MBH-Sent: A Noise-Aware Multi-Branch Hybrid Framework for Sentiment Analysis of Noisy and Dialectal Bangla Text

ABSTRACT

Sentiment analysis is used to identify whether a text expresses a positive, negative, or neutral opinion. Although this research has been widely conducted, research on Bengali remains limited, especially for noisy social media comments. These comments often include spelling mistakes, local dialects, emojis, and unstructured grammar, which make analysis more difficult. This study proposes a noise-aware preprocessing pipeline and a hybrid framework that combines different deep learning models. The system uses convolutional neural networks (CNNs), recurrent neural networks (BiLSTM and GRU), and Transformer-based models to capture different types of information from text. It also brings together multiple feature representations, including TF–IDF features, traditional word embeddings, and contextual embeddings, using an attention-based fusion approach. Several model architectures are evaluated, including standalone recurrent models, CNN–RNN hybrid models, and fine-tuned Transformer models. In addition, a lightweight DistilBERT model designed for mobile devices and low-resource settings is examined. Among all models, the Multi-Branch Hybrid CNN + BiLSTM + GRU + RoBERTa model with full feature fusion achieves the best performance, reaching 94.5% accuracy, 95.1% precision, 93.9% recall, and a 94.4% F1-score on the SentNoB dataset. These results show that combining local patterns, sequential information, and contextual meaning in a single framework leads to more accurate and reliable sentiment analysis for noisy Bangla text. The proposed approach is suitable for real-world applications such as social media analysis, customer feedback evaluation, and public opinion monitoring.

KEYWORDS

Sentiment Analysis, Noisy Bangla Text, Social Media Analytics, Hybrid Deep Learning, Transformer-Based Models, Multi-Branch Fusion, Attention Mechanisms

ACM Reference Format:

. 2026. MBH-Sent: A Noise-Aware Multi-Branch Hybrid Framework for Sentiment Analysis of Noisy and Dialectal Bangla Text . In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Sentiment analysis is the process of extracting and identifying opinions or emotions expressed in text, allowing systems to determine whether the feeling is positive, negative, or neutral [33]. In addition to these three general categories of sentiment, sentiment can also be classified across individual emotions, such as happiness, anger, sadness, fear, sarcasm, and disgust, which can provide deeper insights into human emotions [33]. Linguistic structure, including syntax, morphology, and semantics, provides essential information for reliable sentiment classification in different textual contexts.

Although sentiment analysis has been widely explored in English, it has not yet received sufficient attention in the context of Bengali. This is noteworthy, as Bengali is the seventh most widely spoken language in the world, with around 285 million speakers [24]. The language itself poses significant challenges for natural language processing (NLP). Its morphologically rich nature, including verb forms and compound word structures, makes it unsuitable for traditional NLP methods [16]. In addition, Bengali follows a subject–object–verb (SOV) word order, which adds further complexity when developing computational models. The presence of several regional dialects and the widespread use of informal Bangla text on social media also result in considerable variability [15]. Sentiment classification is further complicated by spelling variations and noisy user-generated content, which makes Bengali sentiment analysis particularly challenging.

To address such problems, most approaches rely on preprocessing pipelines that include tokenization, stemming, stopword removal, and noise reduction. The effectiveness of these steps plays a vital role in model performance [5]. With the rapid growth of online communication, especially in the post-COVID-19 era, the demand for accurate sentiment analytics has increased. Businesses, regulators, and policymakers are now increasingly interested in gaining insights from social media posts, product reviews, and public opinion [19].

However, two major obstacles remain: the scarcity of large and balanced publicly available datasets for Bengali sentiment analysis, and the challenge of modelling the linguistic complexity and regional diversity of noisy text [17]. Traditional machine learning methods, including SVM, Random Forest, and Logistic Regression, as well as deep learning methods such as CNN, LSTM, and BiLSTM, have been used in previous studies. Yet such approaches are not optimal for the compositional structure of Bengali text and may not generalise across domains [3]. Embedding methods such as Word2Vec, GloVe, and FastText, along with Transformer-based models such as BanglaBERT, RoBERTa, and XLM-R, have shown promising results, though they still face difficulties in handling noisy and informal language [3, 10, 19].

In this paper, we present a novel hybrid deep learning model designed for Bengali sentiment analysis. Our method fuses lexical (TF–IDF), semantic (Word2Vec, GloVe, FastText), and contextual representations (BanglaBERT, RoBERTa, XLM-R) through a multi-representation attention fusion mechanism. A noise-sensitive preprocessing pipeline has been employed to handle spelling errors, dialectal variations, and emoji-based sentiment cues, making the framework more robust in real-world scenarios.

The main contributions of this study are as follows:

- A noise-aware preprocessing pipeline for noisy Bangla social media text that applies spelling correction, dialect normalization, sentiment-preserving emoji mapping, and removal of irregular tokens.
- A multi-representation framework that fuses TF–IDF features, static embeddings (Word2Vec, FastText, GloVe), and contextual embeddings (BanglaBERT, RoBERTa, DistilBERT, XLNet) through an attention-based mechanism.
- A systematic evaluation of multiple architectures—RNNs (LSTM, BiLSTM, GRU, BiGRU), hybrid CNN–RNN models, Transformer models, and a multi-branch hybrid approach on the SentNoB dataset.
- An adaptation of a CentralNet-style multi-branch hybrid fusion framework that integrates CNN, BiLSTM, GRU, and RoBERTa representations for noisy and dialectal Bangla sentiment analysis, achieving 94.5% accuracy, 95.1% precision, 93.9% recall, and a 94.4% F1-score on the SentNoB dataset.
- An adaptation of a CentralNet-style multi-branch hybrid fusion framework that integrates CNN, BiLSTM, GRU, and RoBERTa representations for robust sentiment analysis of noisy and dialectal Bangla text.

The subsequent sections of this paper are organised as follows: Section II examines pertinent literature; Section III delineates the proposed model architecture; Section IV addresses experimental outcomes and performance evaluations; and Section V concludes the investigation with recommendations for future research.

