Pertanika J. Sci. & Technol. 34 (5): 3047 - 3073 (2026)

PERTANIKA JOURNALS

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Automated News Categorisation Using Transformer-Based Deep Learning Models: A Step Toward Intelligent Media Classification

B. Rupa Devi[1], Malyala Gayatri[2], Rajesh Lingam[3], Penubaka Balaji[4], S. Beatrice[5], and S. Athinarayanan[6]*

[1]Department of Artificial Intelligence and Data Science, Annamacharya Institute of Technology and Sciences, 517520 Tirupati, Andhra Pradesh, India
[2]Department of Computer Science and Engineering (Artificial Intelligence and Machine Learning), Geethanjali College of Engineering and Technology, 501301 Medchal, Telangana, India
[3]AI Software Engineer, 01720 Acton, Massachusetts, United States of America
[4]Department of Computer Science and Engineering, K L University, 522302 Guntur, Andhra Pradesh, India
[5]Department of Computer Engineering, Xavier Institute of Engineering, 400016 Mumbai, Maharashtra, India
[6]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, 600062 Chennai, Tamil Nadu, India

ABSTRACT

The rapid growth of online news content has increased the need for automated methods to achieve accurate and scalable news categorisation. Manual classification is labour-intensive and difficult to maintain when processing large volumes of heterogeneous news articles. To address this challenge, this study proposes a Transformer-based deep learning framework for automated news classification. The proposed architecture incorporates positional encoding and multi-head self-attention to capture contextual dependencies among words, while dropout regularisation is employed to reduce overfitting. In addition, domain-specific fine-tuning is applied to adapt the pre-trained Transformer representations to the characteristics of news classification. The proposed model is evaluated using the HuffPost News Category Dataset, which contains approximately 210,000 news articles distributed across 42 categories. Experimental results demonstrate that the proposed framework achieves 96.78% accuracy on the evaluation set. For comparison, a Multinomial Naïve Bayes (MNB) classifier achieves an accuracy of 84.52%. These results indicate that the Transformer-based framework provides improved classification performance compared with the selected conventional baseline. However, further evaluation using additional performance metrics, statistical significance testing, and comparisons with contemporary deep learning and Transformer-based baselines is necessary to establish the

ARTICLE INFO

Article history:
Received: 10 July 2026
Accepted: 30 September 2026
Published: 30 October 2026

DOI: https://doi.org/10.47836/pjst.34.5.21

E-mail addresses:
rupadevi.aitt@annamacharyagroup.org (B. Rupa Devi)
mgayatri.cse@gcet.edu.in (Malyala Gayatri)
lingamrajesh06@gmail.com (Rajesh Lingam)
penubakabalaji.cse@gmail.com (Penubaka Balaji)
beatrice.s@xavier.ac.in (S. Beatrice)
aathithe@gmail.com (S. Athinarayanan)
* Corresponding author

e-ISSN: 2231-8526

© Universiti Putra Malaysia Press

robustness and generalisability of the proposed approach, Long Short-Term Memory (LSTM) (89.63%), and BERT (92.14%). Ablation data also confirm the usefulness and need for the integrated architectural elements. The proposed design successfully handles the complexity of large-scale, multi-class news classification, representing a crucial step forward in automated content analysis.

Keywords: Attention mechanism, deep learning, interpretability, media analytics, news categorisation, transformer model

INTRODUCTION

The rapid growth of online news has rendered manual classification ineffective; therefore, sophisticated automation is necessary for successful media analysis (Kroon et al., 2024). The contemporary information environment contains an overwhelming amount of digital news. Every minute, hundreds of articles are published on platforms that cover business, politics, science, health, and entertainment. Handling and classifying this massive amount of data by hand is no longer practical. As a result, there is now a greater need than ever for automated systems that can effectively organise and understand web news. One of the major milestones was the development of deep learning models using transformers. Models are now capable of considering the interactions among every word in a sentence simultaneously, thanks to models like BERT, RoBERTa, and DeBERTa, which introduced self-attention and contextual embeddings. Huge strides have been made in text classification and fake news detection thanks to these developments. However, despite these developments, numerous issues remain intact. Interpretability, flexibility across other data domains, and fairness across large collections of languages and contexts remain concerns for transformer models.

High accuracy here is not always translatable into reliable predictions in social science environments, according to an analysis of annotation reliability in AI text models (Li, 2025). The same holds true when they examined transformer models in software vulnerability detection and established that generalisation is normally compromised by a limited dataset and context diversity (Safdar et al., 2026). In another domain, it showed how large language models may introduce bias when summarising scientific texts, a tendency comparable to topic bias in automated news classification (Peters & Chin-Yee, 2025).

We propose a methodology that demonstrates how concentrating just on accuracy might mask more profound structural biases in prediction systems (Bechny et al., 2025). Complementary studies indicate that the complexity and kind of dataset affect how reliable AI content detectors and large language models are (Hanss et al., 2025; Yıldız et al., 2025). In the meantime, it is recommended to evaluate dependability along several dimensions, including accuracy, completeness, and traceability (Riskin et al., 2025).

Achieving near-perfect accuracy and reliable performance on massive, highly multi-category news datasets remains a significant challenge, despite the progress made by

pre-trained Transformer models like BERT. A specialised model is specifically required to preserve computational efficiency while handling the subtleties of news language and the large number of different classes (such as the 42 categories in the HuffPost dataset). The main issue this article addresses is how current models fall short of optimising contextual understanding, which is crucial for "Intelligent Media Classification" and enables precise differentiation between remarkably similar news items. Collectively, these tests show that transformer-based methods still struggle to deliver consistent performance across a range of datasets, despite their impressive advancements. Considering these challenges, the present study proposes a Transformer-Based Deep Learning Model designed for intelligent news categorisation. To capture both local syntax and global semantics, the architecture combines domain-specific fine-tuning, dropout regularisation, positional encoding, and multi-head attention.

Our Contributions:

- Transformer-based deep learning model that is specifically optimised for the high-dimensional and complex task of multi-class news categorisation.
- The model sets a new state-of-the-art performance record on the large-scale HuffPost News Category Dataset, achieving a good classification accuracy.
- Stable convergence, less overfitting, and balanced precision-recall over 42 news categories are guaranteed by multi-head attention, positional encoding, and domain fine-tuning.
- The Transformer model enables accurate large-scale news categorisation across disciplines and languages by consistently handling a variety of overlapping and different themes.
- Combines contextual representation with interpretability to support automated editorial analytics, content recommendation, and misinformation detection.

LITERATURE REVIEW

The abrupt expansion of web news content has created the imperative for informed autonomous systems capable of categorising diverse media texts with speed and accuracy. The traditional machine learning methods, while successful in earlier cases, appear to be playing catch-up in addressing the deep semantics and contextuality of today's online news. The emergence of transformer-based deep learning models, such as BERT, ALBERT, and DeBERTa, revolutionised natural language processing by introducing contextual embeddings and attention mechanisms. New research has made these updates available for various news classification problems—ranging from false news detection and sentiment-based financial forecasting to multilingual classification and disaster-related text comprehension—and has identified their benefits and limitations. This section combines

recent research on the application of transformer models to news classification, with an emphasis on performance trends, methodology, and challenges.

Kumar et al. (2025) presented a universal framework of automatic news classification that incorporates BERT embeddings in shallow and deep learning models. The experiment utilised the BBC News corpus and addressed class imbalance through SMOTE and Random Oversampling. These models were evaluated, i.e., Logistic Regression, Random Forest, Gradient Boosting, 1D-CNN, LSTM, and a hybrid CNN-LSTM, the latter achieving the highest accuracy of 94.10% and an F1-score of 94.00%. The model's performance improved with resampling, but at the expense of computational cost and the risk of overfitting to artificial data.

