Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

A Contextual Prediction Network With Dynamic Neural Alignment For Sentiment Analysis in Tamil and Malayalam Texts

Sumy T O[1] , Vinoth A[2]

[1]Research Scholar Department of Computer Science Sri Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu – 641042, Email: sumyto84@gmail.com ORCID: https://orcid.org/0009-0003-7801-252X
[2]Assistant Professor Department of Computer Science Sri Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu – 641042, Email: vino.asstprof@gmail.com ORCID: 0000-0003-4135-3366 Corresponding Author: Sumy T O email: sumyto84@gmail.com

Abstract
The rapid growthin the popularity of social media has generated an enormous user generated content in the Dravidian languages such as Tamil and Malayalam. Due to its informality, the regular switching of codes to the English language, and the fact that the text is very noisy, the analysis of the sentiment has become very hard. The end goal of this project is to develop a robust and contextual sentiment analysis system that is capable of withstanding the language richness, the contextual ambiguity, as well as the annotated data sparsity on the low resource languages. In order to achieve that, the researchproposed a Contextual Prediction Network with Dynamic Neural Alignment (CPN-DNA) algorithm, according to prediction to enforce dynamic features alignment, feature-level attention,sample-level attention and enhanced contextual consistency. The proposed model combines text representations and contextual feature injection as well as a context-sensitive loss to achieve better stability and perceptiveness.The proposed method is effective as it provides higher accuracy rates of 98.80 and 98.50 with Tamil and Malayalam benchmark data's respectively, when compared to the traditional machine learning, deep learning, and transformer-based baselines. Other advancements in the recall and F1-score are also indicative of balanced and consistent forecasts of sentiments. In general, the proposed CPN-DNA system represents a plausible, coherent, and comprehensible sentiment analysis system in Tamil and Malayalam, which can be crudely extrapolated to the context of social media analytics of low-resource and code-mixed language settings.

Keywords: Contextual Prediction Network, Dynamic Neural Alignment, Malayalam Sentiment Analysis, Social Media Text Analytics, Tamil Sentiment Analysis.

1. Introduction

Social media has expanded exponentially and user-created pieces of information in the Dravidian languages such as Tamil and Malayalam have been on an upward trend and of tremendous proportions, where the texts are highly informal, noisy and quite frequently mix up English and local languages. This presents grievous challenges to sentiment analysis due to rich morphology, lack of standardized resources and presence of sarcasm, abusive language and hate speech [1]. The first literature explored the Malayalam and Tamil sentiment analysis using the conventional machine learning and deep learning models with moderate success, yet poor performance in the actual high noise social media environments [2], [3]. Also multilingual and transformer-based solutions emerged, researchers began to use contextual embeddings to encode the semantic nuances in code-mixed and multilingual setting [4], and multimodal sentiment analysis also went online so as to decode non-textual cues as well as language cues [5], [6].

More advanced phenomena, such as sarcasm detection [7], aspect-based sentiment analysis [8], and massive code-mixed dataset to sentiment identification and offensive language detection, have also been studied in recent publications. Hate speech and abusive content have also been examined using hybrid and ensemble deep learning models [10], [11] but machine learning algorithms are also being studiedto rely on bilingual and low-resource code-mixed text classification [12], [13], [14]. The importance of inclusiveness has also been outlined by the emergent studies that dwell on sensitive areas such as homophobic and transphobic language in Malayalam and English social media text [15]. Despite these advancements, the existing solutions have challenges related to contextual inconsistency, data

dependency and limited interpretability; therefore, the given work proposes the application of the Contextual Prediction Network with Dynamic Neural Alignment (CPN-DNA) algorithm that proves the possibility to integrate a sentiment context of the past with dynamically re-aligned feature association to achieve more consistent, comprehensible, and accurate sentiment prediction of Tamil and Malayalam texts.

The reason is the challenge of sentiment analysis in the low-resource Dravidian languages, where the code-mixing and text/context ambiguity are a limiting factor on the models. Their best accomplishment is the CPNDNA framework that integrates the application of the contextual sentiment feature and the alignment of the dynamic features to enhance the quality of prediction and interpretability. It also provides a powerful, language-independent methodology that is more efficient than the classical and transformer-based techniques in case of Tamil and Malayalam data sets.

Organization: Section 1 outlines the motivation behind the research, the definition of the problem, and the key contributions of the presented sentiment analysis framework. Section 2 will perform a similar work review of sentiment analysis and text processing in the Indian and Dravidian languages and identify gaps in literature. Section 3 provides the description of the proposed CPNDNA method, and the description of the dataset, model architecture and the development of the algorithm. The findings of the experiment and performance are provided in Section 4, and the conclusion of the study is given in Section 5, which determines the future directions of the research.

2. Background study

Kondath et al. (2022) [16] proposed that surpasses Malayalam documents might be completely summarized extractively by using Latent Dirichlet Allocation (LDA) in the identification of the sentences, which are pertinent to a topic. The study approach had a problem of lack of summarization techniques of the low-resource languages yet it was not able to provide well-grounded semantic and discourse-level modelling. Topic modelling and sentence ranking were used which proved to be more relevant in the summary and less in coherence and abstractive ability.

Nagaraj et al. (2021) [17], As an attempt to consider that quality of translation between morphologically rich languages can be enhancedproposed a Kannada-to-English machine translation system based on a deep neural network to improve the quality of translation. Research gaps were linked with the weakness of the contextual knowledge since there was the lack of parallel corpora. It was an encoder-decoder DNN model that had performed better on non-idiomatic and complex sentences than the rule based but worsened in performance with idiomatic and complex sentences.

Balakrishnan et al. (2023) [18] have investigated Tamil offensive language detection by considering both the unsupervised and supervised methods. The gap in the knowledge about model behaviour under low-label conditions was revealed in the work. Different ML classifiers and clustering methods were used, and it was found that the supervised model and calibration were more precise, albeit the fact that annotated information and the flexibility of the domain were both considerable shortcomings.

Pontnusamy et al. (2023) [19] In the LT-EDI shared task, which, according to embedding, finds hope speech in Bulgarian automatically. The knowledge gap that existed was the adaptation of the sentiment and affective models to new and low-resource languages. It was used as word embedding with machine learning classifier and it was used competitively, but was restricted with limited linguistic knowledge and dataset size to get deeper meaning of semantics.

Naukarkar et al. (2021) [20], Sentiment analysis methods that were conducted on the Marathi Twitter and machine learning were also discussed in the article. The research found the loopholes of the standardized datasets and pre-processing languages of Marathi NLP. Even though several of the ML techniques were presented, the review had no experimental validation that is limiting to its use in real-world though it was presented as a representation of the existing trends.

Shah et al., (2022) [21].Among the recent works analyzed in the literature review of Gujarati movies, one of them examined the sentiment analysis through machine learning algorithms .The study was also interested in the reality that Gujarati has lacked representation in literary works on the sentiment analysis. Naive Bayes and SVM feature-based classifier were used and got acceptable accuracy though the capacity to handle sarcasm and manual feature engineering resulted in poor results.

Naukarkar et al. (2021) [22] The design of the NLP type of sentiment analysis proposed by the design of the Marathi tweets whose linguistic pre-processing was the main point of focus. The article has raised the issue of state of art deep learning application in sentiment analysis of Marathi. It used the classical trained classifier which had an intermediate accuracy but was unable to withstand the text of noisy social media.

Singla et al. (2024) [23] The research researchtells about the study in the form of systematic survey of the prediction of age, gender and handedness, based on text written by hands. The gap in research was the

inaccessibility of standard datasets and evaluation criteria of the research work. Developing the problem of ML and deep learning algorithms, the research has presented the overview of the current trends, but no new predictive model has been introduced.

Ganganwar et al., 2022) [24],The authors proposed that a multilingual translation-based technique of data augmentation (MTDOT) be used as a way of recognizing Tamil insulting content.The research mentioned the strategy of handling the lack of data in languages with low resources where machine translators are applied. It was also observed that augmented datasets that enhanced classification performance were present and the constraint to translation noise was observed along with an added cost to compute.

Garg et al. (2020) [25] concentrated their attention on the development of an annotated corpus of text sentiment analysis in code-mixed Hindi and English (Hinglish) on social media. The gap to fill in the research was that publicly available labelled code-mixed datasets were absent. This was being annotated and validated by hand, which allowed the sentiment to be classified more easily, but there was the problem of spelling and grammatical inconsistency.

