Sinhala stop words.
- Process enhances the precision of sentiment analysis in Sinhala.

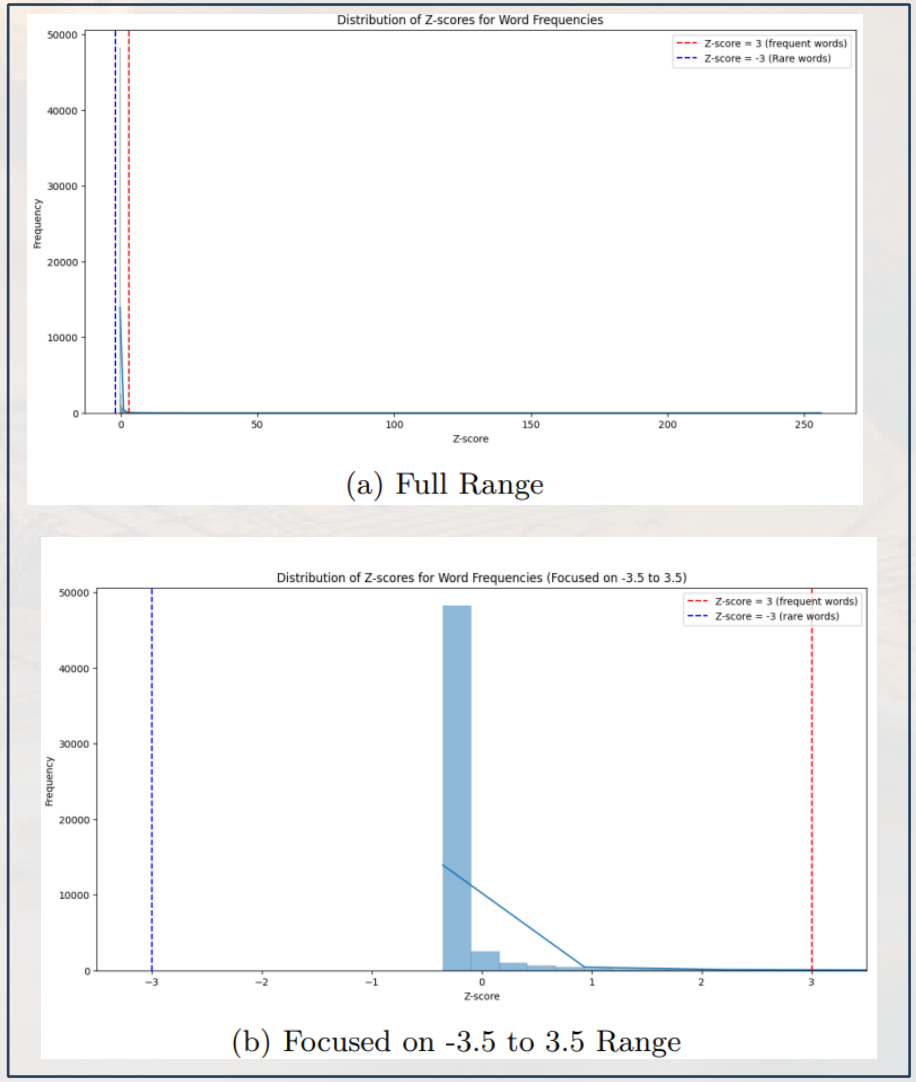


Fig. 3:Distribution of Z Value of Word Frequency

[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

04. Results and Discussion (Cnt.)

4.4 General Data Corpus Construction

- Wikipedia articles on diverse topics enriched the general corpus.
- Included data from Sinhala newspaper articles and government documents.
- Comprehensive word corpus of 121,850 unique words was created.
- Used a mix of One-Hot Encoding and Word2Vec embeddings for visualization.
- Employed t-SNE for dimensionality reduction to observe clustering patterns.
- Larger clusters formed by words from news articles and common words across sources.

