Journal of Theoretical and Applied Information Technology
15st June 2026. Vol.104. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DISENTANGLE ATTENTION AND MASKED CUSTOM LARGE LANGUAGE MODEL BASED FAKE NEWS DETECTION

[1*]K. SARITHA DEVI., [2]DR. M. CHIDAMBARAM

[1*]Research Scholar, Rajah Serfoji Government College (Auto), Thanjavur – 613 005, Tamil Nadu, India.
(Affiliated to Bharathidasan University, Tiruchirappalli – 620 024)
[2]Research Advisor, Associate Professor & Head, Department of Computer Science, Rajah Serfoji Government College (Auto), Thanjavur – 613 005, Tamil Nadu, India.
(Affiliated to Bharathidasan University, Tiruchirappalli – 620 024)
[1*]Corresponding Author Email id: pksdevi81@gmail.com

ABSTRACT

False information presented as reliable news is known as fake news, and it's dissemination undermines democracy and public confidence. Large Language Models (LLMs) have two roles in this field: they can both propagate and identify false information. Although BERT is a popular tool for detecting fake news, many other approaches have trouble with intricate linguistic patterns and don't provide clear justifications for their conclusions. We suggest a BERT-based model, DAE-MCD-LLM, to enhance this. The method consists of three steps: news classification, key term extraction, and data cleaning. The decoder adds more layers to the enhance classification after the encoder captures word meaning and position. When tested on the WELFAKE dataset, the model outperformed two alternative methods in accuracy, precision, F1 score, and speed, demonstrating its efficacy and reliability in detecting fake news.

Keywords: *Fake News Detection, Large Language Model, Disentangle Attention-based Encoder, Masked Custom Decoder, WELFAKE dataset*

1. INTRODUCTION

People get news instantly online, but fake news spreads just as fast, aiming to mislead, provoke emotions, or influence opinions. Researchers use deep learning, hybrid models like GPT+BERT, and analysis of news spread to detect it, but many models are “black boxes” and hard to interpret.

Efforts to improve transparency include explainable models and human expert input, while deep learning remains dominant. Challenges persist in accuracy, speed, and handling new content. The proposed DAE-MCD-LLM model aims to detect fake news more accurately, quickly, and transparently.

1.1. Contributions of the work

The primary contributions of this study are summarized as follows:

- A new system for spotting fake news was created called DAE-MCD-LLM. It brings together BERT, which helps the computer understand what words really mean in a sentence, and powerful language models that are good at generating and understanding text. With both tools working together, the model can explain clearly why it decides that a news story is real or fake, making its answers more transparent and trustworthy for people who use it.
- The Disentangle Attention Encoder carefully looks at both text and images on their own, helping the system find what matters most in each part. It keeps track of where these key details appear in the news. After that, the Masked Custom Decoder, built on BERT, uses all of this information to make smarter and more trustworthy predictions. The training method that uses cross-entropy loss helps the model get better at the task without wasting time or computer power.
- We tested our DAE-MCD-LLM model using the WELFake dataset, a common benchmark. We compared it with other

popular methods and found that our model is more accurate, gives better results, and works faster. This makes it useful for quickly finding fake news as it appears online.

1.2. Work organization

The rest of the paper is structured in this manner. In Section 2, previous studies on the detection of fake news using NLP, BERT, and LLMs are reviewed. In Section 3, the dataset and our suggested model are explained, called the Disentangle Attention Encoder and Masked Custom Decoder (DAE-MCD-LLM), along with its step-by-step pseudo-code. Section 4 discusses how the experiments were set up and what evaluation measures were used. Section 5 gives the results from our tests and shows how well our model performs compared to other advanced methods. It presents clear comparisons, making it easy to see where our system stands out. Section 6 briefly sums up the most important findings from this work and suggests possible areas where future research could make further progress.

2. Related Works

The rapid growth of the internet and social media enables quick sharing of information but also spreads fake news, causing confusion and mistrust [16][17]. Early detection systems use user comments with content via teacher–student models for social media [18], though these are less reliable for formal news. Some methods handle limited data by using minimal text [19] or generating extra training data with LLMs [20], often focusing on surface features. Advanced frameworks combine content, user behavior, and social network patterns [21][22]. Transformers perform well in language tasks, and combining multiple transformers improves performance [23]. Multilingual detection uses sentiment analysis, named entity recognition, and capsule networks [24]. Hybrid models like BERT with Bi-LSTM, Bi-GRU, or CNN layers enhance contextual understanding and accuracy [25][26][28][29][30]. Fuzzy logic–based systems handle ambiguous information better than simple classifiers. These approaches make automated fake news detection increasingly reliable as online news grows [27].

3. Materials and methodology

Online information spreads fast, making it hard to spot fake news. BERT-based LLMs help by understanding context and meaning. This study introduces **DAE-MCD-LLM**, a BERT-based model that focuses on key words and their relationships for more accurate and reliable fake news detection.