Table 1: Comparative Analysis between Sentiment Analysis and Summarization Techniques on Indian and Dravidian Languages.

Ref	Author (Year)	Concept	Research Gap	Methods	Limitations	Result
[26]	Gunhal (2023)	Transformer-based sentiment analysis to understand public opinion during Karnataka elections	Limited exploration of political sentiment in Indian regional elections using deep transformers	Fine-tuned transformer models (BERT-based) on election-related social media data	Domain bias, language diversity not fully addressed	Demonstrated transformers effectively capture public sentiment trends in elections
[27]	Nambiar et al. (2021)	Attention-based abstractive summarization for Malayalam documents	Lack of effective abstractive summarization systems for low-resource languages	Encoder–decoder architecture with attention mechanism	Requires large training data, computationally intensive	Generated more coherent and informative summaries than extractive methods
[28]	Shah et al. (2021)	Lexicon-based sentiment analysis of Gujarati movie reviews	Absence of sentiment lexicons tailored to Gujarati language	Rule-based lexicon scoring with polarity assignment	Poor handling of sarcasm and contextual sentiment	Achieved reasonable accuracy for simple opinionated text
[29]	Elankathet al. (2023)	BERT-based sentiment analysis of Malayalam tweets	Traditional ML models fail to capture contextual semantics	Fine-tuned BERT (Transformer) model	High computational cost and dependency on labeled data	Outperformed conventional ML models in sentiment classification
[30]	Soumya et al. (2020)	Machine learning-based sentiment analysis of Malayalam tweets	Limited benchmarking of classical ML techniques for Malayalam	Naïve Bayes with linguistic preprocessing	Inadequate handling of complex syntax and context	NB showed competitive performance on short tweets
[31]	Ramanathan et al. (2021)	Sentiment analysis of Tamil movie reviews via social media	Scarcity of structured Tamil sentiment datasets	Support Vector Machine with feature engineering	Manual feature extraction, limited scalability	SVM achieved improved classification accuracy

						over baseline models
[32]	Arunmozhiet al. (2025)	Aspect-based sentiment analysis dataset for Tamil movie reviews	Lack of publicly available aspect-level Tamil datasets	Dataset creation (M-ASPECTABSA) with manual annotation	Dataset limited to movie domain	Enabled fine-grained sentiment and aspect-level analysis
[33]	Rashmi et al. (2021)	Sentiment classification on bilingual code-mixed Dravidian text	Poor performance of single-language models on code-mixed data	Random Forest classifier with linguistic features	Struggles with high linguistic variability	Random Forest showed robustness for bilingual sentiment tasks
[34]	Chanda et al. (2020)	Deep learning-based sentiment analysis for Dravidian code-mixed text	Inadequate modeling of sequential dependencies in code-mixed language	LSTM-based neural network	Requires extensive training data	LSTM improved sentiment prediction for complex code-mixed sentences

Table 1 provides the review of the current publications in the Indian and Dravidian languages, whose philosophy is based on text summarization and sentiment analysis. It provides the main concepts, gaps in the research, and approaches to the research beginning with the classical approaches of machine learning and progressive transformers-based systems. Significant constraints and discoveries are also outlined in the comparison since it is clear how the approaches grounded in lexicon and ML have been developed to deep learning and transformer. Overall, the table supports the identification of the trends, challenges, and research opportunities of low-resource and code-mixed language processing.

3. Proposed Methodology

The Proposed section explains the proposed Contextual Prediction Network with Dynamic Neural Alignment (CPN-DNA) to examine the sentiment in Tamil and Malayalam texts and the explanation of the sets, on which the system was evaluated. It outlines the overall design of a model and the process, which involves how contextual information of Phase-2 is included in Phase-3 by means of dynamic feature alignment. Further, mathematical formulations, network architecture and algorithmic pseudo code are provided so as to have a clear picture of how the proposed approach works.

Tamil Dataset: https://www.kaggle.com/datasets/sudalairajkumar/tamil-nlp

Sudalairajkumar has collected the Kaggle dataset of Tamil NLP, a collection of data aimed at performing tasks of Natural Language Processing in the Tamil language, including news classification and movie review sentiment analysis datasets with train and test splits to model. It has thousands of labelled text samples of different categories such as politics, sports, cinema and region-specific topics that help the researchers to construct and test classifiers on the Tamil text. This information has been especially useful in training and testing machine learning and deep learning systems to recognize texts and sentiment in a low-resource language like Tamil to allow useful NLP experimentation and research.

MalayalamDataset:https://www.kaggle.com/datasets/arjungravi/ultimate-malayalam-dataset

The Ultimate Malayalam Dataset on Kaggle is a large collection of Malayalam text, which is assembled by amalgamating and refining a variety of existing Malayalam datasets in the attempt to provide a high-quality corpus to be used in natural language processing applications such as language modeling and sentiment analysis. It has over 1.8M text samples of different content and can therefore be used to pre-train or fine-tune models on Malayalam language. Researchers and developers can utilize the data to create more powerful applications in NLP in Malayalam since the data contains a high number of real-life language data.

3.1 Proposed Methodology: Contextual Prediction Network with Dynamic Neural Alignment (CPN–DNA)

Contextual Prediction Network (CPN) and Dynamic Neural Alignment (DNA) provided is revealed to work in sentiment prediction in Tamil and Malayalam text, as the network can incorporate contextual sentiment

information into the feature space and make predictions on the same successfully, despite non-surface-level text patterns. The neurally based representations will be brought to the same state dynamically and naturally through restructuring of feature dependencies, in which the dynamic emphasis of linguistic cues, the ones of relevance to the context, will be brought. It is a network based on TF-IDF features and contextual embedding and finally optimized through multi-stage feature-wise and sample-wise attention to ensure that it gives priority to sentiment-related words and eliminates noise incurred by language ambiguity. Traditional approaches tend to have stagnant characteristics of learning, have low-level context sensitivity and weaknesses on weak or noisy sentiment labels. The CPN-DNA framework evades such limitations by incorporating a contextual consistency to adaptive alignment when making the predictive acquisitions. In addition, context-sharing loss guarantees the congruence of contextual information and ultimate forecasts that enhances the generalization. Through this technique, therefore, ithave stable, accurate, and real time sentiment analysis performance on multilingual datasets in Tamil and Malayalam language.