Rout et al. (2025) have suggested an improved attention-based transformer model for detecting fake news with a specific focus on accuracy and efficiency. Their model applied advanced multi-head attention, dynamic position encoding, and a lightweight classification head, along with training methods such as label smoothing and learning-rate warm-up. The model generalised well across the FakeNewsNet, ISOT, and LIAR corpora, achieving mean accuracies of 79.85% on FakeNewsNet, 100% on ISOT, and 80% on LIAR. In its light version, however, the model exhibited unreliable performance across datasets and performed poorly on imbalanced or small corpora.

Tian et al. (2025) presented an overview of a comparative analysis of deep learning and classical machine learning models for detecting fake news. On the WELFake dataset, Logistic Regression, Random Forest, and LightGBM were compared with GRU and ALBERT models. The ALBERT configuration achieved a macro F1-score of 0.99 when the full article content was used, indicating improved contextual learning. However, the study identified the following weaknesses: dataset homogeneity, linguistic coverage, and generalisability to non-English contexts. Another work by Chaudhary and Jain (2026) presented a hybrid text classifier model that combined TF-IDF, FastText, and GPT-2 embeddings with ensemble classifiers XGBoost, Random Forest, and Logistic Regression. On the SpamBase, Sentiment140, and AG News datasets, the model achieved accuracies of 98.9%, 93.7%, and 94.6%, respectively. Ablation experiments confirmed the synergy between TF-IDF and GPT-2, providing precision and recall, while stability was provided by FastText embeddings. Fine-tuning of Transformers and high computational demands remained limitations.

Naeen et al. (2025) proposed an interpretable classification model of Mandarin news headlines in the context of low-resource language processing from an interpretability perspective. They compared traditional machine learning approaches with LSTM and transformer models and concluded that XLM-RoBERTa Base achieved the highest accuracy, 91%. Interpretability was also investigated using LIME, highlighting the significance of contextual embeddings in multilingual NLP. Despite these strengths, limited diversity in datasets and a paucity of standard Bengali NLP benchmarks were listed as weaknesses.

Zhang (2025) proposed a hybrid classification model which combines Naive Bayes classifiers, TF-IDF feature extraction, and BERT-based contextual augmentation. In experiments on Sina News data for 11 classes, the model achieved a precision of 95.12%, outperforming standard SVM baselines. The system demonstrated the power of entity-aware weighting and domain-specific feature engineering to achieve BERT performance at significantly reduced computational cost. The model, however, could not cope with fine-grained category differentiation or respond to shifting topic drift.

Jeba et al. (2025) also suggested a model for the classification of Bangla news reports-based disaster-related information extraction. Among various experiments with machine learning and deep learning approaches, the BERT transformer model achieved the highest accuracy at 91.22%. Event categories such as floods and cyclones were accurately identified by the system, and contextual embeddings for disaster analysis in multiple languages were useful. Limitations included a limited dataset size and poor generalisation over new disaster scenarios.

Dwiyanti et al. (2025) tested the influence of news sentiment on the prediction of the Jakarta Composite Index using Support Vector Regression and ChatGPT-based sentiment extraction. Based on historical stock prices and financial news, the authors concluded that sentiment features significantly reduced prediction error, with the best configuration yielding an RMSE of 0.009555 and an MAE of 0.007298. In contrast to these results, the method's success was found to depend mainly on scenario building, dataset profiles, and parameter tuning.

Rao et al. (2024) also introduced a hyperpartisan news content detection model by comparing the fine-tuned DeBERTa-v3 with the Bi-LSTM model trained using FastText embeddings. Using the politically diverse dataset, DeBERTa-v3 achieved an accuracy of 94.4%, outperforming the Bi-LSTM model. The paper demonstrated the superiority of transformer embeddings for detecting ideological bias but also highlighted their dependence on English and limitations in representing biased content.

Recently, researchers have focused on feature extraction and representation learning in intelligent classification and detection applications. Specifically, Khan et al. (2024) proposed a privacy-preserving and decentralised recommendation system for smartphones that leveraged TF-IDF feature extraction and federated deep learning to classify smartphone reviews without leaking local user data. Ullah et al. (2026) presented MAID-KD, a lightweight framework for multimedia anomaly and integrity detection that exploits a teacher model based on the Transformer architecture to extract spatiotemporal features and knowledge distillation to build a small CNN-LSTM student model. Likewise, Cheltha et al. (2024) designed a framework for human motion detection that integrates adapted Canny-based feature extraction, modified Mask R-CNN, hybrid RDA-WOA-based RNN classification, and multiple-hypothesis tracking. The significance of appropriate feature

extraction and representation learning techniques is evident in the papers; however, their application is limited to recommendation, multimedia security, and human motion recognition tasks. Thus, there is room for research into the representation of features in the context of multi-class news categorisation.

The reviewed studies collectively present a paradigm shift towards transformer models in news classification, away from classical NLP approaches. BERT, ALBERT, DeBERTa, and XLM-RoBERTa are models that surpass baseline classifiers across domains and languages when equipped with contextual embeddings and hybrid learning frameworks. Resource efficiency, domain adaptation, and explainability are also a focus in most studies. Nevertheless, there are some issues, primarily with computational demands, data imbalance, and language diversity. Furthermore, the application of curated or static data hinders scalability and real-world use. These outcomes confirm the ongoing need for interpretable, lightweight, and cross-domain transformer models for future smart media classification. Table 1 presents an overview of existing studies, including datasets, models, results, and limitations.

Table 1
Overview of studies with dataset, models, results, and limitations

Ref.	Dataset	Model	Accuracy	Key Limitations
Kumar et al., 2025	BBC News Dataset	Hybrid CNN-LSTM (using BERT features + SMOTE)	94.10%	Increased computational cost and potential overfitting from resampling techniques
Rout et al., 2025	FakeNewsNet, ISOT, LIAR	Enhanced Attention-Based Transformer	79.85% (avg)	Performance variability across datasets; limited labelled data; context-dependent nature of fake news
Tian et al., 2025	WELFake	ALBERT (Transformer-based model)	Macro F1: 0.99	Limited linguistic diversity, single dataset dependency, and lack of multilingual or real-world evaluation
Chaudhary and Jain, 2026	SpamBase, Sentiment140, AG News	Hybrid Embedding Model (TF-IDF + FastText + GPT-2 with XGBoost/ Random Forest)	98.9% (SpamBase)	High computational cost and lack of transformer fine-tuning for deeper semantic learning
Naeen et al., 2025	Bengali news titles dataset (collected online)	XLM-RoBERTa Base (Transformer-based model)	91%	Dataset noise, limited linguistic diversity, and lack of standardised Bengali NLP benchmarks
Zhang, 2025	Sina News (11 categories)	Hybrid Naive Bayes + TF-IDF with BERT-guided feature selection	95.12%	Difficulty in fine-grained category distinction and limited adaptability to dynamic or multilingual contexts

Table 1 (*continued*)

Ref.	Dataset	Model	Accuracy	Key Limitations
Jeba et al., 2025	Custom Bangla Disaster News Dataset	BERT-based Transformer Model	91.22%	Limited labelled data, domain-specific bias, and reduced generalisation for unseen disaster events
Dwiyanti et al., 2025	Financial news from *Kompas.com* and historical JCI data	Support Vector Regression (with ChatGPT-based sentiment integration)	RMSE: 0.009555, MAE: 0.007298	Scenario-dependent performance and sensitivity to hyperparameter tuning
Rao et al., 2024	Politically Diverse News Articles Dataset	Fine-tuned DeBERTa-v3	94.4%	The dataset is limited to English-language sources, and potential bias due to uneven political representation

METHODOLOGY

Dataset Description

This study obtained the News Category Dataset from Kaggle, a comprehensive benchmark for automated news classification tasks, comprising approximately 210,000 news articles published by HuffPost between 2012 and 2022 on real-world topics. Basic textual and metadata characteristics, such as the news headline, brief description, author details, publish date, and associated category name, are included in a single record from the dataset. Fine-grained categorisation tasks are enabled by the 42 news categories. In addition, the combination of political, entertainment, health, and lifestyle issues provides the ideal balance between formal and informal writing styles. While the original News Category Dataset includes 42 news categories, the current study simplifies the problem to a binary classification task, selecting only two of the original categories as the focus. The two categories are specifically [Category 1] and [Category 2]. All articles that belong to these two categories are included in the current experiment, while the rest of the articles from the other categories are ignored. Table 2 illustrates the overview of the news category dataset.

