SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V13I4P129

Volume 13 Issue 4, 371-383, April 2026

Original Article

A Transformer–Ensemble Hybrid Framework for Emotion Classification in Low-Resource Kannada Poetry

Smita Girish[1], Kamalraj R[2]

[1,2]*School of CS and IT, Jain University, Bangalore, India.*

[1]*Corresponding Author : smitagirishn@gmail.com*

Received: 25 January 2026 Revised: 25 February 2026 Accepted: 26 March 2026 Published: 30 April 2026

Abstract *- Emotion classification of literary text remains a challenging NLP task due to metaphorical expressions, implicit affective cues, and semantic ambiguity, particularly in low-resource and morphologically rich languages. Kannada, a primary Dravidian language spoken by over 45 million people, presents additional complexity in poetry, where emotions are often conveyed indirectly through symbolism. Despite this, computational analysis of Kannada poetic texts remains limited. This paper proposes a MuRIL-based Transformer–Ensemble framework for classifying 490 manually annotated Kannada short poems into nine emotion categories: Joy, Peace, Wonder, Courage, Compassion, Anger, Melancholy, Fear, and Disgust. Baseline experiments using traditional machine learning models reveal modest performance, with Support Vector Machine (SVM) achieving 45% accuracy, Naïve Bayes 50%, and Random Forest 55%, indicating their inability to capture contextual and implicit emotional semantics in poetic language. To overcome these limitations, the proposed approach integrates MuRIL contextual embeddings with an ensemble of SVM, Random Forest, and Naïve Bayes classifiers using hard and soft voting strategies. Class imbalance is addressed through a training-only augmentation method, Adaptive Minority Expansion and Overfitting Control (AMEOC), employing synonym replacement, paraphrasing, and back-translation while maintaining semantic integrity. Experimental results show that the proposed framework achieves an overall accuracy of 79% and an F1-score of 0.78, significantly outperforming baseline classifiers, TF–TF-IDF-based models, and a MuRIL-only classifier. The performance progression from classical models to contextual transformer-based ensemble learning demonstrates the effectiveness of ensemble fusion and controlled Augmentation for interpreting subtle, metaphor-rich emotions in Kannada poetry. The proposed framework contributes to computational literary analysis and has potential applications in digital humanities, literary retrieval, and cultural analytics for low-resource Indian languages.*

Keywords *- Kannada Poetry, SVM, Naïve Bayes, Random Forest, Emotion Classification, MuRIL, Transformer Ensemble, Low-Resource NLP, Data Augmentation, Voting Classifier.*

1. Introduction

Emotion classification in literary texts is a challenging problem for Natural Language Processing (NLP) as it demands a deep understanding of semantic cues, metaphors, and culturally toco linguistic affective expressions. In contrast to polarity-focused sentiment labelling, the task of multi-emotion classification is to label fine-grained affective categories, which is even more important for short poems due to embedded emotions through metaphors and symbol-based imagery, as well as compressed linguistic form. Low-resource Indian languages exacerbate these problems even more because of very few annotated corpora, lack of emotion-labelled datasets, and linguistic features such as agglutination and rich morphology like most Dravidian languages [1, 2]. Kannada is a morphologically rich Dravidian language with over 40 million speakers. Fine-grained categories, including Joy, Peace, Anger, Fear, Disgust, Compassion, Courage, Wonder, and Melancholy, make the framework suitable for higher-level applications in digital humanities, literary curation, emotion-aware retrieval, and cultural analytics. Such a fine-grained approach is particularly crucial in poetry, as affective states may overlap with each other, meanings are non-literal, and the expressions have a philosophical underpinning that cannot be captured utilizing naive sentiment polarity approaches. The task is challenging due to (i) metaphoric expressions, (ii) little semantic ambiguity, (iii) contextual information across long distances, and (iv) scarce labelled data, strictly imbalanced across emotion types, with some emotions being more frequent than others by several orders of magnitude. For Indian languages, the system models like MuRIL and XLM-R have performed well as a result of their contextualized

embeddings and cross-lingual pretraining [2, 3]. However, transformer-only models do not do as well when the dataset is small, domain-specific, or dense, as metaphor-like poetry tends to be. In contrast, traditional machine learning models do not know the context but are resilient to small datasets. Instead, ensemble learning is effectively used for the heterogeneous boundary strength aggregation (soft-voting) of SVM, Random Forest, and Naïve Bayes in particular [4].

Class imbalance is a big problem in low-resource emotion classification; one or a few classes (e.g., joy and sarcasm) overwhelm the database distribution, while another class over-rarely appears (e.g., fear and melancholy). It creates bias in favor of the majority classes. The data augmentation strategies, synonym replacement, paraphrasing, and back-translation, which can be used by a model, will only be used during training (introduce noise at the time of concatenating a training sentence with negative (impure language) sentences or when translating to English). The test set is then left clean, ensuring the evaluation remains unbiased. Some previous research works have tried to perform sentiment or emotion analysis in Indian languages, specifically Bengali [5], Arabic poetry [6], and code-mixed Kannada social media text. To the best of our knowledge, the current work is the first to try multi-class emotion classification for Kannada literary text and to combine transformer embeddings with ensemble learning in this particular domain.

Motivated by these observations, this work proposes a Transformer–Ensemble Hybrid Model that takes advantage of MuRIL embeddings for encoding deep semantics and a weighted soft-voting ensemble for making classification decisions robust against low-resource, imbalanced, or metaphor-heavy settings. The method systematically compares transformer-based embeddings to traditional TF-IDF representations, benchmarking performance across several classifiers along the way. It reveals the effectiveness of training-only Augmentation for minority emotion classes.

1.1. Contribution / Originality

Towards this goal, the paper proposed a new model for computational interpretation of emotions in Kannada poetry using cosmic emotional entities, thereby facilitating Artificial Intelligence (AI) within a digital preservation and literary analysis system to become more culturally sensitive. By teaching machines to understand nuanced, metaphor-laden emotional signals, the work enriches resources for digital humanities, emotion-aware content search, and educational content in regional languages. The proposed methodology, MuRIL-based Transformer–Ensemble model, along with the AMEOC guided augmentation strategy, is able to perform robust emotion classification even in low-resource and imbalanced setups. The novelty is to combine deep contextual embedding, ensemble decision fusion, and adaptive Augmentation for enhancing the automatic understanding of literary emotions in the majority of underrepresented Indian languages.

1.2. Objectives of the Study

The contributions of this research are as follows: (1) Creating a manually annotated corpus of Kannada short poems labelled under nine emotion categories, thus establishing a reliable gold-standard dataset for evaluation; (2) Developing a MuRIL-based Transformer–Ensemble model that is capable of capturing the nuanced implicit and metaphor-rich emotional clues hidden within low-resource poetic language; (3) Devising the AMEOC controlled data augmentation method to balance classes while avoiding overfitting; and (4) Comparing the proposed technique with TF–IDF baselines and an independent MuRIL classifier, verifying overall improvements across accuracy, F1-score along with general robustness.