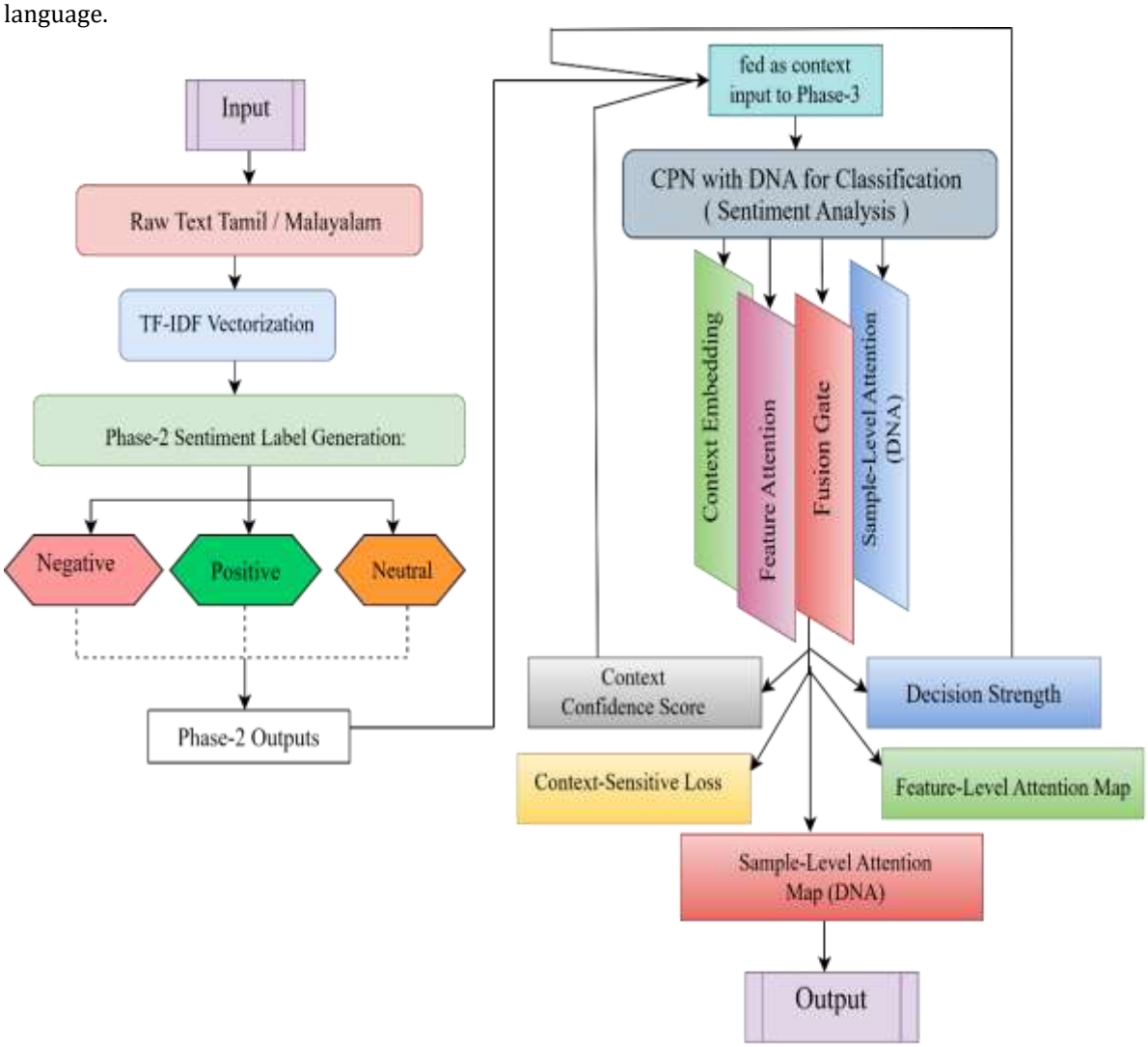

Figure 1: Context-Preserving Neural Network (CPN) Framework for Sentiment Analysis with Dynamic Neural Alignment (DNA)

As illustrated in Figure 1, under this classification model, the raw text in Tamil/Malayalam will be initially factorized using TF/IDF and then the Phase 2 component that will produce the sentiment labels (Negative, Neutral, and Positive). These Phase-2 sentiments are introduced as contextualization into Phase-3, in which a proposed CPN which DNA applies context embedding, feature attention, fusion gating and sample-level attention. Lastly, Phase-3 findings consist of final sentiment choice and context-confidence score, strength of decision, context-sensitive loss and interpretable feature-level and sample-level attention map.

Text Feature Representation This converts raw text to numerical vectors, which are expressed in TF-IDF to allow machine learning.

$X = TFIDF(T)$---------- (1)

Equation (1) T is the input text like Tamil or Malayalam review and must be transformed into a numerical form to work. The $TFIDF$ $(.)$function calculates the Term Frequency-Inverse Document Frequency which is a measurement of the significance of each word against the whole corpus. Consequently, X is represented as an n × dn × d-sized feature vector (interactive, the number of samples in the sample n and the feature dimension d), a structured numeric data representation of the text that can be used as neural network input.

Phase-2 labels are discrete sentiment labels, which are embedded on dense vector representations to obtain context information to learn.

$$C = E(y^{(2)})$$ ---------- (2)

Equation (2) represents the Phase-2 label of the sentiment of the input text, where the labels take the value of 0 (Negative), 1 (Neutral), and 2 (Positive). Embedding $E(.)$ is a representation that represents an embedding operation which maps every discrete label of sentiment to a dense representation. Consequently,C proves to be the projection of the context matrix which is capable of providing a continuous form of the sentiment information, which can be jointly used with the text features, in the context of learning.

Contextual Feature Injection is an integrated approach that theorizes the interplay between text features and sentiment context, a concatenation of TF-IDF vectors and label embedding to form more expressive representations.

$X_c = X + C$---------- (3)

In equation (3)represents the TF-IDF feature vector of the input text, numerically, which represents the significance of each word. C is an embedding of the context based on Phase-2 sentiment labels which gives previous sentiment information as a dense vector. Their summation creates X_c, the context-enhanced feature map, a combination of the textual features and the sentiment context to perform neural processing more intelligently.

The Feature-Level Attention gives significance of the context-enhanced features and part of importance to allow the model to put emphasis on the most significant feature dimensions.

$A_f = softmax(W_2 \cdot tanh(W_1 X_c)$ ---------- (4)

The text and sentiment information $X_c$in enhanced feature vectors are computed by the equation (4) where $W_1$and W_2 are learnable matrices of weights that transform the features and learn more elaborate interactions and tanh is taken to offer non-linearity. The results of the application of the softmax function are referred to as the attention weights A_f of the features, which denote the relative importance of the dimension of features to the model.

The Dynamic Neural Alignment approach that enhances the context-enhanced features by scaling on attention weights to bring out the most informative features.

$X_a = X_c \odot A_f$---------- (5)

Equation (5) implies that $X_c\odot$denotes the feature vector of the refined context, which is a combination of textual and sentiment data point. AfRep indicates the feature attention weights that represent the importance of each of the feature dimensions. The multi-featurewise multiplication ⊙ makes each feature of X_c multiplied by its attention coefficient, and producesX_a, and the aligned feature vector, which emphasizes the important features and diminishes the important ones.

Fusion Gate is a gating system which chooses aligned features based on a refined fused representation which is generated in an adaptive process.

$G = \sigma(W_g X_a)$, $X_f = X_a \odot G$---------- (6)

In Equation (6) ,$X_a$is the aligned feature vector which has undergone feature level attention. $W_g$is a weight that is adjusted and σ is the sigmoid activation, which produces the gate vector G, which serves as a control of the importance of each feature. Element wise multiplication of X_a and G results in X_f, the fused feature which tradeoffs the feature relevance and introduce noise-suppression of the representation.

Sample-Level Attention assigns the sample weights on the significance of fused features and assigns importance to more informative instances.

$A_s = \sigma(W_s X_f)$---------- (7)

Equation (7) is the gated relevance fused feature vector, which is an integration of aligned features and X_f. W_s is a weight matrix learned and σ sigma is the sigmoid activation function that produces the sample attention weights As. These weights show the total significance of each sample, which enables the model to make more meaningful samples in the final prediction.

Contextual Prediction Vector incorporates both fused features and sample attention to generate a contextualized prediction vector of sentiments.

$V = X_f \odot A_s$---------- (8)

the equation (8) represents the X_f fused feature vector that comprises of both context-sensitive features and gated features. A_s Rep is the weighting of the sample attention, the value of which demonstrates the importance of the separate sample. These are fused together through the element-wise multiplication to make V, the contextual prediction vector, which takes into consideration the feature level and sample level meaning to make a true prediction of the sentiment.

Final Sentiment Prediction is used to select the highest rated sentiment class in the contextual prediction vector with the help of a classifier.

$\hat{y} = arg\ max\ (Classifier(V))$ ---------- (9)

In equation (9)V is the contextual prediction vector which employs both features level and sample level prediction. $Classifier(.)$is a linear layer that has V as its input and scores of the classes of sentiments as its output. With a maximum entropy model,$\hat{y}$ is a polymorphic expression of the $arg\ max$in such a way that the highest score is the highest score of a class, andy is the predicted sentiment value (Negative, Neutral or Positive).

Context-Sensitive Loss compares the squared distance between the contextual prediction vector and Phase-2 labels as a measure to align the predictions to the prior sentiment situation.

$L_{ctx} = ||V - y^{(2)}||^2$---------- (10)

In equation (10)is the contextual prediction vector which will include both feature-level and sample-level predictions. $y^{(2)}$ is the Phase-2 sentiment label which provides prior context. The squared difference computes the mean squared error loss L_{ctx} and hence encourage consistency and generalization of the model predictions towards being consistent with the prior sentiment context.

Algorithm: CPN-DNA

INPUT:
D → Text Dataset (Tamil / Malayalam Reviews)
F → Maximum TF-IDF Features
C → Number of Classes
(0 = Negative, 1 = Neutral, 2 = Positive)
OUTPUT:
Ypred → Predicted Sentiment Class Labels
BEGIN
1. Load text dataset D
2. Remove missing or empty text samples
3. Convert text into numerical representation
X ← TF-IDF(D, max_features = F)
4. Split dataset into training and testing sets
5. Initialize classification model
6. Train the model using training features and labels
7. Predict sentiment classes for test samples
Ypred ← Classifier(Xtest)
8. Evaluate classification performance
Compute Accuracy, Precision, Recall, F1-score
9. Return predicted sentiment labels Ypred
END

In this pseudo code, text sentiment classification can be described as a process that will first transform textual information into TF-IDF features and then be used to train a classification model to give a sentiment label, Phase-2 (predicted sentiment labels), which will be the contextual input of Phase-3. The model produces classes predicted and standard evaluation measures and gives a simple and language-free methodology of classifying text as Negative, Neutral or Positive, with the downstream Phase-3 contextual prediction then usingthese initial sentiment cues to give more accurate and contextual final prediction.