2 RELATED WORK

Advanced deep learning and Transformer-based models have recently gained significant attention in sentiment analysis, particularly for multilingual and low-resource languages. Although these models demonstrate strong performance, several challenges remain. User-generated content is often noisy, datasets may suffer from class imbalance, and models frequently struggle to generalize across domains. In addition, morphologically rich languages such as Bangla introduce further complexity for accurate sentiment classification. As summarized in Table 1, existing studies highlight these limitations, indicating the need for more effective and noise-aware sentiment analysis approaches.

Existing sentiment analysis models, including hybrid deep learning methods and advanced Transformer-based architectures, continue to face notable methodological and deployment challenges. Limited computational resources and the prevalence of noisy user-generated text pose significant difficulties for low-resource languages such as Bangla. Moreover, further research is required to enhance model interpretability and to assess the feasibility of real-time and large-scale deployment. To address these challenges, this study proposes a hybrid multi-representation fusion framework that integrates multiple embedding types with a noise-aware preprocessing pipeline and Transformer-based architectures. The proposed approach aims to provide a unified, robust, and scalable solution for sentiment analysis of noisy Bangla text.

3 METHODOLOGY

This section describes the SentNoB dataset, preprocessing techniques, and the proposed model architectures. We implemented a multi-branch hybrid model combining CNNs, RNNs (BiLSTM/GRU), and Transformer models (BanglaBERT, RoBERTa, XLNet) for sentiment analysis of noisy Bangla text. This model captures spatial, temporal, and contextual information, making it suitable for real-world analysis of Bengali. The SentNoB dataset contains Bengali word-for-word comments labeled as positive, negative, or neutral. The data is divided into training, validation, and test sets. We apply a noise-aware preprocessing pipeline, which includes spelling correction, dialect normalization, and emoji mapping to clean and prepare the text. The data is temporally localized and augmented using meaningful substitution and token swapping techniques. The multi-branch hybrid model features a CNN algorithm for discrete features, an RNN (BILSTM/GRU) for capturing temporal networks, and a transformer of long-range relevance. The outputs of each branch are fused using an attention-based blending mechanism, improving robustness and classification performance. The blended features are passed through the middle of the entire edge and are combined. This hybrid model provides an effective solution for the review of lengthy and dialect-mixed Bangla texts, suitable for applications such as feedback and feedback from users of the society.

3.1 Dataset Overview

This paper uses the SentNoB dataset (Sentiment Analysis in Noisy Bangla Texts) itself for experimentation, which is a publicly available dataset on Kaggle [8]. Data was provided from social media and other user-generated platforms across 13 domains (e.g., politics, education, agriculture). For each comment, a sentiment class is assigned, labeled as: 0 = Neutral, 1 = Positive, and 2 = Negative.

The dataset is divided into three predefined splits: 12,575 samples for training, 1,567 samples for validation, and 1,586 samples for testing. The distribution of sentiment classes is presented in Table 2. Overall, the dataset shows a relatively balanced class distribution. Although the Positive class contains slightly more samples, the Neutral and Negative classes are represented in comparable proportions across all splits, reducing the risk of severe class imbalance during model training and evaluation.

3.2 Dataset Preprocessing

The SentNoB dataset consists of noisy Bangla social media comments that include spelling errors, dialectal forms, English words, emojis, and irregular punctuation. We preprocess the text using a structured pipeline to make it suitable for modelling.

Table 1: SUMMARY OF ALL THE RELATED LITERATURE

Author Name	Dataset	Model Name	Accuracy	Limitations
A. A. Syed *et al.* [29]	Skytrax review portal (3 categories: abstractive summarization, domain adaptation, rating-based sentiment classification)	Two-step framework using PEGASUS + BERT	89%	Limited to airline reviews; dataset imbalance; needs augmentation.
P. A. Henriquez *et al.* [13]	Custom dataset of 2.118M tweets (COVID-19, social uprising, wildfires)	Deep Random Vector Functional Link (D-RVFL)	78.30%	Limited generalization; transient nature of events; potential bias from Twitter APIs.
V. KP. *et al.* [30]	Twitter API dataset + SST-2	Ensemble Classifier (RF + SVM + DT), LDA, ConvBiLSTM, Hybrid Lexicon–Naive Bayes	93.42%	High variance due to tree depth; limited scalability across domains.
R. Haque *et al.* [11]	Custom Bengali dataset (42,036 social media comments, 566K words)	CNN–LSTM (CLSTM) + Flask Web App + TF–IDF	Accuracy: 85.85%, F1: 0.86	Overlapping class labels; needs transfer learning (e.g., RoBERTa, DistilBERT, FastText).
N. A. Semary *et al.* [28]	IMDb reviews + US Airline tweets	Hybrid RoBERTa (CNN + LSTM) + SMOTE	IMDb: 96.28%; Twitter: 94.2%	Limited cross-lingual exploration; lacks integration with multilingual Transformer-based architectures.
R. Ahamad *et al.* [1]	Twitter + Amazon reviews	ESIHE_AML framework with CNN–BiLSTM for image-to-text sentiment extraction	90%+	Requires improvements in preprocessing, augmentation, and overall model refinement.
K. I. Islam *et al.* [17]	SentNoB (≈15K noisy Bangla social media comments across 13 domains)	SVM (TF–IDF), BiLSTM + Attention, mBERT (baseline benchmarking)	Best F1 ≈ 64.6%	Primarily focuses on dataset construction and baseline models; neural and Transformer models underperform on noisy text; lacks hybrid architectures and advanced noise-aware fusion.
K. T. Elahi *et al.* [9]	NC-SentNoB (≈15K noisy Bangla texts with 10 noise categories)	SVM, BiLSTM, BanglaBERT, BanglaBERT-Large, Bangla-ELECTRA, MuRIL with noise reduction methods	Best Micro-F1: 0.75 (noisy text), 0.73 (after noise reduction)	Noise reduction methods show limited effectiveness and do not improve sentiment performance on noisy Bangla text.
M. E. Islam *et al.* [18]	SentiGOLD (70K Bangla texts from 30 domains, 5 sentiment classes)	BanglaBERT, BiLSTM, HAN, CNN–BiLSTM with Attention	Macro-F1: 0.62 (intra-dataset), 0.61 (cross-dataset on SentNoB)	Primarily emphasizes dataset construction and benchmarking; does not explore noise-aware or hybrid fusion models for noisy Bangla text.

