



BERTConvNet: A Transformer-Based Framework for Aspect-Based Sentiment Analysis and Fake Review Detection on Self-Created YouTube Review Dataset

Shraddha Kalbhor^{*ID}, Dinesh Goyal^{ID}, Kriti Sankhla^{ID}

Computer Science & Engineering, Poorinma University, Jaipur 303905, India

Corresponding Author Email: shraddha.kalbhor000@gmail.com

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.300622>

ABSTRACT

Received: 5 May 2025

Revised: 8 June 2025

Accepted: 24 June 2025

Available online: 30 June 2025

Keywords:

Aspect-Based Sentiment Analysis, BERTConvNet, fake review detection, transformer models, multimodal fusion, YouTube comments, contextual embedding's

In the era of digital content creation, YouTube has emerged as a dominant platform for information dissemination and user engagement. However, the spread of fake reviews and misleading content makes it hard to trust what's real and honest. Current ABSA methods struggle with contextual nuance in user-generated content; our BERTConvNet framework addresses this by integrating transformer embeddings with CNN-based metadata processing. The approach is built on Aspect-built Sentiment Analysis (ABSA), which makes use of advanced transformer-based models (BERTConvNet). The proposed approach focuses on extracting and analyzing specific aspects of user reviews—such as relevance, grammar, quality, and viewer engagement—by examining raw YouTube Video comments. The process is start with cleaning of comments, then is follows steps like data preparation, feature extraction, language processing, and sentiment classification. Domain-specific details and context are added to boost accuracy and relevance. The proposed approach uses Word Embedding technique to handle comments while label encoding to handle video metadata. Publically available datasets named SemEval along with self-created YouTube ABSA dataset are used for experimental analysis. Experimental evaluations on a diverse dataset demonstrate significant enhancements in sentiment analysis performance compared to existing methods. With the integration of a BERTConvNet model, there has been a notable enhancement in accuracy-per-label and F1-score metrics, exhibiting an approximate increase of ~ 4%.

1. INTRODUCTION

In the contemporary digital era, YouTube has evolved into one of the most influential content-sharing platforms worldwide. With a monthly active user base of over 2.7 billion and over 500 hours of content being uploaded every minute, YouTube has become a core component of digital communication, entertainment, education as well as product marketing. Contributors, users, advertisers, and influencers make it the vibrant ecosystem that drives the worldwide digital economy and the global social conversation [1]. Users are extremely dependent on video reviews, tutorials, unboxings, product comparisons, and expert analysis when making decisions — whether it's about buying a product, deciding on entertainment choices, or even politics. YouTube's comment sections are also a major mode of communication between creators and their audience, as well as feedback, it also allows creators to see audience reaction and what the audience want. Such comments are typically detailed ratings and assessments of a video's content, production quality, factual statements, and relevance. Now, more than ever, brands and content creators need to leverage viewer feedback to measure brand sentiment, inform content programming strategies and build trust with viewers [2].

But the amount and diversity of user-generated content

makes this very hard. The comments are fake; spams, comments of automated bots, deployed misinformation campaigns and fake comments are part of the problem. This has led to a growing demand for intelligent approaches that can derive meaningful and true facts from noisy comment data.

1.1 Fake reviews and deceptive content

As user feedback becomes increasingly influential, the incentive to manipulate it also grows. Platforms such as YouTube are facing a rise in fake reviews, spam comments, and inauthentic engagement strategies [3]. For example, some content creators may purchase fake positive comments to artificially boost their popularity, while others may exploit the system by flooding comment sections with negativity, aiming to spread discord and undermine genuine user interactions. Sometimes, organized misinformation is instead used to lead people to conclusions that may not be true about the quality of product, the merit of a political idea or a social concern. Uncensored comment sections also open up the information that users consume to questions of trust in the information itself but in the platform, as well. Although YouTube has implemented elementary spam filters and terms of service, many such deceptive comments still slip through the cracks,

especially the ones that look the most authentic. These annoyances are not just an issue for internet users, but they also serve to dilute the impact/reporting of well-intentioned content providers and advertisers (who use this honest engagement as a measure of effectiveness [4]). While conventional sentiment analysis approaches are able to classify a total positivity/negativity of text, they are too coarse whenever it comes to analyzing the multidimensional feedback occurring in user comments. A comment can be both positive for one's visual quality and negative for temporal relevance with respect to content. This level of subtle feedback can't be adequately captured using a traditional sentiment analysis, rather, one that is able to recognize unique aspects of a video and evaluate sentiment associated for each of them independently is required.

1.1.1 The need for Aspect-Based Sentiment Analysis (ABSA)

Traditional sentiment analysis methods commonly treat entire comment or review as a monolithic object and provide a single sentiment label such as positive, negative or neutral. But actual feedback on the real world can, especially in YouTube comment sections, be more multidimensional. One comment is also capable of addressing multiple aspects of video, including its relevance to the information, quality of speech, grammar, visual quality, and viewer's overall engagement [5]. These nuances are not captured in standard models. ABSA overcomes this problem by decomposing a text into various aspects and scoring the sentiment for each aspect. This fine-grained sentiment classification allows for a more detailed understanding of user opinions, making it especially valuable for complex platforms like YouTube where content quality cannot be measured using a one-size-fits-all sentiment label [6].

Furthermore, identifying fake or deceptive reviews becomes more feasible when aspect-level sentiment patterns are analyzed. Fake reviews often display unusual sentiment consistency (e.g., overly positive across all aspects) or use generic phrases that can be filtered through aspect-aware language models.

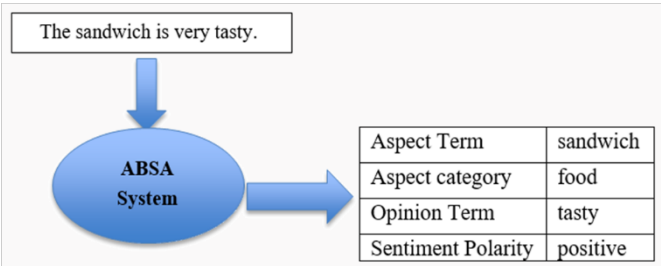


Figure 1. Example of aspect based sentiment analysis

The Figure 1 illustrates the working of an ABSA system. Given the input sentence “The sandwich is very tasty,” the ABSA model identifies key components: the aspect term ("sandwich"), its aspect category ("food"), the opinion term ("tasty"), and the corresponding sentiment polarity ("positive"). This showcases how ABSA provides fine-grained sentiment insights by linking opinions to specific aspects of an entity.

1.2 Proposed contributions

The main contribution of this study is to detect and analyze the aspects of YouTube videos by applying ABSA and

transformer-based techniques on review data. As the consumption of videos on platforms like YouTube continues to grow, the analysis of video reviews becomes an essential task for content creators, marketers, and platform administrators to understand user sentiment, engagement, and feedback. Following are the proposed contribution of our research work:

- To extract and analyse user comments on YouTube videos, identifying sentiment across multiple specific aspects including content relevance, grammatical accuracy, presentation quality, and viewer engagement.
- To enhance contextual sentiment understanding and reduce misclassifications by leveraging advanced transformer architectures, such as BERT and its variants, within ABSA models.
- To develop a classification system capable of distinguishing fake from valid reviews by analysing sentiment patterns, linguistic features, and semantic coherence across multiple aspects.
- To design and implement a comprehensive pre-processing pipeline for cleaning and augmenting user comments, ensuring high-quality input for subsequent sentiment analysis and classification.
- To assess the suggested ABSA-based framework's efficacy through both quantitative metrics and qualitative analysis, contributing to a transparent and reliable content evaluation ecosystem on social media platforms.