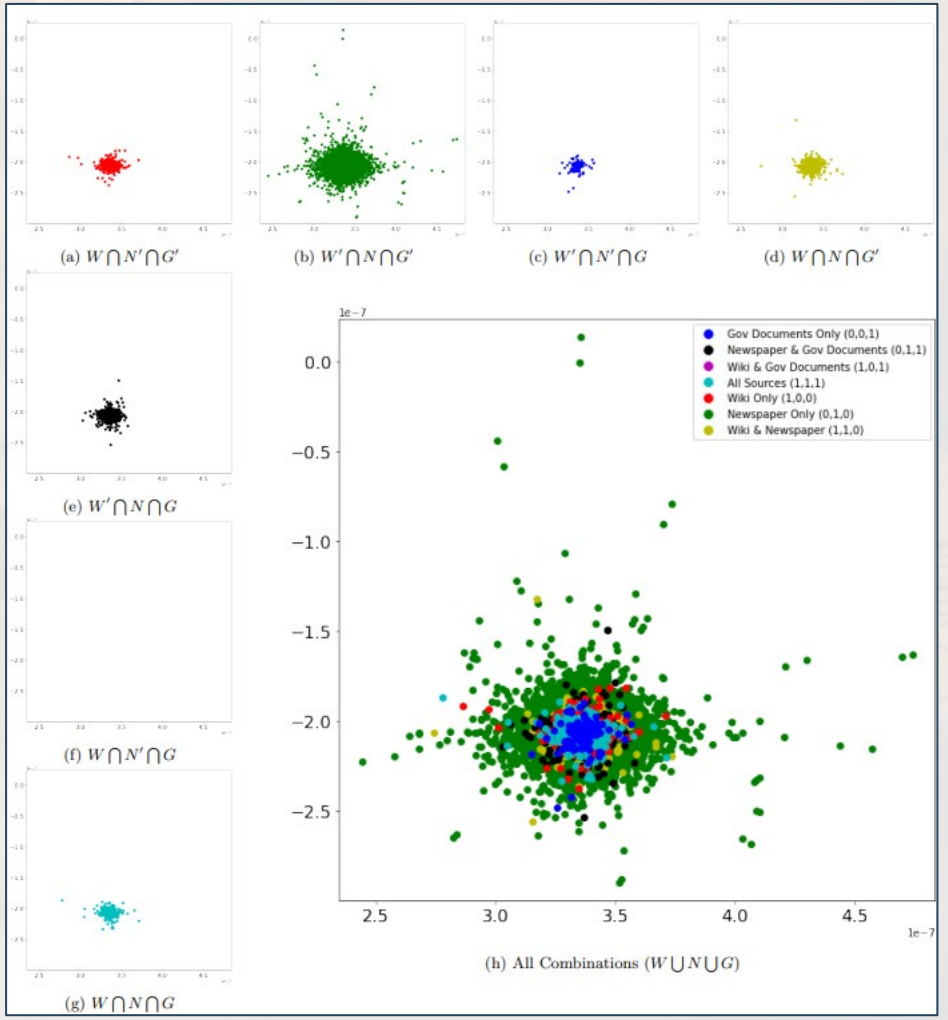


Fig. 4:Word Distribution among Corpora, where the set of words in each of the corpora are denoted as: W= Wikipedia Corpus, N = NEWS Corpus, G= Government Document Corpus

[[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

04. Results and Discussion (Cnt.)

4.5 Comparison of YouTube Comments with General Domain

- 36% overlap of unique words between YouTube and general corpus.
- YouTube domain has distinctive usage patterns.
- Significant overlap suggests representativeness of general usage.
- Identified words more frequent in YouTube domain: “සුපිරි,” “ලස්සනයි.”
- General domain common words less frequent in YouTube: “සඳහා,” “අතර.”
- Z-score analysis revealed domain-specific word usage trends.

