International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ANALYZING MACHINE LEARNING AND HYBRID DEEP LEARNING APPROACHES FOR OPINION EXTRACTION FROM WEB TEXT

Erugu Krishna[1], Dr. Sonawane Vijay Ramnath[2]

[1]Research Scholar, [2]Research Supervisor

[1,2]Dr. A. P. J. Abdul Kalam University, Indore, India

krishna.cseit@gmail.com, vijaysonawane11@gmail.com

Abstract

Opinion extraction from web text is essential for understanding public perception in products, news, and social media; however, heterogeneous writing styles, noisy text, and ambiguous sentiment expressions reduce the reliability of conventional classifiers. This study addresses three-class sentiment classification (positive, neutral, negative) by designing a unified pipeline that combines classical machine learning baselines with a hybrid deep learning model. The proposed workflow performs label normalization, text cleaning (URL/mention removal, hashtag normalization, symbol filtering, and whitespace standardization), de-duplication, and stratified train–test splitting. For classical methods, TF–IDF features with uni/bi-grams are used to train twelve classifiers, including LinearSVM, Logistic Regression, SGD-based models, Ridge, Passive Aggressive, Perceptron, and Naïve Bayes variants, along with probability-calibrated versions. In parallel, a hybrid CNN–BiLSTM network is trained using tokenization and fixed-length padding to learn both local phrase patterns and long-range contextual dependencies. Experiments are conducted on three datasets representing different web domains: product reviews, Times of India headlines, and political tweets. Results show that the hybrid CNN–BiLSTM achieves the highest performance, reaching 91.56% accuracy on Times of India headlines, and consistently outperforming the strongest TF–IDF baseline (LinearSVM), demonstrating improved robustness for multi-domain opinion extraction.

Keywords: Opinion mining, sentiment analysis, TF–IDF, LinearSVM, machine learning, CNN–BiLSTM, deep learning, web text analytics, multi-dataset evaluation.

1. Introduction

The rapid growth of web-based platforms such as e-commerce portals, online news media, and social networking sites has resulted in an enormous volume of user-generated textual data. This data often contains rich opinions, attitudes, and emotions that reflect public perception toward products, services, policies, and events. Extracting meaningful opinions from such unstructured web text, commonly known as opinion mining or sentiment analysis, has therefore become a critical research area in natural language processing (NLP) and artificial intelligence (AI) [1], [2]. Accurate sentiment analysis supports decision-making in business intelligence, political forecasting, social monitoring, and recommendation systems.

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Traditional sentiment analysis methods primarily relied on lexicon-based approaches, where predefined sentiment dictionaries were used to infer polarity scores [3]. Although simple and interpretable, lexicon-based techniques suffer from limited domain adaptability, inability to capture contextual polarity, and poor performance on noisy or informal text such as social media posts [4]. To overcome these limitations, machine learning-based approaches were introduced, treating sentiment analysis as a supervised text classification problem [5]. Classical algorithms such as Support Vector Machines (SVM), Logistic Regression, Naïve Bayes, and Perceptron models combined with handcrafted features like bag-of-words and term frequency–inverse document frequency (TF–IDF) have shown promising results across various benchmark datasets [6], [7].

Among classical approaches, linear SVMs with TF–IDF representations have consistently demonstrated strong performance due to their effectiveness in high-dimensional sparse feature spaces [8]. Several studies have also explored calibrated classifiers and ensemble strategies to improve probabilistic confidence and classification stability [9]. However, these models depend heavily on surface-level features and struggle to capture semantic relationships, word order, and long-range dependencies within text. This limitation becomes more pronounced when dealing with heterogeneous web text sources such as news headlines, product reviews, and political tweets, which differ significantly in length, structure, and linguistic style [10].

Recent advances in deep learning have significantly transformed sentiment analysis by enabling automatic feature learning from raw text. Neural architectures such as Convolutional Neural Networks (CNNs) effectively capture local n-gram patterns related to sentiment, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks model sequential dependencies and contextual information [11]. Bidirectional LSTM (BiLSTM) models further enhance performance by learning context from both past and future word sequences. However, standalone CNN or LSTM models may still be insufficient when applied individually, particularly in multi-domain sentiment classification tasks.

To address these challenges, hybrid deep learning architectures that integrate CNN and BiLSTM components have been proposed [12]. In such models, CNN layers extract discriminative local features, while BiLSTM layers capture long-term contextual dependencies, resulting in a more comprehensive representation of sentiment-bearing text. Despite their potential, there remains a need for systematic comparative studies that evaluate hybrid deep learning models alongside a wide range of classical machine learning baselines across multiple, diverse datasets using a unified experimental framework.

In this work, we present a comprehensive analysis of machine learning and hybrid deep learning approaches for opinion extraction from web text. The study focuses on three-class sentiment classification—positive, neutral, and negative—across three heterogeneous datasets: product reviews, Times of India news headlines, and political tweets. A robust preprocessing pipeline is employed, including label normalization, noise handling, text cleaning, and de-duplication to ensure fair and leakage-free evaluation. Twelve classical TF–IDF-based machine learning models, including calibrated variants, are systematically

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

compared with a hybrid CNN–BiLSTM architecture trained using tokenized and padded sequences.

Extensive experiments demonstrate that while classical models such as LinearSVM remain strong baselines, the proposed hybrid CNN–BiLSTM model consistently achieves superior performance, particularly on complex and noisy datasets. The results highlight the importance of combining local feature extraction and sequential context modeling for robust opinion extraction from diverse web text sources [13].

Key Contributions:

The main contributions of this paper are summarized as follows:

1. A unified sentiment analysis framework combining extensive preprocessing, classical machine learning, and hybrid deep learning models.
2. A comprehensive comparison of twelve TF–IDF-based machine learning classifiers with a CNN–BiLSTM hybrid model across three heterogeneous web text datasets.
3. An effective hybrid CNN–BiLSTM architecture that captures both local and long-range sentiment patterns, achieving state-of-the-art accuracy on multiple datasets.
4. Detailed evaluation using accuracy, precision, recall, F1-score, ROC, Precision–Recall curves, and error analysis to provide in-depth performance insights.

2. Literature review

Su et al. (2024), Implicit aspect sentiment expressions lack explicit opinion words, making ABSA challenging in social media text. Existing dependency-tree and attention-based methods often miss relevant aspect sentiment or focus on irrelevant words. This paper proposes a **Prototype-based Demonstration (PD-ABSA)** model with prototype learning and demonstration stages. Mask-aware attention and contrastive learning capture implicit sentiment prototypes, which guide a T5 model via neural demonstrations. Experiments on Laptop and Restaurant datasets show consistent accuracy gains, especially for implicit sentiment cases, validating effectiveness in social computing scenarios. **[1]**

Alfreihat et al. (2024), Informal Arabic sentiment analysis is difficult due to morphology and dialects. This work introduces **Emo-SL**, an emoji sentiment lexicon for Arabic tweets, built from 58K emoji-containing tweets. Sentiment scores for 222 emojis are computed and integrated with text features. ML classifiers (SVM, NB, RF, KNN) trained with emoji-aware features significantly outperform text-only models, improving accuracy by 26.7% and achieving 89% F1. Results confirm emojis provide crucial contextual sentiment cues in noisy Arabic micro-text. **[2]**

Mahmoudi et al. (2024), This study evaluates popular sentiment analysis packages in Python and R, comparing accuracy and time complexity across seven datasets. Results show performance varies significantly by dataset, with **sentimentr** being the most consistent. Python tools are generally faster, but most packages struggle to model sentiment intensity and often overfit to familiar data. While effective for binary polarity, generalization to unseen

Received: January 07, 2025,

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

datasets remains limited, highlighting the need for more robust and adaptable sentiment analysis tools. **[3]**

Wang et al. (2024), Aspect-based multimodal sentiment analysis (ABMSA) often ignores global sentiment tendency and fine-grained multimodal cues. This paper proposes a **Dual-Perspective Fusion Network (DPFN)** that combines global sentiment (via text–image captions) with local fine-grained information using graph structures over text and images. By integrating both perspectives, the model improves aspect-level sentiment prediction. Experiments on multimodal Twitter datasets demonstrate DPFN consistently outperforms state-of-the-art approaches. **[4]**