Table 2: Number of Samples per Sentiment Class in Each Split

Split	Neutral (0)	Positive (1)	Negative (2)	Total
Train	2,894	5,133	4,548	12,575
Validation	354	623	590	1,567
Test	361	654	571	1,586

The text was first normalized to Unicode, lowercased, and cleaned of links, e-mails, user mentions, and extra punctuation. Hashtags were retained as plain words. The emoticons were mapped to the sentiment-preserving tags (<EMO_POS>, <EMO_NEG>, <EMO>). Common dialectal variations (e.g., হইতেছে → হচ্ছে) and widespread English chat words (e.g., valo → ভালো) were mapped to standard Bangla. Lexicon-based spelling correction accommodated minor variations, and stopwords were stripped except for sentiment-carrying words. After cleaning, the text was tokenized, followed by sentence segmentation where applicable. Standard augmentation techniques such as synonym replacement, token swaps/deletions, and character-level noise were used for data enrichment. For synonym replacement, a manually curated Bangla synonym list was used, and replacements were applied conservatively to non-sentiment-bearing content words to avoid altering the original sentiment polarity. Token swaps and deletions were applied with low probability to preserve semantic coherence. Character-level noise was introduced in a controlled manner to simulate common social media typing variations (e.g., character repetition or minor omissions), rather than arbitrary character insertion or removal.

All preprocessing operations were applied using fixed, rule-based transformations and consistent parameters, ensuring that the same pipeline was deterministically applied across the training, validation, and test splits.

Ambiguous emojis that may convey sarcastic or mixed sentiment were mapped to the neutral placeholder (<EMO>) rather than being forced into positive or negative categories, reducing the risk of sentiment misclassification.

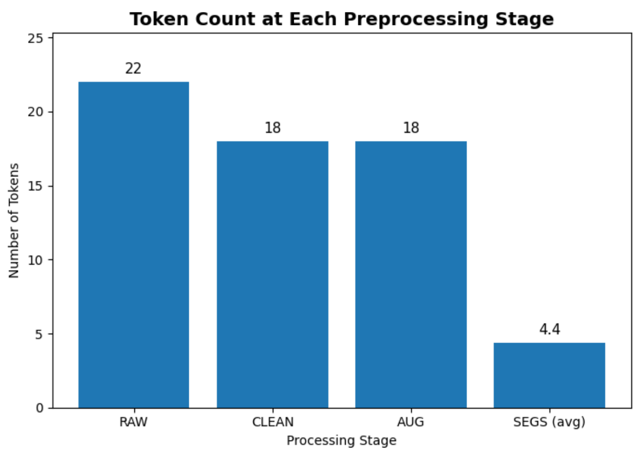

Figure 1: Average token count at each preprocessing stage

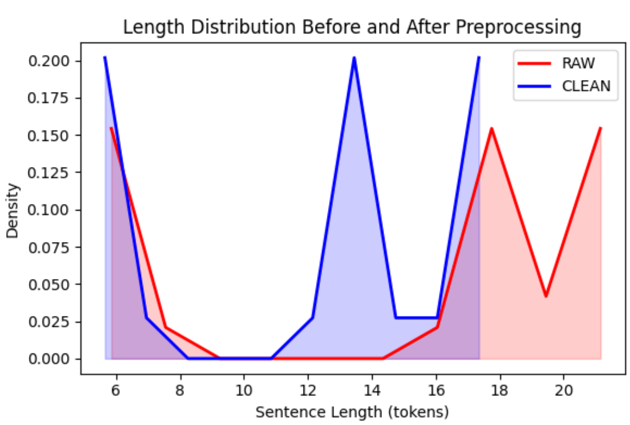

Figure 2: Density-based sentence length distribution before and after preprocessing

Figure 1 shows the reduction in average token counts across stages, while Figure 2 presents sentence length distributions before and after cleaning. Together, these results confirm that preprocessing reduces noise and increases consistency in the text. Representative examples are provided in Table 3, demonstrating how raw inputs are transformed into normalized and augmented variants that preserve sentiment while removing irrelevant noise.

3.3 Feature Extraction

In this work, we employ a multi-representation feature extraction strategy to capture the lexical, semantic, and contextual characteristics of noisy, dialect-mixed Bangla text in the SentNoB dataset. This method integrates TF–IDF features, static word embedding, contextual embedding, and a hybrid feature integration process, resulting in improved sentiment classification performance.

3.3.1 TF–IDF (Word + Character) with Safe Fallbacks. We apply Term Frequency–Inverse Document Frequency (TF–IDF) to both word and character sides to measure token importance relative to the score. The score is defined as:

$$\text{TF–IDF}(t, d) = \text{tf}(t, d) \times \log\left(\frac{N}{1 + \text{df}(t)}\right) \tag{1}$$

where $\text{tf}(t, d)$ is the frequency of term t in document d, N is the total number of documents, and $\text{df}(t)$ is the number of documents containing t [25].

Initial n-gram ranges are set to (1–3) for words and (2–5) for characters, while reduced ranges (1–1 for words and 2–4 for characters) ensure robustness of preprocessing. To reduce sparsity, truncated singular value decomposition (SVD) is applied to reduce the dimensionality to 128 for both word- and character-level features. In the case of very short messages, manually engineered character-counting vectors act as a fallback, providing stability against irregularities in noisy text.

3.3.2 Static Embeddings (Word2Vec, FastText, GloVe): We use pre-trained embeddings Word2Vec [22], FastText [4], and GloVe [23] to encode semantic relations in a vector space. Obtained by averaging pooling over speech-level embedding tokens. If the embedding resource is unavailable (e.g., missing files), zero-vector fallbacks are used to ensure stable performance in the presence of dialectal variation.