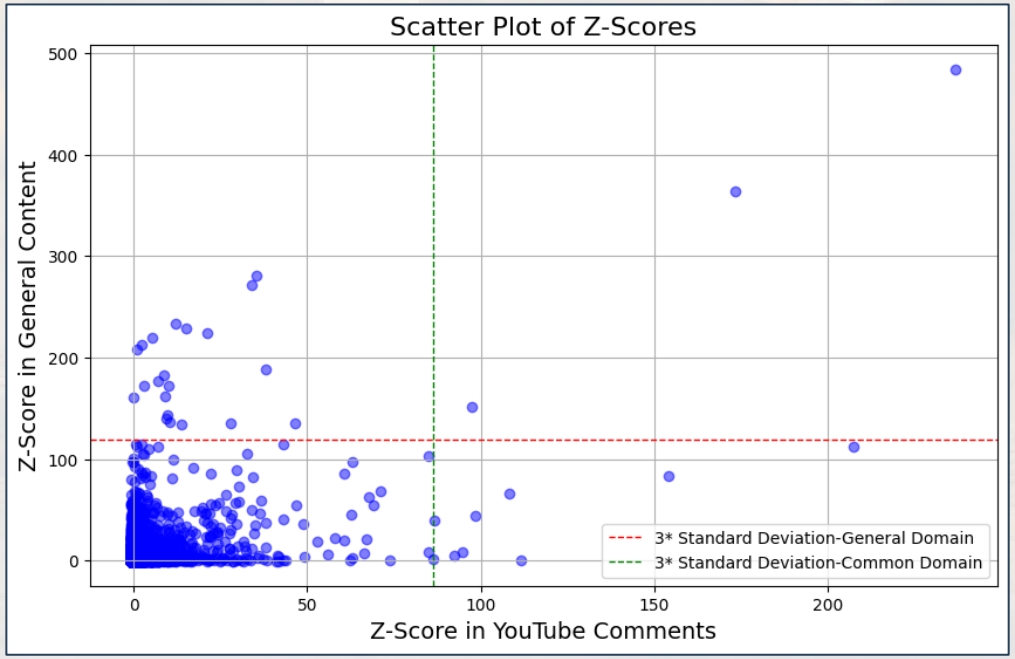


Fig. 5: Scatter Plot of Z-Scores

[15] W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

04. Results and Discussion (Cnt.)

4.6 Performance of Traditional ML Models on Sinhala News Comments

- Evaluated **RF, XGBoost, SBERT, and SinBERT**, demonstrating the benefits of combining traditional and transformer-based representations for Sinhala text classification.
- **Hybrid embeddings** achieved the best performance, with XGBoost + Word2Vec + fastText + SinBERT reaching 80.8% accuracy, highlighting the value of complementary linguistic representations.

4.7 Performance of Deep Learning Models on Sinhala News Commes

- **MLP** achieved the best performance (**Accuracy: 83.7%, F1: 82.0%, ROC-AUC: 83.6%**), outperforming **CNN** and **LSTM**.
- **MLP** also surpassed the best traditional machine learning models, demonstrating the effectiveness of deep neural networks with hybrid embeddings.
- Results indicate that **dense semantic embeddings** capture the most discriminative information, while sequential models provide limited additional benefit.

[17] Y. De Mel and N. de Silva, “GeeSanBhava: Sentiment tagged Sinhala music video comment data set,” in International Conference on Computational Collective Intelligence, pp. 157–171, Springer, 2025.

04. Results and Discussion (Cnt.)

4.7 Performance of Deep Learning Models on Sinhala News Commes (Cnt.)

Model	Accuracy	Precision	Recall	F1 Score	ROC AUC
MLP	0.837	0.850	0.792	**0.820**	**0.836**
LSTM	0.759	0.780	0.678	0.726	0.754
CNN	0.785	0.798	0.727	0.761	0.782

Tab. 3: Performance Metrics of Deep Learning Models

ROC AUC	Layer Sizes	Dropout Rates	Learning Rate
0.836	[512, 256]	[0.3, 0.2]	0.001
0.881	[512, 256]	[0.3, 0.2]	0.001
0.882	[128, 64]	[0.2, 0.1]	0.001
0.882	[512, 256, 128]	[0.4, 0.3, 0.2]	0.0005
0.887	[256, 128, 64]	[0.3, 0.2, 0.1]	0.0003

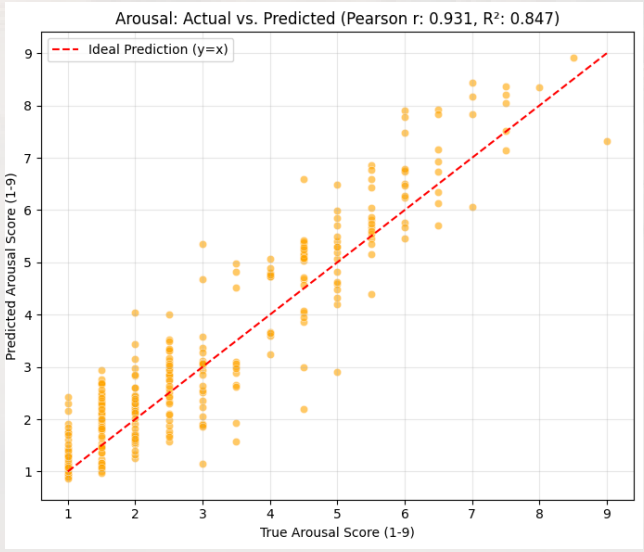
Tab. 4: Performance of Top Five MLP Models