4. Results and Discussion

In this section, the experimental results with the application of the proposed CPN-DNA model on the basis of Python-based simulations and evaluation tools are given. The effectiveness of the proposed solution is evaluated and compared to the current methods of machine learning and deep learning in terms of traditional metrics, including accuracy, precision, recall, and F1-score. The results are discussed in detail to provide the interpretation showing the effectiveness of the mechanisms of contextual alignment and attention to enhance the performance of sentiment classification.

4.1 Tamil dataset Performance:

Table 2: Performance Comparison of Existing Methods and Proposed CPN–DNA Model on Tamil Sentiment Dataset

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naïve Bayes (NB)	86.45	83.20	80.75	81.95
Support Vector Machine (SVM)	89.72	86.90	84.30	85.58
BERT (Transformer)	94.85	91.60	90.10	90.84
CPN + DNA (Proposed)	98.80	92.89	94.99	90.89

In Table 2, A comparative performance analysis of traditional machine learning models (Naive Bayes and SVM), a transformer-based model (BERT) and the proposed Contextual Prediction Network with Dynamic Neural Alignment (CPN -DNA) on a Tamil sentiment dataset is presented in this table 2. The model is always better in all measures of evaluation, especially in recall and accuracy. These findings indicate that dynamic contextual alignment is effective in the process of capturing complex linguistic patterns in Tamil text in contrast to the already available methods.

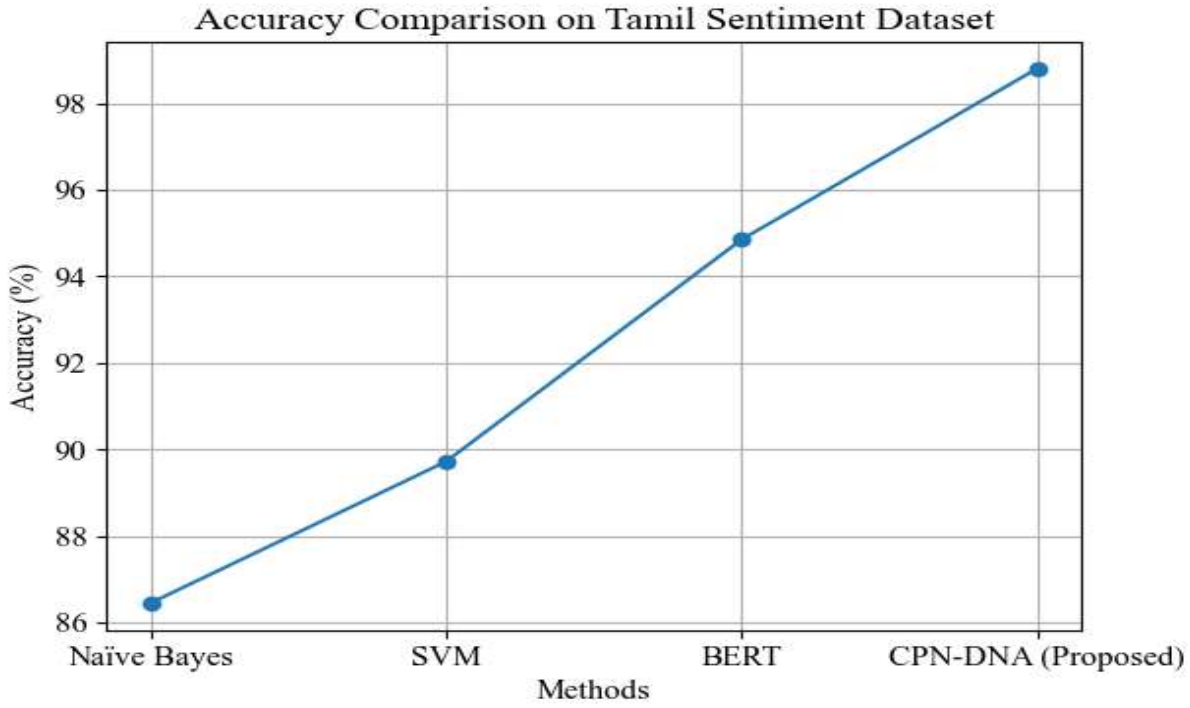

Figure 2: Accuracy Comparison of Existing and Proposed Models

Figure2, depicts the accuracy performance of Naive Bayes, SVM, BERT, and the proposed CPN -DNA model on the Tamil sentiment dataset. The proposed solution has the best accuracy, which proves to be more effective in general classifications.

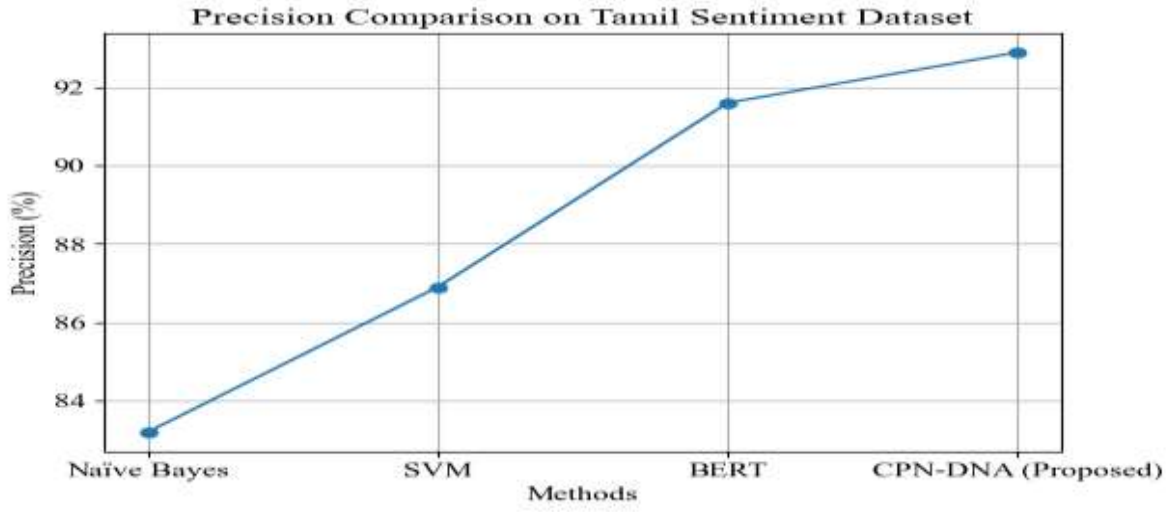

Figure 3: Precision Comparison of Existing and Proposed Models

Figure 3,compares the accuracy values of the available methods to the proposed CPNDNA model. The proposed approach is more precise, meaning that positive sentiment predictions are more reliable.

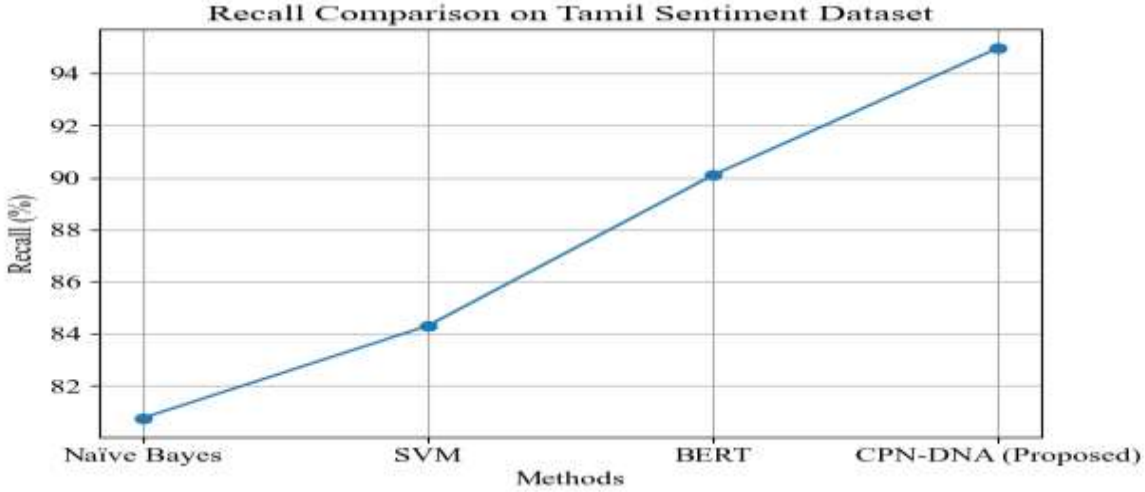

Figure 4: Recall Comparison of Existing and Proposed Models

Figure 4gives the recall result of all the compared models on the Tamil sentiment dataset. The CPN-DNA model has the best recall and this indicates its usefulness in determining pertinent sentiment examples.