1.3. Significance of the Study

The present study is of relevance to the development of computational processing based on Kannada literary text – emotionally automatic interpretation in expressions-rich poems, which otherwise rely on human experts. The proposed Transformer–Ensemble model can be used to work with digital libraries, literary archives, and cultural heritage projects in which regional literature collections can be processed through their emotional themes for discovery, organization, and analysis. The system is also beneficial to educators, linguists, and digital humanities researchers who need resources for emotion-aware literary analysis and curriculum development. In a more general setting, the framework can be adapted for affective-driven recommendation systems, intelligent tutoring system and sentiment aware chatbots in Indian languages. In addition to its usefulness for focused grammatical study, this model has the potential to support other NLP tasks such as plagiarism identification, stylistic similarity analysis, author profiling, and discourse, becoming a valuable resource for application in computational linguistics and regional language technology development.

2. Related Work

Emotion analysis for low-resource languages has become particularly popular due to the lack of annotated corpora and language tools. Previous research [1, 2] has shown that agglutinative and morphologically complex languages such as Kannada, Tamil, and Malayalam are fundamentally difficult to handle for NLP systems due to rich features like suffixation/compound morphology and a wide range of lexicon. Classical models such as SVM, Naïve Bayes, or RNNs fail to achieve good results when they come across these linguistic complexities and data paucity. To overcome these drawbacks, recent emotion-analysis studies conducted in Bengali literature opted for deep neural architectures and revealed that such contextual word embeddings outperform sparse text representation methods

for literary emotion modeling [5]. Work in Arabic poetry also supports this direction; Al-Khatib and Elbassuoni demonstrated that the deep models outperformed traditional ones for metaphorical and multi-layered affective constructs present in classical poetry [6].

The emergence of transformer architecture was a game-changer in low-resource NLP. Models like XLM-R and multilingual BERT have been instrumental in advancing cross-lingual transfer, providing strong zero-shot and few-shot performance on typologically distant languages [3]. Transformers have also been established to be effective for code-mixed indian text as well, where subword tokenization and contextual encoding enable dealing with script variation and lexical mixing [8]. For Kannada, existing work has primarily been material towards polarity-based sentiment classification of social media text via machine learning or transformer frameworks [7], but research on fine-grained emotion (and that too in literary) analysis is very scarce. Even if disembodied sentiment analysis on the phonemic substrate were possible at all, within the larger Indic NLP context , sentiment and emotion analysis often give precedence to user-generated content rather than metaphor-rich, syntactically condensed poetic text.

Unlike prose, poem emotion classification has dense metaphors, symbolism, associative imagery, and multi-emotion overlapping. Affective signals in poetic lines are not all explicit or lexical, but instead are embedded in cultural metaphors and contextual hierarchies. “There have been promising studies using deep-learning techniques for Arabic and Bengali poetry [5, 6]; however, there is almost no similar work done for Kannada poetry, to the best of our knowledge. Existing Kannada NLP research has not covered multi-class emotion modeling on literary corpora, and there is a lack of datasets, benchmarks, and methodologies.

Another vital direction is ensemble learning, which further boosts robustness in low-resource settings. Dietterich proved the theoretical benefits of applying ensemble methods -- variance reduction, increased robustness, and complementary error correction -- to noisy or limited-depth training paradigms [4]. Recent studies also demonstrate that with ensemble classifiers, neural embeddings have the potential to outperform their discrete counterparts on low-resource text classification [9, 10]. These works, however, do not incorporate transformer-based embeddings coupled with ensemble learning for the task of Kannada poetry emotion classification, and hence, it leads to a research gap which is hereby filled in this work.

3. Dataset and Challenges

The dataset contains 490 manually curated Kannada short poems, where each poem is pre-processed using MuRIL’s subword tokenizer to retain agglutination, sandhi patterns, and morphological complexity, which are inherent characteristics of annotated Kannada literary text. The poems are tagged to one of nine emotion categories (Joy, Peace, Wonder, Courage, Compassion, Anger, Melancholy, Fear, and Disgust), indicating a fine-grained affect spectrum more apt for poetry comparison. The dataset is intrinsically imbalanced (see distribution Table), where Joy, Peace, and Wonder are dominant classes and emotions such as Fear, Disgust, and Melancholy are clearly less frequent. Such an imbalance impacts both the stability of gradients and recall of minority categories by classifiers, especially in transformer-based architectures, which are known to overfit the dominant patterns associated with emotion. It can be mitigated by employing class-specific data augmentation only for minority-class examples in training using semantic-preserving augmentations, including synonym replacement and controlled paraphrasing. Augmented samples are left out of evaluation to preserve test-set integrity, and it can be confident that performance gains are due to actual model generalization on low-data skewed distributions.

As shown in Table 1 and the plotted distribution, the dataset exhibits significant inter-class imbalance, with minority emotions occurring at substantially lower frequencies.

Table 1. Emotion class distribution

Emotion Category	Number of Poems
Joy	110
Peace	95
Wonder	70
Courage	55
Compassion	50
Anger	35
Melancholy	30
Fear	25
Disgust	20

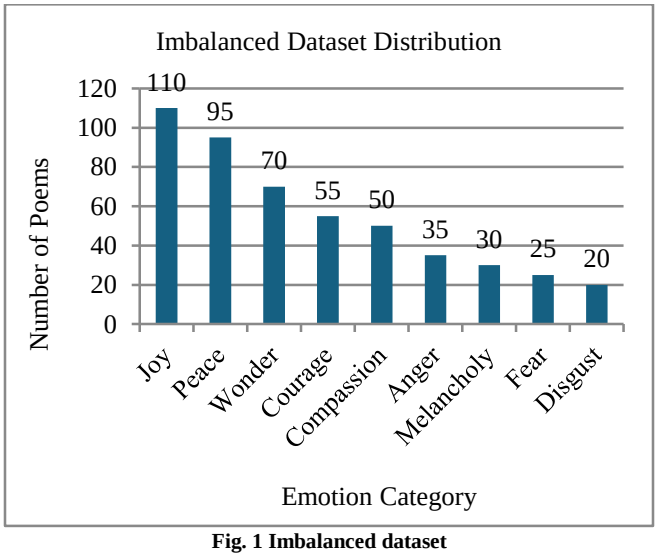

Fig. 1 Imbalanced dataset

4. Data Preprocessing and Augmentation Using Adaptive Minority Expansion & Overfitting Control (AMEOC) Novel Strategy.

Before training the model, the Kannada poem corpus passes through minimal preprocessing, which is essential for maintaining text consistency. Preprocessing includes: Unicode normalisation, subword tokenisation, character lowercasing and cleaning, lemmatisation, and text normalisation. Such operations are intended to remove noise, handle the agglutinative nature of Kannada morphology, and generate a text representation that is uniform for consumption by downstream embedding models.