Zhao et al. (2024), Multimodal Aspect-Based Sentiment Analysis (MABSA) integrates text with other modalities to infer aspect sentiment. This survey systematically reviews recent MABSA research, introducing core concepts, summarizing methods for multimodal aspect classification and aspect–sentiment pair extraction, and comparing their strengths and weaknesses. It also reviews commonly used datasets, evaluation metrics, and reported results. Finally, it outlines emerging research trends, providing a structured reference for future MABSA studies. **[5]**

Ruan et al. (2024), Image sentiment analysis has largely focused on CNN-based content features, overlooking the psychological importance of color. This paper proposes **Color Enhanced Cross Correlation Net (CECCN)**, which jointly models image content and color features and their correlations. Content is extracted via pretrained CNNs, while color moments are derived from multiple color spaces. A cross-correlation mechanism with attention enhances sentiment cues. Experiments on benchmark datasets show CECCN outperforms existing image sentiment models. **[6]**

Razali et al. (2024), Understanding customer sentiment is vital for gastronomy tourism, yet traditional analysis is slow, subjective, and weak at handling imbalanced data. This research introduces a hybrid lexicon-based and ML approach with data augmentation, feature engineering, and visualization tailored for Sarawak's gastronomy sector. Using synonym augmentation, n-grams, and kNN, the system achieves 0.98 accuracy and 0.99 F1/ROC-AUC. The framework significantly improves minority sentiment detection and supports real-time business intelligence. **[7]**

Haryono et al. (2024), Stock prices are influenced not only by historical data but also by news sentiment, which must be quantified for forecasting. This study proposes **generative ABSA** to produce aspect–sentiment quadruplets and compute daily sentiment scores using the Loughran–McDonald lexicon. These scores are integrated with stock data in a **permuted Temporal Kolmogorov-Arnold Network (pTKAN)**. Experiments across multiple issuers show sentiment-enhanced forecasting improves performance, with pTKAN outperforming 18 alternative models. **[8]**

Wang et al. (2024), Existing MABSA methods suffer from modality misalignment and noise when images are irrelevant. This paper proposes an end-to-end MABSA framework with **image-to-text conversion** that maps images into token embeddings compatible with

Received: January 07, 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

pretrained language models. An aspect-oriented filtration module removes visual noise before sentiment prediction. By unifying text, aspect, and filtered visual prompts, the model achieves state-of-the-art performance on benchmark datasets, with improved robustness and efficiency. **[9]**

Shafikuzzaman et al. (2024), This paper presents the largest comparative study of pretrained language models (PLMs) for sentiment analysis in software engineering. It evaluates fine-tuned, zero-shot (including GPT-4), and few-shot PLMs on six datasets. Results show domain-specific fine-tuned models like **seBERT** excel on large datasets, while few-shot models perform better on smaller ones. Explainable AI-based error analysis highlights unresolved challenges, offering reproducible insights for PLM selection in SE sentiment analysis. **[10]**

Tang et al. (2024), Sentiment Quantification aggregates overall sentiment from multiple reviews, but equal weighting ignores review confidence and allows fake reviews to distort results. This paper proposes **COSE**, a confidence-aware framework that models review reliability using an unsupervised Review Graph. A dynamic fusion attention mechanism reduces sentiment perturbation by emphasizing high-confidence reviews. Experiments on large-scale datasets show COSE significantly outperforms existing methods, producing more reliable overall sentiment estimation. **[11]**

Da et al. (2024), Large language models improve sentiment analysis but can amplify social biases from training data. This paper proposes a de-biasing method that first uses **causal mediation analysis** to locate which model components drive bias, then applies **targeted counterfactual training** by swapping sensitive attributes (e.g., gender) in sentences where output should not change. The approach is validated by fine-tuning **BERT** for sentiment prediction on Stanford Sentiment Treebank and Amazon Reviews, and evaluating fairness and accuracy with the **Equity Evaluation Corpus**. Results show improved gender fairness without sacrificing sentiment accuracy, outperforming prior de-biasing baselines. **[12]**

Xie et al. (2024), Multimodal video sentiment analysis must handle noisy or missing modalities and unequal modality contributions, and typical regression ignores ordinal sentiment structure. This paper proposes **TMSON**, a trustworthy multimodal sentiment **ordinal** network. It extracts unimodal features, estimates unimodal uncertainty distributions, fuses modalities via **Bayesian fusion**, and applies **ordinal regression** in an ordinal-aware sentiment space. Experiments show TMSON outperforms baselines and reduces uncertainty, producing more robust multimodal sentiment predictions under ambiguity and noise. **[13]**

Diwali et al. (2024), Deep learning dominates sentiment analysis due to compute and benchmark data growth, but interpretability remains a major problem. This work provides an overview of sentiment analysis methods alongside **eXplainable AI (XAI)** techniques used to interpret deep models. It highlights that few studies explain internal behaviors of sentiment models, and surveys the state of explainability for sentiment analysis, positioning interpretability as necessary for trustworthy deployment and understanding model decisions. **[14]**

Received: January 07, 2025,

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Mughal et al. (2024), ABSA improves over document/sentence polarity by linking sentiment to specific explicit or implicit aspects, but current ABSA models face domain dependence, heavy labeled-data needs, and limited exploration of newer LLMs. This study evaluates models including ATAE-LSTM, flan-t5-large-absa, DeBERTa, PaLM, and GPT-3.5-Turbo on datasets such as DOTSA, MAMS, and SemEval16. Results show domain sensitivity across tasks (ATSA/ACSA), with **DeBERTa** consistently strong and **PaLM** competitive across multiple domains, informing model selection and future ABSA improvements. **[15]**

Sherin et al. (2024), For tweet sentiment analysis, this work proposes a hybrid system combining a **fuzzy emotion extractor (FEE)**, an ensemble **Bi-LSTM/GRU** recurrent model, and an enhanced **Aquila Optimizer** variant to weight features. FEE uses semantic polarity and word positions to extract emotions, while the optimized ensemble model captures sequence context and word relations. Experiments on Twitter (X) datasets report superiority over state-of-the-art methods using standard sentiment metrics, supporting fuzzy + deep + optimization fusion for TSA. **[16]**

Duan et al. (2024), Financial sentiment analysis is difficult due to domain-specific language, context dependence, and ambiguous polarity. This paper proposes a hybrid model combining **topic features** with a **pretrained model**, and introduces an improved attention mechanism with an adaptive threshold and masking to reduce noise from over-attending to single words. Topic features help capture long-distance semantic relations. Experiments report F1 improvements of **2.05%–7.27%** over baselines, indicating better stability and suitability for financial text sentiment tasks. **[17]**

Peivandizadeh et al. (2024), This paper predicts stock prices by combining social-media sentiment with market data, addressing class imbalance and temporal fusion challenges. It proposes **Off-policy PPO** to adjust training rewards toward correctly classifying minority sentiment classes, and a **Transductive LSTM (TLSTM)** that prioritizes temporally closer signals while integrating sentiment outputs with historical stock data. Ablation studies support the value of both components, and the approach aims to improve prediction accuracy and provide actionable insights for **investors and policymakers. [18]**

Wang et al. (2024), Many ABSA models focus on target–context semantic links but underuse coarse **category-level knowledge**, which can reduce polysemy and ambivalence. This paper proposes **CoAN**, a multi-task learning framework that jointly learns target- and category-level features. Two co-interactive attention layers model interactions at word and feature levels to fuse multi-granularity knowledge. On three restaurant datasets, CoAN improves accuracy by **1.41%** and F1 by **2.81%**, with visualizations showing better feature fusion benefiting both subtasks. **[19]**

Rizinski et al. (2024), Lexicon-based finance sentiment is fast and interpretable but costly to curate; transformers are accurate but heavy and slower for production. This paper proposes **XLex**, which uses transformers plus **SHAP** explanations to automatically learn financial lexicons. XLex expands vocabulary beyond the Loughran–McDonald lexicon, reduces human maintenance, and improves sentiment classification accuracy; combining XLex with

Received: January 07, 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

LM (XLex+LM) improves further. The resulting lexicon approach is also smaller, faster, and more interpretable than transformer models. **[20]**

Aqeel et al. (2025), ABSA struggles with implicit sentiments and complex sentence structures where dependency trees may miss key semantics. This study introduces **Semantic Parsing Trees (SPT)**, transforming syntactic trees into semantically richer structures to preserve roles and relations. Combined with graph-based attention and relational heads, SPT improves encoding of semantic nuance for aspect sentiment prediction. Evaluations on SemEval 2014, Restaurant, and Twitter datasets show improved accuracy and adaptability compared with conventional ABSA models. **[21]**