3.3.3 Contextual Embeddings (Transformers): Contextual discourse was extracted using transformer-based models, including Banglabert, RoBERTa, and XLM-R. Speech embeddings are either taken from [CLS] tokens (for BERT-like models) or from hidden mean pooling. This backend supports half-precision (FP16) calculations on the GPU for encoding, target sequence 128, and methods. If model loading is to be written, the zero-vector fallback guarantees positional robustness for different text structures [7][12].

3.3.4 Hybrid Feature Blocks with Attention Fusion. To integrate complementary representations, we design a hybrid architecture consisting of three parallel branches:

- A CNN applied to character-level TF–IDF vectors,
- A BiLSTM applied to static word embeddings, and

Table 3: Dataset Sample After Preprocessing

Label	Raw Text	Cleaned Text	Augmented Text
2 (Neg.)	“এই কুত্তার বাচ্চাদের জন্য দেশটা আজ এমন অবস্তায় ... মেরে পেলা দরকার”	খারাপ বাচ্চাদের দেশটা এমন অবস্তায় তিনটা পুলিশ তরে সবার সামনে মেরে খেলা দরকার	খেলাবাচ্চাদের দেশটা এমন অবস্তায় তিনটা পুলিশ ... খারাপ দরকার
1 (Pos.)	“ভাই আপনার কথাই যাদু রয়েছে”	ভাই আপনার কথাই যাদু রয়েছে	রয়েছে আপনার কথাই যাদু ভাই
0 (Neu.)	“যখন আমরা রাস্তায় বের হই তখন অনেক সময় রাস্তার কুকুর গেউ ... ভাই অনেক কথা বললাম ...”	রাস্তায় বের হই তখন অনেক সময় রাস্তার কুকুর গেউ গেউ ভাই ... ভুল হলে ক্ষমার দৃষ্টিতে দেখবেন	বের হই ভাই অনেক সময় রাস্তার কুকুর গেউ ... ভাই অনেক কথা বললা ...

- Transformer-based contextual embeddings derived from the [CLS] token.

Each branch generates a 128-dimensional feature vector. These vectors are then combined through an attention-based fusion mechanism:

$$H = \sum_{i=1}^{k} \alpha_i H_i \tag{2}$$

where H_i is the feature vector from the i^{th} branch, and α_i is its normalized attention weight. This mechanism dynamically emphasizes informative features, enabling the model to remain robust against noise, spelling variations, dialectal forms, and irregular expressions.

4 MODEL ARCHITECTURES AND FUSION STRATEGY

The study utilises recurrent models, hybrid CNN–RNN models, transformer models, and multi-branch hybrid architectures to manage noisy Bangla texts, which include dialectal variations, spelling errors, and irregular grammar. Hybrid models integrate the advantages of various methodologies, capturing local lexical patterns through CNN layers and long-range contextual dependencies through RNN or transformer layers. Transformer models provide robust context modelling capabilities, while recurrent models, such as LSTM and BiLSTM, focus on sequential information. To more accurately represent the complex characteristics of noisy Bangla text in the SentNoB dataset, multi-branch hybrids employ multiple feature extraction networks in parallel. To enhance the robustness and accuracy of Bangla text processing, the models are implemented with appropriate preprocessing, embeddings, and optimization techniques.

4.1 Recurrent Models

Sequential data, including natural language text, is well-suited for processing by recurrent neural networks (RNNs). In this investigation, we investigate four RNN variants: gated recurrent unit (GRU), bidirectional LSTM (BiLSTM), long short-term memory (LSTM), and bidirectional GRU (BiGRU). These architectures are capable of learning patterns and capturing contextual relationships within text, which allows the models to more accurately manage dialect-mixed content and individual words.

LSTM: The long short-term memory (LSTM) model was created to address the vanishing gradient issue in conventional recurrent neural networks[14]. It is well-suited for sentiment analysis due to its ability to effectively capture long-range dependencies in sequential data. One LSTM layer with 128 units is followed by a dense layer with ReLU activation in our model. In order to mitigate overfitting, an attrition rate of 0.3 is implemented. In order to classify sentiments, the output layer implements softmax activation. The primary evaluation metric is precision, and the model is trained using the Adam optimiser and sparse categorical cross-entropy loss.

BiLSTM: The standard LSTM is expanded by the bidirectional long short-term memory (BiLSTM) model, which processes sequences in both forward and backward directions [27]. This bidirectional approach enables the model to capture more detailed contextual information by taking into account not only the past but also the future words within a sentence. The BiLSTM model in this study is composed of a bidirectional LSTM layer with 128 units, followed by a dense layer with 128 units and ReLU activation. In order to mitigate overfitting, a dropout layer with a rate of 0.3 is implemented. Lastly, the sentiment of the input text is classified into predefined categories using a softmax output layer.

GRU: The gated recurrent unit (GRU) is a more straightforward alternative to the LSTM, as it is designed with fewer gates. This feature allows it to capture long-term dependencies in sequential data while remaining computationally efficient [6]. The GRU model in this study is composed of a GRU layer with 128 units, followed by a dense layer with 128 units and ReLU activation. A dropout layer with a rate of 0.3 is implemented to mitigate overfitting. In order to conduct sentiment classification, the final output layer implements softmax activation.

BiGRU: The GRU is enhanced by the bidirectional gated recurrent unit (BiGRU), which processes the input sequence in both forward and backward directions [6, 27]. This bidirectional design enables the model to enhance sequence representation by capturing contextual information from past and future words. The BiGRU model in this study is composed of a bidirectional GRU layer with 128 units, followed by a dense layer with 128 units and ReLU activation. In order to mitigate overfitting, a dropout layer with a rate of 0.3 is implemented, and the final softmax layer is employed for sentiment classification.