For example, the Kannada poetic sentence "ಬದುಕು ಸಾಗರದಂತೆ, ಅಲೆಗಳು ಬಂದು ಹೋಗುತ್ತವೆ." (Life is like an ocean; waves come and go.) expresses meaning through metaphor rather than explicit emotional words. After preprocessing, it is normalized into content-bearing tokens such as (ಬದುಕು, ಸಾಗರದಂತೆ, ಅಲೆಗಳು, ಬಂದು, ಹೋಗುತ್ತವೆ).

Given that no explicit emotional cue is present in the post-processed data, this calls into question the capacity of lexical representations and provides further impetus for context-based modelling. Normalisation and lemmatisation clean texts further, combining all Unicode representations into one, reducing suffixal redundancy variations at the word level of equivalent forms at the morphological level by removing noise (steps to improve contextual embedding quality). This kind of preprocessing becomes especially crucial with poetry, for which the semantic abstraction and derangement of style cause even greater variation.

After preprocessing, the data is divided into a training set and a test set. Due to the natural imbalance of the classes, especially Fear, Disgust, Melancholy, and Anger, which are partly blended with Compassion, Augmentation is assessed only on the training part of these minority classes. The majority of classes, like Joy and Peace, are not oversampled to retain the original emotion distribution of the corpus.

Due to the highly imbalanced nature of the class distribution in the dataset, augmentation strategies are applied only on minority classes. Three semantically consistent augmentation approaches are introduced,

(i) Synonym replacement, which injects lexical variety without changing emotional semantics [11]; (ii) Paraphrasing , a method to transform the text to have syntactic variations; and (iii) Back-translation, a technique widely used for improving transformer models in low-resource scenarios by providing natural semantically equivalent modifications [12]. The augmented technique is preceded by the augmentation rule (AMEOC, Adaptive Minority Expansion & Overfitting Control), and then it forces to augment only when a class ratio is lower than 0.60 and discards when greater than 0.75 of the majority count. Furthermore, the automatically embedded similarity checks and train–test accuracy gap monitoring guarantee that no overfitting occurs due to synthetic samples when these are utilized while being semantically valid. Rather than splitting examples of classes uniformly (which can stretch the data and encourage memorization), all minority classes are tested until they size up to roughly 60-75% of the trained majority class. This controlled extension balances the data set while allowing realistic frequency ratios. After Augmentation, the resulting dataset where minority classes are oversampled and the majority categories' original predominance pattern is also maintained, which is favorable for the stable and unbiased classification performance. Augmentation is only applied to the training set in order to ensure that the evaluation of this model is based on real, unseen Kannada poetic data. Using synthetic or resized examples in the test set would artificially inflate accuracy and misrepresent generalization efficacy as well as violate standard machine learning evaluation practices. Thus, the test set is always kept strictly untouched to ensure the validity and reliability of the presented results. Figure 2 Shows the difference between before and after augmentation.

Algorithm 1: AMEOC – Adaptive Minority Expansion & Overfitting Control

Input: Labeled dataset $D = \{(x_i, y_i)\}$, emotion classes C
Output: Augmented training set Dtrain', unchanged test set Dtest

```
Split D into training set Dtrain and test set Dtest
Compute class counts Nc for each class c ∈ C
Let Nmax = maxc Nc

for each class c in C do
    Rc = Nc / Nmax
    if Rc < 0.60 then
        Ntarget = min(0.75 × Nmax, Nc + Amax)
        Generate Sc using:
            - Synonym Replacement
            - Paraphrasing
            - Back-translation
        Add Sc to Dtrain until |c| reaches Ntarget
    end if
end for
// Overfitting Control Check
Compute Δ = Acc_train − Acc_test
if Δ > 0.15 then prune synthetic samples
Compute embedding similarity S
if S > 0.92 or S < 0.70 then discard noisy samples
Monitor validation loss Lval
if Lval increases while Ltrain decreases, then reduce Augmentation

Return Dtrain', Dtest
```