This research lays the groundwork for more accurate, interpretable, and trustworthy analysis of user-generated content on video-sharing platforms, addressing the urgent need to combat fake reviews while improving content quality assessment and user experience.

The article includes a literature review (Section 2) on ABSA techniques, following that dataset descriptions (Section 3) covering a self-created YouTube comments dataset and the SemEval 2014 dataset. Section 4 reviews related models, including 1D CNN, BERT, and XLNet. Section 5 introduces the proposed BERTConvNet framework, integrating transformer embedding’s with CNN layers for enhanced sentiment classification. Section 6 covers experimental setup, hyper-parameters, performance metrics, and result analysis. Section 7 concludes with key findings, contributions to ABSA, and future work, with references provided at the end of article.

2. LITERATURE REVIEW

This section elaborately reviews contemporary literature, with a specific focus on the last decade, in relation to recent developments in fake review detection in terms of many feature extraction techniques, NLP, ML algorithms, and ABSA. The Review covers a wide array of academic papers and research, showcasing the new trends, discoveries, and analyses that make up the current research in these fields.

Gruetzemacher and Paradise [7] pointed out the under-representation of transformer language models (TLMs) in Information Systems (IS) research and potential of TLMs to enhance text mining and analysis. Jian et al. [8] built a deep-learning based model called hybrid feature fusion, which fused text and metadata (user behavior) to strengthen the detection of fake restaurant reviews. Both papers highlight the significance of leveraging state-of-the-art text modelling technologies and multiple data sources for more effective analysis.

Sánchez-Reolid et al. [9] used 1D-CNNs to classify arousal states from EDA signals with high accuracy, and demonstrated the potential of real-time emotion recognition. Similarly, Zhou et al. [10] proposed the SetCNN for short text classification that combines semantic extension and pooling, which can handle the sparse text appropriately.

Nagelli and Saleena [11] presented a robust ABSA with Bi-LSTM network that incorporates pre-processing, weighted aspect extraction and sentiment classification to overtake the benchmark performance.

Fregoso et al. [12] tuned CNN models for sign language recognition using PSO in order to enhance accuracy by automating the selection of hyper-parameters. These works validate the versatility and the feasibility of CNNs for different kinds of tasks such as physiological signals, text processing, or gesture recognition.

Budhi et al. [13] proposed a hybrid method that utilizes content-based and behavior based features for fake e-commerce reviews detection. Using the group of 133 features to which they were committed, and machine learning with resampling and parallel cross-validation, their model had a high degree of accuracy. Zhao et al. [14] proposed an enhanced approach to Aspect-Based Sentiment Analysis (ABSA) by integrating BERT-driven context generation with a novel quality filtering mechanism. Their method leverages the contextual richness of BERT to generate more relevant aspect-aware representations, improving the granularity and accuracy of sentiment classification. Additionally, the quality filtering step effectively removes noisy or ambiguous context data, further refining the input for downstream tasks.

Aziz et al. [15] presented a unified model for ABSA that combines BERT with a Multi-Layered Enhanced Graph Convolutional Network (MLEGCN). Their approach captures both deep contextual semantics and syntactic dependencies to jointly handle multiple ABSA subtasks such as aspect–opinion extraction and sentiment classification. By integrating biaffine attention and graph-based linguistic features, the model achieves state-of-the-art results on standard benchmarks, demonstrating improved accuracy and interpretability in fine-grained sentiment analysis.

Chauhan et al. [16] propose an enhanced ABSA framework built upon RoBERTa, optimizing its use for fine-grained sentiment detection toward specific aspects in text. They adapt and fine-tune RoBERTa’s contextualized embeddings, tailoring them to ABSA subtasks, and compare their model’s performance against strong baselines on standard datasets. Their results demonstrate that the enhanced RoBERTa model achieves high accuracy—92.35% on the restaurant dataset and 82.33% on the laptop dataset. By empirically validating RoBERTa’s effectiveness for ABSA, this study underscores the power of transformer-based fine-tuning strategies for improved aspect-level sentiment analysis. These studies demonstrate significant advances in ABSA through deep learning models, enabling more accurate and context-sensitive sentiment analysis.

Table 1 presents the comparative analysis in key aspects of sixteen relevant research papers focused on sentiment analysis, fake review detection, and transformer-based models.

Table 1. Comparison of literature reviews

Ref. No.	Methodology Used	Datasets Used	Advantages	Results
[2]	Transformer-based review + behavior model	Self-curated dataset	Combines content and behavioral features for enhanced detection	Improved fake review detection
[3]	Multilingual Transformers	Code-mixed tweets (Roman Urdu + English)	Handles syntactic ambiguity in code-mixed texts	Outperformed traditional models
[4]	EfficientNet + Vision Transformers	Deepfake video dataset	Low-complexity and high-accuracy model	High performance
[6]	Hierarchical Attention Network (HAN)	Online course comments	Focuses on important sentiment-carrying terms	Improved sentiment classification accuracy
[8]	DNN with hybrid feature fusion	Restaurant reviews	Combines textual and user behavior data	Achieved 93.12% accuracy
[9]	1D-CNNs for arousal classification	EDA signal datasets	Captures temporal signal patterns	High accuracy
[10]	Set-CNN with semantic extension	Short text datasets	Improved classification for sparse texts	Outperformed standard CNNs
[11]	Bi-LSTM with weighted aspect extraction	Benchmark ABSA datasets	Improved ABSA performance	Higher accuracy than existing models
[12]	CNNs optimized with PSO	Sign language datasets	Automatic CNN architecture tuning	Improved gesture recognition accuracy
[13]	Hybrid content + behavior model	E-commerce reviews	Rich feature set; handles class imbalance	Outperformed traditional models
[14]	BERT with novel quality filtering mechanism	SemEval 2014	Removal of Noise or ambiguous context data	Aspect level sentiment classification
[15]	Biaffine attention and graph-based linguistic features	ABSA benchmark datasets	Captures deep contextual semantics and syntactic dependencies to handle ABSA subtasks	Improved accuracy and interpretability in fine-grained sentiment analysis
[16]	Fine-tune RoBERTa’s contextualized embeddings	SemEval 2014	Sentiment detection toward specific aspects in text	Accuracy of 92.35%

3. DATASET DESCRIPTION

In this study, we have utilized two datasets for our analysis.

The first is the publicly available SemEval 2014 dataset, which provides a comprehensive resource for sentiment analysis and related tasks. The second dataset is a self-created YouTube

review dataset, which was specifically curated to capture user-generated content and opinions from YouTube videos. A detailed description of both datasets is provided in Table 2.

3.1 Self-created YouTube review dataset

3.1.1 Data collection and processing

The dataset applied in the previous study generated the output where it showed that the video is spam or not spam (fake or valid). In this study a dataset has been developed where Aspect Based Sentiment Analysis is shown in more enhancing way and the features and aspect are elaborated as follows:

URL Link - This represents the link to a specific YouTube video. It's a unique identifier for each video, typically in the form of https://www.youtube.com/watch?v=VIDEO_ID.

Comment - This is the raw text of a comment left by a viewer on the YouTube video. It includes all the original content, such as text, emojis, links, or any other symbols.

Cleaned comment - This is a version of the Comment after preprocessing steps to remove unnecessary elements like special characters, emojis, links, or irrelevant words. This cleaning process is usually done to prepare the text data for analysis (e.g., sentiment analysis).