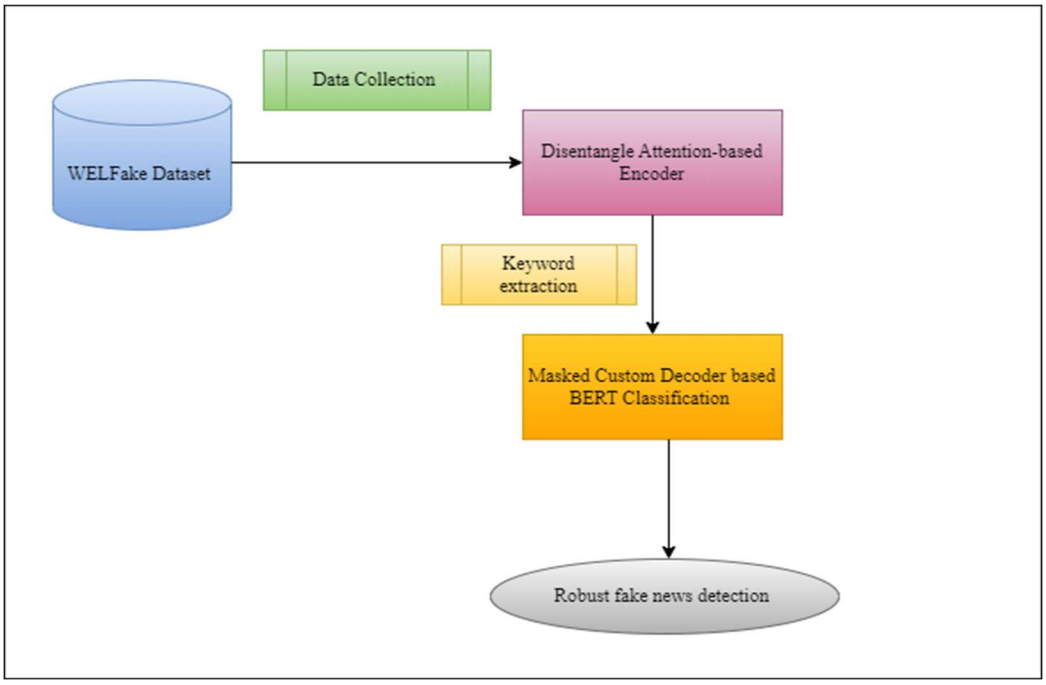


Fig. 1 Architecture For DAE-MCD-LLM Fake News Detection

Figure 1 show the Architecture for DAE-MCD-LLM fake news detection. The DAE-MCD-LLM framework is made up of two key components: the Masked Custom Decoder-based BERT Classification module and the Disentangle Attention Encoder. Starting with data from the WELFake dataset, the system first sends the news text through the Disentangle Attention Encoder, which finds and selects the most important words and context clues. These features are sent to the Masked Custom Decoder–based BERT module, which makes use of them to ascertain whether news is genuine or fake. This two-step architecture improves the model's accuracy and dependability while also strengthening its capacity to recognize bogus input.

3.1. Data collection

The WELFake Fake News Classification dataset, accessible on Kaggle, is used in this work. To dataset contains 72,134 news articles, with 37,106 listed as fake and 35,028 listed as real. It was created by combining information from four reliable sources: BuzzFeed Political, McIntire, Reuters, and Kaggle which contributes to the dataset's increased diversity and balance. A serial number, headline, article text, and label (0 for fake, 1 for true) are all included in each dataset item. Although the original dataset had 78,098 records, only 72,134 were used after cleaning and verification in this study to ensure quality.

3.2. Data Pre-processing

Preprocessing is crucial for fake news detection using NLP. Raw articles often contain links, symbols, and punctuation that do not help the model, so these are removed. The text is converted to lowercase for consistency, and common words like “and” or “the” are discarded. Words are then split, reduced to their base forms (e.g., “running” → “run”), and finally converted into numbers for the model. These steps clean the data and make it easier for the system to distinguish real from fake news.

3.3. BERT

BERT is a pre-trained language model that reads text in both directions, unlike traditional models that read left-to-right or right-to-left. This bidirectional approach helps it better understand word meanings in context. Before processing, sentences are split into smaller units—words, subwords, or characters—to handle rare or new words. BERT also adds special tokens: [CLS] at the start to represent the whole sentence and [SEP] to separate sections. These features allow BERT to create rich text representations, making it highly effective for tasks like fake news detection.

3.3.1. Disentangle Transformer Attention-based Encoder for relevant keyword extraction

Our model uses a Disentangle Transformer Attention-based Encoder to find important details in the input text. This encoder is built from several sets of the same blocks called transformer blocks. Every transformer block consists of two primary components:

The first part is the multi-head self-attention layer.This layer helps the model look at several words in the text at the same time and understand how they relate to each other. It finds which words are important in the sentence and how they connect.

The second part is the feed-forward network. This part takes what the attention layer has learned and improves the information even more, making sure that the key words and context are easy for the next step of the model to use.

The encoder generates clear, detailed characteristics that are prepared for the remainder of the false news detection procedure by running the tokenized text (with [CLS] at the beginning and [SEP] at the end) through these components. The tokenized text input, represented as $S = [CLS], S_1, S_2, \dots, S_m, [SEP]$ includes two special tokens, [CLS] and [SEP], which mark the beginning and end of the sequence, respectively.

First describe the structure of a conventional Transformer encoder before presenting the design of the proposed Disentangle Transformer Attention-based Encoder, which is tailored for effective keyword extraction. The standard Transformer encoder processes the input tokens to generate hidden representations H for each element in the sequence, as formalized below:

$$H = Encode\ (S), H \in \mathbb{R}^{n*d} \quad (1)$$

$$H_l = TransformerBlock\ (H_{l-1}) \quad (2)$$

Using the hidden representations obtained from Equations (1) and (2), The training goal of the model is to minimize the cross-entropy loss while computing attention weights. The sequence length (n) and the number of hidden units (d) are represented in these equations. To conventional attention mechanism transforms the input representations by measuring vector similarity, which is computed using inner product operations,

as described below:

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{Dim_k}}\right)V \quad (3)$$

From the above equation (3), to be more specific, the attention function in this context ascertains 'K' that match query 'Q' and accordingly extracts the subsequent value 'V'. In this manner, the attention function entails to assign weights closer to '1' for words to be identified as important, therefore extracting keywords that specify a sentence or samples. Finally, with the inputs as query 'Q', key 'K' value 'V' and dimensions 'Dim_k' of 'k' according to (3) involves a procedure of measuring the similarity between query 'Q', key 'K' value 'V' through the inner product. Similarly, the transformer encoder improves this procedure by means of multi-head attention mechanism. As seen below, this multi-head attention mechanism requires the parallelization of several scaled dot-product attention samples. Moreover, the basic operation of '$TransformerBlock$' is to evaluate the Multiple-head self-attention as given below.

$$Multi-Head\ (Q,K,V) = Concat(head_1, head_2, \dots, head_k)^{W^O} \quad (4)$$

$$head_k = Attention\left(QW_k^Q, KW_k^K, VW_k^V\right) \quad (5)$$

From Equations (4) and (5), (Q), (K), and (V) represent the query, key, and value matrices, respectively. The matrices (W_k^Q), (W_k^K), and (W_k^V) denote the corresponding projection weight matrices applied to (Q), (K), and (V). In addition, (W^O) represents a learnable output weight matrix used to map the concatenated attention outputs to the model's dimensional space.