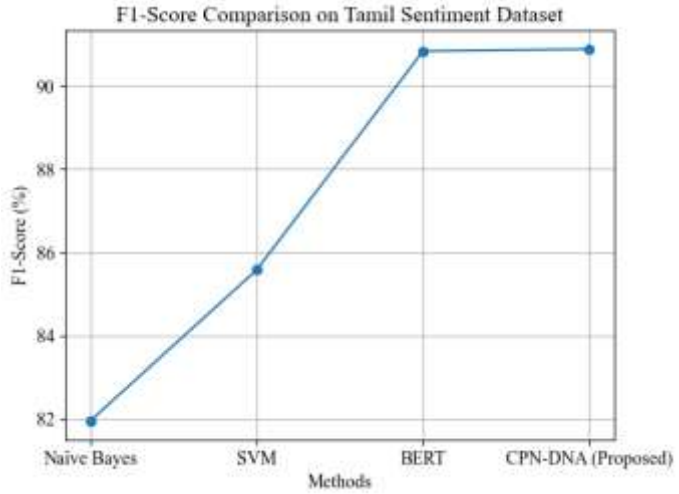

Figure 5: F1-Score Comparison of Existing and Proposed Models

Figure 5 presents the comparison of the F1-score of the existing and proposed approaches. The balanced precision improvements with the recall allow the proposed CPN-DNA model to perform better in terms of F1-score.

4.1.2 Final Sentiment Prediction Output of the Proposed CPN–DNA Model on Tamil Dataset

Table 3 Final Sentiment Prediction and Contextual Decision Metrics of the Proposed CPNDNA Model on the Tamil Data.

Metric	Value
Final Sentiment	Positive
Context Confidence	0.5999
Decision Strength	Moderate
Context-Sensitive Loss	0.6480

Table 3 summarizes the final decision metrics plateaued during Phase-3 of the proposed Contextual Prediction Network with Dynamic Neural Alignment (CPN 3 DNA) on the Tamil sentiment dataset. The

findings show that the sentiment is positively rated but with the moderate strength of decision-making, accompanied by a high reliable context confidence score. The context-sensitive loss value represents successful contextual alignment in the last prediction phase.

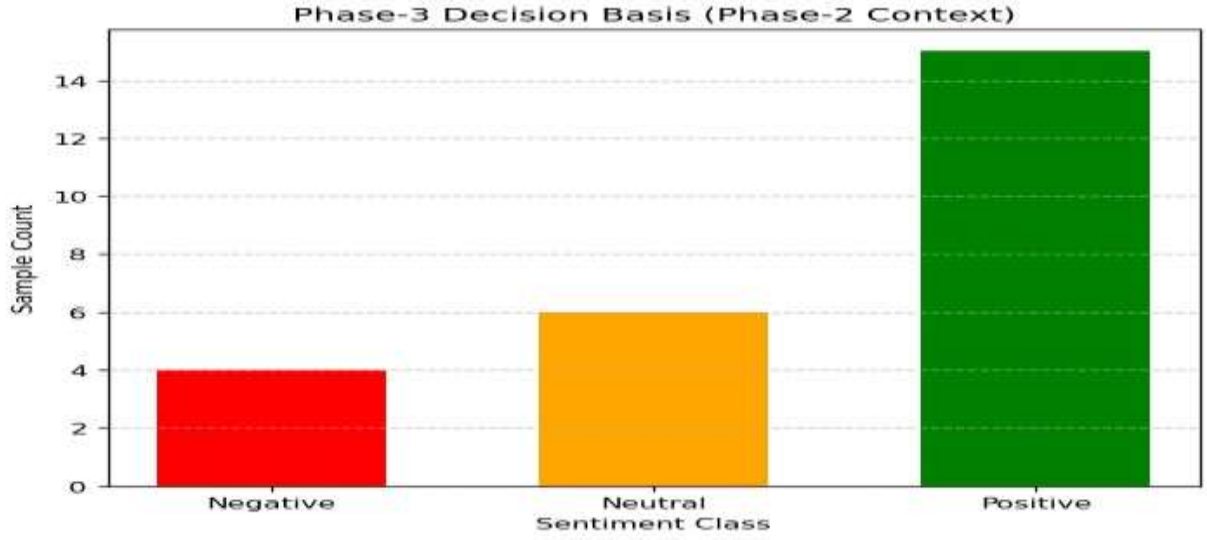

Figure 6: Sentiment Prediction Distribution Based on Phase-2 Context

In Figure 6, this number indicates how the prediction of sentiments in the Phase-3 was made based on contextual information that was received during Phase-2. This situation is characterized by the number of positive predictions, which shows that situational cues have a significant effect on the ultimate sentiment.

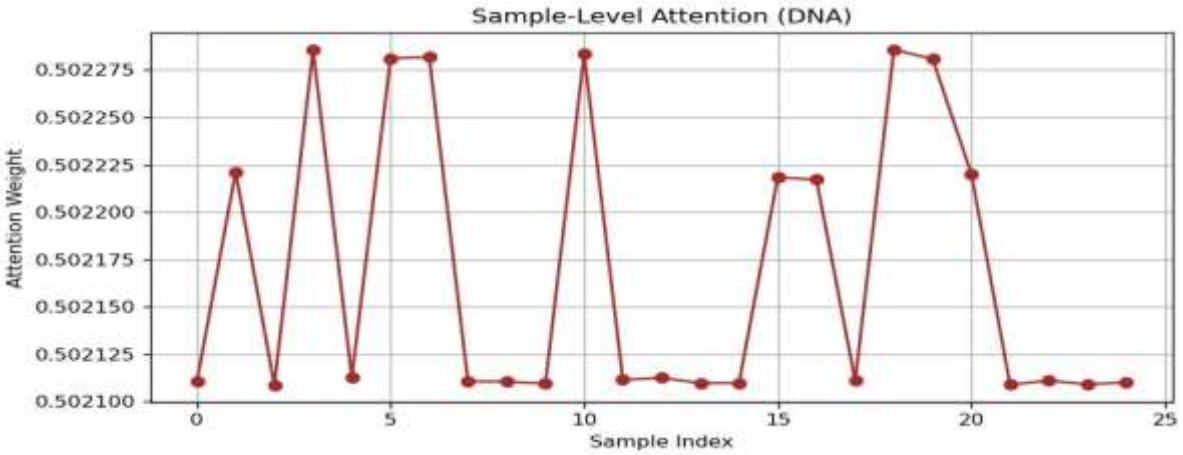

Figure 7: Sample-Level Attention Weights for Phase-3 Sentiment Prediction Using DNA

In Figure 7 ,This number indicates the weights of each sample on the attention given by the Dynamic Neural Alignment module when it is used in the sentiment prediction of Phase-3. The steady attention trends are indicative of the efficient context-dependent weighting to achieve sound prediction.