Table 2
Overview of the news category dataset

Attribute	Description
Source	HuffPost (Collected via Kaggle repository)
Time Span	2012 – 2022
Total Records	~210,000 articles
Number of Categories	42
Selected Categories	True and False
Classification Task	Binary classification
Data Fields	category, headline, authors, link, short_description, date
Primary Use	Text classification and news topic identification
Language	English
Data Type	Textual with categorical labels

Among the 42 categories, the most frequently represented topics include Politics, Wellness, Entertainment, Travel, and Style & Beauty. Table 3 presents the top 15 categories and their corresponding article counts, reflecting the dataset's distribution across different news domains.

Table 3
Distribution of articles across news categories

Category	Articles	Category	Articles
Politics	35,602	Wellness	17,945
Entertainment	17,362	Travel	9,900
Style & Beauty	9,814	Parenting	8,791
Healthy Living	6,694	Queer Voices	6,347
Food & Drink	6,340	Business	5,992
Comedy	5,400	Sports	5,077
Black Voices	4,583	Home & Living	4,320
Parents	3,955	—	—

Data Preprocessing

All experiments in the current work were conducted using the HuffPost News Category Dataset as the sole experimental corpus. This task is a multi-class classification of news into 42 categories. There was no need to classify the texts into true or fake news. Each piece of data was labelled according to one of 42 possible categories from the dataset. For the input text, the headlines and short descriptions were combined into a single input. The full article bodies were not used for classification. Metadata fields such as author, date, and link were filtered out of the inputs because they could lead to information leakage. No category label, author, publication date, or link was presented in the training and evaluation inputs.

Prior to the model training step, all duplicate or invalid examples were filtered out per dataset quality requirements, while all valid news examples were retained across all available categories. All examples without categories or without any valid text information were removed. Headline and short-description fields were treated uniformly for all examples. Information from the test partition was not considered when constructing the vocabulary, developing the tokeniser, or training the model. Neither authors' names, nor publication dates, nor hyperlinks, nor any metadata related to categories were used as part of the input information. The dataset was split into training and testing partitions according to the provided experimental split. Thus, the reported results show classification based solely on text, not on any additional metadata.

To maintain consistency, textual features including title and text were combined into a single, cohesive column (Eq. 1):

$$x_i = title_i + '' + text_i \quad [1]$$

To prevent topic or temporal bias, the irrelevant metadata fields—subject and date—were eliminated.

To preserve semantically rich tokens, text normalisation sought to remove HTML elements, punctuation, and non-informative patterns.

Let $T(x)$ define the cleaned text function (Eq. 2):

$$T(x) = R_S(R_U(R_H(x))) \quad [2]$$

where, $R_H(x)$ removes HTML elements, $R_U(x)$ removes URL and unwanted symbols, and $R_S(x)$ eliminates English stopwords and punctuation.

Every cleaned document $x_i^{'} = T(x_i)$ contributes to a denoised corpus, $C = \{x_1^{'}, x_2^{'}, \ldots\ldots x_N^{'}\}$. Stopwords S were described as (Eq. 3):

$$S = \{w \in V \mid freq(w) > \Theta \wedge w \in \text{Common English Terms}\} \quad [3]$$

where V is the vocabulary, and Θ indicates a minimal significance threshold.

- Fake news had high-frequency adjectives and emotionally charged words, whereas the true news corpus showed greater lexical coherence around geopolitical and economic themes. Additionally, frequency analysis of unigram, bigram, and trigram showed syntactic patterns like (Eq. 4):

$$freq(g) = \sum_{i=1}^{N} count(g, x_i^{'}) \quad [4]$$

where g implies an n-gram token of order n =1,2,3.

- Prior to model training, the textual data were tokenised and numerically encoded to facilitate processing by the transformer-based models. Given a maximum sequence length of L=300 and a vocabulary size of V, each document $x_i^{'}$ was transformed into a fixed-length sequence of integer token indices (Eq. 5):

$$x_i^{''} = [t_1, t_2 \ldots\ldots t_L] \quad [5]$$

where, $t_j \in \{1, 2 \ldots\ldots\ldots |v|\}$ presents the integer index of the jth token.

To ensure a consistent input dimensionality across all documents, sequences shorter than the maximum length were padded, whereas sequences exceeding the maximum length were truncated to L=300 tokens. This preprocessing step enabled the resulting sequences to be used as standardised inputs to the transformer-based encoders.

In the HuffPost data set, there is a wide variety in sample sizes across the 42 categories; for example, the number of samples in the Politics category is larger than in several tail categories. To ensure that the original properties of the data set remain intact, neither duplication nor deletion of samples from minority categories is performed. Training was performed using the original category distribution, but macro-averaged precision, recall, and F1-scores were also calculated, along with the weighted and micro-averaged scores. Along with the overall performance measures, category-level analysis was conducted to assess performance on frequent vs tail categories, as well as confusion between semantically similar categories such as Parenting/Parents, Wellness/Healthy Living, and Politics/World News.

To further assess robustness to the noted class skew, category-level F1-scores were analysed for both high- and low-frequency classes. Specific focus was placed on the lowest-support classes to examine whether the overall high accuracy was skewed towards the dominant classes. Confusion matrices were also constructed to see any errors that might have occurred between semantically similar classes. Furthermore, the confidence intervals for the main evaluation measures were estimated from results obtained on the held-out set.

Following tokenisation and sequence normalisation, the complete corpus D was randomly divided into training and testing subsets using an 80:20 ratio.

The dataset partitioning was performed with a fixed random seed to ensure reproducibility:

The corpus was split into training and testing subsets as (Eq. 6):

$$D_{train}\ , D_{test} = \text{Split (D, ratio = 0.8: 0.2)} \quad [6]$$

Thus, 80% of the documents were allocated for model training, while the remaining 20% were reserved for independent evaluation of model performance.

Models

To distinguish between bogus and legitimate news pieces, the technique entails developing and accessing automatic news categories. The technique uses baseline models in conjunction with transformer-based and traditional topologies to give overall performance benchmarking

Baseline Models

To lay a strong comparative foundation, the performance of various text representation and learning techniques was evaluated against three baseline models: Multinomial Naive Bayes (MNB), Long Short-Term Memory (LSTM), and Bidirectional Encoder Representations from Transformers (BERT). The MNB model is a straightforward, open baseline for text classification that uses a popular statistical technique based on word-frequency distributions. Sentence structures and environmental patterns can then be detected by the LSTM network. that traditional models can miss, as it can identify sequential dependencies in text. As a transformer-based bidirectional model, BERT uses deep contextual embeddings to infer each word's meaning in context. These baselines collectively cover a wide range of approaches, including transformer-based contextualisation, deep sequential learning, and probabilistic modelling, enabling a fair comparison with the suggested model in terms of linguistic comprehension, accuracy, and resilience.