If you want it more concise or more technical, say the word—I'll tune it. With the above conventional model of transformer encoder as base, in this work, Disentangle Attention-based Encoder for relevant keyword extraction is designed in the following sub-sections.

3.3.2. Disentangle Attention-based Encoder for relevant keyword extraction

The disentangled attention-based encoder employed in our work extracts keywords concentrating on separating the impact of content (i.e. keyword) and positional information on word relationships, permitting the method to better understand the context and extract relevant keywords. For a token or sample at position 'i' in a sequence, the sample or the token here is denoted using two vectors '$\{S_i\}$' and '$\{Pos_{i|j}\}$' denoting its keyword and relative position. The Disentangle Attention-based Encoder's structure for extracting pertinent keywords is seen in Figure 2.

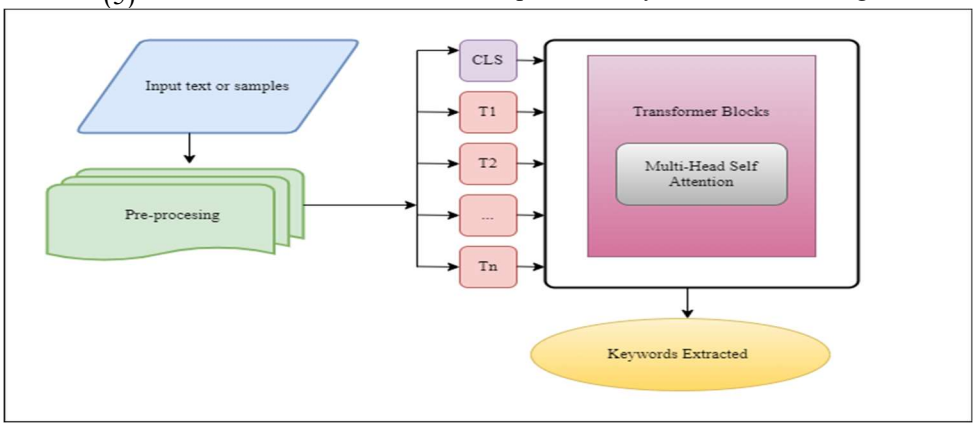


Fig 2: Structure Of Disentangle Attention-Based Encoder For Relevant Keyword Extraction

As shown in the above figure, the Disentangle Attention-based Encoder for relevant keyword extraction process first, obtains the pre-processed text about the news content in the form of tokens (i.e. T1, T2, Tn). The tokens are subjected to transformer block where multi-head self-attention employing disentangle function is applied to extract keywords extensively. The evaluation of cross attention score between tokens or sample 'i' and 'j' is split into four elements as given below.

$$A_{ij} = \{S_i, Pos_{i|j}\} * \{S_j, Pos_{j|i}\}^T = S_i S_j^T + S_i Pos_{j|i}^T + Pos_{i|j} S_j^T + Pos_{i|j} Pos_{j|i}^T \quad (6)$$

From the above formulate (6) the attention score 'A_{ij}' is measured as a sum of four attention scores, i.e., sample to sample '$S_i S_j^T$', sample to position '$S_i Pos_{j|i}^T$', position to sample

'$Pos_{i|j}S_j^T$' and position to position '$Pos_{i|j}Pos_{j|i}^T$' respectively. Also, from the above equation as the attention weight of a word pair depends not only on the keywords (i.e. obtained from samples) but also on their relative positions. Hence the position to position in the fourth element is eliminated for further processing as it has only limited scope. Finally, Disentangle Attention-based Encoder for relevant keyword extraction is mathematically formulated as given below.

$$KE_{ij} = Q_i^S K_j^{S^T} + Q_i^S Pos_{i|j}^T + Pos_j^S Q_{j|i}^T \quad (7)$$

From the above equation (7) the relevant keywords are extracted 'KE_{ij}' from tokens or sample 'i' to tokens or sample 'j' based on the sample keyword to sample keyword '$Q_i^S K_j^{S^T}$', sample keyword to position '$Q_i^S Pos_{i|j}^T$' and position to sample keyword '$Pos_j^S Q_{j|i}^T$' respectively.

3.3.3. Masked Custom Decoder based BERT Classification for fake news detection

After the multi-head self-attention layer, the output passes through a feed-forward network (FFN), which applies non-linear transformations to refine the attention results. During training, the model adjusts its weights to reduce errors, gradually improving its ability to detect fake news. In the final step, the Masked Custom Decoder–based BERT module combines the encoder's keyword selection with precise masked decoding. This design enhances the system's accuracy and reliability in identifying fake news.