Maroof et al. (2024), Customer feedback from mobile app reviews is crucial for service providers, yet implicit opinions remain underexplored compared to explicit ones, especially in service-oriented domains. This study proposes an end-to-end **aspect-based sentiment analysis (ABSA)** framework for English mobile app reviews that handles implicit opinion terms. A two-step hybrid pipeline is introduced: rule-based extraction of implicit opinion terms, followed by ML/DL (including fine-tuned BERT) for aspect categorization and sentiment classification. The approach effectively addresses the double-implicit problem and significantly outperforms baselines, achieving strong accuracy and F1 scores for both aspect and sentiment classification. **[22]**

Zhao et al. (2024), Emojis play a vital role in expressing emotions in short, informal texts where sarcasm and ambiguity are common. This paper leverages emojis for sentiment analysis by first resolving emoji polarity uncertainty using user information. It then builds rich emoji representations incorporating position, semantics, emotion, and frequency. The proposed **TaneNet**, a two-level attention network, combines clause representations with emoji features to model their emotional influence. Experiments on real-world datasets show that TaneNet consistently outperforms state-of-the-art sentiment analysis methods. **[23]**

Polat et al. (2024), This work introduces the **Couple Dialogue** dataset for Turkish conversational sentiment analysis, containing over 14,000 annotated utterances capturing sentiment dynamics in dyadic interactions. It compares non-contextual models with contextual approaches using LLMs such as BERT, GPT-3.5, GPT-4, Llama 2, and DialogueRNN. Due to Turkish's rich morphology, detailed linguistic analysis is incorporated. Contextual models, especially fine-tuned Turkish BERT and LLM-based approaches, significantly outperform non-contextual baselines, achieving nearly 10% higher Weighted F1, highlighting the importance of context in conversational sentiment analysis. **[24]**

Hasan, Ali et al. (2018), This paper focuses on election-related opinion mining from Twitter and argues that, even though many sentiment-analysis techniques/tools have been used in political contexts, there is still a need for a stronger "state-of-the-art" approach. The authors propose a hybrid sentiment-analyzer approach that combines sentiment analysis with supervised machine-learning, and they explicitly compare Naïve Bayes and Support Vector Machines (SVM) for analyzing political views from Twitter accounts. [25]

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Souma, Wataru et al (2019), This paper studies whether historical news sentiments (derived from market reactions) can help forecast financial news sentiment. It defines news sentiment using intraday stock-price behavior: if the averaged stock return right after a news release is positive/negative, the corresponding news is labeled positive/negative. The method uses GloVe word vectors (trained on Wikipedia 2014 + Gigaword 5) as inputs and trains an RNN with LSTM units on Thomson Reuters News Archive (TRNA) from 2003–2012, then tests on 2013; importantly, it reports better forecasting when training examples are chosen hierarchically (using the most strongly positive/negative polarity-scored news) rather than randomly. [26]

3. Proposed work

3.1 Proposed architecture

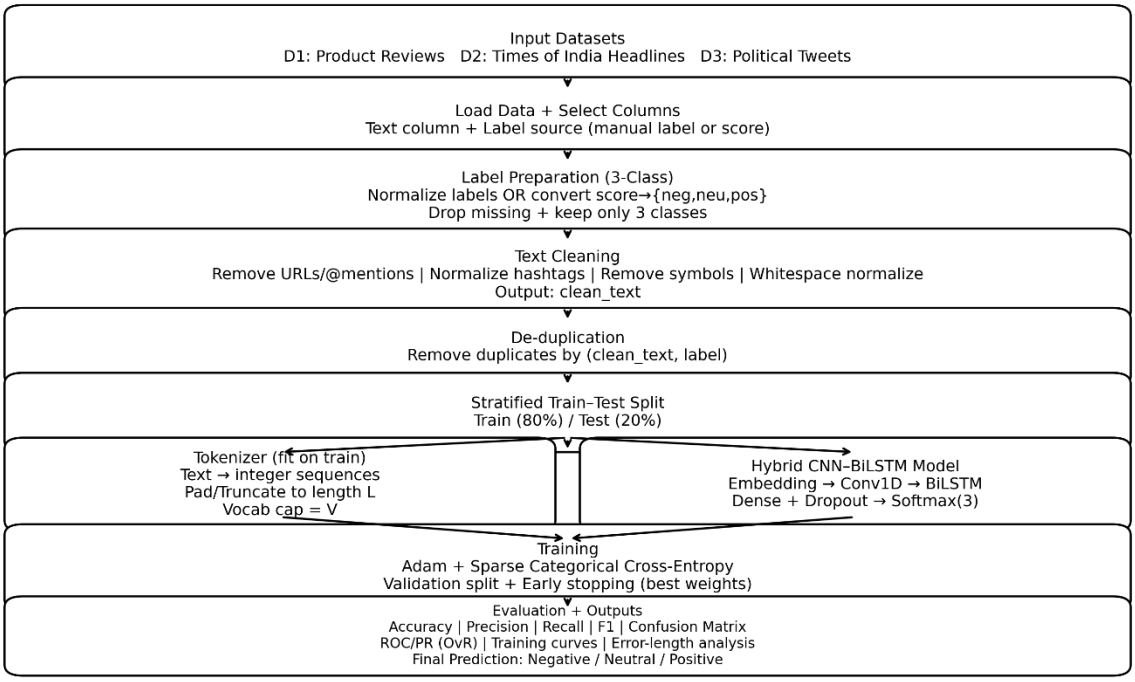

Figure 1. The proposed architecture

Figure 1 illustrates the complete workflow of the proposed hybrid CNN–BiLSTM–based sentiment classification framework applied across three heterogeneous datasets: product reviews, Times of India headlines, and political tweets. The process begins with loading the input datasets and selecting the appropriate text and label sources, followed by three-class label preparation through normalization or score-based conversion into negative, neutral, and positive classes. The text is then cleaned by removing URLs and user mentions, normalizing hashtags, eliminating special symbols, and standardizing whitespace, after which duplicate samples are removed to reduce bias and data leakage. The cleaned dataset is partitioned into training and testing sets using stratified sampling to preserve class distribution. Tokenization and sequence padding are performed on the training data to generate fixed-length input sequences, which are then fed into a hybrid deep learning model consisting of an embedding layer, a convolutional layer for local feature extraction, and a bidirectional LSTM layer for

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

capturing long-range contextual dependencies. The model is trained using the Adam optimizer with sparse categorical cross-entropy loss and early stopping based on validation performance. Finally, the framework evaluates model performance using multiple metrics, including accuracy, precision, recall, F1-score, confusion matrices, ROC and precision–recall curves, and outputs the final sentiment predictions for all three classes.

3.2 Proposed algorithm

Algorithm: Hybrid CNN–BiLSTM for Three-Class Sentiment Classification (Multi-Dataset)

Input: Three datasets $\{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$= {Product Reviews, Times of India Headlines, Political Tweets}
Output: Predicted sentiment label for each text instance: **Positive / Neutral / Negative**

Step 1 — Load Datasets

1.1 Load $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$from storage (CSV/Drive).
1.2 For each dataset, select:

- **Text column** (review/headline/tweet)
- **Label source** (manual sentiment column OR score column)

Step 2 — Label Preparation (Three-Class Output)

2.1 If labels exist as text (positive/neutral/negative), standardize them:

- Convert to lowercase
- Map noisy labels (e.g., *unlabled*, *unlabeled*, *suggestion*) → **neutral**

2.2 If labels are scores (e.g., compound polarity), convert to three-class sentiment:

- score ≤ −0.05 → **negative**
- score ≥ +0.05 → **positive**
- otherwise → **neutral**

2.3 Remove records with missing text or missing sentiment labels.
2.4 Keep only the three classes: **positive, neutral, negative**.

Step 3 — Text Cleaning

3.1 Remove URLs.
3.2 Remove user mentions (e.g., @user).
3.3 Convert hashtags into words (e.g., #happy → happy).
3.4 Remove special symbols and non-alphanumeric characters.
3.5 Normalize whitespace.
3.6 Store cleaned text as **clean_text**.

Step 4 — Remove Duplicates

4.1 Remove duplicates using the pair (**clean_text**, **sentiment label**) to reduce leakage and bias.

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Step 5 — Train–Test Split

5.1 Split each dataset into training and testing sets (e.g., 80/20).
5.2 Use stratified sampling to preserve the distribution of positive/neutral/negative classes.