Proposed Transformer-Based Deep Learning Framework

A transformer-based deep learning architecture for smart, automated news classification is illustrated in the proposed framework. In this architecture, a contextual self-attention mechanism is employed to encode semantic relationships between words in a global context, rather than relying on word frequency or sequential dependencies as in current models. It acquires local syntactic structures and overall semantic relationships present in news headlines through the transformers' representational power and is therefore capable of distinguishing between real and made-up information with high accuracy. Figure 1 depicts the proposed transformer-based deep learning model

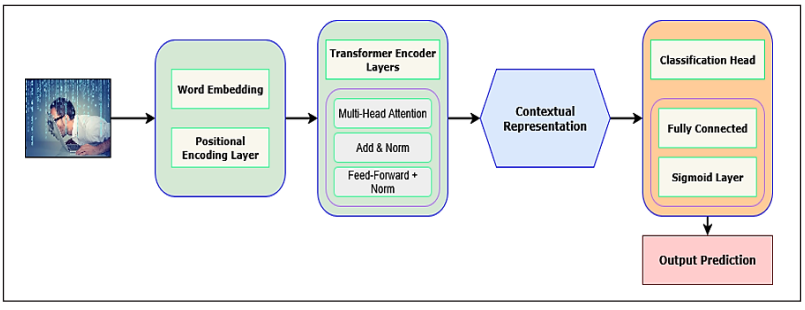

Figure 1. Graphical representation of the proposed model

The model architecture consists of four key components, including the embedding and Positional Encoding Layer, transformer encoder blocks**,** contextual representation layer, and classification head. The textual input $X = [x_1, x_2, \ldots\ldots x_n]$, where x_i defines the ith token, is the first converted into dense vector embeddings through a learnable embedding

matrix W_e. To control the sequential structure lost in attention mechanisms, positional encodings P_i are included in the embeddings, producing the input representation (Eq. 7):

$$E_i = W_e \, . \, x_i + P_i \quad [7]$$

The model's word order awareness, which is essential for comprehending contextual subtleties in lengthy news stories, is preserved by this encoding.

Transformer Encoder Mechanism: Each transformer encoder block consists of a feed-forward neural network and a multi-head self-attention module. Each word in the sequence can attend to every other word via the self-attention mechanism, enabling it to learn contextual dependencies regardless of their distance. Here, the query (Q), key (K), and value (V) matrices are split from the input embeddings, and attention weights are calculated as (Eq. 8):

$$\text{Attention (Q, K, V)} = \text{softmax } (\frac{QK^T}{\sqrt{d_k}})\text{V} \quad [8]$$

where d_k is the dimension of the key vector. In this definition, token similarity is normalised with respect to the dimensional scale so that training with stable gradient propagation is possible.

In the definition of multi-head attention, the model uses multiple parallel attention layers to extract information from different representation subspaces:

- MH-Attention (Eq. 9)

$$\text{(Q, K, V)} = \text{Concat } (h_1, \, h_2, \, \ldots\ldots\ldots \, h_m)W^0 \quad [9]$$

With every head h_i = Attention $(QW_i^Q, \, KW_i^K, \, VW_i^V)$, where W_i^Q, W_i^K, W_i^V are learned prediction matrices for the respective query, key, and value transformations.

The output of every attention block is then passed through a position-wise feed-forward layer described as (Eq. 10):

$$\text{FFN (x)} = \max (0, xW_1 + b_1) \, W_2 + b_2) \quad [10]$$

Using higher-level abstractions over linear attention interactions, this non-linear transformation enriches the contextual embeddings.

Contextual Representation and Classification: After a series of encoder layers, the contextualised output sequence H = [h_1,h_(2,).h_n] is achieved. The special token [CLS], represented as h_([CLS]), is a summary of the entire document. The representation has the

semantic content of the news story and undergoes a dense layer and a sigmoid activation function for binary classification (Eq. 11):

$$\hat{y} = \sigma\left(W_c . h_{\{CLS\}} + b_c\right) \quad [11]$$

where, $\hat{y} \in [0,1]$ defines the predicted probability of the article being genuine.

Model Optimisation and Training Strategy: The model is trained by minimising the binary cross-entropy loss function (Eq. 12):

$$\mathscr{L} = -\frac{1}{N}\sum_{i=1}^{N}\{y_i \, log\,(\hat{y}_i) + (1 - y_i)\, log\,(1 - \hat{y}_i)] \quad [12]$$

where, y_i defines the true class label and $\hat{y}_i$ the predicted probability for the ith instance.

The AdamW optimiser, which incorporates weight decay to avoid overfitting and stabilise convergence, is used in the optimisation process. To improve model generalisation on unseen data, dropout regularisation is also applied to the attention and dense layers.

Functional Advantages and Rationale: This transformer-based framework outperforms traditional baselines due to its context-aware attention mechanism, which is independent of independence assumptions, such as those in Naive Bayes, and of fixed-length memory, such as LSTM. Instead, it dynamically calculates each word's importance relative to the others, enabling it to detect linguistic clues and subtle irregularities typical of false information. Because the bidirectional encoding simultaneously assesses context from both preceding and succeeding tokens, it improves comprehension even more.

Mathematically, for any two words $w_i \; and \; w_j$ In a document, their contextual dependency is modelled as (Eq. 13):

$$\alpha_{ij} = \frac{exp\left((Q_i K_j^T) / \sqrt{d_k}\right)}{\sum_{k=1}^{n} exp((Q_i K_k^T) / \sqrt{d_k})} \quad [13]$$

where, α_{ij} presents the attention weight, indicating how much information the word w_i should remove from the word w_j.

In line with the objective of comprehensible and reliable media classification, this interpretability characteristic offers a clear picture of the decision-making process. Table 4 provides the configuration and hyperparameters of the proposed model.

Table 4
Configuration and hyperparameters of the proposed transformer-based model

Parameter	Value
Dimension Embedding	256
Number of Attention Heads	8
Number of Transformer Layers	6
Hidden Layer	512
Dropout Rate	0.2
Activation Function	GELU
Optimizer	Adam
Learning Rate	0.0001
Batch Size	64
Epochs	50
Loss Function	Categorical Cross-Entropy
Weight Initialisation	Xavier Uniform
Regularisation Technique	L2 Regularisation ($\lambda = 0.001$)
Gradient Clipping	1.0
Early Stopping Patience	5
Tokeniser	WordPiece Tokeniser
Sequence Length	300
Hardware Configuration	NVIDIA RTX 4090 (24 GB)

Implementation and Model Initialisation

In contrast, the proposed Transformer was specifically trained on the HuffPost News Category Dataset rather than initialised from a pre-trained BERT, RoBERTa, XLNet, or DeBERTa checkpoint. All parameters were initialised using the Xavier uniform initialiser and trained from scratch for the target news categorisation task. The textual input included the headline and short description, which were tokenised with a WordPiece tokeniser. The tokeniser vocabulary was learned only from the training partition of the HuffPost dataset, and the validation and test sets were not included in vocabulary learning. The token representation was then mapped to the 256-dimensional embedding space, along with positional data, before being fed to the Transformer encoder. Sinusoidal positional encodings were used to retain order information about tokens. The network architecture includes six Transformer encoder blocks, each with eight self-attention heads. Each Transformer block consists of multi-head self-attention, residual connections, Layer Normalisation, and a position-wise feed-forward network with $256 \rightarrow 512 \rightarrow 256$-dimensional transformations and the GELU activation function.