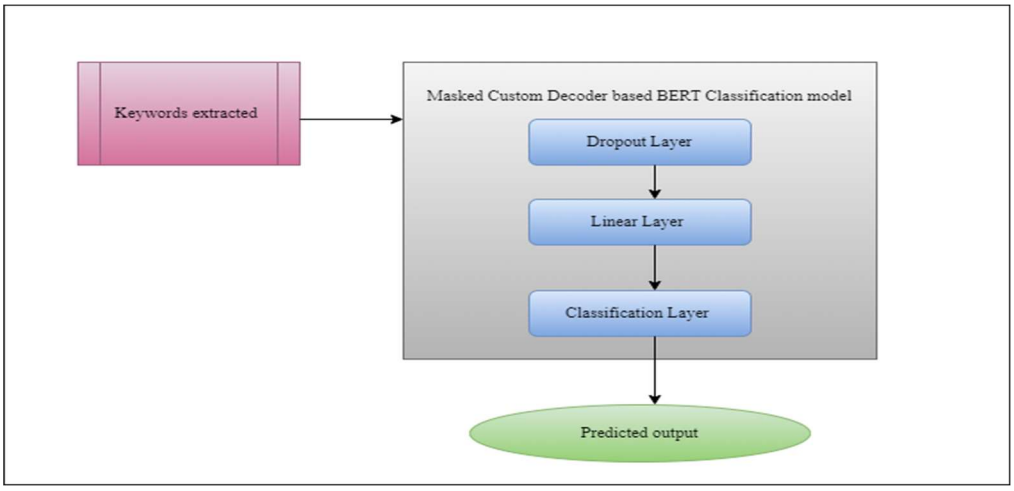


Fig 3: Structure Of Masked Custom Decoder Based BERT Classification Model

As shown in Figure 3, the Masked Custom Decoder–based BERT classification model has three main layers: a dropout layer, a linear layer, and a classification layer. The encoder has already identified key words and their positions in the sentence. Word position is important for understanding meaning. For example, in the sentence "A new warehouse opened beside the new shopping center," if 'warehouse' and 'shopping center' are masked, their roles differ: 'warehouse' is the subject, while 'shopping center' is not. Including position information helps the model correctly interpret such differences.

In our approach, position embeddings are added just before the softmax step, letting the model use both the meaning and position of each word to predict masked tokens. The process happens step by step in the decoder:

- The dropout layer helps stop the model from overfitting, making training more stable.
- The linear layer shrinks the feature size, making the model learn faster and use less memory.
- The classification layer generates the final raw scores for each class to evaluate whether a news report is genuine or fraudulent.

The operation of the dropout function, used within the dropout layer to enhance generalization and prevent overfitting, can be formally expressed as follows:

$$H_{DoL} = \sum_i \sum_j Do(KE_{ij}) \quad (8)$$

From the above equation (8) by applying the dropout function 'Do' aids in learning more robust features not dependent on any single neuron, hence acts as a form of regularization, making the fake news detection process more robust and less prone to overfitting. Following which a linear transformation is performed as given below.

$$H_{LL} = ReLU(W_{LL}H_{DoL} + B_{LL}) \quad (9)$$

From the above equation (9) the linear layer acquires an input vector 'H_{DoL}' and multiplies it by a linear weight matrix 'W_{LL}' and finally adds a linear bias vector 'B_{LL}' resulting in an output matrix 'H_{LL}'. Finally, the transformed linear representation is passed through a classification layer as given below.

$$Logits = W_C H_{LL} + B_C \quad (10)$$

From the above equation (10) the classification layer acquires an input vector 'H_{LL}' and multiplies it by a class weight matrix 'W_C' and finally adds a class bias vector 'B_C' resulting in classified output modeling fake news detection. Finally, to fine-tune the weight, in our work cross entropy loss function is applied so that error can be measured in case. The cross-entropy loss function is expressed mathematically as follows.

$$L = -\sum_i \sum_C Res_{i,C} \log\left(Prob\left(Res_{i,C}|Enc_i[CLS]\right)\right) \quad (11)$$

From the above equation (11) the cross entropy loss function 'L' is generated based on '$Res_{i,C}$' the true label performed over each sample 'i' for class 'C' and predicted probability '$Prob\left(Res_{i,C}|Enc_i[CLS]\right)$' performed over each sample 'i' for class 'C' given with its encoded representation '$Enc_i[CLS]$'. Also, the above cross entropy loss function is evaluated via double summation over each sample 'i' for class 'C' with differentiation between actual and predicted using the logarithmic function 'log'. The pseudo code representation of BERT-based LLM called Disentangle Attention Encoder and Masked Custom Decoder is given below.

Input: Dataset 'DS', Samples '$S = \{S_1, S_2, \dots, S_m\}$', Features '$F = \{F_1, F_2, \dots, F_n\}$'
Output: Robust fake news detection
Step 1: **Initialize** '$m = 20000$', '$n = 4$'
Step 2: **Begin**
Step 3: **For** each Dataset 'DS' with Samples 'S' and Features 'F'
//**Pre-processing**
Step 4: Filter non-alphabet ULR characters
Step 5: Discard numbers and punctuation in the text
Step 6: Remote stop words
Step 7: Convert upper case letters to lower case
//**Transformer encoder**
Step 8: For each Dataset 'DS' with Samples 'S' (i.e. token) and Features 'F'
Step 9: Process token sequence input to generate hidden representations according to (1) and (2)
//**Transformer blocks**
//**Multi-head self-attention keyword extraction**
Step 10: Evaluate attention mechanism by means of inner product measurements according to (3)
Step 11: Evaluate multi-head self-attention according to (4) and (5)
Step 12: Evaluate cross attention score between tokens and split into four elements according to (6)
//**Remove fourth element**
Step 13: Return keyword extracted results according to (7)
//**Feed-forward (custom) network layer**
//**Dropout layer**
Step 14: Perform dropout function according to (8)
//**Linear layer**
Step 15: Perform linear transformation according to (9)
//**Classification layer**
Step 16: Generate classified output according to (10)
Step 17: **If** '$Logits = 0$'
Step 18: **Then** fake news detected
Step 19: **End if**
Step 20: **If** '$logits = 1$'

Step 21: **Then** real news detected
Step 22: **End if**
//Weight fine-tuning
Step 23: Fine tune the weights by measuring cross entropy loss function
Step 24: **End for**
Step 25: **End**

Algorithm 1 BERT-Based LLM Called Disentangle Attention Encoder And Masked Custom Decoder For Identifying False Information

The fake news detection framework has four steps: collect articles from Kaggle, McIntire, Reuters, and BuzzFeed Political; clean and tokenize the text; extract key words with the Disentangle Attention Encoder; and classify using the Masked Custom Decoder–based BERT. Dropout, linear layers, and masking improve precision, enhancing overall performance and F1-score.

4. Experimental Setup

This section describes how the proposed DAE-MCD-LLM model was tested and evaluated. Built on a BERT-based architecture, it was trained and assessed in Python on a PC with an Intel i7 processor and 32 GB RAM. Performance was measured using precision, recall, accuracy, F1-score, and execution time. All models were tested on the same news headlines and articles for fair comparison. DAE-MCD-LLM was compared with three other models: a standard transformer LLM, a multilingual deep learning model, and a method using social context and news features. The results showed notable improvements in accuracy and reliability, demonstrating that the proposed model is effective and dependable for detecting fake news.