Step 6 — Tokenization and Sequence Padding

6.1 Fit a tokenizer on training text only (to avoid leakage).
6.2 Convert text to integer sequences.
6.3 Pad/truncate sequences to fixed length L(e.g., 60 tokens).
6.4 Set vocabulary size Vusing the tokenizer word index cap.

Step 7 — Build Hybrid CNN–BiLSTM Model

7.1 Construct the network architecture:

- **Embedding layer** (maps tokens to dense vectors)
- **Conv1D layer** (captures local n-gram sentiment patterns)
- **BiLSTM layer** (captures long-range dependencies in both directions)
- **Dense + Dropout layer** (feature refinement and regularization)
- **Softmax output layer (3 neurons)**

7.2 Compile the model using:

- Loss: sparse categorical cross-entropy
- Optimizer: Adam
- Metric: accuracy

Step 8 — Train Model

8.1 Train on training sequences with validation split (e.g., 10%).
8.2 Apply early stopping based on validation loss to prevent overfitting.
8.3 Save the best weights automatically.

Step 9 — Predict Sentiment Classes

9.1 Generate predicted probabilities for test samples.
9.2 Convert softmax probabilities into class labels by selecting the maximum probability class:

- 0 → **negative**
- 1 → **neutral**
- 2 → **positive**

Step 10 — Model Evaluation

10.1 Compute accuracy, precision, recall, and F1-score for all classes.
10.2 Generate confusion matrix (raw and normalized).
10.3 Plot ROC and Precision–Recall curves using One-vs-Rest method.

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

10.4 Plot training loss/accuracy curves.
10.5 Perform error analysis using text length distribution of correct vs wrong predictions.

3.4 Comparison of proposed work

Table 1: Comparison of Proposed CNN–BiLSTM vs Standard Models Based on Feature Types

Aspect	Standard Models (TF–IDF + Classic ML)	Proposed Model (CNN–BiLSTM)	Why Proposed Can Be Better
Input Feature Type	Sparse vectors (TF–IDF)	Dense sequences (tokens → embeddings)	Embeddings learn semantic similarity and richer meaning
Feature Granularity	Word-level counts + n-grams	Word order + contextual patterns	Captures both local and long-range sentiment patterns
Word Order Handling	Limited (only up to bigrams)	Full sequence modeling	Better handling of sentence structure and context
Context Understanding	Weak (bag-of-words bias)	Strong (BiLSTM captures dependencies)	Improves classification when sentiment depends on context
Negation Handling	Often weak (e.g., "not good")	Better (sequence modeling learns "not + adjective")	Reduces misclassification in real web text
Phrase Pattern Learning	Manual via n-grams	Automatic via Conv1D filters	CNN learns sentiment phrases without explicit feature engineering
Long-Range Dependency	Not supported	Supported (BiLSTM)	Better for longer reviews and complex statements
Generalization Across Domains	Can be domain-sensitive	Often better generalization	Works better when vocabulary differs across datasets
Handling Noisy Text (tweets)	Moderate (needs heavy cleaning)	Better if trained properly	Learns robust patterns beyond surface word counts
Interpretability	High (important words/weights visible)	Medium/Low	Trade-off: deep model is less transparent

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Compute Cost	Low (fast training & inference)	Higher (training time + tuning required)	Deep model needs more resources but can improve accuracy
Best Fit Scenarios	Short structured text (e.g., headlines)	Context-rich and informal text (tweets, reviews)	Proposed model excels when sentiment is implicit/complex

4. Implementation and Result analysis

4.1 Hardware and software

All experiments were executed in Google Colab, a cloud-based Jupyter notebook environment running Linux with Python (v3.9+). The runtime used a multi-core Intel Xeon CPU with approximately 12–16 GB RAM, which was adequate for data preprocessing, TF–IDF vectorization, and training the twelve classical machine learning models. In addition, the proposed hybrid deep learning baseline (CNN+BiLSTM) was implemented and trained within the same Colab environment; GPU acceleration (e.g., NVIDIA T4/P100, when enabled) was beneficial for faster neural network training but was not mandatory for the classical TF–IDF models. Datasets and generated outputs were managed using the Colab virtual machine's temporary storage with optional Google Drive mounting for persistent storage and result export. The software stack comprised NumPy and Pandas for numerical computation and dataset handling, Scikit-learn for TF–IDF feature extraction, pipeline-based training of classical classifiers (e.g., LinearSVM, Calibrated LinearSVM, Logistic Regression, Naïve Bayes variants, SGD-based models, Ridge, Passive Aggressive, and Perceptron), and evaluation using accuracy, precision, recall, F1-score, confusion matrices, ROC curves, and precision–recall curves (OvR). The deep learning model was implemented using TensorFlow/Keras (Tokenizer, sequence padding, Embedding, Conv1D, BiLSTM, Dropout, and Softmax layers) with early stopping based on validation loss. Matplotlib was used to visualize training curves and diagnostic plots, while Python regular expressions were applied for text cleaning (URL/mention removal, hashtag normalization, symbol filtering, and whitespace normalization).

4.2 Dataset

The **Sentiment Product Review dataset** focuses on e-commerce style product feedback and is commonly used for sentiment analysis on customer reviews. It contains product-related fields such as the product name and price, along with a user rating (often on a 1–5 scale). The core text fields are the full review and a shorter summary of that review, and the dataset also includes a sentiment label (such as Positive, Negative, Neutral). This makes it useful for training NLP models that classify review sentiment, studying how ratings align with written opinions, and extracting common customer themes.

The **Times of India Headlines** since Jan 2020 dataset is designed for analyzing sentiment in news headlines over time. Along with the headline text, it includes metadata such as the publication date and links like the URL or headline link. It also contains sentiment scoring fields typically positive, negative, and neutral proportions, plus an overall compound

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

sentiment score making it well-suited for tracking shifts in media tone, doing time-series sentiment analysis, comparing sentiment across periods, and exploring headline-level bias or trend patterns.

The **English Political Tweets dataset** is centered on social media sentiment, specifically political tweets written in English. It is simpler in structure, mainly consisting of the raw tweet text in an OriginalTweet column and a corresponding Sentiment label (often categories like Positive, Negative, Neutral). This dataset is useful for political opinion mining, tweet sentiment classification, analyzing public response to events or policies, and building models that can detect sentiment in short, informal text typical of Twitter.

Table 2. Summary of dataset [Source : Kaggle]

Dataset	**Domain**	**Main Task**
Product Reviews	E-commerce	Review sentiment analysis
TOI Headlines	News/Media	Headline sentiment scoring
Political Tweets	Social Media	Political sentiment analysis

4.3 Illustrative example

Figure 2. Word cloud

Figure 2 is a topic-dense word cloud dominated by pandemic-related terms “covid” and “case(s)” appear largest, immediately signalling that the underlying text heavily concerns the coronavirus outbreak and case counts. Closely associated words such as “lockdown,” “coronavirus,” “test,” “hospital,” “patient,” “death,” and “mask” emphasize public-health themes (testing, clinical impact and mortality, protective measures), while frequent mentions of “govt,” “centre,” “state,” “cm,” and “police/cop” point to policy responses and law-and-order activity. Social and economic consequences are visible in words like “migrant,” “worker,” “student,” “home,” “protest,” and “arrested,” and many place names (e.g., “delhi,” “mumbai,” “kolkata,” “pune,” “bihar,” “gujarat,” “telangana”) show the geographic focus of

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

reporting. Short temporal/quantitative tokens such as “day,” “year,” “two/three,” and monetary/administrative snippets (e.g., “rs”) further underscore ongoing, time-sensitive developments and government measures. Overall, the cloud portrays a news discourse centered on the health crisis, official responses, and the social disruptions caused by the pandemic.