Relevance sentiment - This is a measure of how relevant a comment is to the video or topic. (0 – Not Relevant, 1 - Relevant)

Grammar sentiment - This focuses on the English Grammar aspects of the comment. (0 – Grammatically Incorrect, 1 - Grammatically Correct)

Quality sentiment - This represents the perceived quality of the comment, often judged by its constructiveness, clarity, informativeness, or appropriateness. (0 - Negative, 1 - Positive)

Viewers engagement sentiment - This metric captures the overall sentiment or emotional response of viewers based on

their engagement (likes, dislikes, replies, etc.) with the comment. (0 - Negative, 1 - Positive)

Class fake valid - This indicates whether the Youtube link's content is classified as "fake" or "valid" based on the Title of the video. (Fake - 0, Valid - 1)

Our methodology begins with a specialized technique for collecting YouTube video metadata and associated user comments, utilizing the YouTube API to obtain comprehensive details such as video titles, descriptions, view counts, upload dates, and viewer feedback. The reviews will be passed through a pre-processing pipeline that includes text cleaning and feature extraction using Natural Language Processing techniques to find key phrases for sentiment analysis. Every video in the dataset is first manually labeled either as "fake" or "not fake," while further annotations were generally made across four key aspects of a video: 'Relevance', 'Grammar', 'Quality', and 'Viewer Engagement.' These aspects are scored based on sentiment; for positive indications, the score is kept at 1, while for negative indications, the score is 0 and neutral as 2. This will be helpful in deriving finer details regarding video quality and viewer response patterns. It further verifies the reliability of the dataset and enhances the accuracy of detection by applying various ML algorithms, including logistic regression, SVM, and ensemble methods. Figure 2 shows the self-created YouTube review dataset sample.

The constructed custom YouTube comments dataset consist of approximately ~22,000 comments from 2,200 videos, ensuring a balanced sampling of 10 comments per video. After applying several pre-processing steps, including stemming, stop-word removal, and emoji processing the comments containing fewer than three words were discarded to eliminate noise and non-informative entries. After pre-processing, the dataset was refined to ~ 20,000 high-quality comments, which were then used as input for subsequent sentiment analysis tasks. For model training and evaluation, we maintained an 80:20 train-test split to ensure fair performance assessment.

	video_id	trending_date	title	channel_title	category_id	tags	views	likes
0	2kyS6SvSYSE	17.14.11	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	SHANtell martin	748374	57527
1	1ZAPwfrtAFY	17.14.11	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	last week tonight trump presidency "last week ...	2418783	97185
2	5qpjK5DgCt4	17.14.11	Racist Superman Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	superman "rudy" "mancuso" "king" "bach"... racist	3191434	146033

Figure 2. Self-created YouTube review dataset sample

Table 2. SemEval 2014 dataset description

Name	Restaurant Review Dataset	Laptop Review Dataset
Size	3693/6 (row/features)	2359/6 (row/features)
Parameters	id, sentences, polarity, from, to, label	id, sentences, polarity, from, to, label
Features	Dataset includes: - aspect term polarities, aspect category-specific polarities, sentence splitting mistakes, and annotations for aspect terms that appear in the sentences	Over three thousand English sentences have been retrieved from customer reviews of laptops and are included in the dataset.
Polarity	Positive, negative, neutral	Positive, negative, neutral
Dataset link	https://www.kaggle.com/datasets/charitarth/semeval-2014-task-4-aspectbasedsentimentanalysis?resource=download	https://www.kaggle.com/datasets/charitarth/semeval-2014-task-4-aspectbasedsentimentanalysis?resource=download

3.2 SemEval-2014 dataset [17]

The fields of analysis were extracted as sentence identification, text, aspect word, aspect category and aspect polarity;

Sentence id – It reflects the special numeric analysis id.

Text – The written food/restaurant/service consumer analysis. aspect term – a rating given by the customer for the product.

Aspect category – Depending on the summary text, each review includes a minimal category of 1 to a maximum of 5 factors. Dataset of five aspects: menu, food price, anecdotes/subtlety and services are listed in restaurant analysis.

Aspect polarity – It shows a sense of the element of the category.

4. RELATED WORK

In recent years, advanced DL architectures have significantly enhanced the accuracy and robustness of ABSA systems. This section reviews key models relevant to our study, including 1D CNN, BERT, and XLNet, are explained in detail. Together, these models provide a robust foundation for developing a comprehensive sentiment analysis system.

4.1 dimensional convolutional neural network (1D CNN)

A CNN [18] is a DL algorithm designed for image data but also widely used for text and other data. It extracts key features through convolution operations. Figure 3 shows the 1D CNN architecture.

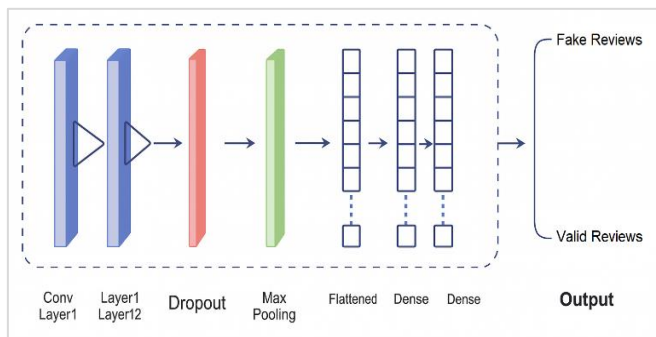


Figure 3. 1D CNN architecture

4.1.1 Input layer

The input consists of structured data, such as one-hot encoded vectors or numerical features.

One-hot encoded data - This is typically a sparse vector representing categorical variables, where each category is encoded as a binary vector.

Numeric data - These could be continuous or discrete features representing attributes or measurements from the data.

4.1.2 Convolution layer

1D convolutional filters are applied over the sequence of features (one-hot or numeric) to capture local dependencies or patterns between adjacent features. For one-hot encoded data, this allows the network to learn spatial relationships between categorical variables.

4.1.3 ReLU activation

The ReLU activation function is used after the convolution layer to introduce non-linearity: $f(x) = \max(0, x)$.

4.1.4 Pooling layer

A reduction in the dimensionality of the feature map is achieved through the pooling layer. In order to determine which attributes are the most significant, Max Pooling is utilized.

4.1.5 Flattening

The output from the pooling layer (which is a 1D feature map) is flattened into a 1D vector. This step converts the multi-dimensional feature map into a single vector that can be fed into the fully connected layers.

4.1.6 Fully connected layer

The fully connected layer (dense layer) processes the flattened 1D vector. It learns to classify or predict based on the features extracted by the convolutional layers.

4.1.7 Softmax / output layer

In order to generate probabilities for each class, the softmax activation is utilized in the output layer when classification tasks are being performed.

4.2 Bidirectional Encoder Representations from Transformers (BERT)

A machine learning paradigm for NLP is referred to as BERT, which is an abbreviation for "Bidirectional Encoder Representations from Transformers" [19].

Steps in BERT Algorithm:

Step 1. Input representation

- Converts input text into tokens.
- Adds special tokens like [CLS] (classification) and [SEP] (separator).
- Embeds tokens using:
 - Token Embedding
 - Segment Embedding
 - Positional Embedding

Step 2. Transformer layers

- A stack of Transformer Encoders processes the input. Each encoder uses:
 - Self-Attention to compute relationships between all words in a sentence.
 - Feed-Forward Neural Networks for deeper feature extraction.

Step 3. Output representation

- The [CLS] token represents the entire sentence and is used for classification tasks.
- Token-level embedding is used for token-based tasks like Named Entity Recognition (NER) [2].

Figure 4 illustrates the BERT masked language model architecture. In this example, the input sentence "how are [MASK] doing today" is fed into BERT, with the word "you" masked. BERT processes the sentence and predicts the most probable word for the masked position based on its contextual understanding. The output shows that "you" has the highest probability among possible candidates like "you", "they", or "your".