5. Implementation and Discussion

In this study, propose a reliable BERT-based LLM-based fake news detection method., termed the Disentangle Attention Encoder and Masked Custom Decoder (DAE-MCD-LLM). The framework reduces training time while achieving improved precision, recall, and accuracy. The three primary stages of pre-processing, keyword extraction, and classification the DAE-MCD-LLM architecture.

- The performance of DAE-MCD-LLM is compared against three baseline approaches Transformer-based LLM [1], Multilingual Deep Learning Model [2], and Fake News Detection via News Content and Social Context (FND-NS) [3]—using same dataset to ensure a fair and consistent evaluation.
- Initially, raw input samples are collected and subjected to pre-processing, this involves tokenization and data cleaning to get the text ready for additional analysis.
- The Disentangle Attention-based Encoder is then applied to the processed dataset to extract contextually relevant keywords. This step allows the model to capture complex contextual interactions by concurrently taking positional and content information into account. By computing cross-attention scores between tokens, the encoder performs precise keyword extraction. This disentangled attention mechanism also contributes to reduced training time and improved precision and accuracy in fake news classification.
- Next, the Masked Custom Decoder-based BERT Classification module classifies the extracted keywords. To improve the network, this step incorporates three special layers: dropout, linear, and classification. Higher F1-scores are obtained by the model using a cross-entropy loss function, indicating its efficacy in robust fake news identification.
- Finally, a Deep Neural Kernel Perceptron-based classifier is applied to the extracted features to further enhance precision and recall, yielding significant improvements in overall detection performance.

Following these implementation steps, five distinct evaluation metrics are used in the next subsection to comprehensively assess the model's performance.

5.1. Performance analysis of precision, recall and accuracy

When it comes to identifying fake news, precision measures the proportion of news articles that are accurately classified as fake. The primary focus of precision is the accuracy of positive predictions, or the model's ability to correctly identify a news article as fraudulent. When the system determines that a news story is fraudulent, a high precision

rating means that the forecast is very accurate, reducing false positives. The mathematical formulation of precision is expressed as follows:

$$Pre = \frac{TP}{TP+FP} \quad (12)$$

From Equation (12), precision (Pre) is calculated using the TP and FP rates. In terms of fake news detection, recall assesses the model's ability to correctly identify each relevant instance of fake news in the dataset. Together with precision, recall serves as a key performance metric, providing insight into detecting system's overall efficacy. The mathematical formulation of recall is given below:

$$Rec = \frac{TP}{TP+FN} \quad (13)$$

From Equation (13), recall (Rec) is calculated using the TP and FN rates. Accuracy, evaluates the overall accuracy of the model's predictions by taking into account both real and fake news instances. Although a high accuracy typically indicates that the model can differentiate between real and fake news, it does not directly reflect the model's effectiveness in recognizing fake news on its own. To mathematical expression for accuracy is given below:

$$Acc = \frac{TP+TN}{TP+F\ +TN+FN} \quad (14)$$

From Equation (14), accuracy (Acc) is calculated using the TP, FP, TN, and false negative (FN) rates. A comparison of accuracy, precision, and recall using four distinct methods is shown in Table 1: DAE-MCD-LLM, Transformer-based LLM [1], Multilingual Deep Learning Model [2], and Fake News Detection using News Content and Social Context (FND-NS) [3].

Table 1: Tabulation of precision, recall and accuracy

Samples 2000	Precision				Recall				Accuracy			
	DAE-MCD-LLM	Transformer-based LLM [1]	multilingual deep learning [2]	FND-NS [3]	DAE-MCD-LLM	Transformer-based LLM [1]	multilingual deep learning [2]	FND-NS [3]	DAE-MCD-LLM	Transformer-based LLM [1]	multilingual deep learning [2]	FND-NS [3]
2000	0.99	0.98	0.96	0.97	0.99	0.99	0.99	0.99	0.99	0.98	0.96	0.97
4000	0.98	0.93	0.9	0.88	0.97	0.92	0.87	0.84	0.98	0.88	0.83	0.8
6000	0.97	0.92	0.89	0.87	0.95	0.9	0.85	0.82	0.95	0.85	0.8	0.77
8000	0.96	0.91	0.88	0.86	0.94	0.89	0.84	0.81	0.92	0.82	0.77	0.74
10000	0.95	0.9	0.87	0.85	0.92	0.87	0.82	0.79	0.91	0.81	0.76	0.73
12000	0.96	0.91	0.88	0.86	0.95	0.9	0.85	0.82	0.9	0.8	0.75	0.72
14000	0.97	0.92	0.89	0.87	0.97	0.92	0.87	0.84	0.93	0.83	0.78	0.75
16000	0.98	0.93	0.9	0.8	0.93	0.88	0.83	0.8	0.95	0.85	0.8	0.77
18000	0.97	0.92	0.89	0.87	0.95	0.9	0.85	0.82	0.97	0.87	0.82	0.79
20000	0.98	0.93	0.9	0.8	0.98	0.93	0.88	0.85	0.98	0.88	0.83	0.8