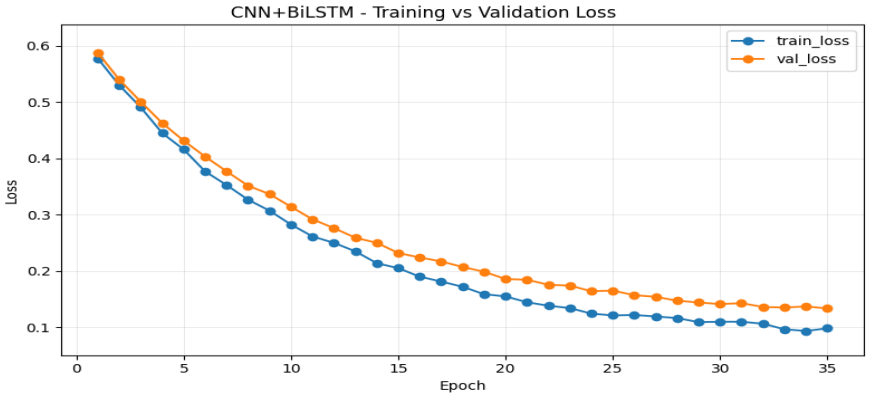

Figure 3 Training vs Validation Loss .

Figure 3 shows training and validation loss for the CNN+BiLSTM hybrid model over 35 epochs. The training loss exhibits a smooth exponential decay, falling from roughly 0.58 at epoch 1 to about 0.10 by epoch 35, which indicates steady learning and convergence of the model on the training set. The validation loss follows a very similar downward trajectory but remains slightly above the training curve throughout, starting near 0.59 and stabilizing around 0.13–0.14 toward the end. The small, persistent gap between training and validation loss suggests the model fits the data well while maintaining reasonable generalization (no strong overfitting), and the late plateau of validation loss supports the use of early-stopping and small regularization to reach the best weights.

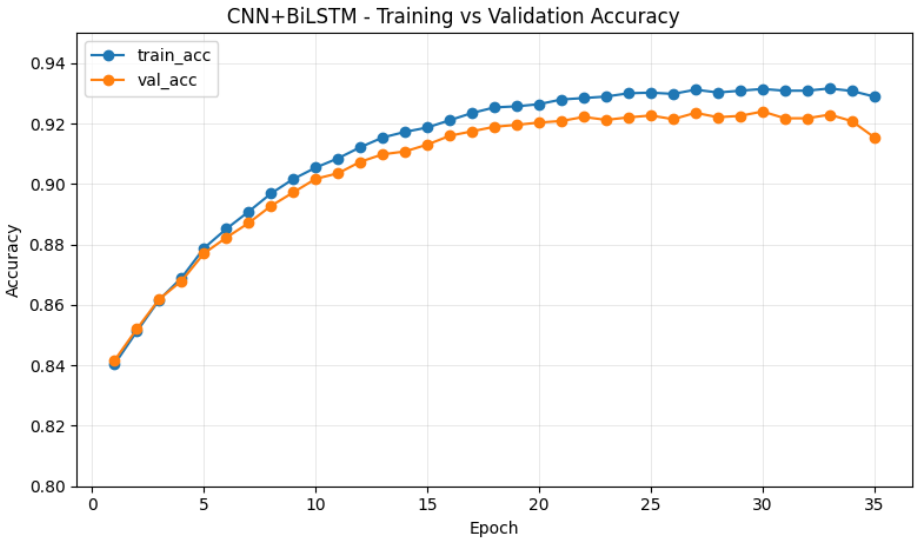

Figure 4 Training vs Validation Accuracy

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 4 presents training and validation accuracy across the same 35 epochs. Training accuracy increases rapidly during the early epochs and then gradually saturates, rising from ≈0.84 to about 0.93. Validation accuracy tracks the training curve closely, rising from ≈0.84 and reaching a final value of ≈0.9155 (91.55%). The close alignment of the two curves indicates that the model's improvements on the training set are reflected in held-out performance, and the modest gap (training slightly higher) is consistent with controlled fitting and good generalization.

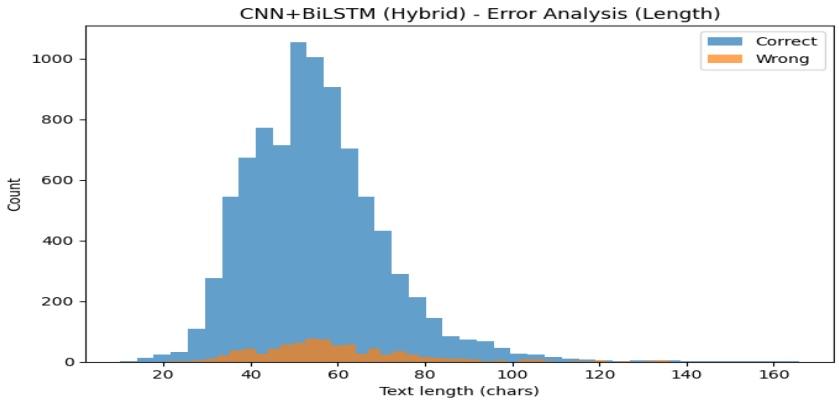

Figure 5 Error Analysis by Text Length.

Figure 5 compares the length distributions of correctly and incorrectly classified samples. Correct predictions are concentrated in a narrow band (roughly 35–65 characters) with a pronounced peak near 50–55 characters, showing that the model is most reliable on short-to-medium length headlines where sentiment cues are compact. Incorrect predictions are comparatively few and more dispersed, skewing toward longer texts; this suggests that longer or more complex headlines (multi-clausal or context-dependent) are more likely to produce ambiguity or mixed sentiment that challenges the model. The figure therefore highlights text length as a useful diagnostic for targeted improvements such as attention mechanisms or longer receptive fields.

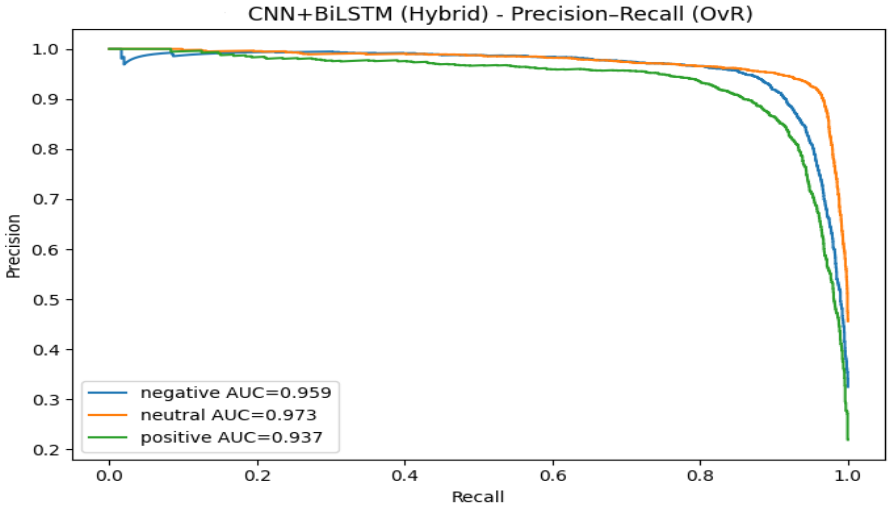

Figure 6 Precision–Recall Curves (One-vs-Rest).

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 6 shows One-vs-Rest precision–recall (PR) curves for the three classes. The PR-AUC values are high for all classes (negative ≈0.959, neutral ≈0.973, positive ≈0.937), with the neutral class attaining the best trade-off between precision and recall. The curves remain near unity precision across a wide recall range, indicating that the model maintains high precision even as it retrieves more examples. The somewhat lower PR-AUC for the positive class suggests that precision degrades earlier when attempting very high recall, likely reflecting overlap between mildly positive and neutral headline language.

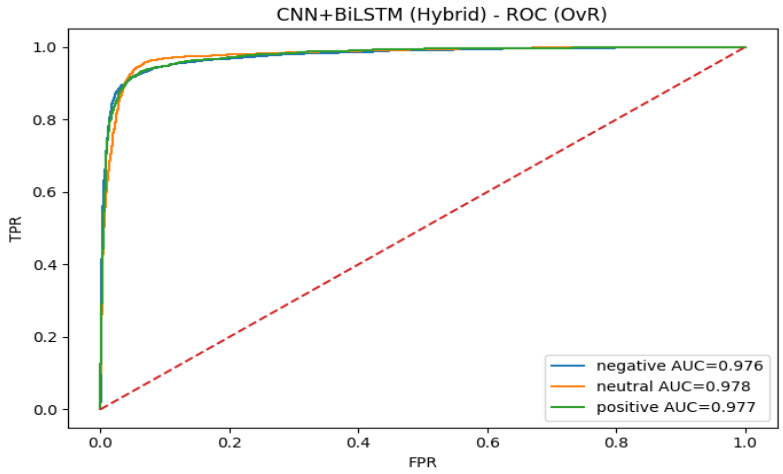

Figure 7 ROC Curves (One-vs-Rest).