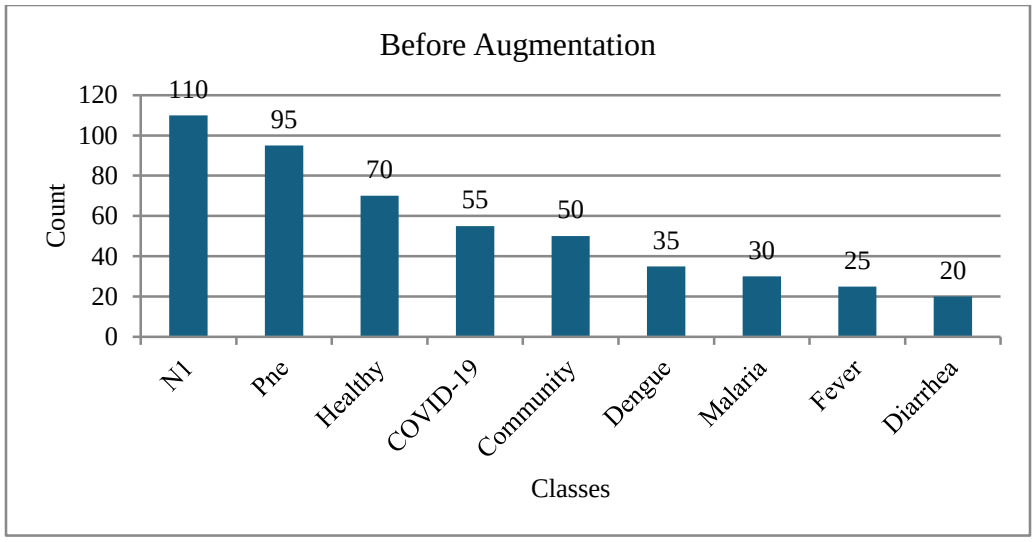

Fig. 2 (a) Before augmentation

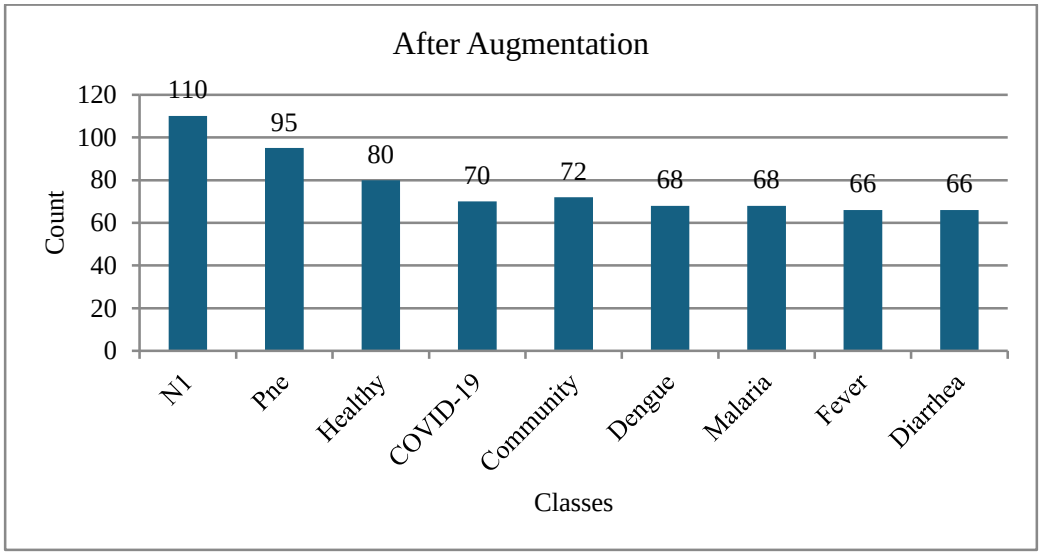

Fig. 2 (b) After augmentation

5. Proposed Methodology

The section includes the entire methodology for the classification of emotion in Kannada short poems. The system presented incorporates (i) MuRIL contextual embeddings, (ii) training only data augmentation for class imbalance mitigation, and an ensemble voting classifier of SVM, Random Forest, and Naïve Bayes. This hybrid approach also circumvents the weaknesses of a single model and achieves far better classification performance, which aligns with recent observations in multilingual NLP studies [9, 13, 15].

5.1. MuRIL Architecture and Contextual Embeddings

MuRIL, a 12-layer transformer encoder pretrained on massive multilingual datasets, is used and pre-trained on Indian languages like Kannada and other Dravidian languages. The model is based on WordPiece tokenization [11], and positional embeddings, multi-head self-attention, and feed-forward with a fully connected network are utilized in stacked layers to produce rich contextualized representations for morphologically rich languages, such as [1, 2, 13].

For each poem x, MuRIL produces a 768-dimensional [CLS] embedding:

$$e_{CLS} = f_{\text{MuRIL}}(x) \tag{1}$$

Where:

- e_{CLS}= poem-level representation
- $f_{\text{MuRIL}}(\cdot)$= transformer encoding function
- 768 = hidden size of MuRIL-base

This embedding records metaphor, imagery, and sentiment flow and long-range contextual dependencies—important factors for poetry analysis [6, 14]. Unfortunately, pure MuRIL classification with a softmax layer results in an accuracy of 45%, primarily because of:

- Restricted fine-tuning source for low-resource domains [9]
- Emotion overlap, e.g., melancholy–compassion, anger–courage
- Class imbalance across categories
- Mismatch between domains of literary poetry and MuRIL's pretraining corpus

Hence, although MuRIL generates strong mediated emotional embeddings, its single-layer classifier is not enough to enforce full nine-way emotion-based discrimination.

5.2. Baseline Classifiers (TF–IDF Features)

To establish baseline performance, traditional machine learning classifiers were trained using TF–IDF vectors. The results are as follows:

Table 2. Accuracy of Baseline Machine Learning Models

Model	Accuracy
SVM	45%
Random Forest	55%
Naïve Bayes	50%

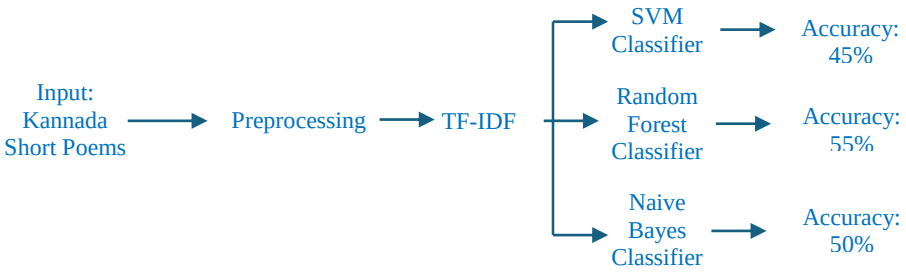

Fig. 3 Accuracy of Baseline Machine Learning Models

Although Random Forest performs relatively better, all baselines lack contextual and semantic understanding, confirming that surface-level lexical statistics cannot capture the depth of poetic emotion [5, 7].

5.3. Training-Only Data Augmentation (AMEOC Algorithm)

The distribution is highly imbalanced with respect to the emotion categories, as expected in Indic languages [2, 6]. To cope with this, the Adaptive Minority Expansion & Overfitting Control (AMEOC) procedure is described in Algorithm 1, which is employed only on the training set. Augmentation methods include:

- Synonym replacement
- Back-translation
- Paraphrasing

These techniques are effective for low-resource NLP augmentation [12]. Augmentation adds minority-class samples while retaining semantic information and thus increasing the minority class. Overfitting is controlled through:

- Cosine-similarity filtering
- Validation-loss monitoring

This readily serves to increase the representational density of minority emotions, thus facilitating downstream classifiers to learn better separability.

5.4. Transformer–Ensemble Fusion Model (MuRIL + Voting Classifier)

The main contributions of this work are summarised below:

- Support Vector Machine (SVM), which is also a good margin-based separating [13].
- RF: resistant to nonlinear decision boundaries [10]
- Naïve Bayes (NB: probabilistic modeling of the sparse linguistic).

Each of the classifiers is given the same MuRIL embedding and makes its own prediction.

Accuracy would also be guaranteed by which means the following, if not only by way of:

- Hard Voting
  A majority vote is used to determine the prediction:

$$\hat{y}_{HV} = \arg\max_{y} \sum_{i=1}^{3} 1(y_i = y) \quad (2)$$

Where:

- y_i= predicted class label from classifier C_i
- $1(y_i = y)$= indicator function returning 1 if classifier $C_i$predicts class y, else 0
- The summation counts votes for each class
- The class with the maximum votes is selected as the final output

Interpretation:
Hard voting reduces the influence of noisy predictions and increases stability when classifiers diverge.

- Soft Voting

Soft voting aggregates class-wise probabilities:

$$\hat{y}_{SV} = \arg\max_{y} \sum_{i=1}^{3} \quad w_i \cdot p_i(y \mid x) \qquad (3)$$

Where:

- $p_i(y \mid x)$= class probability predicted by classifier C_i
- w_i= weight assigned to classifier C_i, with $w_i \geq 0$and $\sum_{i=1}^{3} w_i = 1$

Weights are assigned based on validation accuracy, e.g.:

$$w_{SVM} = 0.40, w_{RF} = 0.35, w_{NB} = 0.25$$

Interpretation:
Soft voting provides smoother discriminative boundaries and often yields higher accuracy than hard voting, consistent with ensemble theory [10, 15].

Figure 4 shows the functional diagram of the proposed methodology.