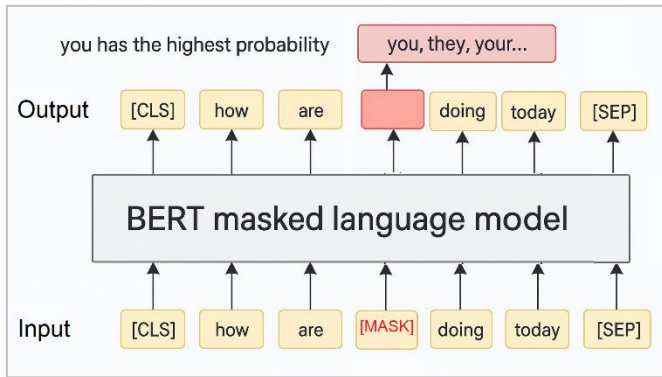


Figure 4. BERT algorithm working flow

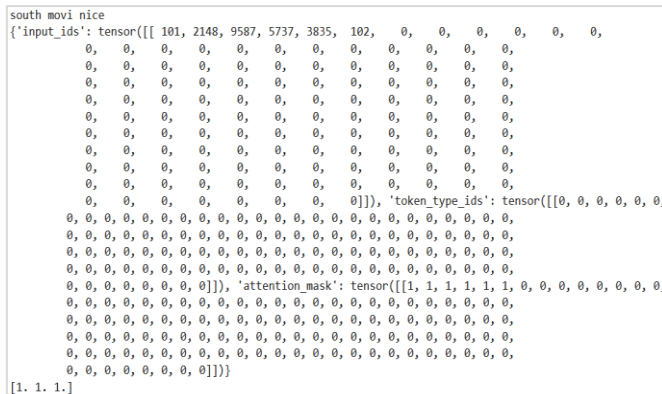


Figure 5. BERT algorithm output

Figure 5 shows the tokenization and encoding output of a BERT tokenizer applied to the input text "south movie nice". It includes three main components:

input_ids. The tokenized representation of the input sentence using BERT's vocabulary, with special tokens [CLS] (101) and [SEP] (102) included.

token_type_ids. Distinguishes sentence pairs, with all values set to 0 for a single-sentence input.

attention_mask. highlights which tokens should be attended to (1) and which are padding (0).

The final array [1. 1. 1.] represents predicted sentiment scores (e.g., relevance, grammar, and quality), showing positive sentiment for each aspect. This format is commonly used in preparing text for transformer-based models in sentiment analysis tasks.

4.3 XLNet

XLNet, short for Generalized Autoregressive Pretraining for Language Understanding, is a Machine Learning (ML) model designed to improve upon the limitations of previous language models like BERT. It combines the strengths of autoregressive models and permutation-based training to achieve cutting edge performance in various NLP tasks [20].

Steps in XLNet Algorithm:

Step 1. Input representation

- Converts input text into token sequences (subword units, e.g., WordPiece or SentencePiece).
- Introduces permutation-based masking, allowing the model to consider different factorization orders (permutations) of token sequences.
- Embeds tokens using:
 - Token Embedding

- Segment Embedding (to distinguish sentence pairs)
- Positional Encoding (to preserve token order)

Step 2. Permutation language modeling

- Unlike BERT's masked language model (MLM), XLNet permutes the factorization order of the input sequence, meaning it considers different ways of predicting tokens based on various preceding contexts.
- For each permutation:
 - Autoregressive modeling predicts each token based on the context of its preceding tokens in that permutation.
 - Two-Stream Self-Attention:
 - One stream for content (token embeddings).
 - One stream for query (predictive context).
 - Enables bidirectional context learning while maintaining autoregressive predictability.

Step 3. Transformer layers

- A stack of Transformer Encoder blocks processes the permuted input.
- Uses self-attention mechanisms to capture relationships between tokens across permutations.
- Feed-forward layers extract high-level features for each token.

Step 4. Output representation

- For token-level tasks (like NER), XLNet uses token embeddings from the final layer.
- For sentence-level tasks (like classification), it aggregates token representations.
- Unlike BERT, XLNet doesn't rely on a special [CLS] token but instead learns contextual relationships from permutations.

5. PROPOSED METHODOLOGY

Figure 6 shows the system architecture of a comprehensive framework designed for ABSA using a transformer-driven approach integrated with deep learning techniques. The pipeline begins with an input dataset, which is split into two main components: comment data and meta data. Each component undergoes specialized preprocessing. Comment data is processed using NLP techniques including stop-word removal, stemming, and other textual cleaning operations. Simultaneously, metadata such as likes/dislikes, view counts, and other engagement metrics are cleaned and encoded. Following preprocessing, the textual data is transformed into vector representations through word embedding, while the metadata is encoded into suitable numerical formats. The core of the system lies in the proposed transformer-based model, which processes text using BERT for contextual embedding and applies CNNs to metadata and supplementary features. These dual outputs are combined at a fusion layer, ensuring that both linguistic and behavioral cues are considered in sentiment classification.

The fused representation is then passed to the ABSA classifier, which tags specific aspects of video content with their corresponding sentiments. Finally, a performance analysis module evaluates the effectiveness of the model based on metrics such as accuracy, precision, and recall, supplying information into system reliability and effectiveness.

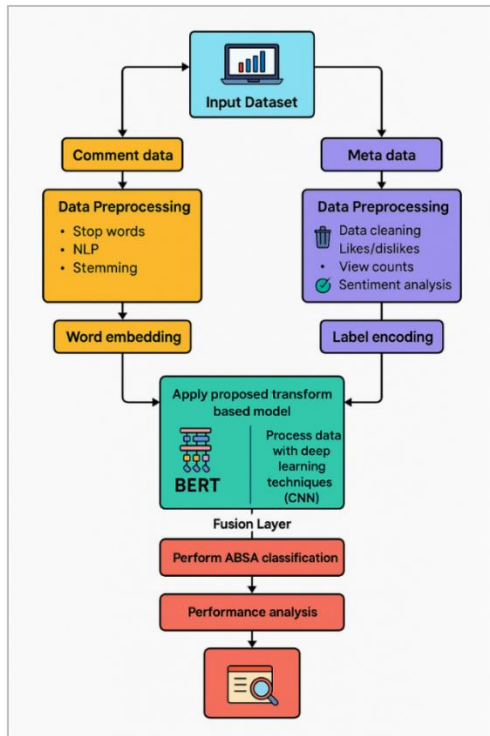


Figure 6. System architecture

5.1 Pre-processing pipeline (YouTube video comments processing)

To ensure high-quality data, user reviews undergo a pre-processing pipeline. This includes:

Text cleaning. Removal of irrelevant characters, stop words, emoji's, and normalization of text.

Feature extraction. Implementing NLP techniques to refine and structure the text data. This involves tokenization, stemming, and the identification of key phrases that contribute

to sentiment analysis.

Manual labelling and sentiment analysis. (YouTube video Comments), Figure 7 shows the YouTube Video ABSA Labelling.

5.2 Data preprocessing (metadata processing)

Figure 8 shows a sample of YouTube video metadata, including the URL, title, channel title, video ID, likes, dislikes, upload date, views and comment count. This metadata serves as a critical component in the preprocessing stage of sentiment and fake review analysis, helping assess viewer interaction and content reach. Several steps are applied for metadata processing including;

Data cleaning. Ensures consistency by handling missing values, duplicates, and outliers.

Likes/dislikes & view counts. Converts numerical data into standardized formats or bins.