4.2 Hybrid CNN–RNN Models

The hybrid CNN-RNN model combines an attention mechanism with Recurrent Neural Networks (RNNs), specifically Bidirectional Long Short-Term Memory (BiLSTM) or Gated Recurrent Unit (GRU), and Convolutional Neural Networks (CNNs). This design is well-suited for tasks that necessitate both local feature extraction and long-term sequence modelling, such as sentiment analysis.

Local features are initially extracted from the input sequence by CNN layers using convolutional filters, and dimensionality is reduced and salient patterns are emphasised through max-pooling. The BiLSTM or GRU layer processes the sequence in forward and backward directions to capture contextual information, and the resulting feature maps are transmitted to it. The accuracy and interpretability of the sequence are enhanced by an attention mechanism that emphasises the most informative components. The final sentiment classification is conducted using a sigmoid or softmax activation layer, after the attended features are transmitted through dense layers with dropout regularisation [20].

Algorithm 1 Proposed Algorithm: Hybrid CNN–RNN Sentiment Classifier

Require: Dataset D with noisy Bangla text; batch size = 16; epochs $E = 50$

Ensure: Sentiment $\in$ {Positive, Neutral, Negative}

1. Pretrained word vectors are used to initialise the embedding layer.
2. Utilise convolutional layers to extract local n-gram features.
3. Use max-pooling to reduce dimensionality and highlight salient patterns.
4. Analyse the sequence in both forward and backward orientations.
5. Apply an attention mechanism to concentrate on the most pertinent terms.
6. Generate a fixed-length representation by employing global pooling.
7. Transfer the representation to dense layers that utilise dropout regularisation.
8. Utilise a softmax layer to categorise sentiment.

4.3 Transformer Models

Our research employs transformer models, including BanglaBERT, RoBERTa, DistilBERT, and XLNet, to solve natural language processing (NLP) problems in the Bengali language. These models are capable of capturing long-term dependencies within text and enabling parallel processing, which improves efficiency in comparison to conventional recurrent networks. Our objective is to enhance the classification accuracy and overall performance of sentiment analysis and other NLP applications by fine-tuning these models on Bengali data [2, 21, 26, 32].

BanglaBERT: BanglaBERT is a BERT-based language model pre-trained on a large Bengali corpus, enabling it to capture the linguistic structure and nuances of the language [2]. In this research, BanglaBERT was fine-tuned for sentiment analysis, named entity recognition (NER), and question answering, where it demonstrated strong performance in handling noisy and dialect-rich Bangla text. Its deep contextual understanding makes it particularly effective for analyzing informal and noisy Bangla text.

RoBERTa: [21], RoBERTa is an optimised version of BERT that leverages a longer training duration and a larger amount of training data, while also eliminating constraints such as Next Sentence Prediction (NSP). When customised to Bangla-specific data, RoBERTa exhibited superior results in sequence labelling and text classification tasks. In comparison to traditional BERT models, its ability to capture more detailed semantic information allows it to achieve greater precision in sentiment classification and entity recognition.

DistilBERT: DistilBERT is a compressed and faster variant of BERT that reduces computational cost while preserving a significant portion of its language understanding ability [26]. In this regard, it is appropriate for real-time applications, including sentiment analysis in social media and chatbot responses. DistilBERT was tailored to Bangla text in this investigation, achieving a satisfactory equilibrium between speed and precision. Its diminutive size enables deployment on mobile devices and peripheral platforms without incurring substantial performance degradation.

XLNet: XLNet is a generalised autoregressive model that enhances BERT by utilising a permutation-based training objective, which allows it to more effectively incorporate bidirectional context [32]. XLNet demonstrated exceptional performance in tasks that necessitated long-range dependency modelling, including document-level sentiment analysis and summarisation, after being fine-tuned on Bangla text. It is particularly effective in comprehending the overall structure of extended text due to its capacity to capture global context.

4.4 Multi-Branch Hybrid Models

We suggest a multi-branch composite model that employs CentralNet-style fusion [31] to integrate Transformer-based models, Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN). The CNN branch extracts local features, such as n-gram patterns, while the RNN branch (BiLSTM/GRU) models temporal dependencies by processing sequences in both directions. The Transformer branch (e.g., BERT, RoBERTa, XLNet) captures long-range dependencies and global context using self-attention. Each branch contributes complementary strengths.

A comprehensive feature representation is generated by combining the outputs of the three branches at a central layer. This fusion allows the model to simultaneously utilise spatial, sequential, and contextual information. The final output is generated using softmax (for multi-class classification) or sigmoid (for binary tasks) after the combined features are transmitted through dense layers with dropout for high-level learning. The multi-branch hybrid approach obtains robust performance in noisy Bangla text sentiment analysis and other sequence classification tasks by utilising multiple perspectives.

All hyperparameter values are reported in Table 4 and were selected using the predefined validation split without applying cross-validation.

Algorithm 2 Proposed Algorithm: Multi-Branch Hybrid Sentiment Classifier

Require: Dataset D (noisy Bangla); batch size = 16; epochs E = 50
Ensure: Sentiment $s \in$ {Positive, Neutral, Negative}
1: Build three branches: CNN, RNN (BiLSTM/GRU), Transformer (BanglaBERT/RoBERTa/XLNet)
2: Tokenize inputs; create embeddings (shared or branch-specific)
3: **CNN branch:** conv + max-pool → c (local n-gram/char features)
4: **RNN branch:** BiLSTM/GRU (bi-directional) + attention → r
5: **Transformer branch:** self-attention encoder; [CLS]/mean-pool → t
6: **Central fusion:** $\mathrm{z} = f(\mathrm{c}, \mathrm{r}, \mathrm{t})$ (CentralNet-style) [31]
7: Dense layer(s) with dropout on z
8: Softmax layer → sentiment prediction s
9: Train with AdamW; loss = cross-entropy; report accuracy, macro-F1