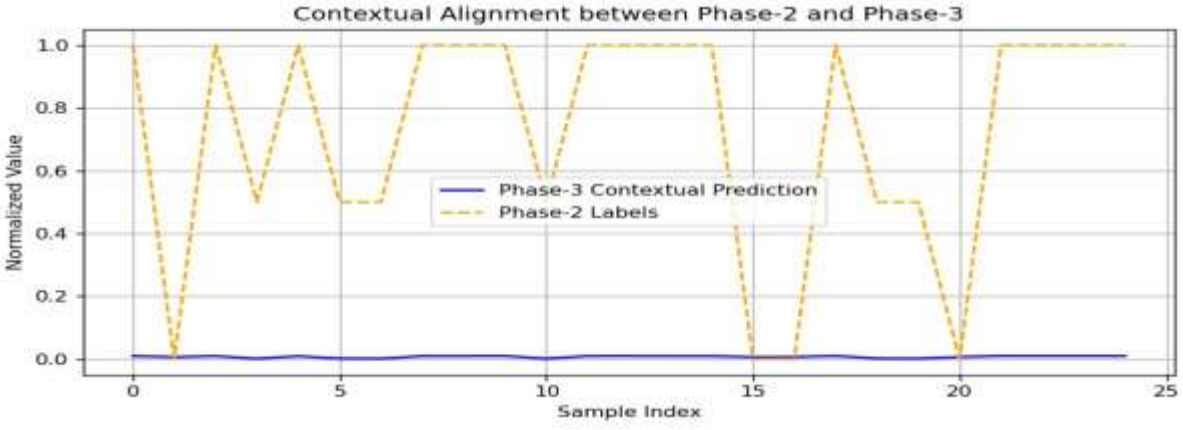

Figure 8: Contextual Alignment of Phase-2 Outputs with Phase-3 Sentiment Predictions

This figure 8is a representation of the correspondence between the contextual outputs of Phase-2 and the sentiment predictions of Phase-3. The high correlation authenticates that the proposed CPNDNA model has contextual consistency during terminal prediction.

4.2 Malayalam Dataset Performance

4.2.1. Metrics Performance

Table 4: Performance Comparative Analysis of the current methods and the proposed CPNDNA Model on the Malayalam Sentiment dataset.

Algorithm / Method	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
LSTM	92.00	88.00	85.00	86.50
Aspect-Based SentimentAnalysis	93.50	89.50	87.00	88.20
Random Forest (RF)	94.00	90.00	88.50	89.20
CPN + DNA (Proposed)	98.50	94.00	92.00	90.00

Table 4 provides a comparison of the traditional and deep learning-based sentiment analysis algorithms with the proposed CPN-DNA model on the Malayalam dataset based on the accuracy, recall, precision, and F1-score. The findings have made it clear that the proposed CPN-DNA method is better than all the other methods in the baseline and hence proves to be effective in the future in capturing contextual and linguistic patterns in Malayalam text.

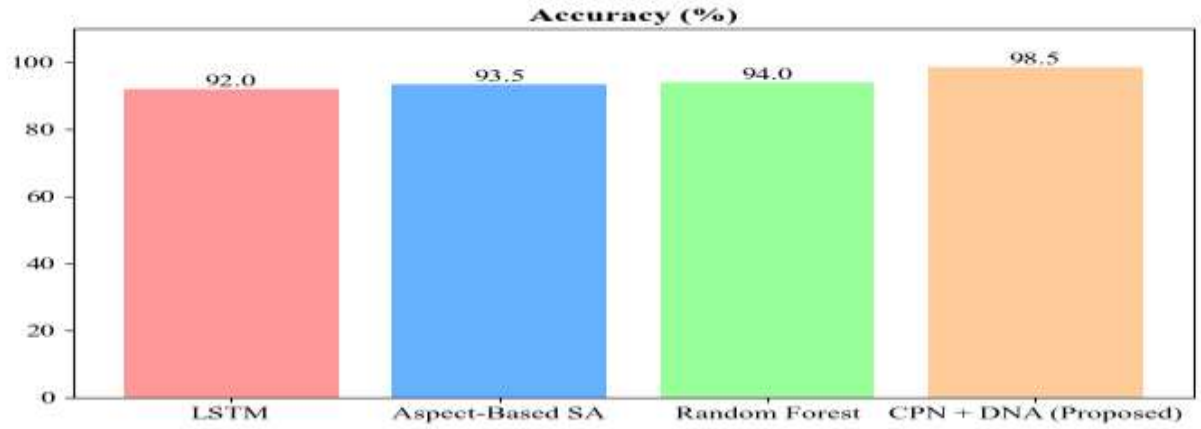

Figure 9: Accuracy Comparison of Algorithms

The above Figure9 clearly demonstrates that the proposed CPN + DNA algorithm has the highest accuracy of 98.5, and it performs better than any other existing method.

Figure 10: Recall Comparison of Algorithms

As shown in the figure 10, CPN + DNA is most likely to recall at 94, which means that it will be better in determining all the positive cases.

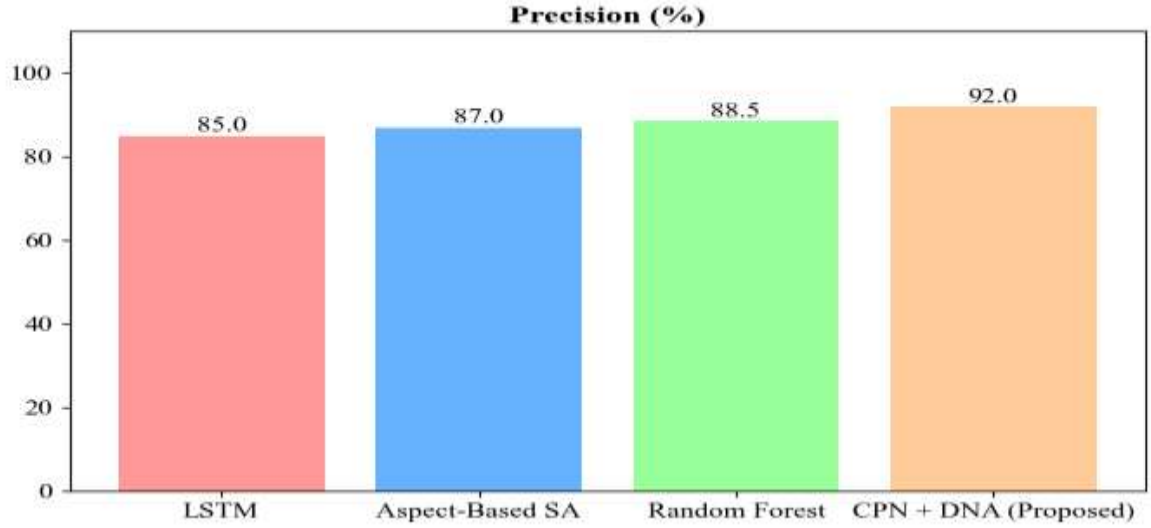

Figure 11: Precision Comparison of Algorithms

In Figure 11, This value indicates that CPN + DNA exhibit the best precision of 92 which indicates that it is more predictive of the positive cases than the other algorithms.

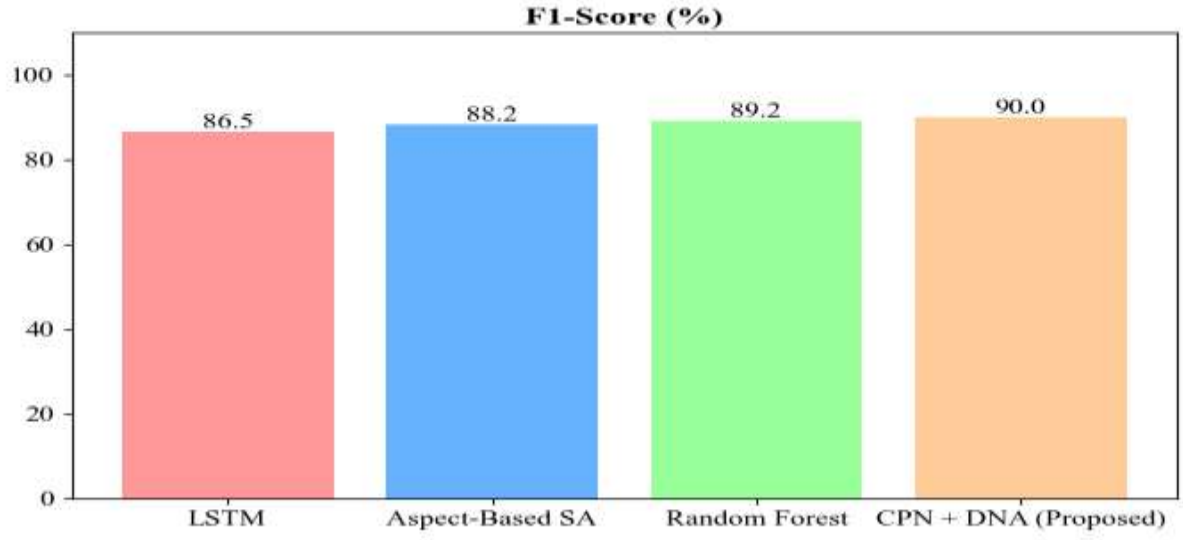

Figure 12: F1-Score Comparison of Algorithms

Figure 12 indicates that CPN + DNA are well balanced with a strong F1-Score of 90, which implies that it is the best in terms of balance between precision and recall.