[17] Y. De Mel and N. de Silva, “GeeSanBhava: Sentiment tagged Sinhala music video comment data set,” in International Conference on Computational Collective Intelligence, pp. 157–171, Springer, 2025.

04. Results and Discussion (Cnt.)

4.8 Multi-task learning framework on English FaceBook Comments

- **Proposed multi-task transformer** achieved strong performance on the English Facebook benchmark **(R^2: 0.653/0.847**, Pearson: 0.829/0.931 for Valence/Arousal).
- **Outperformed** previously reported lexicon-based and Bag-of-Words baselines, validating the effectiveness of multi-task learning.
- **Valence and Arousal prediction** plots closely followed the ideal 45° line, confirming strong model reliability.



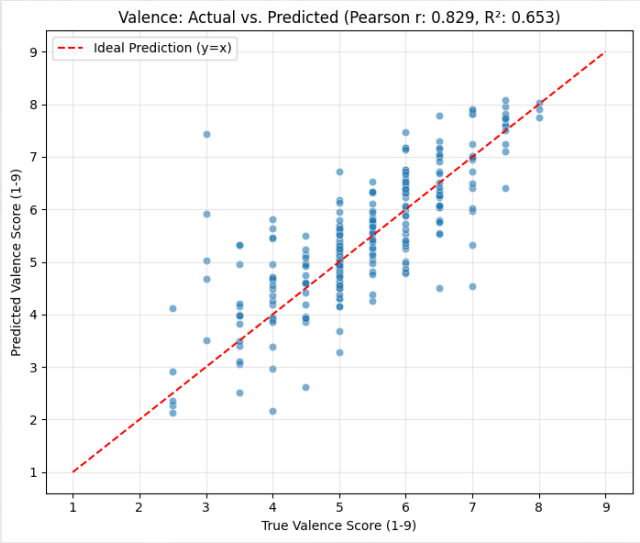
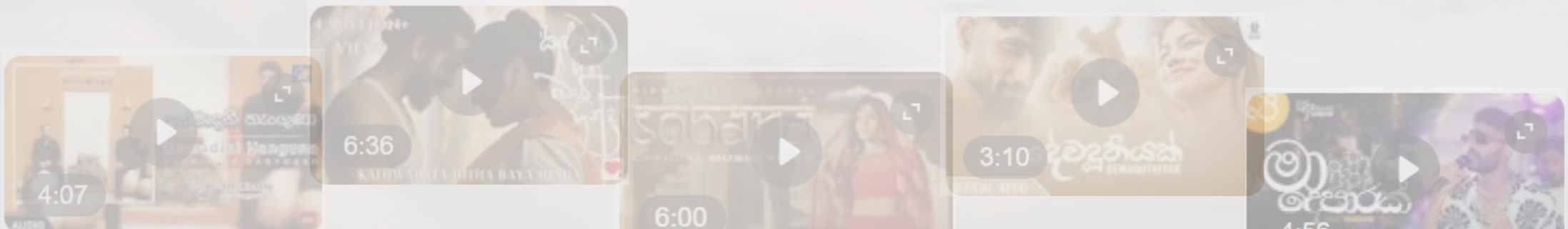


Fig. 6: Predicted Vs Actual valance value for English FaceBook comment dataset using multi-task architecture.

04. Results and Discussion (Cnt.)

4.9 Identifying the Optimal Representation for Sinhala Affect Modeling

- **SinBERT** significantly outperformed **XLM-RoBERTa (Avg. R^2: 0.798 vs. 0.219)**, highlighting the value of language-specific transformers for Sinhala.
- **Ablation** study evaluated **eight embedding configurations** under identical training conditions.
- **All embedding configurations** outperformed the **text-only baseline (Avg. R^2: 0.6716)**, demonstrating the benefit of static semantic embeddings.
- **FastText + SinBERT-vec + SinLLaMA** achieved the best performance **(Avg. R^2: 0.7085)**, showing the effectiveness of combining complementary embeddings.



