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

Emotion Modelling in Social Media: Current Limitations

- Most NLP systems treat emotion as discrete classification
- Psychological theory: Emotion is continuous
  - Valence (positive ↔ negative)
  - Arousal (calm ↔ excited)
- Discretization causes:
  - Information loss
  - Unstable boundaries
  - Poor generalization
- Low-resource languages (Sinhala) remain underexplored

Core Research Questions

- Does discretized V–A classification degrade performance?
- Does continuous regression improve alignment with human annotations?
- Can hybrid representation learning improve affect prediction in Sinhala?
- What embedding combination yields optimal performance?
- How much can systematic hyperparameter tuning improve results?

Related Work

Valence–Arousal Modeling

- Russell (1980) — Circumplex Model of Affect
- Bradley & Lang (1999), Warriner et al. (2013)
    - Lexicon-based V–A modeling
- Preoțiuc-Pietro et al. (2016)
    - Continuous regression > classification
- Beck et al. (2014), Buechel et al. (2018)
    - Multi-task regression improves stability

Transformer-Based Emotion Modeling

- BERT, RoBERTa, XLM-R
- Multi-task learning improves affect modeling
- Correlation-aware objectives improve alignment
- Sentence-BERT & FastText improve robustness
- SinBERT (Dhananjaya et al., 2022)
- SinLLaMA (Aravinda et al., 2025)

Initial Baseline: Discrete Classification

- Discretized V–A Prediction
  - V–A space divided into joint bins
  - Hybrid model trained to predict categories
- Results
  - R^2 Valence = 0.56
  - R^2 Arousal = 0.58
  - F1 Valence = 0.44
  - F1 Arousal = 0.52
- Interpretation
  - Moderate regression alignment
  - Poor classification discrimination
  - Discretization introduces information loss

Multi-Task Continuous V–A Modeling

- Predict:
  - Continuous valence
  - Continuous arousal
- Auxiliary:
  - Discretized bin supervision
- Loss function:
  - Huber regression
  - Pearson correlation loss
  - Cross-entropy (auxiliary)

Validation on Facebook V–A Dataset (English)

Results

- Pearson:
  - Valence = 0.829
  - Arousal = 0.931
- R^2:
  - Valence = 0.653
  - Arousal = 0.847

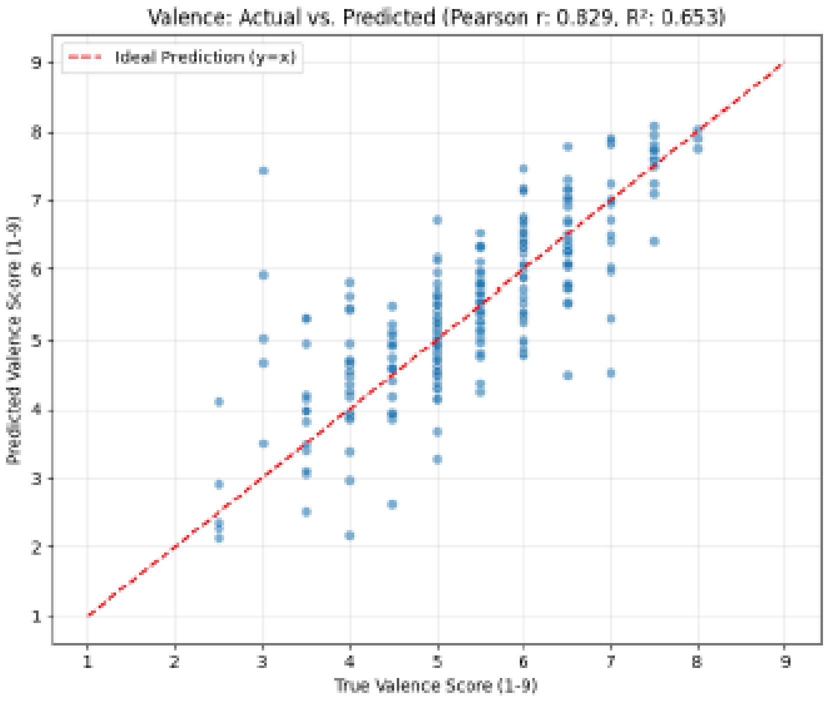

Fig1:Predicted Vs Actual valance value for English FaceBook comment dataset (Preo¸tiuc-Pietro et al., 2016) using multi-task architecture.