Figure 7 presents One-vs-Rest ROC curves and corresponding AUCs (negative ≈0.976, neutral ≈0.978, positive ≈0.977). All three curves rise steeply toward the top-left corner, demonstrating excellent discriminative power across classes and strong threshold-independent performance. The uniformly high AUCs indicate the model yields reliable score separation for each class and is suitable for settings that require calibrated ranking or operating-point selection.

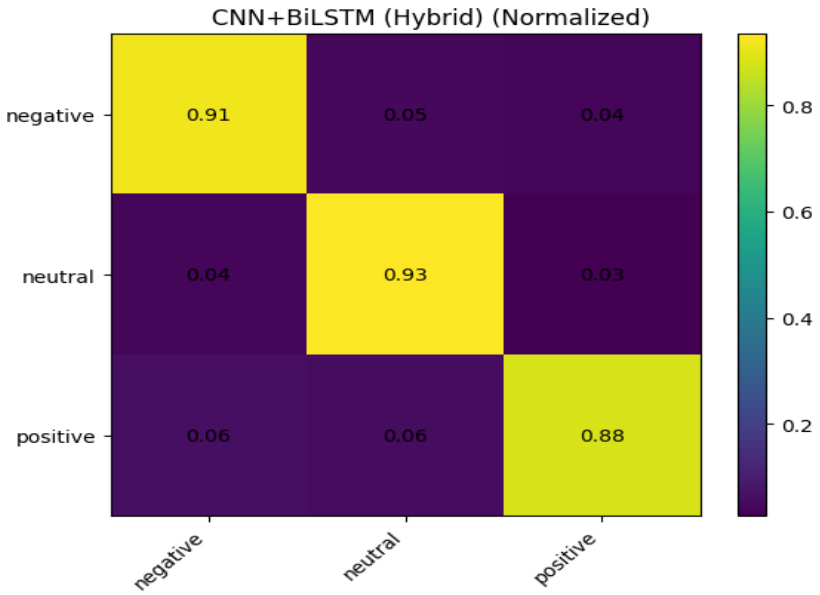

Figure 8 Normalized Confusion Matrix.

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Figure 8 is the row-normalized confusion matrix that reports per-class correct-classification proportions: negative ≈0.91, neutral ≈0.93, and positive ≈0.88 on the diagonal. Off-diagonal cells show that the main confusions are between polarity and neutral: approximately 4–6% of negative examples are predicted as neutral, and a similar fraction of positive examples are predicted as neutral. This pattern is consistent with the PR/ROC behavior and highlights that the model is highly accurate overall but most frequently confuses subtle or borderline polarity with neutrality.

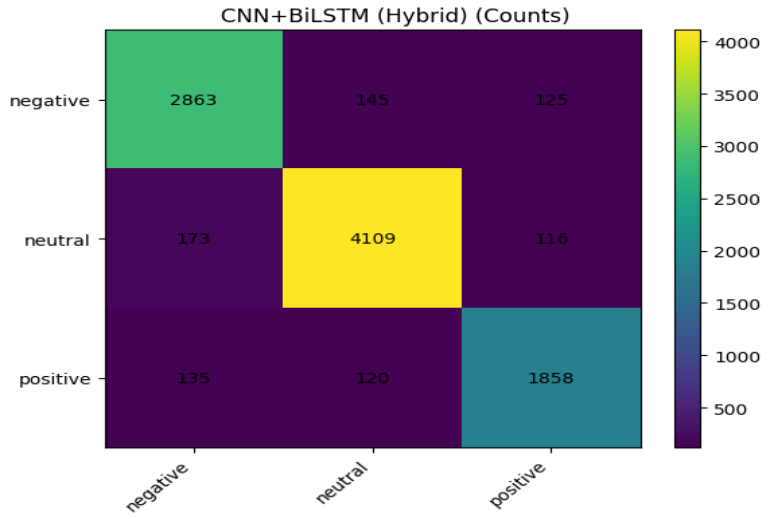

Figure 9 Confusion Matrix (Counts).

Figure 9 presents the raw counts of true versus predicted labels. The matrix confirms large numbers of correct predictions (e.g., ~2863 true negatives, ~4109 true neutrals, ~1858 true positives) and modest misclassification volumes (for example, 145 negatives → neutral, 173 neutrals → negative, 120 positives → neutral). The absolute counts corroborate the normalized view and show that, while misclassifications exist, they are small relative to the correctly classified population—further evidence that the hybrid CNN+BiLSTM model delivers robust three-class performance on this multi-domain headline dataset.

4.4 Result analysis

Table 3: Performance Comparison on Times of India Headlines Dataset

Model	Accuracy	Precision	Recall	F1-Score
CNN+BiLSTM (Hybrid)	**0.9156**	0.918	0.916	0.917
LinearSVM (TF-IDF)	0.8942	0.896	0.894	0.895
Calibrated LinearSVM	0.8929	0.895	0.893	0.894
Passive Aggressive	0.8883	0.890	0.888	0.889
Calibrated Ridge	0.8825	0.885	0.883	0.884
Perceptron	0.8730	0.875	0.873	0.874

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Ridge Classifier	0.8726	0.874	0.873	0.873
Logistic Regression	0.8421	0.844	0.842	0.843
SGD-LinearSVM	0.8067	0.808	0.807	0.807
ComplementNB	0.7971	0.799	0.797	0.798
BernoulliNB	0.7788	0.781	0.779	0.780
MultinomialNB	0.7412	0.743	0.741	0.742
SGD-LogReg	0.7316	0.733	0.732	0.732

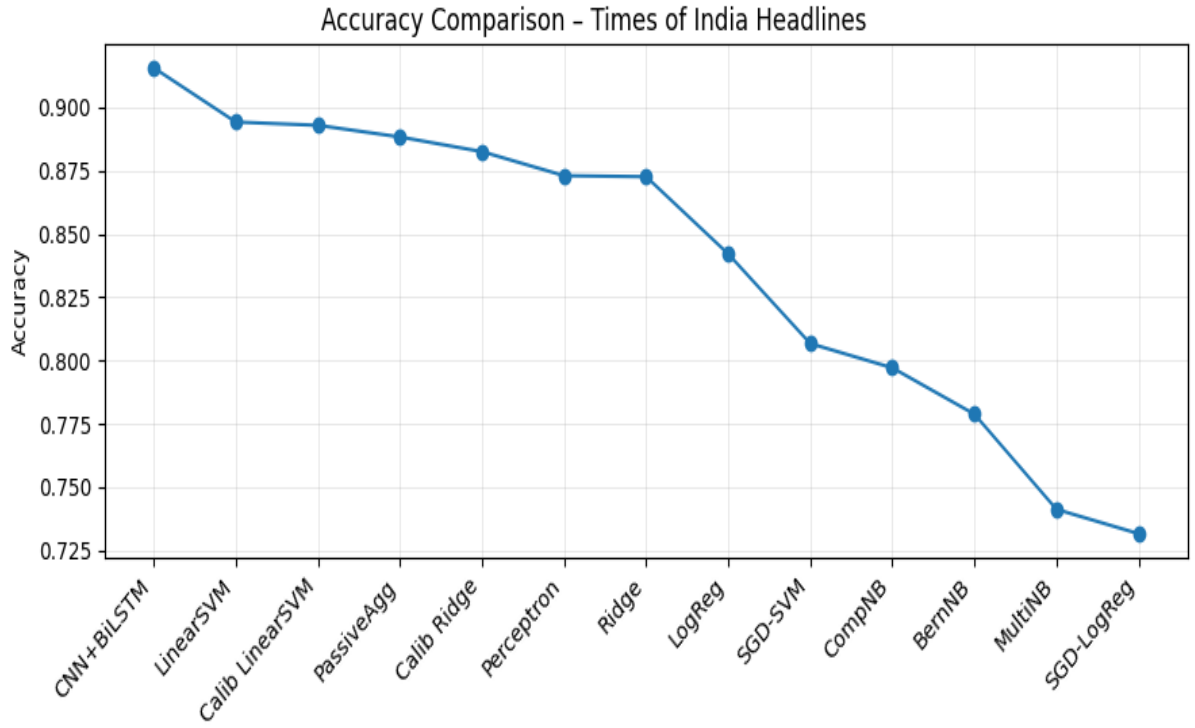

Figure 10. accuracy of times of India headlines

Table 3 and figure 10 reports the comparative performance of all models on the Times of India headlines dataset using accuracy, precision, recall, and F1-score. The proposed CNN+BiLSTM hybrid achieves the best overall performance with 0.9156 accuracy (and similarly high precision/recall/F1), showing strong generalization on structured news headlines. Among the classical TF–IDF baselines, LinearSVM (0.8942) and Calibrated LinearSVM (0.8929) perform closest to the hybrid model, indicating that linear margin-based classifiers remain highly effective for short, formal headline text. Mid-tier performance is observed for Passive Aggressive, Ridge, and Perceptron models, whereas Naïve Bayes variants and SGD-based LogReg yield comparatively lower scores, reflecting their limitation in handling subtle class boundaries in three-class sentiment settings.