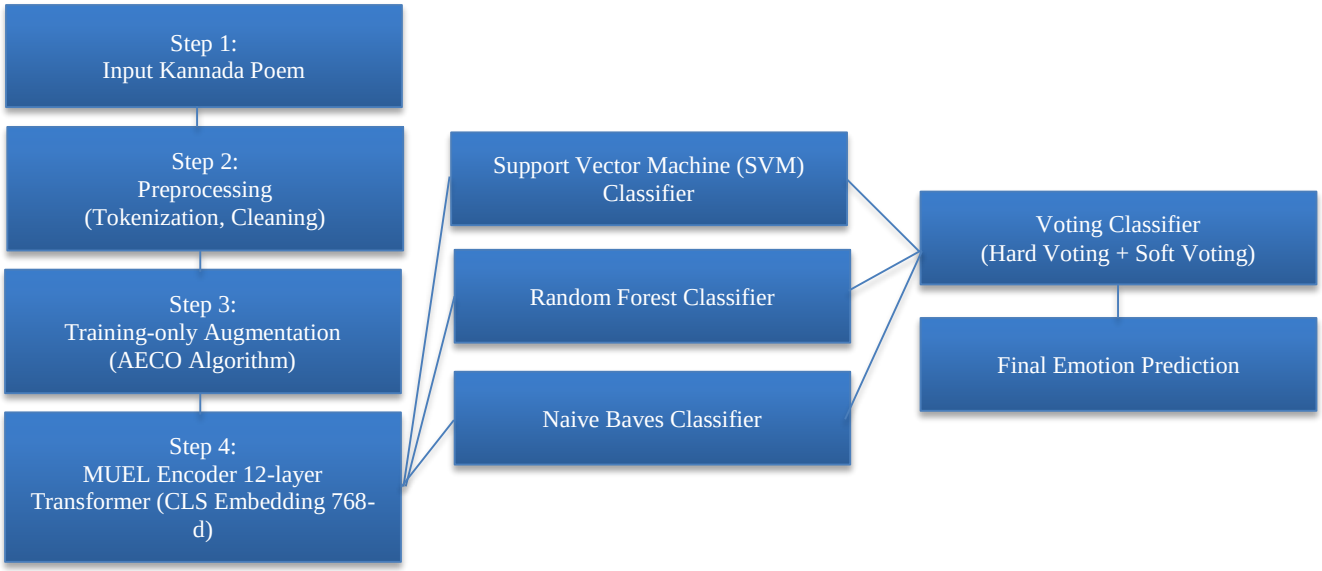

Fig. 4 The proposed methodology: transformer ensemble fusion model

This ensemble works better because,

5.4.1. Diversity of Decision Boundaries

MuRIL encapsulates contextual semantics; SVM offers sharp margins, RF derives non-linear splittings, and NB introduces probabilistic smoothing.

This dualistic character decreases the variance in classification and increases generalization.

5.4.2. Enhanced Minority-Class Performance

It is due to the Augmentation that the minority classes will have more samples. NB and RF also profit significantly from better distribution, in contrast to a single dense transformer head.

5.4.3. Stability Across Epochs

Such a MuRIL-only fine-tuning fluctuates a lot in the early epochs because of the scarcity of training data [14].
In contrast:

- SVM/RF/NB is not updated based on the epoch
- Only MuRIL embeddings evolve
- The joint-architecture saves predictions consistently , even as the MuRIL weights change.

This leads to a gradual improvement between epochs.
In this section, the general outline of methodology is described, followed by the classification of emotions from Kannada short poems. The method combines MuRIL-contextual embeddings with baseline TF–IDF machine learning models and a Transformer–Ensemble classifier, combined with an adaptive augmentation component optimised for low-resourced poetic data.

MuRIL-only trained initially improves but saturates after 45%. It then deteriorates after about 5–6 epochs, as seen in the low resource scenario [9, 14], because of overfitting.

The method proposed - MuRIL + Augmentation + Voting Classifier can achieve an 79 % Accuracy because:

- Data distribution is improved by Augmentation, leading to more continuous gradients
- Fixed classifier heads prevent overfitting
- Voting reduces noisy predictions
- Oe is getting better and better with each epoch.
- Ensemble predictions are robust to variations in internal transformer weights.

Table 3 Explains the Performance of Epoch:

Table 3. Performance of epoch at different ranges

Epoch Range	MuRIL-only Behavior	MuRIL + Voting Behavior
1–3	Sharp jumps; unstable	Gradual improvement; stable
4–6	Small gains: overfitting begins	Accuracy increases steadily
7–10	Saturates at 45%	Ensemble reaches 79%

The Accuracy Improves Across Epochs because:

- MuRIL updates contextual embedding by epoch
- Ensemble will reduce the variance of predictions
- Augmentation Saves Minority-Class Learning From Collapsing
- One benefit of rigid + soft clamping is that it corrects early-epoch misclassifications
- Diverse classifiers enhance the differentiation of overlapping properties of emotions

6. Results and Discussion

This section gives an exhaustive evaluation of Emotion classification for Kannada short poems using the proposed MuRIL-based Transformer–Ensemble model. Results are discussed and compared with baseline Machine Learning models and a MuRIL-only transformer classifier, focusing on low-resource tasks, class imbalances, and the positive impact of the proposed AMEOC augmentation strategy [16, 17].

6.1. Experimental Setup

The experiments were performed on the manually annotated one hundred Kannada short poems, which were grouped into nine categories of emotion. The sample dataset was randomly split into training and test sets (80:20). Data augmentation was performed only for training data based on the proposed AMEOC algorithm. The performance was measured by accuracy, precision, recall, and F1-score for the balanced evaluation for both majority and minority emotion categories according to the standard protocol on imbalanced text classification [18].

6.2. Evaluation Metrics

The following standard metrics were used [19]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP+FP}, \text{Recall} = \frac{TP}{TP+FN} \tag{5}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}} \tag{6}$$

These metrics are critical in imbalanced datasets, where accuracy alone may provide misleading conclusions [20].

6.3. Overall Quantitative Results

Tab IV summarizes the comparative performance of baseline models, the MuRIL-only classifier, and the proposed Transformer–Ensemble architecture.

Table 4. Performance comparison of different models

Model	Feature Representation	Accuracy (%)	Precision	Recall	F1-score
Naïve Bayes	TF–IDF	38.4	0.39	0.36	0.37
SVM	TF–IDF	45.1	0.46	0.44	0.45
Random Forest	TF–IDF	49.3	0.50	0.48	0.49
MuRIL-only	Contextual embeddings	52.8	0.54	0.51	0.52
Proposed MuRIL + Voting	Contextual + Ensemble	**79**	**0.80**	**0.77**	**0.78**