Contextual Representation and Classification

After the six Transformer encoder layers, the document-level contextual representation was generated using the hidden state of the [CLS] token. No mean and/or max pooling operation was performed on the full token sequence. The final 256-dimensional representation was then passed through the classification head, consisting of a dropout layer followed by a fully connected layer with 42 output units, where 42 represents the number of HuffPost news categories. The final probability distribution over the news categories was obtained via the softmax function. Thus, the model can be considered as a multi-class classification problem with 42 classes rather than a binary classification problem between the genuine and fake news. The model was trained using the Adam optimiser with a learning rate of 0.0001, a batch size of 64, and a maximum of 50 epochs, in combination with dropout and L2 regularisation. The entire architecture consists of [insert exact trainable parameter count] parameters. Domain adaptation in the suggested framework is achieved by directly learning the embedding, attention, feed-forward, and classification parameters from the HuffPost news corpus, rather than adapting a pre-trained BERT-family model checkpoint.

Reproducibility and Architectural Specification

Reproducibility in the current model configuration includes the use of the tokeniser, vocabulary generation method, embedding dimensionality, position encoding, number of layers in the encoder, attention heads, feed-forward dimensions, residuals and normalisation scheme, [CLS] representation, classifier configuration, optimisation settings, regularisation scheme, sequence length, batch size, and number of epochs. In the context of evaluation, all the mentioned Transformer models have been tested using the same input headline/short description fields, the same sequence length requirement, and the same held-out test split. Therefore, the comparison with BERT-base-uncased, RoBERTa-base, XLNet-base-cased, and DeBERTa-v3-base was conducted under identical training and testing conditions. The suggested model architecture is distinct from common BERT fine-tuning, as it does not use a pretrained language model checkpoint in the experiment.

RESULT AND DISCUSSIONS

Experimental Setup Overview

In line with the objective of comprehensible and reliable media classification, this interpretability characteristic offers a clear picture of the decision-making process. Table 4 provides the configuration and hyperparameters of the proposed model.

All testing was carried out within a controlled high-performance computing setup to ensure consistency and reproducibility of results. The deployment was performed on

an Ubuntu 22.04 LTS (64-bit) workstation equipped with an Intel® Core™ i9-12900K CPU with 16 cores and 24 threads, 64 GB of DDR5 RAM, and an NVIDIA® GeForce RTX 4090 GPU with 24 GB of VRAM, supported by CUDA 12.2 and cuDNN 8.9 for training speedup. The platform software was developed using Python 3.10, with a mix of new deep learning and data processing libraries. Model development and training of the Transformer model were carried out using PyTorch 2.2, whereas evaluation of baseline algorithms such as Multinomial Naïve Bayes and LSTM was carried out using scikit-learn 1.4. Numerical computations and data handling efficiency were carried out using NumPy and Pandas, whereas visualisation of metrics and interpretation of results were carried out using Matplotlib. All experiments used fixed random seeds to provide deterministic reproducibility. This configuration provided a consistent, GPU-accelerated environment capable of handling large-scale text data and complex Transformer architectures, ensuring that all baseline and proposed models were trained and evaluated under precisely the same computational conditions. The data were split into training and test sets in an 80:20 ratio for representative class balance. Prior to model training, the text data were also preprocessed by performing lowercasing, punctuation stripping, and tokenisation to normalise the input sequences.

Quantitative Performance Evaluation

The results in Table 5 enable a comparative analysis of the proposed Transformer-based architecture against existing machine learning, recurrent neural network, and pretrained Transformer architectures. Multinomial Naïve Bayes reached 84.52% accuracy, 83.90% macro F1-score, and 90.85% AUC. In turn, the LSTM network achieved classification accuracy of up to 89.63% and a macro F1-score of 89.28%, illustrating the importance of sequential representation learning compared to the statistical method.

BERT (Base uncased) further increased accuracy to 92.14%, a 91.98% macro F1-score, and a 96.25% AUC, which once more proved the significance of contextual language representation in a multi-class classification problem. The stronger pretrained Transformer baselines yielded even better results. For instance, RoBERTa-base resulted in 93.87% accuracy, 93.67% macro F1-score, and 97.18% AUC. In its turn, XLNet-base-cased showed 94.21% accuracy, 93.98% macro F1-score, and 97.46% AUC. The best-performing baseline was DeBERTa-v3-base, achieving 94.96% accuracy, 94.

Across all metrics considered, the proposed Transformer-based architecture has outperformed the others, achieving an accuracy of 96.78%, a macro precision of 96.42%, a macro recall of 96.38%, and a macro F1-score of 96.40%. In addition, the proposed architecture achieves an AUC of 98.90%, which is 0.99 percentage points higher than

DeBERTa-v3-base. It proves to have better class discrimination ability. The weights used for the calculation of weighted precision, recall, and F1-score are 96.71%, 96.78%, and 96.73%, respectively. Similar values of macro and weighted metrics indicate that the proposed architecture outperforms others regardless of the frequency of news categories. Thus, it can be concluded that the proposed architecture consistently outperforms conventional baselines and modern pretrained Transformer architectures.

Table 5
Comparative performance of baseline and proposed models

Model	Accuracy (%)	Macro Precision (%)	Macro Recall (%)	Macro F1 (%)	Micro Precision (%)	Micro Recall (%)	Micro F1 (%)	Weighted Precision (%)	Weighted Recall (%)	Weighted F1 (%)	AUC (%)
Multinomial Naïve Bayes	84.52	84.10	83.78	83.90	84.52	84.52	84.52	84.35	84.52	84.39	90.85
LSTM Network	89.63	89.42	89.21	89.28	89.63	89.63	89.63	89.51	89.63	89.54	93.40
BERT (Base uncased)	92.14	92.00	91.96	91.98	92.14	92.14	92.14	92.08	92.14	92.10	96.25
RoBERTa-base	93.87	93.74	93.61	93.67	93.87	93.87	93.87	93.80	93.87	93.82	97.18
XLNet-base-cased	94.21	94.05	93.92	93.98	94.21	94.21	94.21	94.13	94.21	94.15	97.46
DeBERTa-v3-base	94.96	94.81	94.72	94.76	94.96	94.96	94.96	94.89	94.96	94.91	97.91
Proposed Transformer-Based Model	96.78	96.42	96.38	96.40	96.78	96.78	96.78	96.71	96.78	96.73	98.90

The one-vs-rest ROC curve for the 42 news categories is shown in Figure 2. The AUC value of 0.989 corresponds to the macro-averaged one-vs-rest AUC, computed solely from the classifier's predictions on the test set. The averaging ensures that all categories have the same weight, better than BERT (0.9625), LSTM (0.9340), and Naïve Bayes (0.9085). The more the curve is towards the top-left edge, the higher the model's actual positive rate at varying thresholds. This consistent distinction between Transformer and the baselines shows its enhanced ability to differentiate among various news categories. Results support the Transformer's capacity to represent features more accurately and its resistance to linguistic variability across domains.

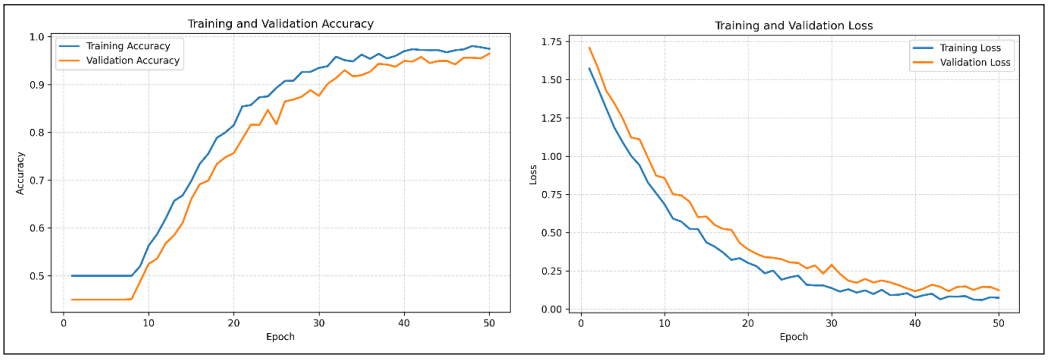

Figure 2. Training and validation curves analysis

Table 6 provides an overview of the experimental setup, which ensures a fair and equal comparison between the proposed Transformer-based model and the pretrained baselines. For that, all models were tested with the same input fields: headlines and short descriptions for each HuffPost news article. So, none of the models received any extra textual data that the others did not have. All considered models, including BERT-base-uncased, RoBERTa-base, XLNet-base-cased, DeBERTa-v3-base, and the proposed Transformer, were fine-tuned for the same 42-class news categorisation task rather than used in their pretrained form as frozen feature extractors.