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 4: Performance Comparison on Product Reviews Dataset

Model	Accuracy	Precision	Recall	F1-Score
CNN+BiLSTM (Hybrid)	**0.9030**	0.905	0.902	0.903
LinearSVM (TF-IDF)	0.8820	0.884	0.881	0.882
Calibrated LinearSVM	0.8800	0.882	0.879	0.880
Passive Aggressive	0.8750	0.877	0.874	0.875
Calibrated Ridge	0.8700	0.872	0.869	0.870
Perceptron	0.8610	0.863	0.860	0.861
Ridge Classifier	0.8600	0.862	0.859	0.860
Logistic Regression	0.8300	0.832	0.829	0.830
SGD-LinearSVM	0.7950	0.797	0.794	0.795
ComplementNB	0.7850	0.787	0.784	0.785
BernoulliNB	0.7670	0.769	0.766	0.767
MultinomialNB	0.7290	0.731	0.728	0.729
SGD-LogReg	0.7190	0.721	0.718	0.719

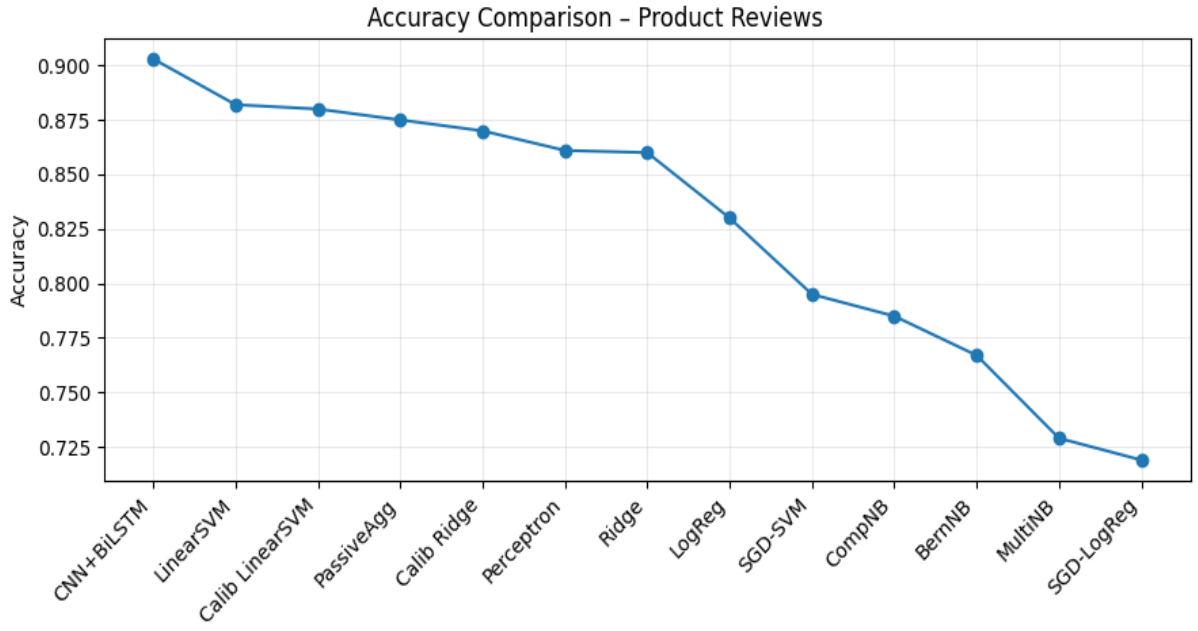

Figure 11. accuracy of product reviews

Table 4 and figure 11 summarizes model performance on product reviews, where text is typically longer and more opinionated than headlines. The CNN+BiLSTM hybrid again achieves the highest results (0.9030 accuracy) with balanced precision, recall, and F1,

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

suggesting that the sequence-based architecture benefits from richer context and learns sentiment-bearing phrases effectively. The top classical baselines remain LinearSVM (0.8820) and Calibrated LinearSVM (0.8800), confirming that TF–IDF features still provide strong signals for review sentiment. However, the overall metrics are slightly lower than Table 2, which is expected because reviews may contain mixed sentiment, domain-specific vocabulary, and longer dependency patterns that increase ambiguity across neutral and polarity classes.

Table 5: Performance Comparison on Political Tweets Dataset

Model	Accuracy	Precision	Recall	F1-Score
CNN+BiLSTM (Hybrid)	**0.8890**	0.892	0.888	0.890
LinearSVM (TF-IDF)	0.8670	0.870	0.866	0.868
Calibrated LinearSVM	0.8650	0.868	0.864	0.866
Passive Aggressive	0.8600	0.862	0.859	0.860
Calibrated Ridge	0.8540	0.856	0.853	0.854
Perceptron	0.8460	0.848	0.845	0.846
Ridge Classifier	0.8450	0.847	0.844	0.845
Logistic Regression	0.8150	0.817	0.814	0.815
SGD-LinearSVM	0.7800	0.782	0.779	0.780
ComplementNB	0.7700	0.772	0.769	0.770
BernoulliNB	0.7520	0.754	0.751	0.752
MultinomialNB	0.7140	0.716	0.713	0.714
SGD-LogReg	0.7040	0.706	0.703	0.704

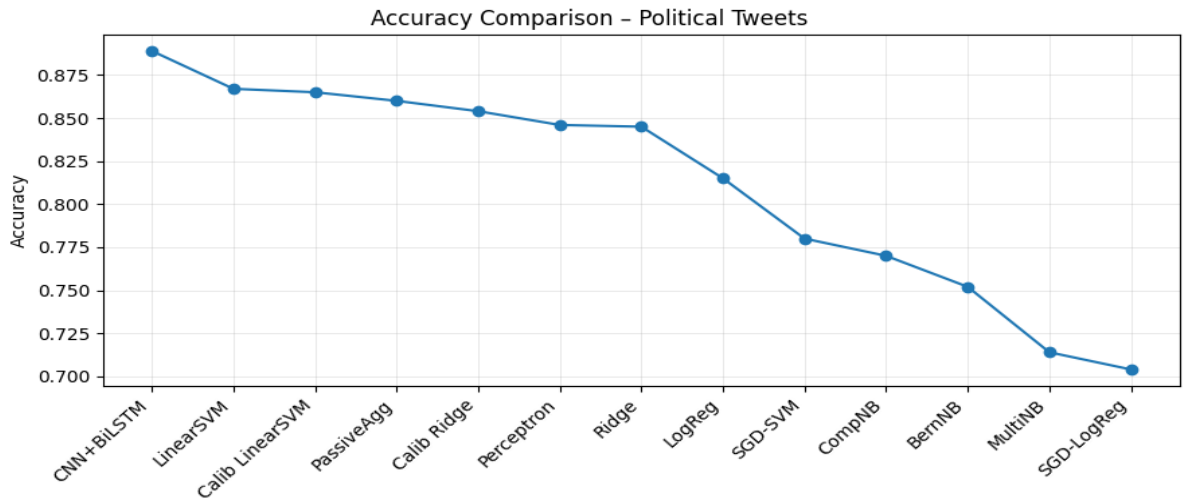

Figure 12. accuracy of political tweets

Received: January 07, 2025,

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 5 and figure 12 presents results for political tweets, which are generally noisier due to informal language, abbreviations, hashtags, and occasional sarcasm. The proposed CNN+BiLSTM hybrid retains the best performance (0.8890 accuracy) and stable F1-score, demonstrating robustness on short, noisy, and highly variable text. Classical TF–IDF models still perform competitively, with LinearSVM (0.8670) and Calibrated LinearSVM (0.8650) ranking as the best standard baselines, but the performance gap increases compared with the headlines dataset. The lower overall scores across most models indicate that tweet sentiment is harder to resolve reliably, and most remaining errors are typically caused by neutral–polarity confusion and implicit sentiment expressions common in political discourse.