5.3 Feature extraction and merging

Word Embedding: The preprocessed comment data undergoes word embedding to convert text into numerical representations. Methods like Word2Vec and contextual embedding (e.g., BERT) can be used [16].

Label encoding. Converts categorical labels (e.g., sentiment scores) into numerical formats, which are necessary for model training.

Example: Positive → 1, Neutral → 0, Negative → 2.

Figure 9 shows the sentiment-labeled dataset of YouTube comments linked to a specific video ID. Each comment is evaluated across four distinct sentiment aspects: relevance, grammar, quality, and viewer engagement. Additional metadata such as views and likes is also included, providing context for viewer interaction. ABSA models require this organized dataset in order to be trained and evaluated efficiently.

	trailer_link comment	Relevance_sentiment	Grammar_sentiment	Quality_sentiment	Viewer Engagement_sentiment	CLASS_FAKE_VALID
1	dlW1Dzkje bore movi	2	2	2	2	1
2	dlW1Dzkje rememb actress montana sky old	1	2	1	1	1
3	dlW1Dzkje highli process food	1	1	1	1	1
4	dlW1Dzkje feel like horror movi imagin want	1	1	1	1	1
5	dlW1Dzkje videomark 1930	1	2	1	1	1
6	dlW1Dzkje need button much aggress	1	2	1	2	1
7	dlW1Dzkje nice movi post upload choppi and	1	1	1	1	1
8	dlW1Dzkje nice movi star actress talk much f	1	1	1	1	1
9	dlW1Dzkje as tri teach evrybodi religion end	1	2	1	2	1

Figure 7. YouTube video ABSA labelling

	URL	Video ID	Title	Channel Title	Date	Views	Likes	Dislikes	Comment Count
0	https://www.youtube.com/watch?v=JOHOezhX-W0	JOHOezhX-W0	NEW MOVIE TRAILERS 2024	MovieGasm.com	2024-07-31T16:00:02Z	2333936	7894	0	543
1	https://www.youtube.com/watch?v=dd0yy_RJLLA	dd0yy_RJLLA	10 MASTERPIECE Movies	Movies Jockey	2023-09-15T12:58:16Z	4688756	205814	0	1693
2	https://www.youtube.com/watch?v=wSgZVpHmILM	wSgZVpHmILM	THE SAINT BEST CIA SPY MOVIE FULL LENGTH ...	Hundaaf Tech	2024-07-19T17:32:39Z	1086252	5754	0	188

Figure 8. Metadata of video URL (CSV)

	Video ID	comment	Relevance_sentiment	Grammar_sentiment	Quality_sentiment	Engagement_sentiment	Viewer	Views	Likes
0	MKg4smUaDj8	nadeem sahodias e	1		1		1	237	243
1	MKg4smUaDj8	bahut imot movi hai	1		2	1	1	237	243
2	MKg4smUaDj8	2024 kon kon dekh raha hai	1		1	1	1	237	243

Figure 9. Merge aspect and metadata

5.4 Proposed transformer-based model (BERTConvNet)

In order to carry out Aspect-Based Sentiment Analysis, the pipeline incorporates both textual and meta data, making use of both cutting-edge transformer-based models (BERT) and more conventional deep learning techniques. The final output is a detailed analysis of sentiment for individual aspects, which is validated through performance metrics. Figure 10 shows the BERTConvNet architecture diagram and detail layer description is provided in Table 3.

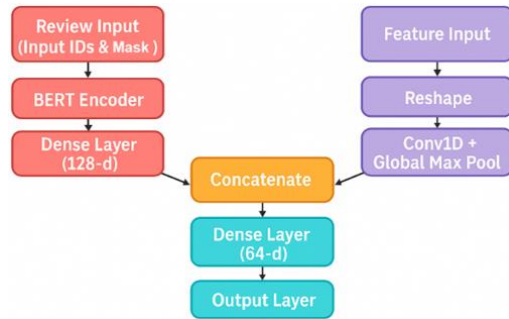


Figure 10. Proposed BERTConvNet model architecture

BERT encoder. Uses BERT to extract contextual embeddings from text data. This step leverages pre-trained BERT models to understand context and nuances in sentences.

Conv1D + global max pooling. Process Data with Deep Learning Techniques: Meta data is further processed with 1D CNN: Captures local patterns in text data.

Fusion layer. Combines outputs from BERT, CNN into a unified representation for classification (BERTConvNet).

Dense layer. The fused data is passed to a classification

layer for ABSA. Identifies sentiment for specific aspects (e.g., "grammar_sentiment: positive").

The study makes use of 1D CNN, XLNet, BERT, and BERTConvNet architectures in order to identify the sentiment of evaluations and to categorize reviews as either genuine or fraudulent. This demonstrates the efficacy of these algorithms in perform tasks related to sentiment analysis and review validation.

6. RESULT

In this section, the results of the experimental investigation we performed on the XLNet, BERT, and BERTConvNet models are reported and a discussion on these results is provided. We used the performance matrices accuracy, accuracy per label, precision, F1-score and recall to assess how a model was doing. Sentiment Analysis Sentiment analysis and fake review detection are performed using these metrics. The comparison makes the merits and demerits of each strategy appear.

6.1 Experimental setup

The experimental setup was built on Python as the core programming language and a number of DL and NLP libraries including TensorFlow, Keras, and Transformers. Training and testing models were performed on Google Colab supported with GPU. The architectures featured 1D convolutional layers for the CNN model and the transformer layers of the models BERT and XLNet were fine-tuned on pre-trained transformer models. The hardware and software configurations employed for implementing and testing the proposed model are presented in Table 4.

Table 3. BERTConvNet model configuration details

Input Layers	Operator	Output Shape / Dimensions	Description
Review Input	Tokenizer (Input IDs & Mask)	(Max_seq_len)	Tokenized textual input for BERT
BERT Encoder	Pre-trained Transformer	(Max_seq_len, hidden_dim)	Outputs contextual embeddings from the input text
Dense Layer	ReLU	(128)	Reduces BERT output to 128-dimensional representation
Feature Input	Numerical / Structured Input	(n_features)	Auxiliary input data (e.g., metadata, numerical features)
Reshape	-	(n_features, 1)	Reshaped to fit Conv1D input requirements
Conv1D Layer	ReLU	(feature_maps)	Applies 1D convolution over reshaped features
Global Max Pooling	-	(1)	Reduces each feature map to its max value
Concatenate	-	(128 + pooled_features)	Concatenates BERT and CNN branches
Dense Layer	ReLU	(64)	Fully connected layer after fusion
Output Layer	Softmax / Sigmoid	(n_classes / 1)	Final classification output

Table 4. Hardware / software configuration

Sr. No	Hardware / Software	Specification
1.	OS	Google Colab (Cloud-based)
2.	CPU	Hosted on Google Cloud (GPU backend)
3.	RAM	~28 GB (Colab environment)
4.	Python	Version 3.10
5.	TensorFlow	Version 2.15
7.	Transformers	Version 4.38
8.	Platform	Google Colab

Tool descriptions:

Python 3.10. A versatile, high-level programming language known for its readability and extensive support for scientific and AI libraries. It simplifies model development and integration for machine learning tasks.

Google Colab. A cloud-based development environment that supports GPU/TPU acceleration and allows real-time collaboration. It is ideal for deep learning prototyping and experimentation.

TensorFlow. A ML framework by Google that is open-source and developed by the company. Applications that are powered by machine learning can be deployed with the help of this ecosystem, which offers a comprehensive and versatile collection of tools, libraries, and community resources.