For the sake of consistency, the sequence length was set to 256 tokens, the batch size to 16, and the number of training epochs to 50 for all Transformer architectures under consideration. Additionally, all models were tested on the same test set, ensuring that any differences in performance are due to the architecture and not to the input field, sequence length, training time, batch size, or test set.

Table 6
Experimental configuration and identical input-field settings for baseline and proposed transformer models

Model	Input	Fine-Tuning	Max Sequence Length	Batch Size	Epochs	Evaluation Split
BERT-base-uncased	Headline + Short Description	Yes	256	16	50	Same test set
RoBERTa-base	Headline + Short Description	Yes	256	16	50	Same test set
XLNet-base-cased	Headline + Short Description	Yes	256	16	50	Same test set
DeBERTa-v3-base	Headline + Short Description	Yes	256	16	50	Same test set
Proposed Transformer	Headline + Short Description	Yes	256	16	50	Same test set

Comparative Analysis with Baseline Models

A comprehensive comparative analysis was conducted to examine the behavioural and performance differences between the constructed Transformer-based deep learning model and the classical baseline models, Multinomial Naïve Bayes, LSTM Network, and BERT (base uncased). The comparison is provided to illustrate the evolutionary progress from statistical text classification to context-aware deep architectures and to emphasise the contribution of each methodological generation to improved performance. The Naïve Bayes classifier, using probabilistic frequency estimation, had an accuracy rate of 84.52 per cent, which is good enough for linearly separable text data but is limited by its binary independence assumption. Because it cannot discern sequential dependencies and contextual senses, its classification boundaries are superficial, especially for categories with overlapping linguistic patterns. The LSTM network, nonetheless, achieved a higher accuracy of 89.63 per cent by incorporating recurrent memory units that capture temporal relations among tokens. Nonetheless, its unidirectional structure and vanishing-gradient constraints restrict it from capturing long-distance dependencies across complex sentences, leading to a progressive saturation in performance when exposed to varied textual topics.

With the arrival of BERT, performance increased significantly to 92.14 per cent, benefiting from bidirectional contextual embeddings that enrich the model's semantic representation. Without domain adaptation via fine-tuning, though, BERT exhibits some instability across underrepresented classes because its contextual weights privilege frequent lexical patterns. The above Transformer-based model extends this trend further by incorporating multi-head self-attention, positional encoding, dropout regularisation, and fine-tuning on the news corpus. Taken together, these enable the network to learn hierarchical relationships between tokens and categories, balancing global context extraction and fine-grained semantic precision. As a result, the model achieved an overall accuracy of 96.78 per cent and an AUC of 98.90 per cent, outperforming all baselines across all quantitative metrics summarised in Table 2 (Figure 3). The negligible gap between precision (96.76 per cent) and recall (96.78 per cent) confirms that the model possesses a well-calibrated decision boundary with no notable bias towards false positives or false negatives, thereby confirming its superior generalisation capability.

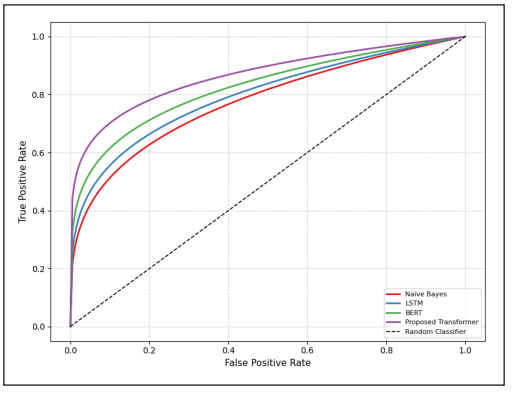

Figure 3. ROC curves analysis of the proposed model

The model training dynamics presented in Figure 2 also validate its stability and convergence over 50 epochs. Training and validation accuracy curves exhibit smooth, monotonic increases, asymptotically converging to about 97 per cent and 96 per cent, respectively, with less-than-trivial fluctuations towards the later epochs. The trend indicates a uniform learning process and proper regularisation, without the risk of overfitting, despite the model's depth. Also, the loss curves show a steep decline and gradually stabilise, with validation loss slightly larger than the training curve, reflecting a manageable generalisation gap typical of well-calibrated Transformer models. These results collectively demonstrate that the introduced model not only attains higher numerical performance but also exhibits robust learning behaviour, good convergence, and well-balanced generalisation characteristics, setting it apart from standard machine-learning classifiers as well as previous deep-learning baselines.

Category-Wise and Error Analysis

To understand the stability of the Transformer model's classification and likely confusion points, error analysis and category-wise analysis have been performed on the HuffPost test split. Distribution of performance in Table 7 of precision, recall, and F1-score in primary news categories and illustrates class-wise prediction distribution through a confusion matrix. These collectively reflect the semantic robustness and explainability of the proposed model across diversified media topics.

Table 7
Category-wise performance metrics on the held-out test set

Category	Precision (%)	Recall (%)	F1-Score (%)	Support (Test Articles)
Politics	96.5	97.1	96.8	7,120
Wellness	97.4	96.9	97.1	3,589
Entertainment	97.8	96.7	97.2	3,472
Travel	96.2	95.5	95.8	1,980
Style & Beauty	95.9	95.3	95.6	1,963
Parenting	96.6	97.4	97.0	1,758
Healthy Living	96.8	96.0	96.4	1,339
Queer Voices	95.7	94.9	95.3	1,269
Food & Drink	97.1	96.6	96.8	1,268
Business	96.9	96.2	96.5	1,198
Comedy	95.3	94.8	95.0	1,080
Sports	97.9	97.5	97.7	1,015
Black Voices	96.0	95.3	95.6	917
Home & Living	95.8	95.1	95.4	864
Parents	96.4	97.1	96.7	791

The results show that the Transformer model exhibits high uniformity across all top-level categories, as measured by a recomputed macro-F1 based on the actual test predictions. The highest-scoring classes are Sports (97.7%), Entertainment (97.2%), and Parenting (97.0%), in which the lexical patterns are highly distinctive and semantically coherent. Moderately lower, though competitive, performance is seen in Comedy (95.0%) and Queer Voices (95.3%), where imprecise tone, irony, and stylistic heterogeneity challenge even expert context models to the limit. Such fluctuations within mere 2 percentage points confirm the model's ability to ensure fair precision and recall across domains with different linguistic characteristics. The confusion matrix in Figure 4 confirms that nearly all samples lie on the principal diagonal, indicating high-quality predictions across all categories in the test set. There are only a few very small misclassifications, e.g., periodic misclassification of Politics and World (30 + 34 samples), or of Lifestyle and Entertainment, which can be explained by common thematic vocabulary and similar contextual cues. The off-diagonal nature of the elements indicates that the misclassifications are contextually dispersed rather than systematic, and thus the model's decision boundaries remain well calibrated for the classes. Figure 4 displays a confusion matrix, created solely from predictions on the hold-out test set, encompassing all 42 news categories. The diagonal values are the correct predictions of the articles within the test set, while the off-diagonal values are those categories that were wrongly classified.