04. Results and Discussion (Cnt.)

4.9 Identifying the Optimal Representation for Sinhala Affect Modeling (Cont.)

Rank	Static Features	R^2_{val}	R^2_{aro}	Avg
1	FastText + SinBERT-vec + SinLLaMA	0.7095	0.7074	0.7085
2	FastText + SinLLaMA	0.6994	0.6871	0.6933
3	FastText + SinBERT-vec	0.6970	0.6807	0.6888
4	FastText	0.6889	0.6882	0.6885
5	SinLLaMA	0.6975	0.6788	0.6881
6	SinBERT-vec + SinLLaMA	0.6918	0.6821	0.6869
7	SinBERT-vec	0.6856	0.6815	0.6835
8	No static Encoding	0.6746	0.6685	0.6716

Tab. 5:Static encoding ablation results on Sinhala YouTube comments, ranked by the average of valence and arousal R2 scores

04. Results and Discussion (Cnt.)

4.10 Optuna Optimization and Final Model Performance

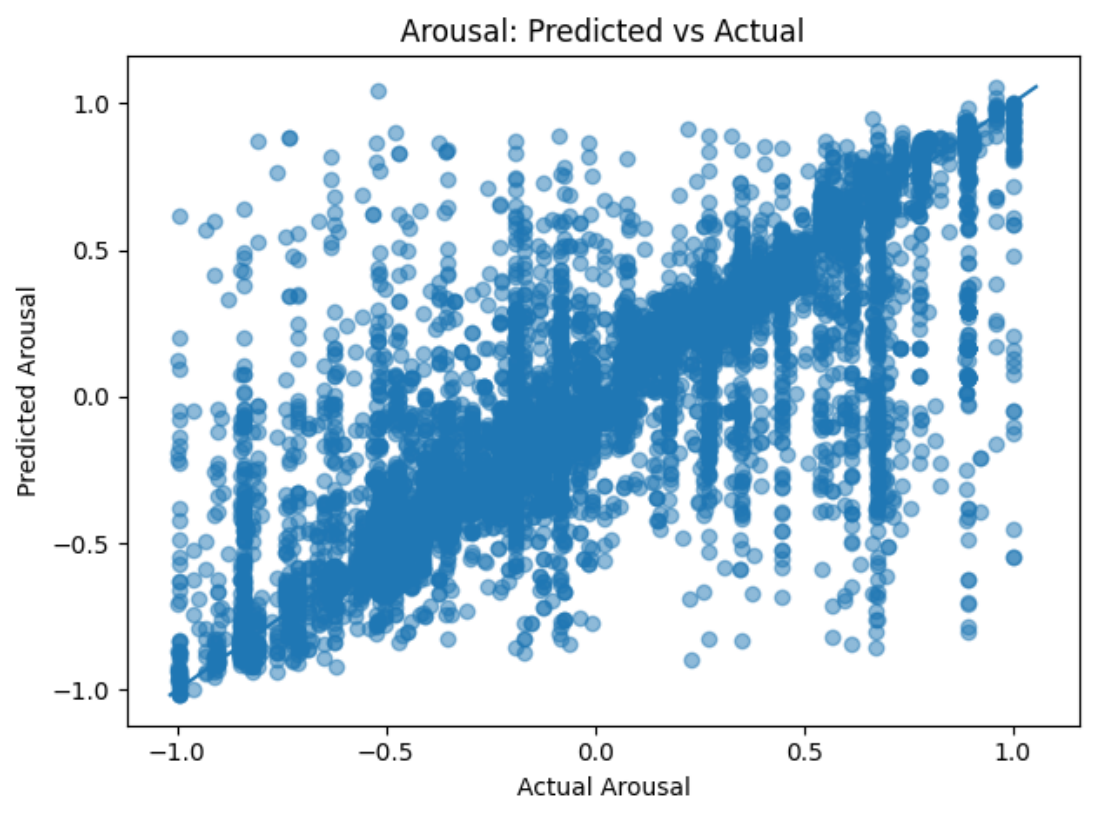
- **Optuna optimization** identified the optimal hyperparameters, improving alignment with human valence–arousal annotations.
- **Optimized hybrid model** achieved **R^2 = 0.798 (Valence)** and **R^2 = 0.797 (Arousal)**, demonstrating strong continuous emotion prediction.
- **High agreement with human annotations**: Pearson = **0.896/0.896**, Spearman = **0.858/0.893** (Valence/Arousal).
- **Prediction and residual analyses** confirmed high accuracy, low error, and strong generalization on unseen Sinhala YouTube comments.