Transformers. A state-of-the-art library providing thousands of pre-trained models for natural NLU and generation (NLG). It supports architectures like BERT, GPT, etc., and is highly optimized for deep learning NLP pipelines.

6.2 Result analysis

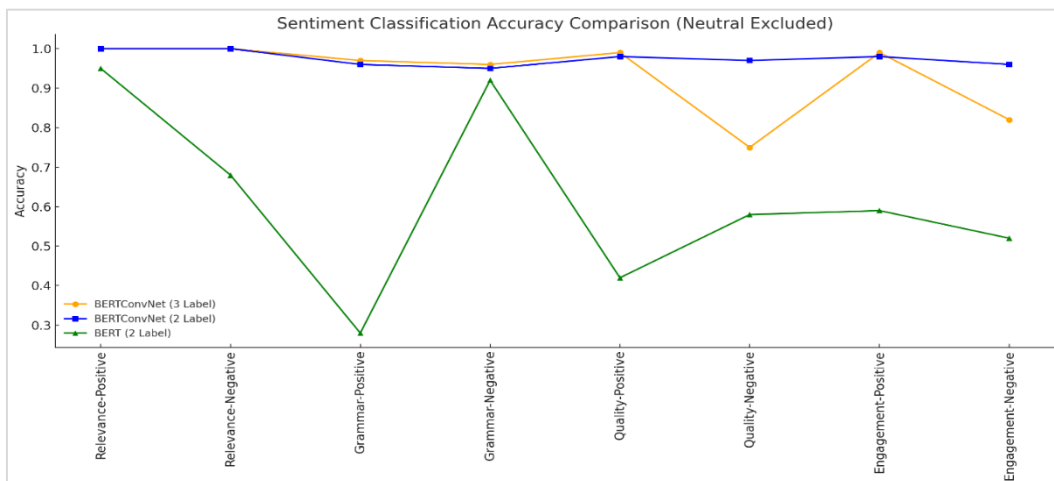
The experimental results are based on two datasets: a self-created YouTube ABSA dataset and the SemEval 2014 dataset, which includes Restaurant and Laptop reviews. These datasets were used to assess and compare the performance of three models—BERT, XLNet, and the proposed BERTConvNet—under both binary (positive, negative) and three-class (positive, negative, neutral) sentiment classification settings. The results highlight the models' effectiveness in Aspect-Based Sentiment Analysis across diverse domains and sentiment structures.

6.2.1 Result of self-created YouTube ABSA dataset

Table 5 presents the binary and three-label based sentiment classification accuracy (per label) for all classes in the YouTube ABSA dataset, covering the aspects as Relevance, Grammar, Quality, and Viewer Engagement. Accuracy is reported separately for each sentiment label per aspect to ensure a detailed evaluation of model performance in both binary and multi-class scenarios. The table also compares the performance of BERT and the suggested model (BERTConvNet). The Proposed Model consistently outperforms BERT across all aspects and sentiment labels. It achieves perfect accuracy (1.0) in Relevance, and near-perfect scores in Grammar (0.97 positive / 0.96 negative), Quality (0.99 / 0.98), and Viewer Engagement (0.99 / 0.98), significantly surpassing BERT's performance.

Table 5. Accuracy per label for binary and three-class classification on the YouTube ABSA dataset

Label	YouTube ABSA Dataset		
	BERTConvNet_Accuracy (3 Label)	BERTConvNet_Accuracy (2 Label)	BERT_Accuracy (2 Label)
Relevance_sentiment-Positive	1.0	1.0	0.95
Relevance_sentiment-Negative	1.0	1.0	0.68
Relevance_sentiment-Neutral	1.0	-	-
Grammar_sentiment-Positive	0.97	0.96	0.28
Grammar_sentiment-Negative	0.96	0.95	0.92
Grammar_sentiment-Neutral	0.95	-	-
Quality_sentiment-Positive	0.99	0.98	0.42
Quality_sentiment-Negative	0.75	0.97	0.58
Quality_sentiment-Neutral	0.86	-	-
Viewer Engagement_sentiment-Positive	0.99	0.98	0.59
Viewer Engagement_sentiment-Negative	0.82	0.96	0.52
Viewer Engagement_sentiment-Neutral	0.87	-	-

**Figure 11.** Models accuracy per label comparison graph on self-created YouTube ABSA dataset

6.2.2 Results of restaurant review dataset

The macro-average accuracy across all sentiment labels for BERTConvNet_Accuracy (3 Label) is 0.93, indicating consistently better classification performance. This balanced average supports the model’s generalizability and suggests that the perfect scores are not isolated artifacts but part of a broader, robust performance.

Notably, neutral sentiments exhibited higher misclassification rates due to their inherently ambiguous phrasing, which often lacks strong emotional or contextual cues, posing a greater challenge for accurate classification. Figure 11 shows the corresponding line graph comparison is

presented.

Table 6 compares the performance of two models, XLNet and BERT, across different sentiment classes. The sentiment categories include "Food", "Price", "Service", "Ambience", and "Anecdotes/miscellaneous", each evaluated for both positive and negative sentiments. The accuracy scores for a restaurant dataset across multiple sentiment categories with both 3-label and 2-label classifications. In most categories, BERT slightly outperforms XLNet. Overall, both XLNet and BERT shows strong and comparable results across various sentiment classes. Figure 12 shows the corresponding line graph comparison.

Table 6. Accuracy per label for binary and three-class classification on the restaurant dataset

Label	BERT_Accuracy (3 Label)	BER_Accuracy (2 Label)	XLNet_Accuracy (2 Label)
FOOD-Positive	0.93	0.96	0.95
FOOD-Negative	0.93	0.93	0.91
FOOD-Neutral	0.66	-	-
Positive-PRICE	0.89	0.97	0.98
Negative-PRICE	0.90	0.98	0.97
Neutral-PRICE	0.61	-	-
SERVICE-Positive	0.96	0.98	0.96
SERVICE-Negative	0.91	0.94	0.92
SERVICE-Neutral	0.71	-	-
AMBIENCE-Positive	0.89	0.94	0.93
AMBIENCE-Negative	0.89	0.94	0.96
AMBIENCE-Neutral	0.75	-	-
ANECDOTES_MISCELLANEOUS-Positive	0.90	0.93	0.93
ANECDOTES_MISCELLANEOUS-Negative	0.86	0.85	0.85
ANECDOTES_MISCELLANEOUS-Neutral	0.90	-	-

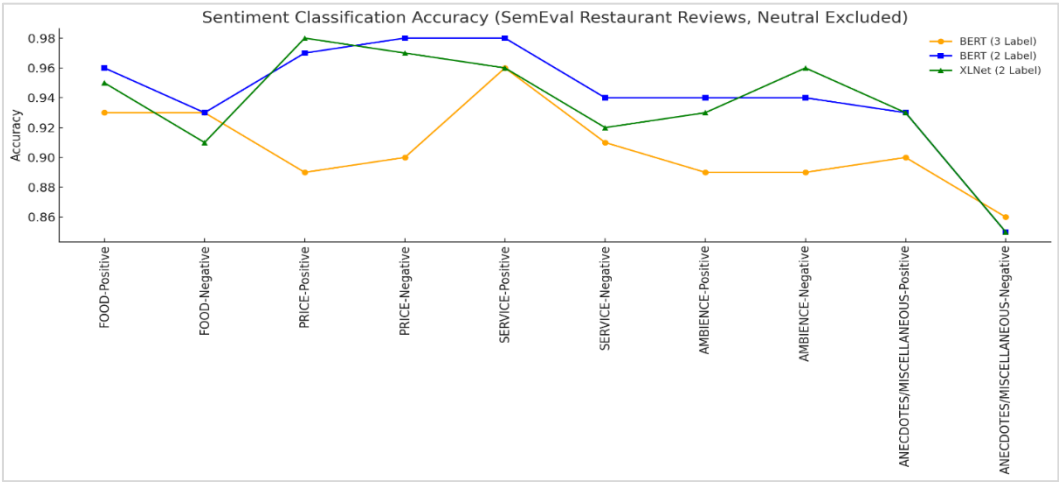


Figure 12. Models accuracy per label comparison graph on semeval (restaurant review) dataset