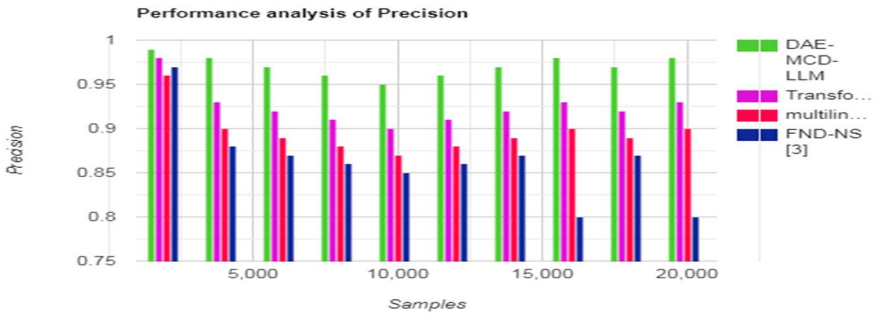
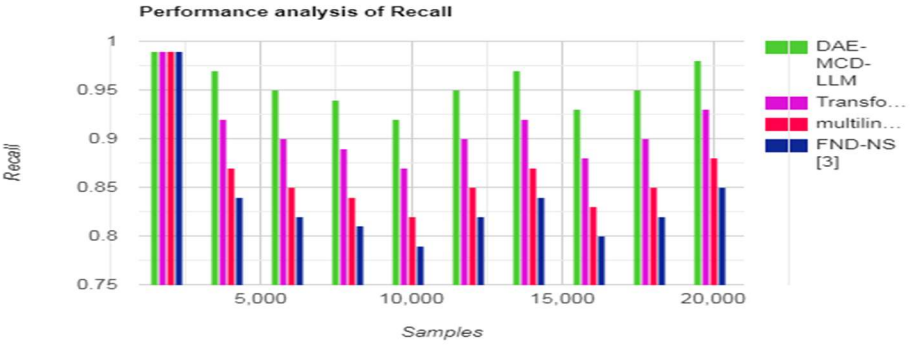


Fig 4: Precision Versus Samples

Figure 4 shows precision results for 20,000 news samples across four methods: DAE-MCD-LLM, Transformer-based LLM [1], Multilingual Deep Learning Model [2], and FND-NS [3]. For 2,000 sample tests, true positives were 1,970, 1,950, 1,920, and 1,935, while false positives were 10, 30, 60, and 45, respectively. This yielded precision values of 0.99, 0.98, 0.96, and 0.97, showing that DAE-MCD-LLM achieves the highest precision and outperforms the other methods in accurately detecting fake news



.

Fig 5: Recall Versus Samples

Figure 5 illustrates the recall rates obtained from 20,000 news samples in the fake news detection dataset, evaluated for the proposed DAE-MCD-LLM method and three baseline approaches [1], [2], and [3]. From the simulations conducted on 2,000 samples, the true positive (TP) counts for the four methods were 1,970, 1,950, 1,920, and 1,935, respectively, while the false negative (FN) counts were 1, 2, 4, and 5, respectively. As a result, all four methods achieved an overall recall of 0.99, demonstrating strong performance in accurately identifying fake news across the different approaches

.

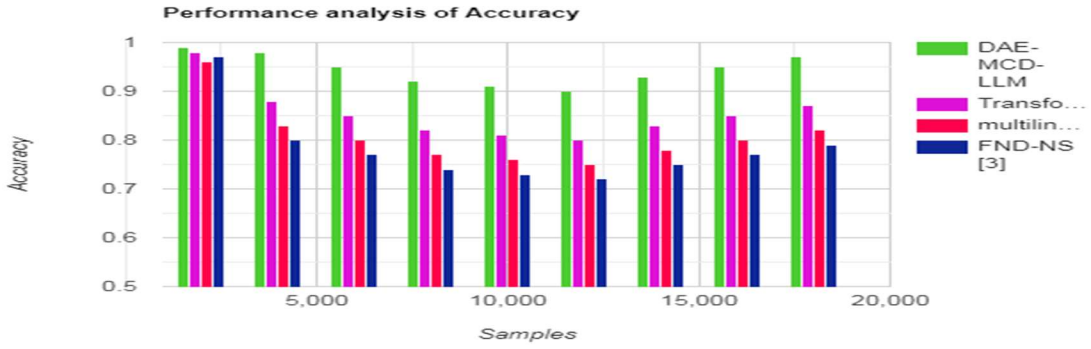


Fig 6: Accuracy Versus Samples

Figure 6 presents accuracy results for the four methods, with sample size on the vertical axis and precision on the horizontal. For 2,000 samples, accuracy was 0.99 for DAE-MCD-LLM, compared to 0.98, 0.96, and 0.97 for [1], [2], and [3], showing its superior performance. DAE-MCD-LLM outperformed other models across precision, recall, and accuracy. The Masked Custom Decoder–based BERT module iteratively processes keywords and fine-tunes weights using cross-entropy loss, improving precision by 5%, 8%, and 12% over [1], [2], and [3]. The Disentangle Attention Encoder incorporates content and word position, reducing false negatives and boosting recall by 5%, 9%, and 12%. These enhancements resulted in overall accuracy gains of 10%, 15%, and 17%, highlighting the framework's effectiveness and robustness.

5.2. Performance analysis of F1-core

The F1-score is a crucial performance statistic in false news detection that strikes a compromise between recall and precision, offering a thorough assessment of the model's overall efficacy. The F1-score has a range of 0 to 1, with 1 being the best performance and 0 denoting the worst. A higher F1-score therefore reflects a model that is more accurate and reliable in correctly classifying and detecting fake news instances.

$$F1 - score = 2 * \frac{Pre*Rec}{Pre+Rec} \quad (15)$$

Table 2: Tabulation of F1-score

Samples	F1-score			
	DAE-MCD-LLM	Transformer-based LLM [1]	multilingual deep learning [2]	FND-NS [3]
2000	0.984975	0.984975	0.974769	0.979898
4000	0.954346	0.924973	0.884746	0.859535
6000	0.944339	0.90989	0.86954	0.84426
8000	0.934332	0.899889	0.859535	0.834251
10000	0.924324	0.884746	0.84426	0.818902
12000	0.934332	0.904972	0.86474	0.839524
14000	0.944339	0.92	0.879886	0.854737
16000	0.954346	0.904309	0.863584	0.8
18000	0.944339	0.90989	0.86954	0.84426
20000	0.954346	0.93	0.889888	0.824242