Table 4: Hyperparameter Settings Used in Implementation

Hyperparameter	Value
Batch size	16
Epochs	50
Optimizer	Adam / AdamW
Loss function	Sparse categorical cross-entropy
Evaluation metrics	Accuracy, Precision, Recall, Macro-F1
Hidden units (RNN)	128
Dropout rate	0.3
TF–IDF n-grams (word)	(1–3)
TF–IDF n-grams (char)	(2–5)
SVD components	128
Embedding types	Word2Vec, FastText, GloVe
Transformer models	BanglaBERT, RoBERTa, DistilBERT, XLNet
Max sequence length	128
Fusion method	CentralNet-style attention fusion
Output activation	Softmax

5 RESULT ANALYSIS

The performance evaluation of our multi-branch hybrid model for sentiment analysis on noisy and dialect-mixed Bangla text is presented in this section. The method integrates a variety of deep learning architectures, such as recurrent models (LSTM, BiLSTM, GRU, BiGRU), hybrid CNN–recurrent models, transformer models (BanglaBERT, RoBERTa, DistilBERT, XLNet), and a multi-branch hybrid model that is based on CentralNet-style fusion. The models were evaluated on the SentNoB dataset, which comprises chaotic social media comments that include informal expressions, noisy Bangla text, and dialectal variations.

Evaluation protocol: All reported F1 scores correspond to the macro-averaged F1 (Macro-F1), computed as the unweighted mean of per-class F1 scores across the Positive, Neutral, and Negative classes. Precision and recall are also macro-averaged. Hyperparameter tuning was conducted using the predefined validation split of the SentNoB dataset. No cross-validation was employed, as the dataset provides fixed training, validation, and test partitions, and this protocol is followed to ensure fair and reproducible comparison with prior work.

5.1 Results of Recurrent Models

We begin with the evaluation of recurrent neural networks (RNNs), including LSTM, BiLSTM, GRU, and BiGRU, for sentiment classification. These models are well suited for sequential data as they capture temporal dependencies, which are essential in sentiment analysis.

Table 5: Performance of Recurrent Models

Model	Accuracy	Precision	Recall	F1-score
LSTM	86.1%	86.5%	85.7%	86.0%
BiLSTM	**87.8%**	**88.3%**	**87.0%**	**87.6%**
GRU	85.4%	85.9%	84.8%	85.3%
BiGRU	86.9%	87.5%	86.0%	86.7%

The BiLSTM model demonstrated the highest performance among all recurrent models, as evidenced by the F1-score of 87.6%, precision of 88.3%, recall of 87.0%, and accuracy of 87.8%, as illustrated in Table 5. The model's bidirectional structure allows it to extract context from both past and future tokens, which is especially beneficial for comprehending noisy and dialect-mixed Bangla text. Our results demonstrate a significant enhancement in comparison to those of related studies, as indicated in Table 1. For instance, Henriquez et al. reported an accuracy of 78.3%, while Haque et al. achieved 85.85%. Our fundamental LSTM model attained an accuracy of 86.1%. This illustrates that sentiment classification for chaotic Bangla social media data can be more reliable and accurate when hyperparameters are meticulously tuned and recurrent architectures are implemented.

5.2 Results of Hybrid CNN–RNN Models

The proposed hybrid CNN–RNN models were evaluated on the SentNoB dataset. Table 6 summarises their performance in terms of accuracy, precision, recall, and F1-score.

Table 6: Performance of Hybrid CNN–Recurrent Models

Model	Accuracy	Precision	Recall	F1-score
CNN + BiLSTM	88.6%	89.0%	87.8%	88.4%
CNN + GRU	87.9%	88.5%	86.9%	87.6%
CNN + BiLSTM + Attention	**91.2%**	**91.8%**	**90.6%**	**91.1%**

Our hybrid CNN-RNN models, as illustrated in Table 6, outperformed previously published works in the literature. The accuracy of the CNN + BiLSTM model was 88.6%, which is higher than the 85.85% reported by Haque et al. In the same vein, the CNN + GRU model achieved 87.9%. CNN + BiLSTM + Attention yielded the most favourable outcome, with an F1-score of 91.1% and an accuracy of 91.2%.

This performance exceeds the majority of conventional baselines, such as Henriquez et al. (78.3%) and Haque et al. (85.85%), and is at least as effective as the CNN–BiLSTM framework of Ahamad et al. (90%). The findings verify that the integration of CNN, RNN, and attention layers results in a more robust and potent architecture for sentiment analysis on noisy Bangla text.

5.3 Results of Transformer Models

On the SentNoB dataset, we optimised numerous Transformer-based architectures, such as BanglaBERT, RoBERTa, DistilBERT, and XLNet. Because their self-attention mechanism incorporates long-range dependencies, these models are effective for sentiment analysis, as they are capable of handling the complex linguistic structures that are prevalent in noisy Bangla text.

Table 7: Performance of Transformer Models

Model	Accuracy	Precision	Recall	F1-score
BanglaBERT	90.1%	90.7%	89.6%	90.1%
RoBERTa	91.0%	91.5%	90.2%	90.8%
DistilBERT	89.2%	89.9%	88.3%	89.1%
XLNet	**92.3%**	**92.8%**	**91.5%**	**92.1%**

The highest performance was attained by XLNet, as demonstrated in Table 7, with an F1-score of 92.1% and an accuracy of 92.3%. This outcome exceeds the scores reported in previous studies, which are summarised in Table **??**. The majority of models maintained a score between 78% and 90%. For instance, Haque et al. reported an accuracy of 85.85%, while Ahamad et al. achieved an accuracy marginally above 90% using a CNN–BiLSTM approach. In contrast, our Transformer models consistently surpass these benchmarks, with RoBERTa achieving 91.0% and BanglaBERT reaching 90.1%. These results indicate that Transformer architectures offer a more dependable and precise solution for sentiment analysis of chaotic Bangla social media data.

5.4 Results of Multi-Branch Hybrid Models

Finally, we assess the performance of the multi-branch hybrid model with CentralNet-style fusion. This approach combines the outputs from the CNN, RNN (BiLSTM or GRU), and Transformer branches through an attention-based fusion mechanism. This design allows the model to capitalise on spatial, temporal, and contextual features, resulting in improved performance across all metrics.