Metric	Valence	Arousal
RMSE	0.1207	0.2223
MAE	0.0436	0.1054
R^2	0.7978	0.7973
Pearson	0.8962	0.8959
Spearman	0.8575	0.8930

Tab. 6:Final test performance of the Optuna-optimized hybrid multi-task regression model using FastText + SinBERT-vec + SinLLaMA static encodings.

04. Results and Discussion (Cnt.)

4.10 Optuna Optimization and Final Model Performance (Cont.)



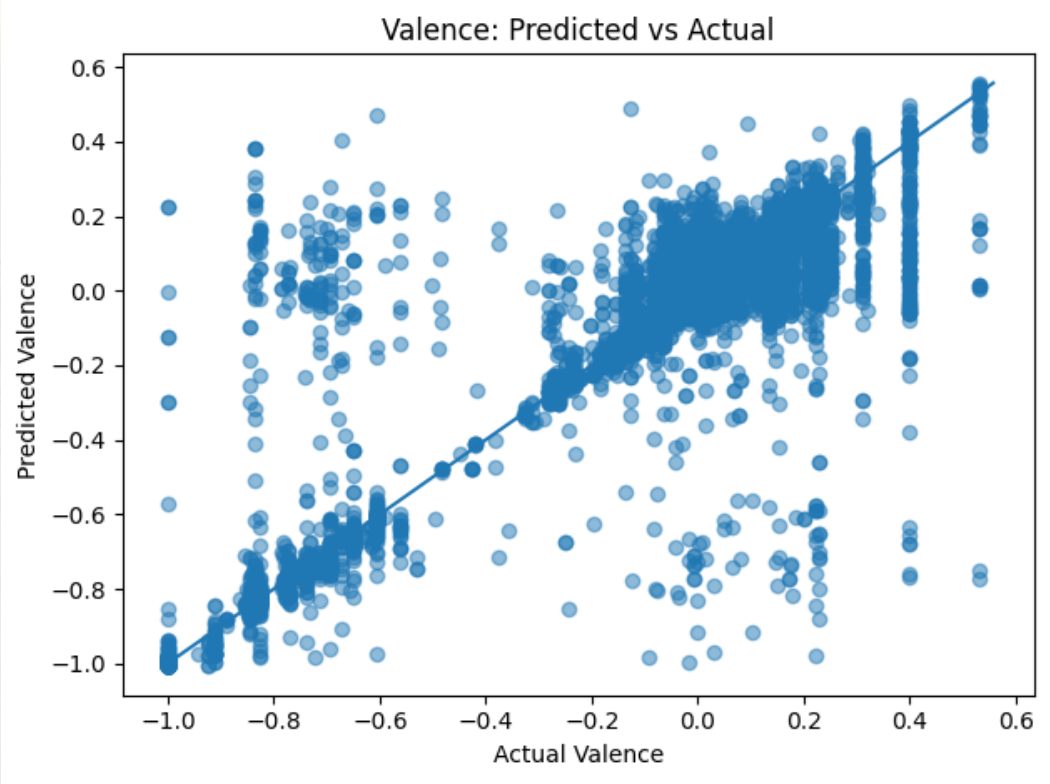
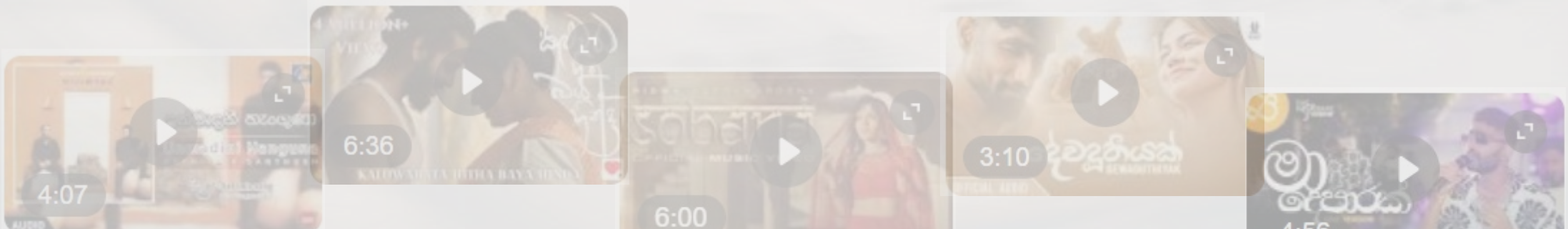