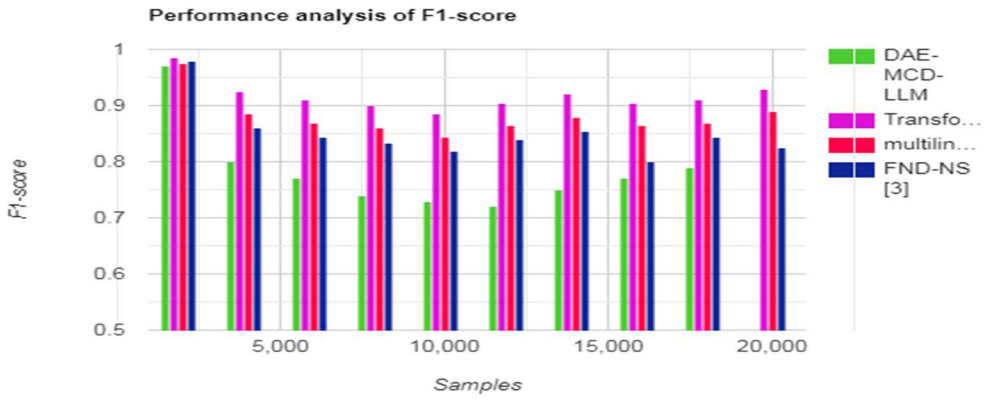


Fig 7: F1-Score Versus Samples

Figure 7 presents F1-score results for 20,000 samples across four methods. The baseline models show stable F1-scores, largely unaffected by sample size, while DAE-MCD-LLM demonstrates a clear improvement over [1], [2], and [3]. This gain comes from the Masked Custom Decoder–based BERT module, which uses a dropout layer to prevent overfitting, a linear layer

to reduce feature dimensionality, and a classification layer to generate logits. Overall, DAE-MCD-LLM raises F1-score by roughly 9% over [1] and [2] and 8% over [3], highlighting its superior effectiveness and reliability in fake news detection.

5.3. Performance analysis of execution time

Finally, execution time in fake news detection refers to the overall amount of time needed for the model to detect instances of bogus news. Depending on the particular technique or strategy used, this measure may change. Execution time can be quantified and calculated using the following expression:

$$ET = \sum_{i=1}^{n} S_i * Time\ (Logits) \quad (16)$$

From the above equation, the execution time (ET) is determined based on the number of input samples (S_i) and the time required to detect fake news using the transformed linear representations (Time (Logits)). The execution time is measured in milliseconds (ms). The simulation results of execution time for the fake news detection dataset using four different approaches are shown in Table 3.: DAE-MCD-LLM, Transformer-based LLM [1], Multilingual Deep Learning Model [2], and (FND-NS) [3].

Table 3: Tabulation Of Execution Time

Samples	Execution time (ms)			
	DAE-MCD-LLM	Transformer-based LLM [1]	multilingual deep learning [2]	FND-NS [3]
2000	220	260	300	360
4000	245	285	315	375
6000	270	300	335	390
8000	295	315	355	405
10000	315	328	385	415
12000	335	355	400	435
14000	300	330	375	415
16000	285	315	355	400
18000	255	300	335	385
20000	275	309	345	355

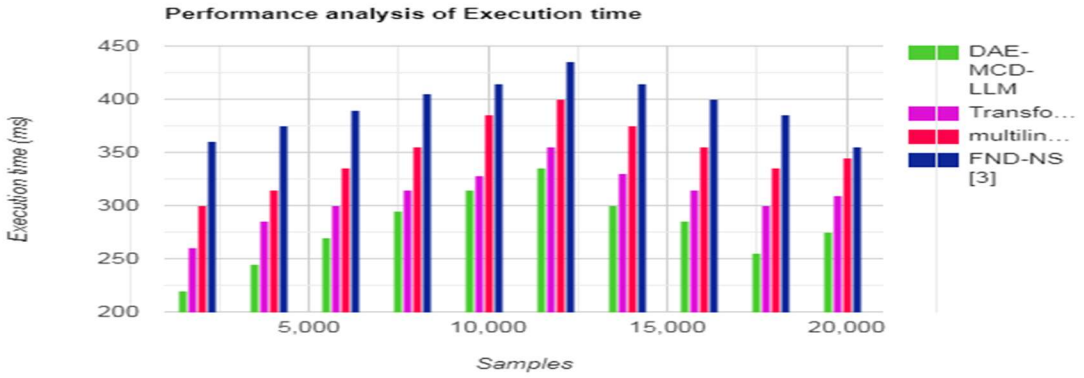


Fig 8: Execution Times Versus Samples

Figure 8 illustrates execution time for detecting fake news across eight sets of 20,000 items. DAE-MCD-LLM is significantly faster than [1], [2], and [3], reducing average time by about 11%, 26%, and 42% over ten runs. While baseline models are slower in the first six iterations, their times drop in later runs, showing that DAE-MCD-LLM accelerates both training and classification. This speedup is largely due to the Disentangle Attention Encoder, which efficiently extracts key words by combining content and word position. This improves feature extraction, making classification more efficient and reducing overall detection time.

6. RESEARCH PROBLEMS AND OPEN ISSUES

Despite significant progress in fake news detection, several challenges remain unresolved.

Current models often struggle to understand complex language, including sarcasm, irony, and subtle forms of misinformation, which can easily mislead both humans and automated systems. Many existing approaches also lack explainability, acting as black boxes that provide little insight into why a news item is classified as fake or real, which reduces user trust. In addition, real-world data is frequently noisy, incomplete, or biased, which can compromise the accuracy and reliability of predictions. Models trained on one dataset may fail to generalize to other datasets or domains, limiting their adaptability. The rapid evolution of misinformation strategies further complicates detection, as new types of fake news can quickly render models outdated. Finally, bias and fairness remain major concerns, as models may inherit biases from their training data, leading to unfair or skewed predictions. These challenges highlight the need for robust, interpretable, and adaptable systems capable of maintaining high performance across diverse contexts while providing transparent and trustworthy results.

7. CONCLUSION

With news spreading rapidly on social media, distinguishing real from fake information is increasingly difficult. The DAE-MCD-LLM model addresses this by first cleaning and preparing articles, then using a specialized encoder to extract key words, and finally classifying news with a BERT-based system. Tests show that DAE-MCD-LLM outperforms other methods, improving accuracy by up to 23%, precision by 20%, and cutting training time by over half. These results demonstrate its effectiveness in detecting and preventing the spread of false information online.

ACKNOWLEDGMENTS

The authors would like to thank the Deanship of Rajah Serfoji Government College (Auto), Thanjavur for supporting this work.