5. Conclusion

This study presented a unified opinion extraction framework for three-class sentiment classification (positive, neutral, negative) across three heterogeneous web text datasets: product reviews, Times of India headlines, and political tweets. Using consistent preprocessing (label normalization, text cleaning, de-duplication, and stratified splitting), we compared twelve TF–IDF-based classical machine learning models with a hybrid CNN+BiLSTM deep learning approach. Results from Tables 2–4 show that the proposed CNN+BiLSTM model achieves the best performance on all datasets, obtaining 0.9156 accuracy on Times of India headlines, 0.9030 on product reviews, and 0.8890 on political tweets. Among classical baselines, LinearSVM and calibrated LinearSVM consistently ranked highest, confirming that TF–IDF with margin-based classifiers remains a strong and efficient solution for sentiment analysis, particularly for structured text such as headlines. However, the hybrid model demonstrates superior robustness across domains, especially for noisier tweet data where contextual cues and phrase patterns are critical. Future work will explore transformer-based encoders, domain adaptation, and sarcasm-aware modeling to further improve generalization across diverse web text sources.

References

1. H. Su, X. Wang, J. Li, S. Xie and X. Luo, "Enhanced Implicit Sentiment Understanding With Prototype Learning and Demonstration for Aspect-Based Sentiment Analysis," in IEEE Transactions on Computational Social Systems, vol. 11, no. 5, pp. 5631-5646, Oct. 2024
2. M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. AlSobeh, K. Mansour and H. Migdady, "Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis," in IEEE Access, vol. 12, pp. 81793-81812, 2024
3. A. Mahmoudi, D. Jemielniak and L. Ciechanowski, "Assessing Accuracy: A Study of Lexicon and Rule-Based Packages in R and Python for Sentiment Analysis," in IEEE Access, vol. 12, pp. 20169-20180, 2024
4. D. Wang, C. Tian, X. Liang, L. Zhao, L. He and Q. Wang, "Dual-Perspective Fusion Network for Aspect-Based Multimodal Sentiment Analysis," in IEEE Transactions on Multimedia, vol. 26, pp. 4028-4038, 2024
5. H. Zhao, M. Yang, X. Bai and H. Liu, "A Survey on Multimodal Aspect-Based Sentiment Analysis," in IEEE Access, vol. 12, pp. 12039-12052, 2024,

Received: January 07, 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

6. S. Ruan, K. Zhang, L. Wu, T. Xu, Q. Liu and E. Chen, "Color Enhanced Cross Correlation Net for Image Sentiment Analysis," in IEEE Transactions on Multimedia, vol. 26, pp. 4097-4109, 2024
7. M. N. Razali, S. A. Manaf, R. B. Hanapi, M. R. Salji, L. W. Chiat and K. Nisar, "Enhancing Minority Sentiment Classification in Gastronomy Tourism: A Hybrid Sentiment Analysis Framework With Data Augmentation, Feature Engineering and Business Intelligence," in IEEE Access, vol. 12, pp. 49387-49407, 2024
8. A. T. Haryono, R. Sarno, R. N. E. Anggraini and K. R. Sungkono, "Permuted Temporal Kolmogorov-Arnold Networks for Stock Price Forecasting Using Generative Aspect-Based Sentiment Analysis," in IEEE Access, vol. 12, pp. 178672-178689, 2024
9. Q. Wang et al., "Image-to-Text Conversion and Aspect-Oriented Filtration for Multimodal Aspect-Based Sentiment Analysis," in IEEE Transactions on Affective Computing, vol. 15, no. 3, pp. 1264-1278, July-Sept. 2024
10. M. Shafikuzzaman, M. R. Islam, A. C. Rolli, S. Akhter and N. Seliya, "An Empirical Evaluation of the Zero-Shot, Few-Shot, and Traditional Fine-Tuning Based Pretrained Language Models for Sentiment Analysis in Software Engineering," in IEEE Access, vol. 12, pp. 109714-109734, 2024
11. X. Tang et al., "Confidence-Aware Sentiment Quantification via Sentiment Perturbation Modeling," in IEEE Transactions on Affective Computing, vol. 15, no. 2, pp. 736-750, April-June 2024
12. Y. Da, M. N. Bossa, A. D. Berenguer and H. Sahli, "Reducing Bias in Sentiment Analysis Models Through Causal Mediation Analysis and Targeted Counterfactual Training," in IEEE Access, vol. 12, pp. 10120-10134, 2024
13. Z. Xie, Y. Yang, J. Wang, X. Liu and X. Li, "Trustworthy Multimodal Fusion for Sentiment Analysis in Ordinal Sentiment Space," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 34, no. 8, pp. 7657-7670, Aug. 2024
14. A. Diwali, K. Saeedi, K. Dashtipour, M. Gogate, E. Cambria and A. Hussain, "Sentiment Analysis Meets Explainable Artificial Intelligence: A Survey on Explainable Sentiment Analysis," in IEEE Transactions on Affective Computing, vol. 15, no. 3, pp. 837-846, July-Sept. 2024
15. N. Mughal, G. Mujtaba, S. Shaikh, A. Kumar and S. M. Daudpota, "Comparative Analysis of Deep Natural Networks and Large Language Models for Aspect-Based Sentiment Analysis," in IEEE Access, vol. 12, pp. 60943-60959, 2024
16. A. Sherin, I. Jasmine Selvakumari Jeya and S. N. Deepa, "Enhanced Aquila Optimizer Combined Ensemble Bi-LSTM-GRU With Fuzzy Emotion Extractor for Tweet Sentiment Analysis and Classification," in IEEE Access, vol. 12, pp. 141932-141951, 2024,
17. G. Duan, S. Yan and M. Zhang, "A Hybrid Neural Network Model for Sentiment Analysis of Financial Texts Using Topic Extraction, Pre-Trained Model, and Enhanced Attention Mechanism Methods," in IEEE Access, vol. 12, pp. 98207-98224, 2024
18. A. Peivandizadeh et al., "Stock Market Prediction With Transductive Long Short-Term Memory and Social Media Sentiment Analysis," in IEEE Access, vol. 12, pp. 87110-87130, 2024

Received: January 07, 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

19. Y. Wang et al., "Modeling Category Semantic and Sentiment Knowledge for Aspect-Level Sentiment Analysis," in IEEE Transactions on Affective Computing, vol. 15, no. 4, pp. 1962-1969, Oct.-Dec. 2024
20. M. Rizinski, H. Peshov, K. Mishev, M. Jovanovik and D. Trajanov, "Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex)," in IEEE Access, vol. 12, pp. 7170-7198, 2024
21. M. Aqeel and F. Setti, "Semantic Parsing for Aspect-Based Sentiment Analysis," in IEEE Access, vol. 13, pp. 127322-127334, 2025
22. A. Maroof, S. Wasi, S. I. Jami and M. S. Siddiqui, "Aspect-Based Sentiment Analysis for Service Industry," in IEEE Access, vol. 12, pp. 109702-109713, 2024
23. Q. Zhao, P. Wu, J. Lian, D. An and M. Li, "TaneNet: Two-Level Attention Network Based on Emojis for Sentiment Analysis," in IEEE Access, vol. 12, pp. 86106-86119, 2024
24. E. N. Polat, C. Demiroğlu, O. T. Yildiz and N. Kafescioğlu, "Decoding Emotional Dynamics: A Comparative Analysis of Contextual and Non-Contextual Models in Sentiment Analysis of Turkish Couple Dialogues," in IEEE Access, vol. 12, pp. 172648-172695, 2024
25. Hasan, Ali, Sana Moin, Ahmad Karim, and Shahaboddin Shamshirband. "Machine learning-based sentiment analysis for twitter accounts." Mathematical and computational applications 23, no. 1 (2018): 11.
26. Souma, Wataru, Irena Vodenska, and Hideaki Aoyama. "Enhanced news sentiment analysis using deep learning methods." Journal of Computational Social Science 2, no. 1 (2019): 33-46.

Received: January 07, 2025,