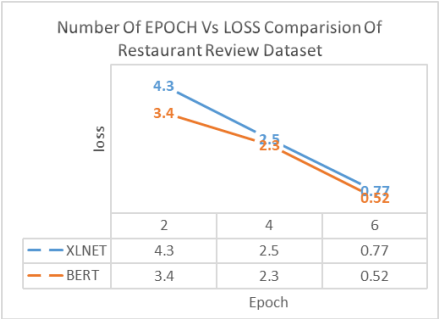


Figure 13. Loss comparison graph of models on restaurant review dataset

The Loss comparison graph of two models, XLNet and BERT, across three epochs (2, 4, and 6) on the Restaurant Review Dataset is shown in Figure 13. Overall, the BERT model consistently outperforms XLNet, particularly at later epochs, indicating its superior optimization and faster convergence on this dataset.

6.2.3 Results of laptop review dataset

Table 7 provides a comparative analysis of XLNet and BERT models on a laptop review dataset across various product-related sentiment categories, including General, Price, Quality, Design Features, Operation Performance, Usability, Connectivity, Portability and Miscellaneous. BERT generally

performs slightly better in positive sentiment detection, while XLNet shows marginal advantages in some negative sentiment categories. In the three-label setting, the "Neutral" label tends to have lower accuracy across most categories compared to the "Positive" and "Negative" labels. Overall, both XLNet and BERT exhibit solid and comparable

performance across sentiment classes, with minor differences in specific categories and sentiment types, reflecting the nuanced behavior of sentiment classification across various laptop features. Figure 14 shows the corresponding line graph comparison.

Table 7. Accuracy per label for binary classification on the laptop dataset

Laptop Dataset			
Label	BERT_Accuracy (3 Label)	BERT_Accuracy (2 Label)	XLNet_Accuracy (2 Label)
GENERAL-Positive	0.88	0.88	0.79
GENERAL-Negative	0.83	0.79	0.80
GENERAL-Neutral	0.43	-	-
OPERATION_PERFORMANCE-Positive	0.76	0.80	0.72
OPERATION_PERFORMANCE-Negative	0.77	0.83	0.84
OPERATION_PERFORMANCE-Neutral	0.34	-	-
DESIGN_FEATURES-Positive	0.91	0.90	0.83
DESIGN_FEATURES-Negative	0.54	0.62	0.62
DESIGN_FEATURES-Neutral	0.12	-	-
USABILITY-Positive	0.76	0.92	0.86
USABILITY-Negative	0.36	0.72	0.69
USABILITY-Neutral	0.05	-	-
PORTABILITY-Positive	0.88	0.84	0.78
PORTABILITY-Negative	0.54	0.36	0.47
PORTABILITY-Neutral	0.47	-	-
PRICE-Positive	0.72	0.65	0.65
PRICE-Negative	0.47	0.36	0.36
PRICE-Neutral	0.86	-	-
QUALITY-Positive	0.72	0.77	0.79
QUALITY-Negative	0.87	0.87	0.55
QUALITY-Neutral	0.86	-	-
MISCELLANEOUS-Positive	0.76	0.85	0.84
MISCELLANEOUS-Negative	0.68	0.72	0.67
MISCELLANEOUS-Neutral	0.12	-	-
CONNECTIVITY-Positive	0.66	0.74	0.74
CONNECTIVITY-Negative	0.85	0.44	0.72
CONNECTIVITY-Neutral	0.30	-	-

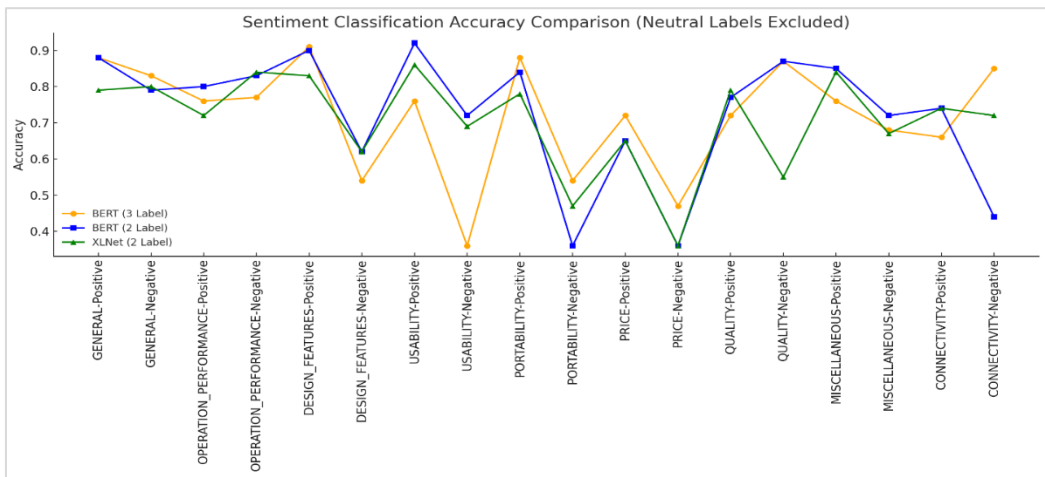


Figure 14. Models accuracy per label comparison graph on semeval (laptop review) dataset

Figure 15 compares the loss values of XLNet and BERT, across three epochs (2, 4, and 6) on the Laptop Review Dataset. The loss values for both models decrease as the number of epochs increases, indicating improved performance.

6.3 Discussion

The evaluation across three datasets—the self-created

YouTube ABSA dataset and the SemEval 2014 Restaurant and Laptop review datasets—presents that the proposed BERTConvNet model outperforms traditional models in sentiment classification tasks.

On the YouTube dataset, BERTConvNet achieved consistently higher accuracy than BERT across all aspects, particularly excelling in binary sentiment classification. Figure 11 confirms its superior performance visually. In the Restaurant Review dataset, both BERT and XLNet showed

strong and comparable results. BERT slightly outperformed XLNet in positive sentiment detection, while XLNet had minor advantages in some negative sentiment categories. Similarly, in the Laptop Review dataset, BERT generally led in positive sentiment detection, whereas XLNet performed slightly better in certain negative categories. Neutral sentiment remained the most challenging for both models. Overall, BERTConvNet is highly effective for user-generated content, while BERT and XLNet are competitive for structured reviews.

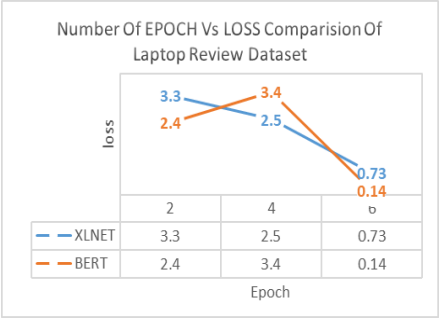


Figure 15. Loss comparison graph laptop review dataset

While the proposed BERTConvNet model shows improved performance in fake review detection, its computational efficiency remains a concern. Experiments on Google Colab revealed that BERTConvNet required about 20% more training time than the baseline BERT model, mainly due to the added convolutional layers. This trade-off between accuracy and processing overhead is important for deployment in resource-limited settings.