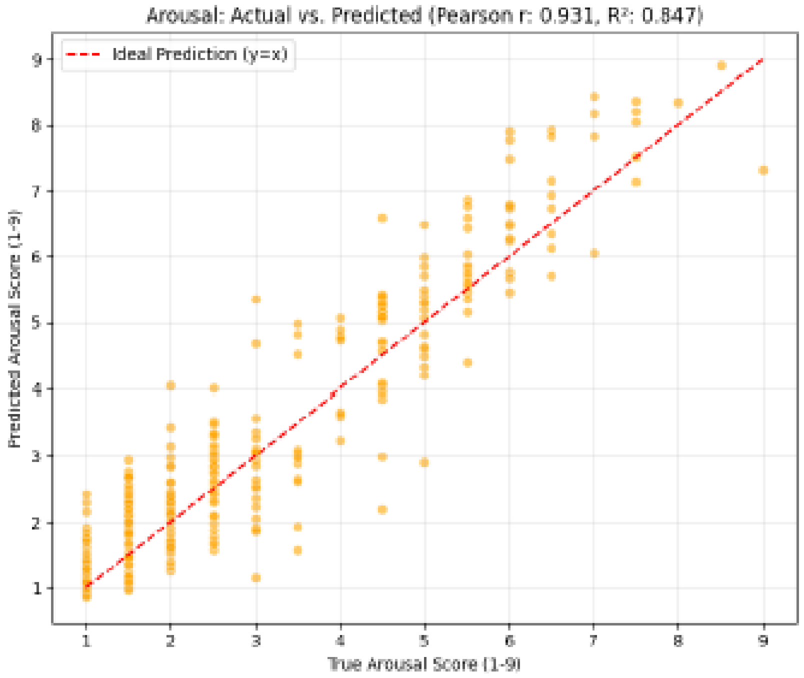

Fig2: Predicted Vs Actual arousal value for English FaceBook comment dataset (Preo¸tiuc-Pietro et al., 2016) using multi-task architecture.

Sinhala YouTube Comment Dataset

- Challenges:
  - Informal syntax
  - Code mixing
  - Lexical variation
  - Limited annotated data
  - Imbalanced V–A distribution
- Approach:
  - Hybrid modeling
  - Targeted oversampling
  - Multi-task objective retained

Hybrid Architecture

- Architecture Components:
- SinBERT-large contextual encoder
- Static embedding branch:
  - FastText
  - Sentence-level SinBERT
  - SinLLaMA
- Gated fusion mechanism
- Shared regression + auxiliary heads

Hybrid Architecture Cont.

Fig1:

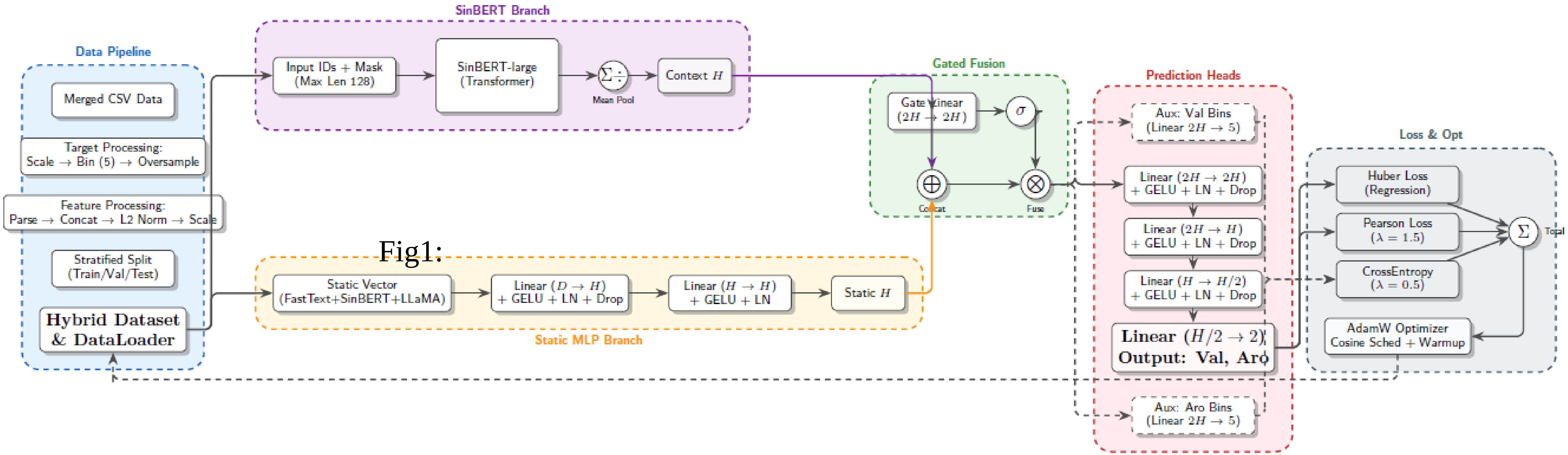

Fig3: Hybrid multi-task architecture for valence–arousal prediction in Sinhala social media comments. The model integrates contextual representations from SinBERT with static external embeddings (FastText, SinBERT sentence vectors, and SinLLaMA embeddings) using gated fusion, and jointly optimizes regression and auxiliary classification objectives.