The CNN + BiLSTM + GRU + RoBERTa full-fusion model achieved the highest accuracy of 94.5%, with a precision of 95.1%, recall of 93.9%, and F1-score of 94.4%. This performance surpasses that of all other models examined in this study, such as isolated recurrent, hybrid CNN–RNN, and Transformer-based architectures.

Table 8: Performance of Multi-Branch Hybrid Models

Model	Accuracy	Precision	Recall	F1-score
CNN + RNN + Transformer (Fusion)	93.1%	93.6%	92.4%	93.0%
CNN + BiLSTM + GRU + RoBERTa (Full Fusion)	**94.5%**	**95.1%**	**93.9%**	**94.4%**

Our multi-branch hybrid model achieves state-of-the-art performance on the SentNoB dataset for Bangla sentiment analysis in comparison to prior studies (Table 1). For example, Haque et al. reported an accuracy of 85.85%, while Ahamad et al. achieved an accuracy just above 90% by employing a CNN–BiLSTM framework. In our own experiments, even sophisticated Transformer-based methods, such as XLNet (92.3%), are less effective than the multi-branch fusion approach. The results of these studies confirm that the integration of CNN, RNN, and Transformer branches through CentralNet-style fusion results in a more robust and potent model for the classification of sentiment in noisy Bangla text.

5.5 Comparison with Previous Work

The following table provides a contextual comparison between the proposed multi-branch hybrid model and representative sentiment analysis approaches reported in prior studies. These works employ a range of architectures, including traditional machine learning models, hybrid deep learning frameworks, and Transformer-based methods, evaluated under different datasets and experimental settings. Accordingly, the comparison is intended to situate the proposed approach within the broader sentiment analysis literature rather than serve as a direct dataset-level benchmark.

It should be noted that several of the compared studies report results on different datasets and languages; therefore, the values shown are not directly comparable to our results, which are obtained exclusively on the SentNoB Bangla dataset.

As shown in Table 9, the proposed Multi-Branch Hybrid CNN + BiLSTM + GRU + RoBERTa (Full Fusion) model demonstrates strong performance on noisy Bangla sentiment analysis, achieving 94.5% accuracy, 95.1% precision, 93.9% recall, and an F1-score of 94.4%. Compared with the baseline models reported for the SentNoB dataset by Islam *et al.* (F1-score $\approx$64.6%), the proposed approach shows a substantial improvement. The model also outperforms earlier Bangla-focused methods such as the CNN–LSTM model of Haque *et al.* (85.85%) and the CNN–BiLSTM framework of Ahamad *et al.* (around 90%). While some prior studies report higher accuracy on non-Bangla datasets, these results are included for contextual reference only. Overall, the comparison indicates that the proposed full-fusion strategy enhances robustness and scalability for sentiment analysis of noisy Bangla text.

Table 9: Comparison of Our Model with Previous Works

Previous Work	Model	Accuracy	Precision	Recall	F1-score
K. I. Islam *et al.* [17]	SVM (TF–IDF), BiLSTM + Attention, mBERT (baselines)	–	–	–	≈64.6%
K. T. Elahi *et al.* [9]	SVM, BiLSTM, BanglaBERT, ELECTRA, MuRIL	–	–	–	0.75 (Micro-F1)
M. E. Islam *et al.* [18]	BanglaBERT, BiLSTM, HAN, CNN–BiLSTM	–	–	–	0.62 (Macro-F1)
R. Haque *et al.* [11]	CNN–LSTM (CLSTM) + TF–IDF	85.85%	86.0%	–	86.0%
R. Ahamad *et al.* [1]	CNN–BiLSTM (ESIHE_AML)	90.0%+	–	–	–
A. A. Syed *et al.* [29]	PEGASUS + BERT (two-step)	89.0%	–	–	–
P. A. Henriquez *et al.* [13]	Deep RVFL (D-RVFL)	78.3%	–	–	–
V. K. P. *et al.* [30]	Ensemble (RF + SVM + DT + ConvBiLSTM)	93.42%	–	–	–
Our Proposed Model	**Multi-Branch Hybrid CNN + BiLSTM + GRU + RoBERTa (Full Fusion)**	**94.5%**	**95.1%**	**93.9%**	**94.4%**

6 CONCLUSION AND FUTURE WORK

This study investigated the challenge of sentiment analysis in noisy Bangla text, where spelling errors, dialectal expressions, informal writing, emojis, and irregular grammar reduce model effectiveness. To address these issues, we proposed a noise-aware preprocessing pipeline together with a hybrid fusion framework that integrates CNNs, RNNs (BiLSTM and GRU), and Transformer models. The best-performing system, the Multi-Branch Hybrid CNN + BiLSTM + GRU + RoBERTa (Full Fusion), achieved 94.5% accuracy, 95.1% precision, 93.9% recall, and a 94.4% F1-score, clearly outperforming baseline methods and previously reported approaches. These results demonstrate that combining spatial, sequential, and contextual features through full fusion provides a robust and scalable solution for sentiment analysis of noisy Bangla text.

For future work, we plan to extend the dataset with a broader range of regional dialects and informal writing styles to improve adaptability to real-world social media content. We also intend to design lighter and more efficient architectures that can be deployed on mobile and other low-resource devices. The modular design of the proposed framework further enables selective branch removal, allowing flexible deployment under constrained computational budgets. In addition, enhancing interpretability through advanced explainable AI methods and maintaining a dynamic lexicon that evolves with slang and abbreviations will help ensure the system remains practical, transparent, and sustainable for large-scale applications.

REFERENCES

[1] Rayees Ahamad and Kamta Nath Mishra. 2025. Exploring sentiment analysis in handwritten and E-text documents using advanced machine learning techniques: a novel approach. *Journal of Big Data* 12, 1 (2025), 11.

[2] Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. 2022. BanglaBERT: Language Model Pretraining and Benchmarks for Low-Resource Language Understanding Evaluation in Bangla. In *Findings of the Association for Computational Linguistics: NAACL 2022*. Association for Computational Linguistics, 1318–1327.