REFERENCES

[1] LekshmiAmmal HR, Madasamy AK. A reasoning based explainable multimodal fake news detection for low resource language using large language models and transformers. Journal of Big Data. 2025 Feb 23;12(1):46.
[2] Mohawesh R, Maqsood S, Althebyan Q. Multilingual deep learning framework for fake news detection using capsule neural network. Journal of Intelligent Information Systems. 2023 Jun;60(3):655-71.
[3] Raza S, Ding C. Fake news detection based on news content and social contexts: a transformer-based approach. International Journal of Data Science and Analytics. 2022 May;13(4):335-62.
[4] Dhiman P, Kaur A, Gupta D, Juneja S, Nauman A, Muhammad G. GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection. Heliyon. 2024 Aug 30;10(16).
[5] Rai N, Kumar D, Kaushik N, Raj C, Ali A. Fake News Classification using transformer based enhanced LSTM and BERT. International Journal of Cognitive Computing in Engineering. 2022 Jun 1; 3:98-105.
[6] Mallick C, Mishra S, Senapati MR. A cooperative deep learning model for fake news detection in online social networks. Journal of Ambient Intelligence and Humanized Computing. 2023 Apr;14(4):4451-60.
[7] Samadi M, Momtazi S. Fake news detection: deep semantic representation with enhanced feature engineering. International Journal of Data Science and Analytics. 2025 Aug;20(2):325-36.
[8] Szczepański M, Pawlicki M, Kozik R, Choraś M. New explainability method for BERT-based model in fake news detection. Scientific reports. 2021 Dec 8;11(1):23705.
[9] E. Almandouh M, Alrahmawy MF, Eisa M, Elhoseny M, Tolba AS. Ensemble based high performance deep learning models for fake news detection. Scientific Reports. 2024 Nov 4;14(1):26591.
[10] Harris S, Hadi HJ, Ahmad N, Alshara MA. Multi-domain Urdu fake news detection using pre-trained ensemble model. Scientific Reports. 2025 Mar 13;15(1):8705.
[11] Wierzbicki A, Shupta A, Barmak O. Synthesis of model features for fake news detection using large language models. InComputational Linguistics Workshop at CoLInS 2024 (pp. 50-65).
[12] Koka S, Vuong A, Kataria A. Evaluating the efficacy of large language models in detecting fake news: a comparative analysis. arXiv preprint arXiv:2406.06584. 2024 Jun 5.
[13] Alnabhan MQ, Branco P. Fake news detection using deep learning: A systematic literature review. IEEe Access. 2024 Jul 29.

[14] Alghamdi J, Luo S, Lin Y. A comprehensive survey on machine learning approaches for fake news detection. Multimedia Tools and Applications. 2024 May;83(17):51009-67.
[15] Thapa S, Shiwakoti S, Shah SB, Adhikari S, Veeramani H, Nasim M, Naseem U. Large language models (llm) in computational social science: prospects, current state, and challenges. Social Network Analysis and Mining. 2025 Dec;15(1):1-30.
[16] Mridha MF, Keya AJ, Hamid MA, Monowar MM, Rahman MS. A comprehensive review on fake news detection with deep learning. IEEE access. 2021 Nov 18; 9:156151-70.
[17] Wang J, Zhu Z, Liu C, Li R, Wu X. LLM-Enhanced multimodal detection of fake news. PloS one. 2024 Oct 24;19(10): e0312240.
[18] Nan Q, Sheng Q, Cao J, Zhu Y, Wang D, Yang G, Li J. Exploiting user comments for early detection of fake news prior to users' commenting. Frontiers of Computer Science. 2025 Oct;19(10):1910354.
[19] Zhou X, Jain A, Phoha VV, Zafarani R. Fake news early detection: A theory-driven model. Digital Threats: Research and Practice. 2020 Jun 11;1(2):1-25.
[20] Lai J, Yang X, Luo W, Zhou L, Li L, Wang Y, Shi X. Rumorllm: A rumor large language model-based fake-news-detection data-augmentation approach. Applied Sciences. 2024 Apr 22;14(8):3532.
[21] Park M, Chai S. Constructing a user-centered fake news detection model by using classification algorithms in machine learning techniques. IEEE Access. 2023 Jul 12; 11:71517-27.
[22] Al-Zahrani L, Al-Yahya M. Pre-trained language model ensemble for Arabic fake news detection. Mathematics. 2024 Sep 1;12(18):1-7.
[23] Rustam F, Aljedaani W, Jurcut AD, Alfarhood S, Safran M, Ashraf I. Fake news detection using enhanced features through text to image transformation with customized models. Discover Computing. 2024 Dec 28;27(1):54.
[24] Mohawesh R, Maqsood S, Althebyan Q. Multilingual deep learning framework for fake news detection using capsule neural network. Journal of Intelligent Information Systems. 2023 Jun;60(3):655-71.
[25] Koru GK, Uluyol Ç. Detection of Turkish fake news from tweets with BERT models. IEEE Access. 2024 Jan 15; 12:14918-31.
[26] Xu C, Kechadi MT. An enhanced fake news detection system with fuzzy deep learning. IEEE Access. 2024 Jun 24; 12:88006-21.
[27] Tan M, Bakır H. Fake News Detection Using BERT and Bi-LSTM with Grid Search Hyperparameter Optimization. Bilişim Teknolojileri Dergisi. 2025 Jan 1;18(1):11-28.
[28] Al-Quayed F, Javed D, Jhanjhi NZ, Humayun M, Alnusairi TS. A hybrid transformer-based model for optimizing fake news detection. IEEE Access. 2024; 12:160822-34.
[29] Mohawesh R, Salameh HB, Jararweh Y, Alkhalaileh M, Maqsood S. Fake review detection using transformer-based enhanced LSTM and RoBERTa. International Journal of Cognitive Computing in Engineering. 2024 Jan 1; 5:250-8.
[30] Verma PK, Agrawal P, Amorim I, Prodan R. WELFake: word embedding over linguistic features for fake news detection. IEEE Transactions on Computational Social Systems. 2021 Apr 5;8(4):881-93.