Fig. 7: Predicted Vs Actual Valance and Arousal value for Sinhala Youtube comment dataset using Hybrid multi-task fine tuned model.

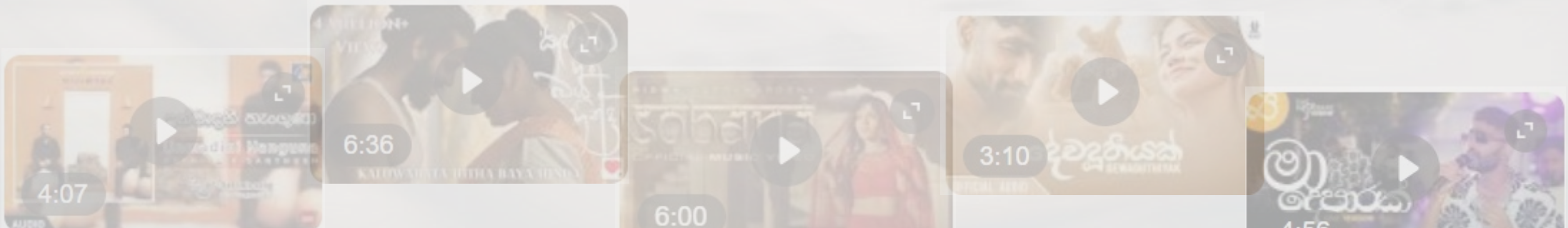
05. Conclusion

- Developed a comprehensive Sinhala NLP pipeline encompassing transliteration, linguistic analysis, and continuous emotion recognition, addressing several fundamental challenges in low-resource language processing.
- Constructed the first large-scale, human-annotated Sinhala YouTube emotion dataset (63,471 comments), providing a reliable benchmark for future research in Sinhala NLP, Music Information Retrieval (MIR), and Music Emotion Recognition (MER).
- Demonstrated that Sinhala social media comments exhibit rich linguistic and affective characteristics, establishing YouTube comments as a valuable resource for linguistic, cultural, and emotion analysis.



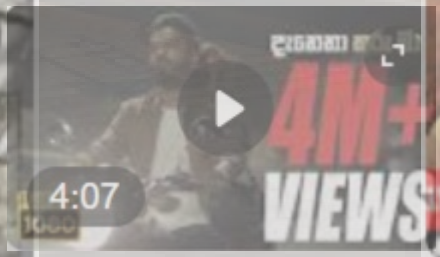
05. Conclusion (Cnt.)

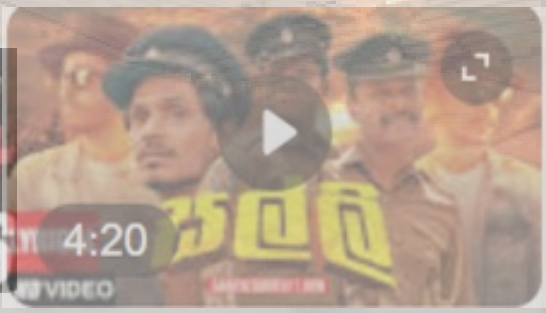
- Proposed a hybrid multi-task continuous Valence–Arousal regression framework that integrates contextual transformers with complementary static embeddings, significantly outperforming traditional classification-based emotion modeling.
- Established a strong foundation for future Sinhala affective computing by releasing valuable datasets, validating continuous emotion modeling, and advancing state-of-the-art methodologies for low-resource language understanding.