Controlled Ablation Study

- 8 configurations evaluated

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

Hyperparameter Optimization

- Optuna-Based Optimization was used.
- Tuned: Learning rate, Layer-wise decay, Dropout, Loss weights, Batch size, Warmup, Optimizer
- Best Configuration: AdamW, $\lambda corr = 2.38$, $\lambda cls = 0.20$, Dropout = 0.26

Final Model Performance

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

Final Model Performance Cont.

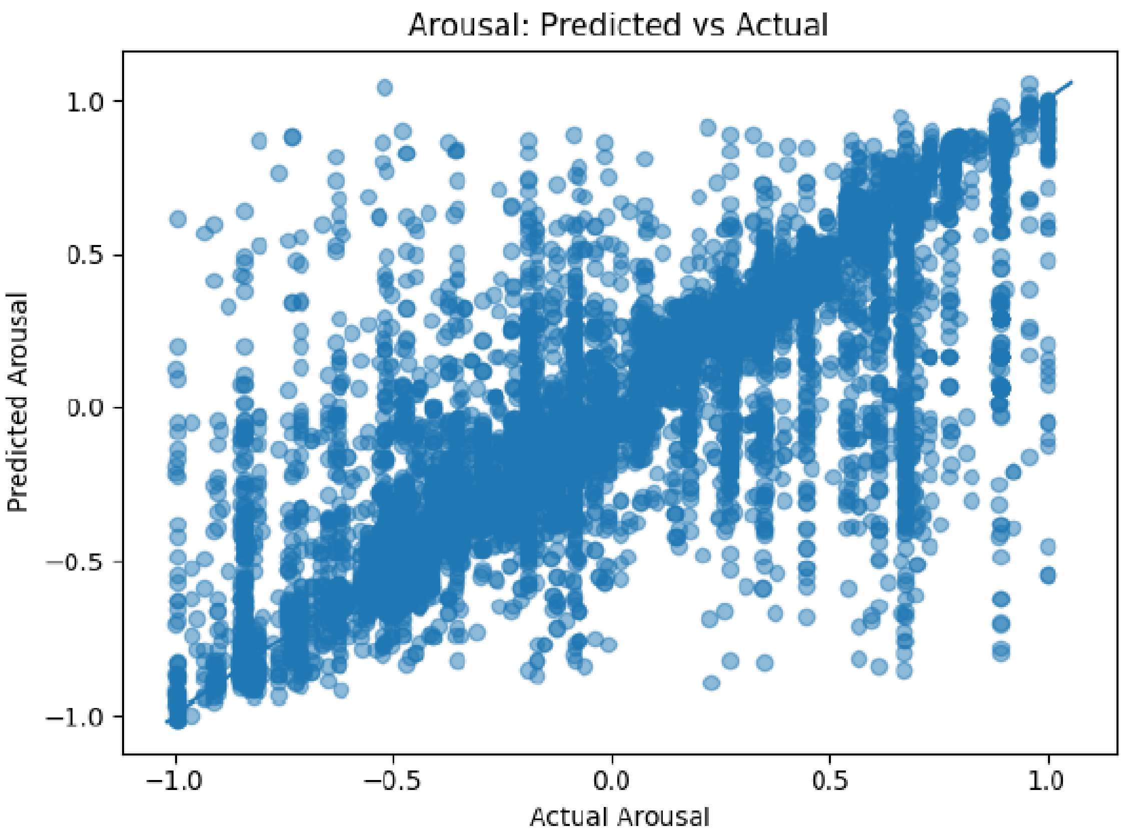

Fig4: Predicted Vs Actual arousal value for Sinhala Youtube comment dataset using Hybrid multi-task finetuned model.