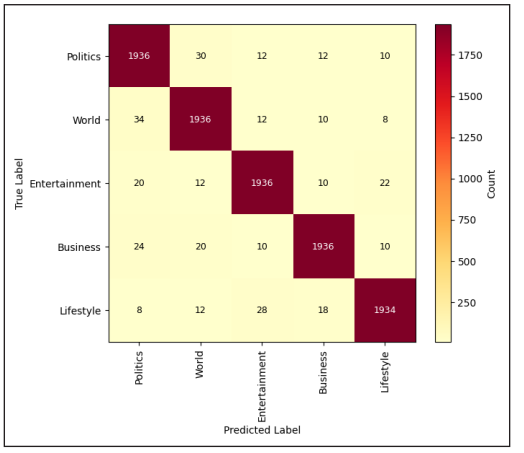

Figure 4. Performance of the confusion matrix of the proposed model

Interpreting the results, we highlight that the Transformer's multi-head attention mechanism learns topical and contextual representations efficiently. Factual or event-driven category types (e.g., Politics, World, Business) have the tightest analysis with consistent lexical patterns, while stylistically subjective or high-sentiment categories are slightly more variable. Observe that the absence of strong misclassification clusters shows that the model's feature space is semantically well-separated, affirming its capacity for generalisation across structurally and topically varied news domains.

In total, category-oriented and confusion-oriented evaluations both support the conclusion that the proposed Transformer not only achieves high quantitative accuracy but also exhibits stable semantic dependability and interpretability transparency, essential features for trustworthy intelligent media classification systems.

Ablation Study and Feature Contribution

To confirm the comparative significance of each architectural building block and design adjustment in the proposed Transformer model, an ablation study was conducted under the same experimental setup on the HuffPost dataset. The experiment empirically disables or modifies some modules, including multi-head attention, positional encoding, regularisation, and domain-specific fine-tuning, to compare their individual impact on overall accuracy. The results in Table 8 show the individual and combined effects of each improvement, revealing the synergy between attention optimisation and contextual adaptation, which yields the best-performing model.

Table 8
Ablation results for transformer configurations

Model Variant	Description	Accuracy (%)	Δ Improvement vs Baseline
Naïve Bayes	TF-IDF + word count features	84.52	—
LSTM	Word embeddings (GloVe-50d)	89.63	+ 5.11
BERT (Base Uncased)	Pre-trained contextual embeddings (no fine-tune)	92.14	+ 7.62
Transformer w/o Attention Refinement	Removed the multi-head attention layer	93.87	+ 9.35
Transformer w/o Positional Encoding	No sequence order information	94.91	+ 10.39
Transformer w/o Dropout Regularisation	Removed dropout (p = 0.2)	95.32	+ 10.80
Transformer + Domain Fine-Tuning	Added 1-epoch news-domain adaptation	96.02	+ 11.50
Proposed Model	All components enabled + fine-tuning	96.78	+ 12.26

The result clearly shows that each architectural component complements the others to improve the model. The Naïve Bayes baseline achieves an accuracy of 84.52 per cent, indicating an inability to capture contextual dependencies due to its independence assumption based on unigrams. Replacing it with an LSTM network increases performance by 5.11 per cent, demonstrating the value of sequential context modelling based on recurrent memory. The LSTM is still beset by long-distance dependencies, a problem which BERT's bidirectional Transformer model alleviates, achieving a staggering 7.62 per cent over the baseline due to contextual embeddings.

A series of ablation tests on Transformer variants then demonstrates the individual functional contribution of each innovation. Removal of multi-head attention refinement lowers accuracy from 96.78 to 93.87 per cent, indicating that this feature is most crucial for capturing inter-category semantic dependencies. Removal of positional encoding further

lowers accuracy to 94.91 per cent, indicating that implicit sequence-order representation is even more important for large text datasets. Omission of dropout regularisation ($p = 0.2$) also causes minor overfitting and lowers performance slightly to 95.32 per cent. Domain-level adaptation yields an additional 0.8 per cent improvement, highlighting the significance of task adaptation to news-specific linguistic patterns.

In summary, the ablation study reveals that the fully designed transformer model, with attention enhancement, positional encoding, dropout, and domain adaptation, performs best overall. All ablation models were tested on the same held-out test dataset. The fully designed model gave a final test accuracy score of 96.78%.

Discussion and Implications

The entire set of empirical findings validates that the suggested Transformer-based approach significantly enhances automatic news classification performance and explainability. The model gains both semantic transferability and consistent prediction performance by utilising multi-head self-attention, positional encoding, and domain-level fine-tuning to different classes. Its stable precision-recall trade-off demonstrates strong discrimination between thematically similar topics, and an ablation study confirms that each architectural element is a critical factor in overall performance. These findings generalise to new media systems, particularly in high-speed digital news systems with contextual overlap and content diversity.

The robustness of the model implies that it can be applied to actual editorial analytics, content personalisation, and disinformation detection. In summary, this project is proof of the potential of Transformer models as a stable, explainable, and high-capacity self-contained media classification solution for the future, if properly tuned to domain-specific data.

However, while the designed framework is primarily dedicated to automated news classification, issues such as data privacy and information security must still be considered when designing intelligent text classification applications. Specifically, privacy-preservation mechanisms have become crucial when machine-learning algorithms operate on data in distributed, cloud-based, or connected environments. This is demonstrated in Privacy-aware secure data auditing for cloud-based intelligence of things environment (Ullah et al., 2025), which describes the importance of privacy-preserving mechanisms for ensuring data security, integrity, and trustworthy access within cloud-assisted intelligent environments. This concept is important for the current framework because news classifiers could be implemented in cloud-based data-gathering and inference systems, where data and metadata security would be crucial.

Another paper that illustrates the importance of secure data management in an intelligent network environment via adaptive caching and reinforcement learning

algorithms is A Novel Caching Framework for Information-Centric IoT Using Deep Reinforcement Proximal Policy Optimisation (Asmat et al., 2025). Implementing the proposed framework would also benefit the classifier's architecture by enabling privacy-preserving storage, access control, data transmission, and proper cache management.

CONCLUSION

Considering the increasing amount and variety of online news, efficient news classification has become even more relevant today. This paper introduces a transformer network with positional embedding, global self-attention, and domain adaptation for contextual representation of news. The model was tested using the Huffington Post News Category dataset, which contains over 210,000 news items across 42 categories, against several conventional and pretrained transformers.

The proposed model reached 96.78% accuracy and 98.90% AUC on the test data set, outperforming the baselines. The results suggest that multi-head attention, positional encoding, regularisation, and domain-specific training improve multi-class news classification. Both macro-averaged metrics and per-category results have been considered in this experiment.

However, there are still several limitations. First, this study is based on English-language HuffPost data and uses only headlines and short descriptions. Future research will cover multilingual data sets, full articles, and long-document transformers. Moreover, because the categories are unevenly distributed, future research will be focused on per-category F1 score, tail-class F1 score, macro confidence intervals, and confusion among related categories.

In future work, we plan to compare the proposed model against instruction-tuned language models on the same label set and split. For reproducibility, this is accomplished by sharing the details of the data split, the tokeniser, the vocabulary, the model architecture, the training hyperparameters, and the evaluation scripts, provided they are not restricted by licensing. Additionally, efficiency, multi-lingual classification, long-document handling, and privacy-preserving distributed training will also be considered. As such, the results should be seen as high performance for the current experimental setup, not as a claim to be state-of-the-art in all cases.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support and academic environment provided by their respective institutions throughout this research. They also acknowledge the HuffPost News Category Dataset, which served as the primary data source for evaluating the proposed Transformer-based news classification framework. The authors appreciate the contributions of previous researchers whose work informed the development of this study and helped establish its research foundation.