[3] Anirban Bhowmick and Abhik Jana. 2021. Sentiment analysis for Bengali using transformer based models. In *Proceedings of the 18th International Conference on Natural Language Processing (ICON)*. 481–486.

[4] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching Word Vectors with Subword Information. In *Transactions of the Association for Computational Linguistics*, Vol. 5. MIT Press, 135–146.

[5] Sovon Chakraborty, Protiva Das, Shakib Mahmud Dipto, Md Aktaruzzaman Pramanik, and Jannatun Noor. 2025. An Analytical Review of Preprocessing Techniques in Bengali Natural Language Processing. *IEEE Access* (2025).

[6] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).

[7] Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. *Advances in neural information processing systems* 32 (2019).

[8] CryptexCode. 2021. SentNoB: Sentiment Analysis in Noisy Bangla Texts. https://www.kaggle.com/datasets/cryptexcode/sentnob-sentiment-analysis-in-noisy-bangla-texts. Accessed: 2025-09-03.

[9] Kazi Toufique Elahi, Tasnuva Binte Rahman, Shakil Shahriar, Samir Sarker, Md. Tanvir Rouf Shawon, and G. M. Shahariar. 2024. A Comparative Analysis of Noise Reduction Methods in Sentiment Analysis on Noisy Bangla Texts. *arXiv* abs/2401.14360 (2024). https://arxiv.org/abs/2401.14360

[10] Afifa Farzana and Sukanto Roy. 2021. A Comparative Study between English and Bangla: The Perspective of Phonemics, Morphology and Syntax. (2021).

[11] Rezaul Haque, Naimul Islam, Mayisha Tasneem, and Amit Kumar Das. 2023. Multi-class Sentiment Classification on Bengali Social Media Comments Using Machine Learning. *International Journal of Cognitive Computing in Engineering* 4 (2023), 21–35.

[12] Asif Hassan, Mohammad Rashedul Amin, N Mohammed, and AKA Azad. 2016. Sentiment analysis on bangla and romanized bangla text (BRBT) using deep recurrent models. *arXiv preprint arXiv:1610.00369* (2016).

[13] Pablo A. Henríquez and Francisco Alessandri. 2024. Analyzing Digital Societal Interactions and Sentiment Classification in Twitter (X) During Critical Events in Chile. *Heliyon* 10, 12 (2024).

[14] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[15] Fahima Hossain, Palash Saha, and Ismail Hossain. [n.d.]. Leveraging NLP and ML for Sentiment Analysis of In-formal Bangla Text: A Social Media Comment Analysis Tool for Page Owners. ([n. d.]).

[16] Tanzir Hossain, Ar-Rafi Islam, Md Humaion Kabir Mehedi, Annajiat Alim Rasel, M Abdullah-AL-Wadud, and Jia Uddin. 2025. BanglaNewsClassifier: A machine learning approach for news classification in Bangla Newspapers using hybrid stacking classifiers. *PLoS One* 20, 6 (2025), e0321291.

[17] Khondoker Ittehadul Islam, Sudipta Kar, Md Saiful Islam, and Mohammad Ruhul Amin. 2021. SentNoB: A dataset for analysing sentiment on noisy Bangla texts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. 3265–3271.

[18] Md. Ekramul Islam, Labib Chowdhury, Faisal Ahamed Khan, Shazzad Hossain, Md. Sourave Hossain, Mohammad Mamun Or Rashid, Nabeel Mohammed, and

Mohammad Ruhul Amin. 2023. SentiGOLD: A Large Bangla Gold Standard Multi-Domain Sentiment Analysis Dataset and Its Evaluation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, 4207–4218.

[19] Md Shymon Islam and Kazi Masudul Alam. 2024. Sentiment analysis of Bangla language using a new comprehensive dataset BangDSA and the novel feature metric SkipBangla-BERT. *Natural Language Processing Journal* 7 (2024), 100069.

[20] Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent Convolutional Neural Networks for Text Classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*. AAAI Press, 2267–2273.

[21] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

[22] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR*. https://arxiv.org/abs/1301.3781

[23] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 1532–1543.

[24] Amartya Roy, Kamal Sarkar, and Chintan Kumar Mandal. 2023. Bengali text classification: A new multi-class dataset and performance evaluation of machine learning and deep learning models. (2023).

[25] Gerard Salton and Christopher Buckley. 1988. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* 24, 5 (1988), 513–523.

[26] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *arXiv preprint arXiv:1910.01108* (2019).

[27] Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 11 (1997), 2673–2681.

[28] Noura A. Semary, Wesam Ahmed, Khalid Amin, Paweł Pławiak, and Mohamed Hammad. 2023. Improving Sentiment Classification Using a RoBERTa-Based Hybrid Model. *Frontiers in Human Neuroscience* 17 (2023), 1292010.

[29] Ayesha Ayub Syed, Ford Lumban Gaol, Alfred Boediman, and Widodo Budiharto. 2024. Airline Reviews Processing: Abstractive Summarization and Rating-Based Sentiment Classification Using Deep Transfer Learning. *International Journal of Information Management Data Insights* 4, 2 (2024), 100238.

[30] K. P. Vidyashree, A. B. Rajendra, H. L. Gururaj, et al. 2024. A Tweet Sentiment Classification Approach Using an Ensemble Classifier. *International Journal of Cognitive Computing in Engineering* 5 (2024), 170–177.

[31] Valentin Vielzeuf, Alexis Lechervy, Stéphane Pateux, and Frédéric Jurie. 2018. CentralNet: A Multilayer Approach for Multimodal Fusion. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Springer.

[32] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized Autoregressive Pretraining for Language Understanding. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS)*. Curran Associates Inc., 5753–5763.

[33] Chiara Zucco, Barbara Calabrese, Giuseppe Agapito, Pietro H Guzzi, and Mario Cannataro. 2020. Sentiment analysis for mining texts and social networks data: Methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 10, 1 (2020), e1333.