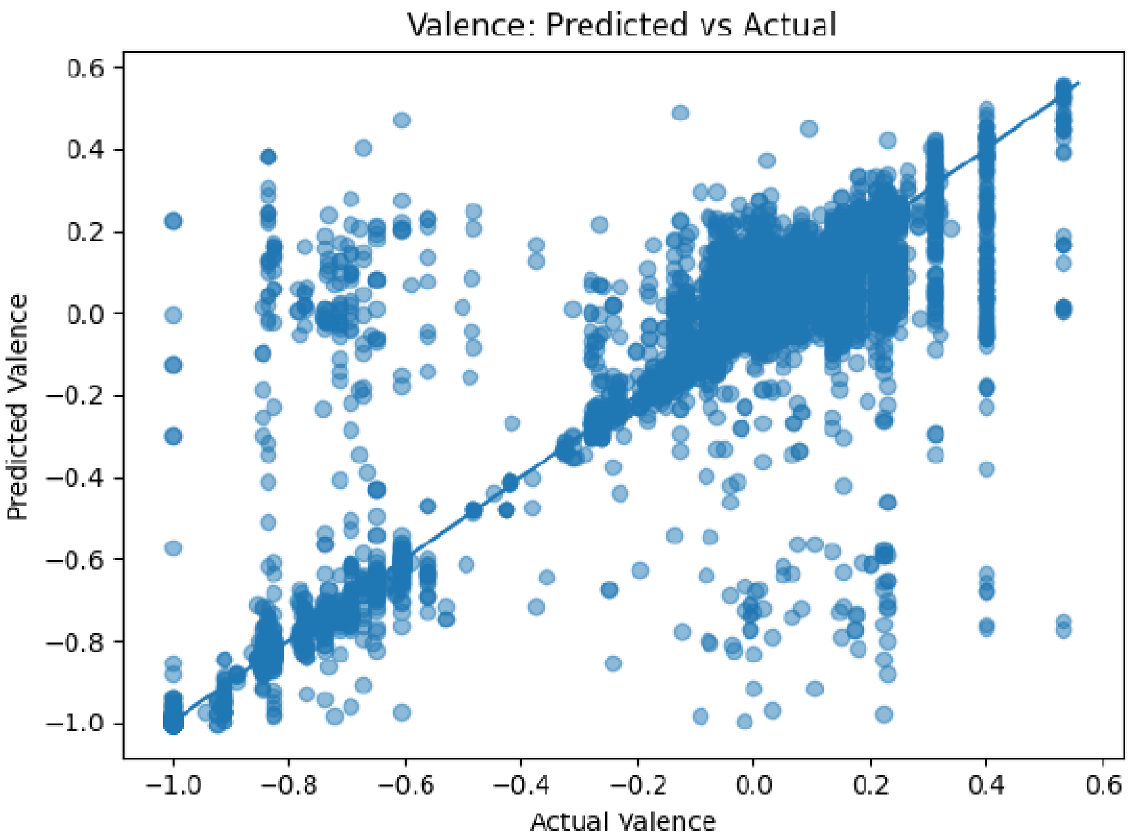

Fig5: Predicted Vs Actual valance value for Sinhala Youtube comment dataset using Hybrid multi-task finetuned model.

Key Contributions

- Demonstrated instability of discretized V–A classification
- Established continuous regression framework for Sinhala
- Proposed hybrid gated fusion architecture
- Systematically evaluated static embedding combinations
- Achieved $R^2 \approx 0.80$ in low-resource setting
- Provided strong baseline for Sinhala affective computing

Future Work

- Cross-domain evaluation (news, forums)
- Multilingual transfer experiments
- Domain adaptation
- Larger-scale annotation expansion
- Lightweight deployment variants

References

Russell, James A. 1980. A circumplex model of affect. Journal of Personality and Social Psychology, 39(6):1161–1178.

Margaret M. Bradley and Peter J. Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Behavior Research Methods, pages 25–46.

Amy Beth Warriner, Victor Kuperman, and Marc Brysbaert. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behavior Research Methods, pages 1191–1207.

Daniel Preoțiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. Modelling valence and arousal in Facebook posts. In Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 9–15. Association for Computational Linguistics.

Daniel Beck, Trevor Cohn, and Lucia Specia. 2014. Joint emotion analysis via multi-task Gaussian processes. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1798–1803. Association for Computational Linguistics.

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. arXiv preprint arXiv:1808.10399.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of NAACL-HLT 2019, pages 4171–4186. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In Proceedings of ACL 2020, pages 8440–8451.

Vinura Dhananjaya, Piyumal Demotte, Surangika Ranathunga, and Sanath Jayasena. 2022. BERTifying Sinhala: A comprehensive analysis of pre-trained language models for Sinhala text classification. In Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022), pages 7377–7385.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5:135–146.

References

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5:135–146.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using siamese BERT-networks. In Proceedings of the EMNLP Workshop on Representation Learning, pages 1–10.

HWK Aravinda, Rashad Sirajudeen, Samith Karunathilake, Nisansa de Silva, Rishemjit Kaur, and Surangika Ranathunga. 2025. SinLLaMA: A large language model for Sinhala. In Proceedings of the 2025 Moratuwa Engineering Research Conference (MERCon), pages 617–622. IEEE.

Yomal De Mel and Nisansa de Silva. 2025. Linguistic analysis of Sinhala YouTube comments on Sinhala music videos: A dataset study. arXiv preprint arXiv:2501.18633.

Yomal De Mel and Nisansa de Silva. 2025. GeeSanBhava: Sentiment tagged Sinhala music video comment dataset. In International Conference on Computational Collective Intelligence, pages 157–171. Springer.