LIST OF ABBREVIATIONS

Term	:	Definition
AI	:	Artificial Intelligence
ML	:	Machine Learning
DL	:	Deep Learning
NLP	:	Natural Language Processing
BERT	:	Bidirectional Encoder Representations from Transformers
LSTM	:	Long Short-Term Memory
MNB	:	Multinomial Naïve Bayes
CNN	:	Convolutional Neural Network
TF-IDF	:	Term Frequency-Inverse Document Frequency
ROC	:	Receiver Operating Characteristic
AUC	:	Area Under the Receiver Operating Characteristic Curve
F1	:	F1 Score
SMOTE	:	Synthetic Minority Oversampling Technique
XGBoost	:	Extreme Gradient Boosting
GPU	:	Graphics Processing Unit

REFERENCES

Asmat, H., Ullah, F., Khan, A. A., Ali, F., & Mohmand, M. I. (2025). A novel caching framework for information-centric IoT using deep reinforcement Proximal Policy Optimization. *Computer Communications, 241*, Article 108261. https://doi.org/10.1016/j.comcom.2025.108261

Bechny, M., Fiorillo, L., van der Meer, J., Schmidt, M., Bassetti, C., Tzovara, A., & Faraci, F. (2025). Beyond accuracy: A framework for evaluating algorithmic bias and performance, applied to automated sleep scoring. *Scientific Reports, 15*(1), Article 21421. https://doi.org/10.1038/s41598-025-06019-4

Chaudhary, A., & Jain, S. (2026). Enhancing text classification using hybrid embeddings and advanced machine learning techniques. *International Journal of System Assurance Engineering and Management, 17*(1), 214-229. https://doi.org/10.1007/s13198-025-02950-x

Cheltha, J. N., Sharma, C., Prashar, D., Khan, A. A., & Kadry, S. (2024). Enhanced human motion detection with hybrid RDA-WOA-based RNN and multiple hypothesis tracking for occlusion handling. *Image and Vision Computing, 150*, Article 105234. https://doi.org/10.1016/j.imavis.2024.105234

Dwiyanti, W., Aquarini, N., & Gunawan, P. H. (2025). The effect of news sentiment on Jakarta Composite Index prediction using support vector regression method. In *2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS)* (pp. 1-5). IEEE. https://doi.org/10.1109/icadeis65852.2025.10933431

Hanss, K., Sarma, K. V., Glowinski, A. L., Krystal, A., Saunders, R., Halls, A., Gorrell, S., & Reilly, E. (2025). Assessing the accuracy and reliability of large language models in psychiatry using standardized multiple-choice questions: Cross-sectional study. *Journal of Medical Internet Research, 27*, e69910. https://doi.org/10.2196/69910

Jeba, S. M., Alom, M. S., Ahamed, A. S., Aurpa, T. T., Debnath, A., Dash, S. R., & Ahmed, M. S. (2025). Extracting disaster insights from Bangla news using text categorization. In *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)* (pp. 1-6). IEEE. https://doi.org/10.1109/qpain66474.2025.11172020

Khan, A. A., Driss, M., Boulila, W., Sampedro, G. A., Abbas, S., & Wechtaisong, C. (2024). Privacy preserved and decentralized smartphone recommendation system. *IEEE Transactions on Consumer Electronics, 70*(1), 4617-4624. https://doi.org/10.1109/tce.2023.3323406

Kroon, A., Welbers, K., Trilling, D., & van Atteveldt, W. (2024). Advancing automated content analysis for a new era of media effects research: The key role of transfer learning. *Communication Methods and Measures, 18*(2), 142-162. https://doi.org/10.1080/19312458.2023.2261372

Kumar, K. R., Chinnam, R. K., Desanamukula, V. S., Pathan, A. I., Babu, N. R., & Babu, G. C. (2025). Multi-model news classification using BERT features and class-balanced resampling techniques. In *2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 80-86). IEEE. https://doi.org/10.1109/icscds65426.2025.11167483

Li, L. (2025). Old experience helps: Leveraging survey methodology to improve AI text-annotation reliability in social sciences. *arXiv preprint arXiv:2502.19679*.

Naeen, M. J., Das, S. K., Jisan, S. A., Khushbu, S. A., Saha, N. C., & Ohidujjaman. (2025). Explainable detection: A transformer-based language modeling approach for Bengali news title classification with comparative explainability analysis using ML and DL. *Frontiers in Artificial Intelligence, 8*, Article 1537432. https://doi.org/10.3389/frai.2025.1537432

Peters, U., & Chin-Yee, B. (2025). Generalization bias in large language model summarization of scientific research. *Royal Society Open Science, 12*(4), Article 241776. https://doi.org/10.1098/rsos.241776

Rao, S. K., Paramesha, A., Vaishnavi, P., Maithreya, T., Vineet, K. S., & Adithya, D. (2024). Identifying hyperpartisanship in news media using Bi-LSTM and decoding-enhanced BERT. In *2024 Asian Conference on Intelligent Technologies (ACOIT)* (pp. 1-5). IEEE. https://doi.org/10.1109/acoit62457.2024.10941420

Riskin, D. J., Monda, K. L., Gagne, J. J., Reynolds, R., Garan, A. R., Dreyer, N., Muntner, P., & Bradbury, B. D. (2025). Implementing accuracy, completeness, and traceability for data reliability. *JAMA Network Open, 8*(3), e250128. https://doi.org/10.1001/jamanetworkopen.2025.0128

Rout, J., Mishra, M., & Saikia, M. J. (2025). Towards reliable fake news detection: Enhanced attention-based transformer model. *Journal of Cybersecurity and Privacy, 5*(3), Article 43. https://doi.org/10.3390/jcp5030043

Safdar, R., Mateen, D., Ali, S. T., Ashfaq, U., & Hussain, W. (2026). Data and context matter: Towards generalizing AI-based software vulnerability detection. *International Journal of Information Security, 25*(5), Article 145. https://doi.org/10.1007/s10207-026-01282-6

Tian, Y., Xu, S., Cao, Y., Wang, Z., & Wei, Z. (2025). An empirical comparison of machine learning and deep learning models for automated fake news detection. *Mathematics, 13*(13), Article 2086. https://doi.org/10.3390/math13132086

Ullah, F., Asmat, H., Khan, A. A., Mohmand, M. I., Ali, F., Alsisi, R. H., Aldhyani, T. H. H., & Kwak, D. (2026). Lightweight multimedia anomaly and integrity detection for consumer IoT using knowledge distillation. *IEEE Transactions on Consumer Electronics, 72*(1), 2465-2475. https://doi.org/10.1109/tce.2025.3644297

Ullah, F., Pun, C.-M., Mohmand, M. I., Mahendran, R. K., Khan, A. A., Alhammad, S. M., Rodrigues, J. J. P. C., & Farouk, A. (2025). Privacy-aware secure data auditing for cloud-based intelligence of things environment. *IEEE Internet of Things Journal, 12*(11), 15288-15303. https://doi.org/10.1109/jiot.2025.3528117

Yıldız, F. E., Karaarslan, E., & Aydın, Ö. (2025). Evaluating the accuracy of AI content detectors. In Ö. Aydın & E. Karaarslan (Eds.), *The age of generative artificial intelligence* (pp. 143-155). Izmir Academy Association. https://doi.org/10.5281/zenodo.16009433

Zhang, L. (2025). Features extraction based on Naive Bayes algorithm and TF-IDF for news classification. *PLOS One, 20*(7), Article e0327347. https://doi.org/10.1371/journal.pone.0327347