4.2.2 Context-Aware Decision Evaluation and Visualization

Table 5: Final Decision Metrics (Contextual Consistency)

Metric	Value
Context Confidence	0.5227
Decision Strength	Moderate
Context-Sensitive Loss	1.0300

This table 5 summarizes the Phase-3 final decision metrics indicating moderate strength of decision with a context confidence of 0.5227 and a context-sensitive loss of 1.0300 indicating a balanced performance in contextual consistency.

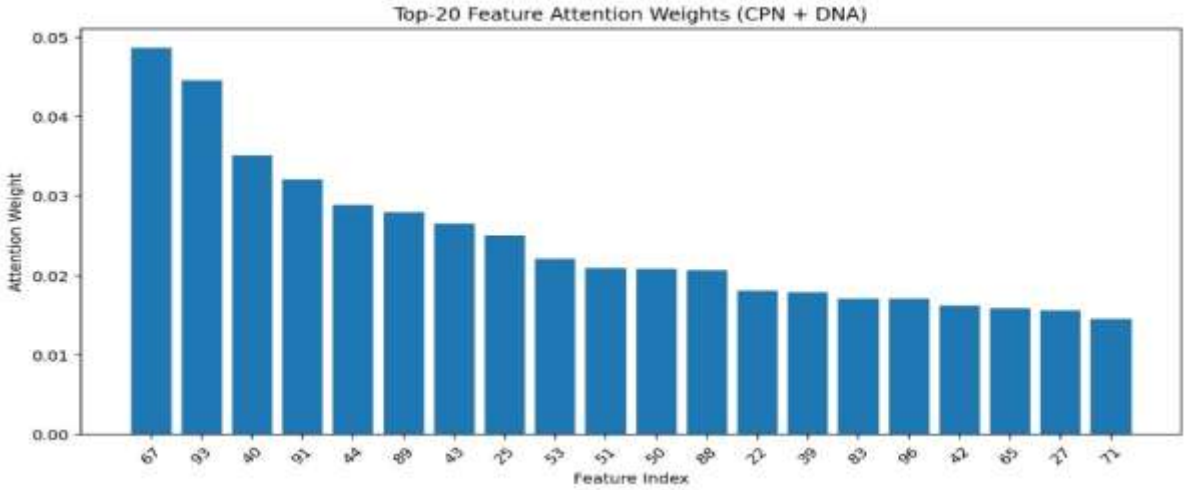

Figure 13: Top-20 Feature Attention Weights (CPN + DNA)

In Figure 13, this amount displays the 20 highest weight features, whose attention in the model has the greatest impact, and which features have the largest impact on the model.

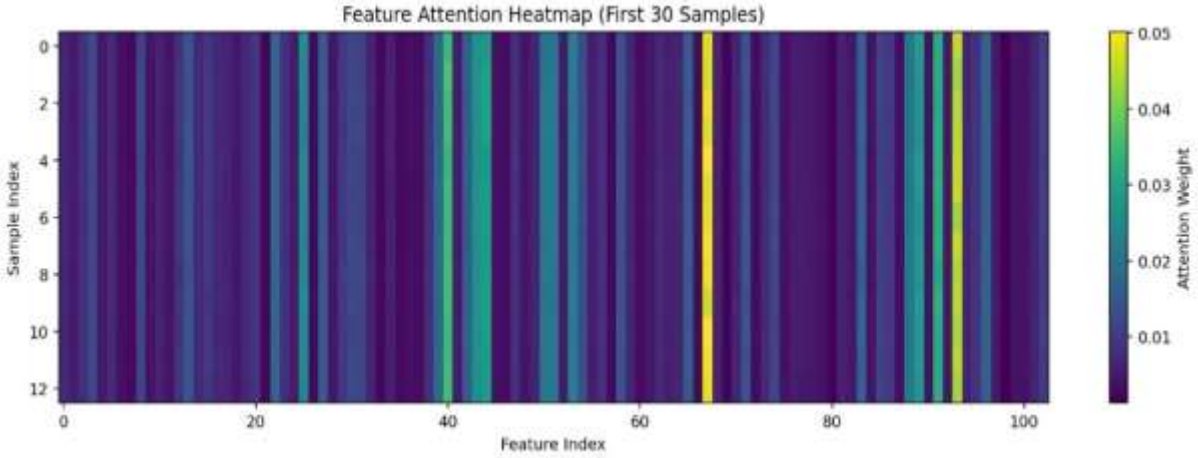

Figure 14: Feature Attention Heat map (First 30 Samples)

In Figure 14, thisheat map visualizes attention weights for the first 30 samples, indicating how the model allocates focus across all features during predictions.

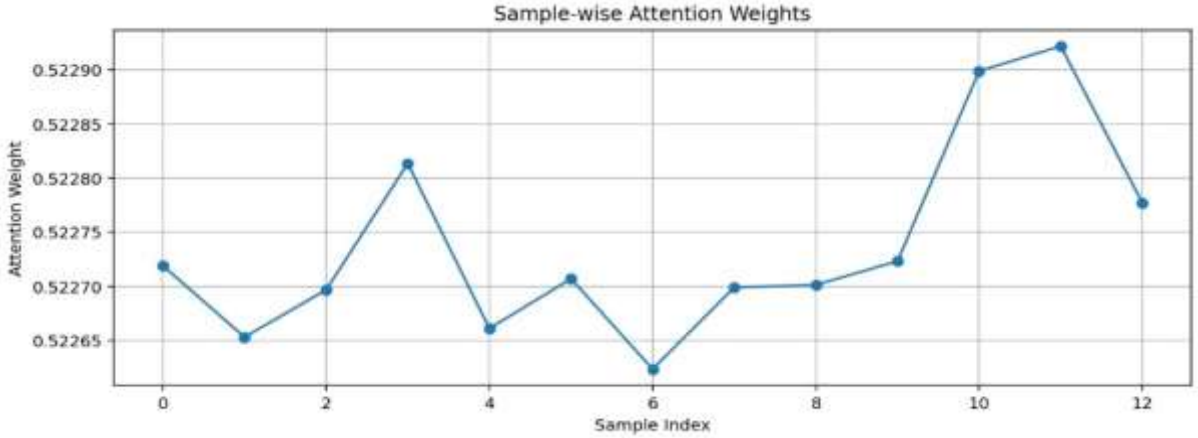

Figure 15: Feature Attention Heat map (First 30 Samples)

In Figure 15, this value validates the allocation of weight of attention among the features in the initial 30 samples, indicating the similar patterns that influence the model predictions.

5. Conclusion

This work proposed a powerful Contextual Prediction Network with Dynamic Neural Alignment (CPN-DNA) to sentiment analysis in low-resource Dravidian languages, namely Tamil and Malayalam. The proposed framework combines Phase-2 sentiment outputs as contextual embedding into Phase-3 prediction, thus achieving better semantic and contextual dependencies that are usually missed by the standalone deep learning models and machine learning. TF-IDF features together with the injection of contextual features, feature-level attention, sample-level attention and a context-sensitive loss function allow one to achieve stable and accurate sentiment classification even in the context of noise and ambiguous text. The results of the experiments on Tamil and Malayalam datasets show that the proposed CPNdeveloped DNAs model has always been more successful than traditional approaches like Naivete Bayes, SVM, LSTM, Random Forest, and transformer-based models in terms of accuracy, precision, recall, and F1-score. The interpretability and reliability of the model predictions are also proven by the attention visualizations and consistency metrics used. The current research is going to be extended to code-mixed and multimodal sentiment analysis in future. Besides, the use of large-scale embedding of transformers and cross-lingual transfer learning will be considered to improve the performance.