7. CONCLUSION

This study presents a robust and scalable framework for evaluating YouTube content using ABSA powered by transformer-based deep learning models. By integrating the proposed BERTConvNet architecture with word embedding and metadata encoding, the system effectively classifies sentiments across four crucial aspects: relevance, grammar, quality, and viewer engagement. Experimental results on both a self-created YouTube ABSA dataset and benchmark SemEval 2014 datasets (Restaurant and Laptop reviews) demonstrate the superiority of BERTConvNet over traditional models like BERT and XLNet. Specifically, BERTConvNet achieved near-perfect accuracy in binary sentiment classification and showed notable improvements in F1-score—up to 4% higher than baseline models—especially in handling subtle, aspect-specific sentiments. The findings confirm that BERTConvNet not only enhances the accuracy of content evaluation but also supports the reliable identification of fake or deceptive reviews. In comparison, BERT and XLNet performed competitively on structured review data, but struggled more with the complexity of user-generated content in the YouTube dataset. The proposed method’s adaptability makes it well-suited for deployment across various platforms to support content moderation and sentiment authenticity. Overall, this research highlights the transformative impact of combining advanced NLP techniques with deep learning for sentiment-driven content validation. The proposed approach contributes to increasing transparency, improving user trust, and supporting content creators and

platform administrators in managing authentic engagement. Future work will explore enhanced detection of negative sentiments and expanding dataset coverage to enable broader, cross-platform applicability.

A key limitation of the current system lies in its reliance on a self-created YouTube dataset, which may introduce biases due to its English-centric content and lack of demographic diversity among commenters. Future work will focus on expanding the dataset to include multilingual and demographically varied user comments, while also incorporating privacy safeguards and ethical data handling practices.

REFERENCES

- [1] Wang, J., Huang, J.X., Tu, X., Wang, J., Huang, A.J., Laskar, M.T.R., Bhuiyan, A. (2024). Utilizing bert for information retrieval: Survey, applications, resources, and challenges. *ACM Computing Surveys*, 56(7): 185. <https://doi.org/10.1145/3648471>
- [2] Sun, P., Bi, W., Zhang, Y., Wang, Q., Kou, F., Lu, T., Chen, J. (2024). Fake review detection model based on comment content and review behavior. *Electronics*, 13(21): 4322. <https://doi.org/10.3390/electronics13214322>
- [3] Hashmi, E., Yayilgan, S.Y., Shaikh, S. (2024). Augmenting sentiment prediction capabilities for code-mixed tweets with multilingual transformers. *Social Network Analysis and Mining*, 14(1): 86. <https://doi.org/10.1007/s13278-024-01245-6>
- [4] Zhang, Y., Zhang, X., Li, B. (2023). EfficientNet-based multi-dimensional network optimization for Deepfake video detection. In *Proceedings of the 2023 6th International Conference on Artificial Intelligence and Pattern Recognition*, Xiamen, China, pp. 433-438. <https://doi.org/10.1145/3641584.3641648>
- [5] Thilagavathy, A., Therasa, P.R., Jasmine, J.J., Sneha, M., Lakshmi, R.S., Yuvanthika, S. (2023). Fake product review detection and elimination using opinion mining. In *2023 World Conference on Communication & Computing (WCONF)*, RAIPUR, India, pp. 1-5. <https://doi.org/10.1109/WCONF58270.2023.10234996>
- [6] Su, B., Peng, J. (2023). Sentiment analysis of comment texts on online courses based on hierarchical attention mechanism. *Applied Sciences*, 13(7): 4204. <https://doi.org/10.3390/app13074204>
- [7] Gruetzemacher, R., Paradice, D. (2022). Deep transfer learning & beyond: Transformer language models in information systems research. *ACM Computing Surveys (CSUR)*, 54(10s): 204. <https://doi.org/10.1145/3505245>
- [8] Jian, Y., Chen, X., Wang, H. (2022). Fake restaurant review detection using deep neural networks with hybrid feature fusion method. In *International Conference on Database Systems for Advanced Applications*, pp. 133-148. https://doi.org/10.1007/978-3-031-00129-1_9
- [9] Sánchez-Reolid, R., de la Rosa, F.L., López, M.T., Fernández-Caballero, A. (2022). One-dimensional convolutional neural networks for low/high arousal classification from electrodermal activity. *Biomedical Signal Processing and Control*, 71: 103203. <https://doi.org/10.1016/j.bspc.2021.103203>
- [10] Zhou, Y., Li, J., Chi, J., Tang, W., Zheng, Y. (2022). Set-CNN: A text convolutional neural network based on

- semantic extension for short text classification. *Knowledge-Based Systems*, 257: 109948. <https://doi.org/10.1016/j.knosys.2022.109948>
- [11] Nagelli, A., Saleena, B. (2022). Optimal trained Bi-long short term memory for aspect based sentiment analysis with weighted aspect extraction. *Journal of Web Engineering*, 21(7): 2115-2148. <https://doi.org/10.13052/jwe1540-9589.2176>
- [12] Fregoso, J., Gonzalez, C.I., Martinez, G.E. (2021). Optimization of convolutional neural networks architectures using PSO for sign language recognition. *Axioms*, 10(3): 139. <https://doi.org/10.3390/axioms10030139>
- [13] Budhi, G.S., Chiong, R., Wang, Z., Dhakal, S. (2021). (2021). Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. *Electronic Commerce Research and Applications*, 47: 101048. <https://doi.org/10.1016/j.elerap.2021.101048>
- [14] Zhao, C., Feng, R., Sun, X., Shen, L., Gao, J., Wang, Y. (2024). Enhancing aspect-based sentiment analysis with BERT-driven context generation and quality filtering. *Natural Language Processing Journal*, 7: 100077. <https://doi.org/10.1016/j.nlp.2024.100077>
- [15] Aziz, K., Ji, D., Chakrabarti, P., Chakrabarti, T., Iqbal, M.S., Abbasi, R. (2024). Unifying aspect-based sentiment analysis BERT and multi-layered graph convolutional networks for comprehensive sentiment dissection. *Scientific Reports*, 14(1): 14646. <https://doi.org/10.1038/s41598-024-61886-7>
- [16] Chauhan, A., Sharma, A., Mohana, R. (2025). An enhanced aspect-based sentiment analysis model based on RoBERTa For text sentiment analysis. *Informatica*, 49(14): 193-202. <https://doi.org/10.31449/inf.v49i14.5423>
- [17] Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S. (2016). Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2016)*, Dublin, Ireland, pp. 27-35. <https://doi.org/10.3115/v1/S14-2004>
- [18] Rana, M.R.R., Nawaz, A., Ali, T., Alattas, A.S., Abdelminaam, D.S. (2024). Sentiment analysis of product reviews using transformer enhanced 1D-CNN and BiLSTM. *Cybernetics Inform. Technol*, 24(3): 112-131. <https://doi.org/10.2478/cait-2024-0028>
- [19] Katya, E., Rahman, S.R. (2024). Applications of natural language processing in social media sentiment analysis. *International Journal of Recent Advances in Engineering and Technology*, 13(1): 13-19. <https://journals.mriindia.com/index.php/ijraet/article/view/56>
- [20] Xu, S., Jia, Y., Zhang, Z., Xiang, Y. (2023). Research on aspect-based sentiment analysis based on XLNet-GCN. In *Proceedings of the 2023 12th International Conference on Computing and Pattern Recognition*, Qingdao, China, pp. 491-495. <https://doi.org/10.1145/3633637.3633714>