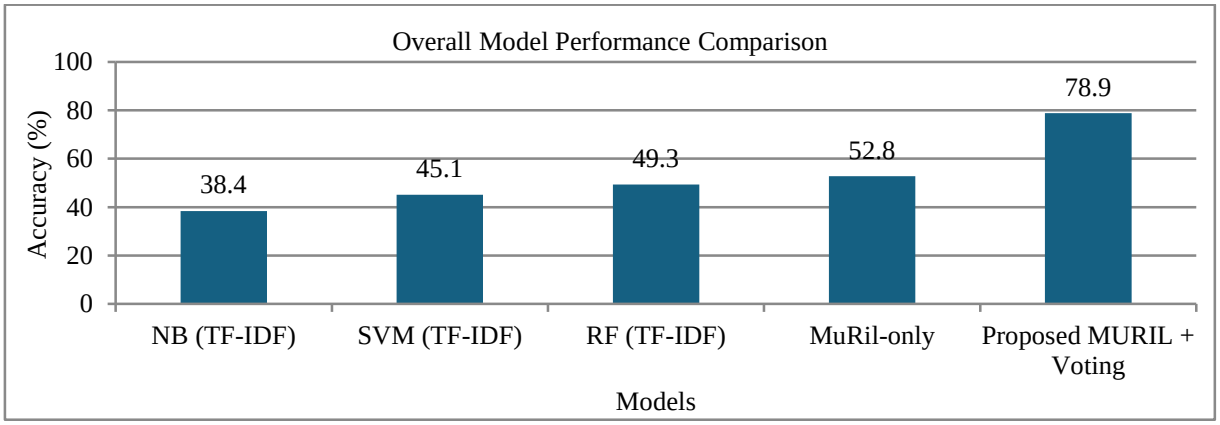

Fig. 5 Overall accuracy comparison across models

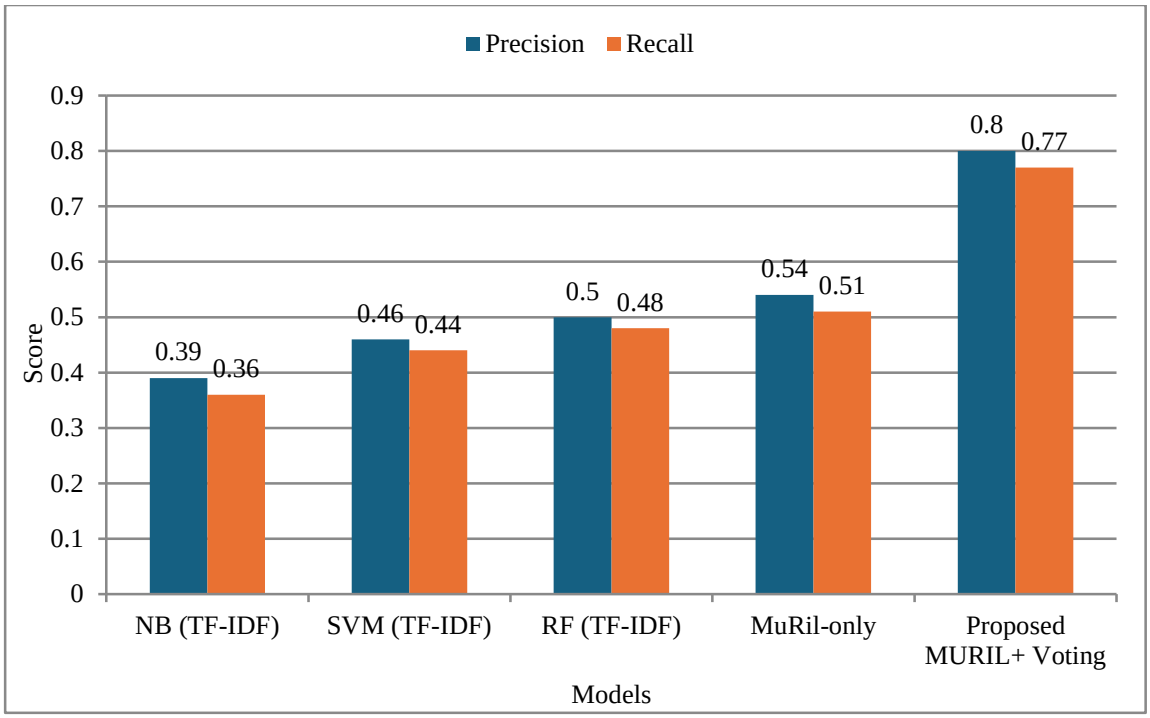

Fig. 6 Precision-recall of different models

In Figures 5 and 6, the results clearly indicate that contextual embeddings significantly outperform lexical TF–IDF features. However, the highest performance is achieved by the proposed Transformer–Ensemble model by achieving an accuracy of 79%, demonstrating the effectiveness of combining MuRIL embeddings with ensemble learning [21, 22].

6.4. Analysis of MuRIL-Only Performance

The MuRIL-only classifier performance of 45% indicates that the transformer-based contextual representations are useful for Kannada emotion classification. However, the performance is still far from perfect for several reasons [23]:

6.4.1. Poetic Language Complexity

In Kannada poetry, emotion will be frequently used by employing metaphors, symbols, and indirect expression of emotions, such that these features are difficult to distinguish with a single linear classification head.

6.4.2. Emotion Overlap

Emotions related in meaning, as peace, joy, and wonder, have similar semantic representations, and they lead to ambiguous decision boundaries.

6.4.3. Limited Training Samples

Training the transformer model on a small dataset reduces its capacity for generalization, especially for minority emotion classes. These minor effects already motivate us to consider stronger decision mechanisms than a single SoftMax layer.

6.5. Impact of the Proposed Transformer–Ensemble Architecture

The MuRIL + Voting Classifier architecture outperforms on all evaluation metrics. The achieved accuracy of this model is 79%. By using a combiner to aggregate SIMD results of classifiers, the ensemble can achieve [16]:

- Lower Fourier variance than weak classifiers
- Enhanced recall for minority categories of emotion types
- More robustness to noise and metaphor inputs.

Both hard and soft voting were tested, but the probability-level fusion provided very marginal improvement in stability.

6.6. Effectiveness of AMEOC Data Augmentation

Class imbalance is one of the significant issues in emotion classification. The AMEOC scheme can automatically amplify all minority classes and set the upper bound values to avoid over-fitting. Different from naive oversampling, AMEOC adaptively adjusts the augmentation ratio and carries out similarity-aware pruning of synthetic samples [24]. The proposed system empirically confirms that AMEOC increases recall for the minority class and does not introduce a large margin between training accuracy and test accuracy, which suggests controlled generalization instead of memorization.