Thank

Q & A





References

1. N. de Silva, “Survey on publicly available sinhala natural language processing tools and research,” arXivpreprint arXiv:1906.02358, 2019.

2. D. Upeksha, C. Wijayarathna, M. Siriwardena, L. Lasandun, C. Wimalasuriya, N. De Silva, and G. Dias, “Implementing a corpus for sinhala language,” in Symposium on Language Technology for South Asia, vol. 2015, 2015, p. 3.

3. N. de Silva, “Sinhala text classification: observations from the perspective of a resource poor language,” ResearchGate, 2015.

4. Y. Wijeratne and N. de Silva, “Sinhala language corpora and stopwords from a decade of sri lankan facebook,” arXiv preprint arXiv:2007.07884, 2020.

5. H. Hettiarachchi, D. Premasiri, L. Uyangodage, and T. Ranasinghe, “Nsina: A news corpus for sinhala,” arXiv preprint arXiv:2403.16571, 2024.

6. M. A. Casey, R. Veltkamp, M. Goto, M. Leman,C. Rhodes, and M. Slaney, “Content-based music information retrieval: Current directions and future challenges,” Proceedings of the IEEE, vol. 96, no. 4,pp. 668–696, 2008.

7. F. Simonetta, S. Ntalampiras, and F. Avanzini, “Multimodal music information processing and retrieval:Survey and future challenges,” in 2019 international workshop on multilayer music representation and processing (MMRP). IEEE, 2019, pp. 10–18.

8. S. Hizlisoy, S. Yildirim, and Z. Tufekci, “Music emotion recognition using convolutional long short termmemory deep neural networks,” Engineering Science and Technology, an International Journal, vol. 24,no. 3, pp. 760–767, 2021.

References

9. Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton,P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, “Music emotion recognition: A state of the art review,” in Proc. ismir, vol. 86, 2010, pp. 937–952.

10. Y.-H. Yang and H. H. Chen, “Machine recognition of music emotion: A review,” ACM Transactions on Intelligent Systems and Technology (TIST), vol. 3,no. 3, pp. 1–30, 2012.

11. S. Koelsch, “Investigating emotion with music: neuroscientific approaches,” Annals of the New York Academy of Sciences, vol. 1060, no. 1, pp. 412–418,2005.

12. L.-C. Yu, J. Wang, K. R. Lai, and X. Zhang, “Refining word embeddings using intensity scores for sentiment analysis,” IEEE/ACM Transactions on Audio,Speech, and Language Processing, vol. 26, no. 3, pp.671–681, 2017.

13. T. Mikolov, K. Chen, G. Corrado, and J. Dean,“Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

14. S. Lai, L. Xu, K. Liu, and J. Zhao, “Recurrent convolutional neural networks for text classification,”in Proceedings of the AAAI conference on artificial intelligence, vol. 29, no. 1, 2015.

15 W. M. De Mel and N. de Silva, “Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study,” arXiv preprint arXiv:2501.18633, 2025.

16 Y. De Mel, K. Wickramasinghe, N. de Silva, and S. Ranathunga, “Sinhala transliteration: A comparative analysis between rule-based and Seq2Seq approaches,” in Proceedings of the First Workshop on Natural Language Processing for Indo-Aryan and Dravidian Languages, pp. 166–173, 2025.

References

17 Y. De Mel and N. de Silva, “GeeSanBhava: Sentiment tagged Sinhala music video comment data set,” in International Conference on Computational Collective Intelligence, pp. 157–171, Springer, 2025.

18 S. Ranathunga and I. U. Liyanage, “Sentiment analysis of Sinhala news comments,” Transactions on Asian and Low-Resource Language Information Processing, vol. 20, no. 4, pp. 1–23, 2021.

19 T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv preprint arXiv:1301.3781, 2013.

20 P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, “Enriching word vectors with subword information,” Transactions of the Association for Computational Linguistics, vol. 5, pp. 135–146, 2017.

21 V. Dhananjaya, P. Demotte, S. Ranathunga, and S. Jayasena, “Bertifying Sinhala: A comprehensive analysis of pre-trained language models for Sinhala text classification,” in Proceedings of the Thirteenth Language Resources and Evaluation Conference, pp. 7377–7385, 2022.

22 N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 3982–3992, 2019.