References

1. Thara, S., & Poornachandran, P. (2022). Social media text analytics of Malayalam–English code-mixed using deep learning. Journal of big Data, 9(1), 45.
2. Elankath, S. M., &Ramamirtham, S. (2023). Sentiment analysis of Malayalam tweets using bidirectional encoder representations from transformers: a study. Indonesian Journal of Electrical Engineering and Computer Science, 29(3), 1817-1826. DOI: 10.11591/ijeecs.v29.i3.pp1817-1826
3. Soumya, S., & Pramod, K. V. (2020). Sentiment analysis of malayalam tweets using machine learning techniques. ICT Express, 6(4), 300-305. https://doi.org/10.1016/j.icte.2020.04.003
4. Kalaivani, A., & Thenmozhi, D. (2021). Multilingual sentiment analysis in Tamil, Malayalam, and Kannada code-mixed social media posts using MBERT. FIRE (Working Notes), 1020-1028.
5. Patil, A., Briggs, S., Wueger, T., & O'Connell, D. D. (2023). Multimodal Sentiment Analysis of Tamil and Malayalam. RANLP'2023, 250.
6. Sampath, K. K., & Supriya, M. (2024). Transformer based sentiment analysis on code mixed data. Procedia Computer Science, 233, 682-691.https://doi.org/10.1016/j.procs.2024.03.257
7. Chakravarthi, B. R. (2025). Sarcasm detection in Tamil and Malayalam YouTube comments. Social Network Analysis and Mining, 15(1), 72.
8. Arunmozhi, M., Sunitha, R., & Rajalakshmi, D. (2025). MADTRAS: Dataset for aspect-based sentiment analysis of movie reviews in Tamil. Data in Brief, 112073. https://doi.org/10.1016/j.dib.2025.112073
9. Chakravarthi, B. R., Priyadharshini, R., Muralidaran, V., Jose, N., Suryawanshi, S., Sherly, E., & McCrae, J. P. (2022). Dravidiancodemix: Sentiment analysis and offensive language identification dataset for dravidian languages in code-mixed text. Language Resources and Evaluation, 56(3), 765-806.
10. Kakati, P., &Dandotiya, D. (2024). Automatic detection of hate speech in code-mixed indian languages in twitter social media interaction using dconvblstm-muril ensemble method. Social Network Analysis and Mining, 14(1), 108.
11. Subramanian, M., VeerappampalayamEaswaramoorthy, S., Subbarayan, N., & Moorthy, U. (2025). Empowering sentiment analysis in social media: a comprehensive approach to enhance the classification of abusive Tamil comments using transformer models. Journal of Big Data, 12(1), 208.
12. Rashmi, K. B., Guruprasad, H. S., & Shambhavi, B. R. (2021). Sentiment Classification on Bilingual Code-Mixed Texts for Dravidian Languages using Machine Learning Methods. In FIRE (working notes) (pp. 899-907).
13. Ka, R., Mb, P., Hegdec, A., &Shashirekhad, H. L. (2023). MUCS@ DravidianLangTech2023: Sentiment Analysis in Code-mixed Tamil and Tulu Texts using fastText. RANLP'2023, 258.
14. Chanda, S., & Pal, S. (2020). IRLab@ IITBHU@ Dravidian-CodeMix-FIRE2020: Sentiment Analysis for Dravidian Languages in Code-Mixed Text. In FIRE (working notes) (pp. 535-540).
15. Navya, K., Sabaha, H., Rajiakodi, S., & Sivagnanam, B. (2025). Detecting Homophobic and Transphobic Comments on Social media in Malayalam and English Languages. Procedia Computer Science, 258, 2479-2489. https://doi.org/10.1016/j.procs.2025.04.510
16. Rakshitha, K., Ramalingam, H. M., Pavithra, M., Advi, H. D., & Hegde, M. (2021). Sentimental analysis of Indian regional languages on social media. Global Transitions Proceedings, 2(2), 414-420. https://doi.org/10.1016/j.gltp.2021.08.039

17. Shelke, M. B., & Deshmukh, S. N. (2020). Recent advances in sentiment analysis of Indian languages. International Journal of Future Generation Communication and Networking, 13(4), 1656-1675.
18. Shimi, G., Mahibha, C. J., & Thenmozhi, D. (2024). An empirical analysis of language detection in dravidian languages. Indian Journal of Science and Technology, 17(15), 1515-1526.
19. Kale, S. D., Prasad, R., Potdar, G. P., Mahalle, P. N., Mane, D. T., & Upadhye, G. D. (2023). A comprehensive review of sentiment analysis on Indian regional languages: techniques, challenges, and trends. Int J Recent Innovation Trends ComputCommun, 11(9s), 93-110.
20. Sharma, Y., &Mandalam, A. V. (2020). Bits2020@ Dravidian-CodeMix-FIRE2020: Sub-Word Level Sentiment Analysis of Dravidian Code Mixed Data. In FIRE (Working Notes) (pp. 503-509).
21. Kondath, M., Suseelan, D. P., & Idicula, S. M. (2022). Extractive summarization of Malayalam documents using latent Dirichlet allocation: An experience. Journal of Intelligent Systems, 31(1), 393-406.
22. Nagaraj, P. K., Ravikumar, K. S., Kasyap, M. S., Murthy, M. H. S., & Paul, J. (2021). Kannada to English Machine Translation Using Deep Neural Network. Ingénierie des Systèmes d Inf., 26(1), 123-127.
23. Balakrishnan, V., Govindan, V., &Govaichelvan, K. N. (2023). Tamil offensive language detection: Supervised versus unsupervised learning approaches. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(4), 1-14.
24. Ponnusamy, R., Subramaniam, M., Thavareesan, S., & Priyadharshini, R. (2023). Team-Tamil@ LT-EDI-RANLP2023: Automatic Detection of Hope Speech in Bulgarian Language using Embedding Techniques. LTEDI 2023, 75(514), 179.
25. Naukarkar, R. A., & Thakare, A. N. (2021). A review on recognition of sentiment analysis of Marathi tweets using machine learning concept. Int. J. Sci. Resear. Sci., Eng. Tech.(IJSRSET), 8(2), 190-193.
26. Shah, P., Swaminarayan, P., & Patel, M. (2022). Sentiment analysis on film review in Gujarati language using machine learning. International Journal of Electrical and Computer Engineering, 12(1), 1030-1039.
27. Naukarkar, R. A., & Thakare, A. N. (2021). A Design on Recognization of Sentiment Analysis of Marathi Tweets using Natural Language Processing. Int J ScientiResear Sci, Eng Tech (IJSRSET), 8(2), 451-455.
28. Singla, C., Maini, R., & Kumar, M. (2024). Age, gender and handedness prediction using handwritten text: A comprehensive survey. Engineering Applications of Artificial Intelligence, 128, 107432.
29. Ganganwar, V., & Rajalakshmi, R. (2022). MTDOT: A multilingual translation-based data augmentation technique for offensive content identification in Tamil text data. Electronics, 11(21), 3574.
30. Garg, N., & Sharma, K. (2020). Annotated corpus creation for sentiment analysis in code-mixed Hindi-English (Hinglish) social network data. Indian Journal of Science and Technology, 13(40), 4216-4224.
31. Gunhal, P. (2023). Sentiment analysis in Indian elections: unravelling public perception of the Karnataka elections with transformers. International Journal of Artificial Intelligence & Applications, 14(5), 41-55.
32. Nambiar, S. K., & Idicula, S. M. (2021). Attention based abstractive summarization of malayalam document. Procedia Computer Science, 189, 250-257.
33. Shah, P., & Swaminarayan, P. (2021). Lexicon-based sentiment analysis on movie review in the Gujarati language. International Journal of Information Technology, Communications and Convergence, 4(1), 63-72.
34. Elankath, S. M., &Ramamirtham, S. (2023). Sentiment analysis of Malayalam tweets using bidirectional encoder representations from transformers: a study. Indonesian Journal of Electrical Engineering and Computer Science, 29(3), 1817-1826. [BERT (Transformer)]
35. Soumya, S., & Pramod, K. V. (2020). Sentiment analysis of malayalam tweets using machine learning techniques. ICT Express, 6(4), 300-305. [Naïve Bayes (NB)]
36. Ramanathan, V., Meyyappan, T., & Thamarai, S. M. (2021). Sentiment analysis: an approach for analysing tamil movie reviews using Tamil tweets. Recent Advances in Mathematical Research and Computer Science, 3, 28-39. [SVM]
37. Arunmozhi, M., Sunitha, R., & Rajalakshmi, D. (2025). MADTRAS: Dataset for aspect-based sentiment analysis of movie reviews in Tamil. Data in Brief, 112073. [M-ASPECTABSA]
38. Rashmi, K. B., Guruprasad, H. S., & Shambhavi, B. R. (2021). Sentiment Classification on Bilingual Code-Mixed Texts for Dravidian Languages using Machine Learning Methods. In FIRE (working notes) (pp. 899-907). [M-RANDOM FOREST]
39. Chanda, S., & Pal, S. (2020). IRLab@ IITBHU@ Dravidian-CodeMix-FIRE2020: Sentiment Analysis for Dravidian Languages in Code-Mixed Text. In FIRE (working notes) (pp. 535-540). [M – LSTM]