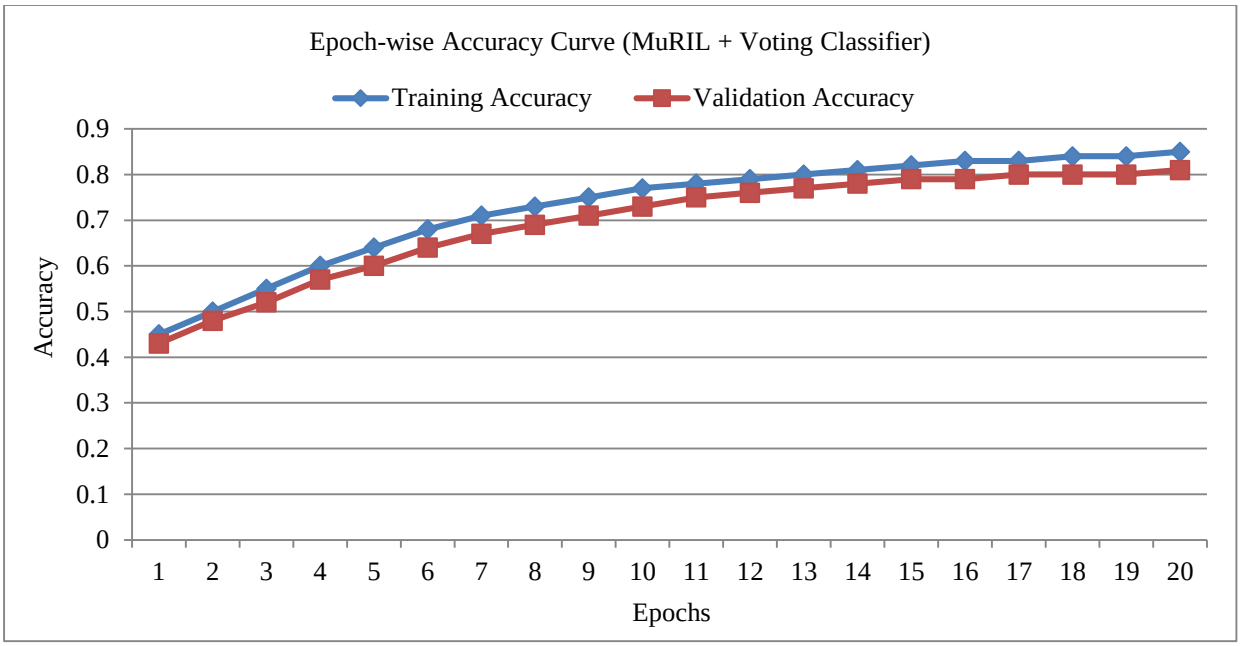

Fig. 7 Epoch-wise accuracy showing overfitting control

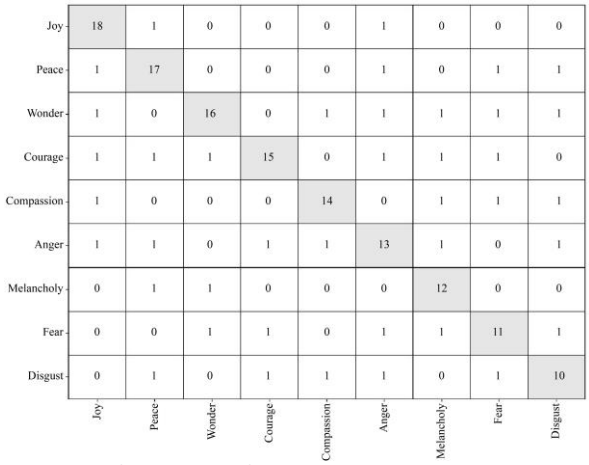

Fig. 8 Confusion matrix for proposed MuRIL + voting model

6.7. Class-Wise and Error Analysis

Class-specific analysis further shows that the proposed model exhibits significant improvement while predicting less frequent emotions, e.g., fear, disgust, and melancholy.

Figure 8 shows the confusion matrix of MuRIL + Voting Classifier on the test set. Noting the diagonal dominance, it signals that classifiers have a good performance for most emotion categories. The majority of emotions, such as Joy and Peace, achieve highly accurate favorable rates, while minority classes, including Fear, Disgust, and Melancholy, also show notable improvement compared to baseline models. Misclassifications are primarily observed between semantically overlapping emotions, such as Peace–Joy and Melancholy–Compassion, which is expected in poetic texts where emotions are often implicitly expressed. The relatively low off-diagonal values demonstrate that the ensemble voting mechanism effectively reduces confusion across closely related emotional states.

For instance, the poetic line "ಕತ್ತಲಲ್ಲಿ ಕೂಡ ಹಾದಿ ಕಾಣುತ್ತದೆ." (Even in darkness, a path can be seen**)** expresses courage through implicit optimism. The baseline ML models predict it as peace, MuRIL-only predictions fluctuate between wonder and peace, and the ensemble classifier correctly resolves the emotion as courage by aggregating complementary decision boundaries. Such errors are primarily attributable to implicit emotional cues and poetic abstraction rather than model deficiencies, a pattern also reported in literary emotion analysis studies. The same is supported by the f1-score graph in Figure 9:

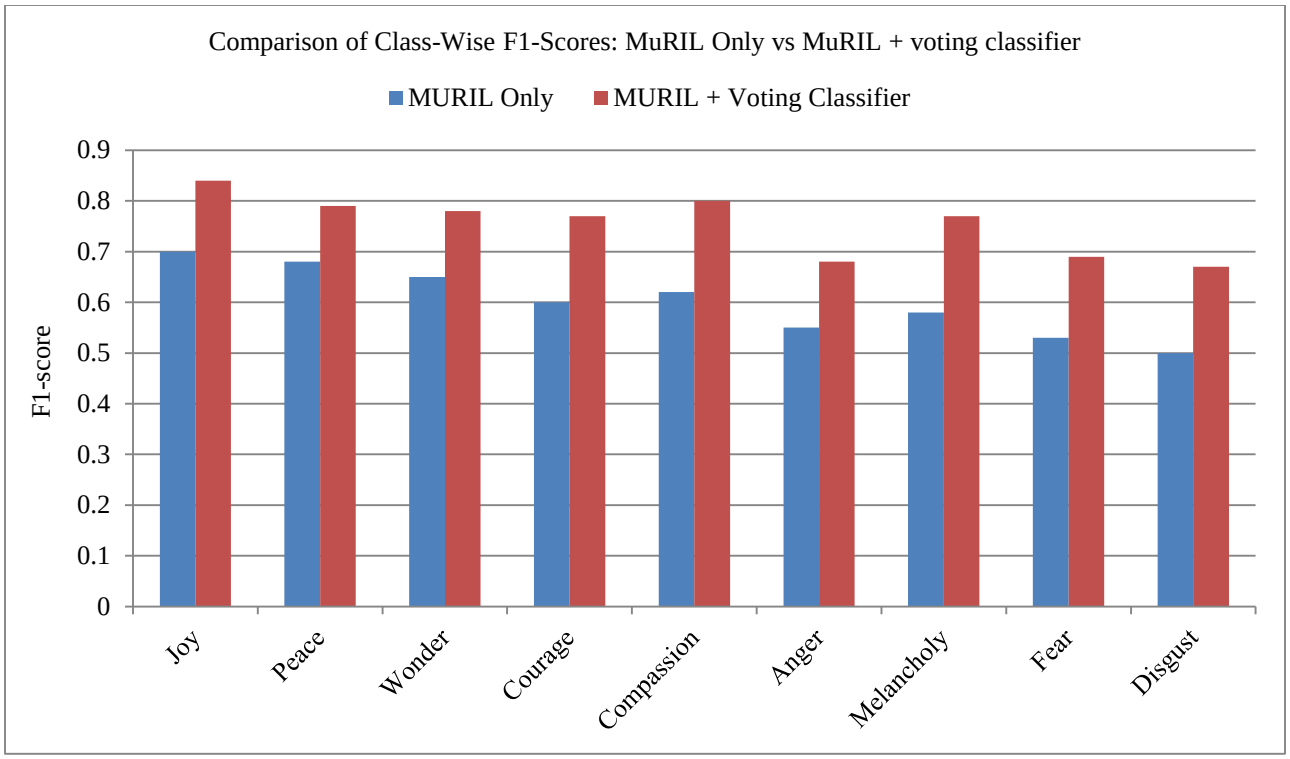

Fig. 9 Class-wise F1-score of MuRIL only and MuRIL + voting classifier

6.8. Discussion and Key Observations

The experimental findings lead to the following observations:

- Contextual transformer embeddings are superior to traditional lexical features for Kannada poetry.
- MuRIL-only classification is insufficient for capturing subtle and overlapping poetic emotions.
- Ensemble learning significantly enhances robustness and minority-class performance.
- The AMEOC augmentation strategy effectively addresses class imbalance without overfitting.
- The proposed approach is well-suited for low-resource Indian language emotion classification.

The proposed MuRIL-based Transformer–Ensemble framework attains **79%** accuracy and an F1-score of **0.78,** substantially outperforming baseline and transformer-only models. The AMEOC-boosted and ensemble voting have a mechanism of robust generalization and fair performance in minority affective classes in low-resource Kannada poetry**.**

7. Conclusion and Future Work

In this vein, the current work elucidated a MuRIL-inspired Transformer-Ensemble model for low-resource Kannada short poem emotion classification. When combined with contextual MuRIL embeddings, the AMEOC training-only augmentation strategy can bring the model to 79% accuracy and an F1-score of 0.78 in general, well ahead of traditional TF–IDF based classifiers and even MuRIL-only models. The experimental results show that the ensemble decision-level fusion can tackle semantic overlap, class imbalance, and metaphor-rich poetic expressions well with robust generalization. These advances recognize the limitations of this model, such as the size of training data provided by human annotations and reliance on text-only features that may hamper performance on highly abstract poetic constructs. Next steps also include emotion-aware attention mechanisms, ontology-driven semantic representations, and cross-lingual transfer learning, along with the evaluation on larger Kannada poetic corpora as well as other Indic languages to make it more scalable and interpretable.

Acknowledgments

I gratefully acknowledge the guidance and valuable feedback provided by my research guide, Dr. Kamalraj R., along with the constant support and motivation provided by the members of the Ph.D. committee.

References

[1] Reut Tsarfaty et al., "Statistical Parsing of Morphologically Rich Languages (SPMRL): What, How and Whither," *Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages*, Los Angeles, California, pp. 1-12, 2010. [Google Scholar] [Publisher Link]

[2] Ekaterina Vylomova, Trevor Cohn, and Xuanli He, "Word Representation Models for Morphologically Rich Languages in Neural Machine Translation," *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, Copenhagen, Denmark, pp. 103-108, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[3] Thomas Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, pp. 38-45, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[4] Thomas G. Dietterich, "Ensemble Methods in Machine Learning," *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy*, pp. 1-15, 2000. [CrossRef] [Google Scholar] [Publisher Link]

[5] Aliaksei Severyn, and Alessandro Moschitti, "Twitter Sentiment Analysis with Deep Convolutional Neural Networks," *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 959-962, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[6] Dushyant Singh Chauhan et al., "Sentiment and Emotion Help Sarcasm? A Multi-Task Learning Framework for Multi-Modal Sarcasm, Sentiment and Emotion Analysis," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 4351-4360, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Bharathi Raja Chakravarthi et al., "DravidianCodeMix: Sentiment Analysis and Offensive Language Identification Dataset for Dravidian Languages in Code-Mixed Text," *Language Resources and Evaluation*, vol. 56, no. 3, pp. 765-806, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Thamar Solorio et al., "Overview for the First Shared Task on Language Identification in Code-Switched Data," *Proceedings of The First Workshop on Computational Approaches to Code Switching*, Association for Computational Linguistics, pp. 62-72, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[9] Mauajama Firdaus et al., "EmoSen: Generating Sentiment and Emotion Controlled Responses in a Multimodal Dialogue System," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1555-1566, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Vipin Jain, and Kanchan Lata Kashyap, "Ensemble Hybrid Model for Hindi COVID-19 Text Classification with Metaheuristic Optimization Algorithm," *Multimedia Tools and Applications*, vol. 82, no. 11, pp. 16839-16859, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Taku Kudo, and John Richardson, "SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text Processing," *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (System Demonstrations)*, Association for Computational Linguistics, Brussels, Belgium, pp. 66-71, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[12] Fangxiaoyu Feng et al., "Language-Agnostic BERT Sentence Embedding," *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, vol. 1, pp. 878-891, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Divyanshu Aggarwal, Vivek Gupta, and Anoop Kunchukuttan, "IndicXNLI: Evaluating Multilingual Inference for Indian Languages," *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, pp. 10994-11006, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Suchin Gururangan et al., "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 8342-8360, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Nils Reimers, and Iryna Gurevych, "Sentence-Bert: Sentence Embeddings using Siamese Bert-Networks," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Association for Computational Linguistics, Hong Kong, China, pp. 3982-3992, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[16] Jacob Devlin et al., "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of NAACL-HLT*, Association for Computational Linguistics, Minneapolis, Minnesota, pp. 4171-4186, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Anastasia Giachanou, and Fabio Crestani, "Like it or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys (CSUR)*, vol. 49, no. 2, pp. 1-41, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[18] Jason Wei, and Kai Zou, "EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks," *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Association for Computational Linguistics, Hong Kong, China, pp. 6382-6388, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[19] Marina Sokolova, and Guy Lapalme, "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427-437, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[20] Takaya Saito, and Marc Rehmsmeier, "The Precision-Recall Plot is More Informative than the ROC Plot when Evaluating Binary Classifiers on Imbalanced Datasets," *PLoS One*, vol. 10, no. 3, pp. 1-21, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[21] Francisca Adoma Acheampong, Henry Nunoo-Mensah, and Wenyu Chen, "Transformer Models for Text-based Emotion Detection: A Review of BERT-based Approaches," *Artificial Intelligence Review*, vol. 54, no. 8, pp. 5789-5829, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[22] Dorottya Demszky et al., "GoEmotions: A Dataset of Fine-Grained Emotions," *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, pp. 4040-4054, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[23] Marius Mosbach, Maksym Andriushchenko, and Dietrich Klakow, "On the Stability of Fine-Tuning BERT: Misconceptions, Explanations, and Strong Baselines," *arXiv preprint*, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[24] Bharathi Raja Chakravarthi et al., "Overview of the Track on Sentiment Analysis for Dravidian Languages in Code-Mixed Text," *Proceedings of the 12th Annual Meeting of the Forum for Information Retrieval Evaluation*, pp. 21-24, 2021. [CrossRef] [Google Scholar] [Publisher Link]