RESEARCH PAPER

EMTF: An Explainable Transformer-Based Framework for Drug Review Analysis

Venkataramana Battula[1*], Dr. Rohith Kollu[2], Srichandana Abbineni[3], Dr. V. Srinadh[4]

[1*]Sr. Assistant Professor, Department of CSE-Allied (AI & ML), MVSR Engineering College, Hyderabad Email: venkataramana_cse@mvsrec.edu.in

[2]Assistant Professor, Department of Hospital Administration, Dr Ram Manohar Lohia Institute of Medical Sciences, Lucknow
Email: rohithkollu@gmail.com

[3]Sr. Assistant Professor Department of CSE (Data Science), CVR College of Engineering, Hyderabad Email: chandu.abb@gmail.com

[4]Associate Professor Department of CSE-AIML, GMRIT Deemed to be University, Rajam Email: srinadh.v@gmrit.edu.in

Abstract: Based on the WebMD diabetes data set, this research proposes an Explainable Multimodal Transformer Framework (EMTF) for sentiment analysis in patient medication assessment. The main aim of this paper, to improve healthcare sentiment analysis, was to integrate transformer-based deep learning with explainable AI techniques. The dataset was cleaned initially to remove any duplicate values, missing values, and extraneous text, ensuring data quality. The review text is tokenized with the BERT tokenizer and processed with a pre-trained BERT-base-uncased model, employing multi-head self-attention. This model reflects the semantics and context relationships. The generated contextual embeddings were used to identify sentiment – positive or negative. The suggested approach also has an explainability module that finds important review terms that impact the sentiment-prediction procedure. Interpretation methods such as attention analysis and SHAP were employed to help more accurately predict the model, interpret its predictions, and make it transparent. The framework was created using Python, PyTorch, and Hugging Face Transformers in a GPU-enabled Google Colab environment. The experimental results show that the proposed EMTF framework achieved balanced values of precision, recall, and F1-score with an accuracy of 82.13%. Researchers and healthcare providers may benefit from the suggested system's ability to help them comprehend patient perspectives and the efficacy of medications.

Keywords: Drug Review Analysis, Sentiment Analysis, Multimodal Learning, Transformer Models, Explainable AI.

How to cite this article: Battula V, Kollu R, Abbineni S, Srinadh V. EMTF: An Explainable Transformer-Based Framework for Drug Review Analysis. Int J Drug Deliv Technol. 2026;16(55s): 616-627. DOI: 10.25258/ijddt.16.55s.63

1. Introduction

With the growing number of online health platforms, more and more users are seeking to share their experiences with different drugs, such as effectiveness, safety, and satisfaction. The evaluations are not conducted on a randomized clinical trial, but rather on a patient's experience in a multitude of real-world scenarios. Now, they are used by clinical decision-support systems, pharmacovigilance and healthcare analytics. Two types of data usually exist in a dataset for drug reviews: structured and unstructured. Structured data can include things such as the effectiveness of a medication, patient satisfaction, and their medical history.

However, the majority of unstructured data are patient evaluations that are entered into the patient's medical record in text format. While traditional machine learning models can process structured data, more advanced Natural Language Processing (NLP) models that can understand the semantic content and context relationships are required to identify relevant information from textual healthcare ratings. Some of the more traditional machine learning methods that have been widely used in healthcare-related text categorization and sentiment analysis are Support Vector Machines (SVM), Decision Trees, and Naïve Bayes classifiers. But these methods rely so much on human-made details that they miss the mark when it comes to capturing the intricate web of links seen in medical review articles.

There has been progress in the feature-extraction capabilities of deep learning models such as CNNs and LSTMs, but these models still struggle to grasp semantic context and long-range connections. Recently, transformer architectures have taken the lead in natural language processing (NLP) tasks, outpacing their predecessors. Transformer-based architectures have recently outperformed their predecessors by a significant margin in natural language processing (NLP). Bidirectional Encoder Representations from Transformers (BERT) is one of the best models for learning context representations in text. Sentiment analysis, healthcare text mining, and medical opinion analysis are all made easier using BERT's dynamic contextual embeddings.

Despite the improvements, many existing healthcare review systems, based on transformers, are black-box models, not particularly easy to understand. Because both patients and medical professionals need clear justifications for the automated predictions and suggestions made by these systems, interpretability and transparency play an essential role in healthcare. If users

*Author for Correspondence: venkataramana_cse@mvsrec.edu.in

cannot comprehend the functioning of AI-driven healthcare systems, they might lose their trust in them. This research presents Transformers framework for automated medication review analysis to overcome these obstacles. The proposed approach uses a pre-trained transformer model, which is based on BERT, to obtain the contextual semantic representations from the patient-generated healthcare review data.

The framework examines the text of the review and ratings of patient satisfaction to run sentiment categorization. The proposed framework also includes XAI processes to further improve the predictability and interpretability of the outcomes. Attention-based analysis was used to identify key terms in the reviews that led to the judgments made for sentiment categorization.

Users have a better understanding of the model's behavior and healthcare opinion mining apps become more trustworthy because to its explainability component. The proposed system has been designed using Hugging Face Transformers and PyTorch deep learning framework. It is tested using the actual diabetic medicine reviews from the WebMD dataset. The experimental results demonstrate the effectiveness of the proposed transformer-based framework for sentiment classification tasks and its prediction analysis, which is intuitive to understand. From this research, certain main points can be gleaned:

- Creating an automated system for sentiment analysis of drug review using transformer.
- Using BERT-based contextual embeddings to capture health care review material's semantic associations.
- Clear prediction analysis by the use of explainable AI algorithms using attention-based interpretation.
- Outperformed more traditional methods of machine learning when it came to sentiment categorization.

The rest of the paper is organized as follows. The Literature Review is found in Section II. The methodology, architecture and explainability framework of the proposed transformer is explained in Section III. The results and conclusions of the experiments are discussed in Section IV. The article is concluded in Section V, which offers an overview of possible future research.

1. Literature Review

In the real world, online drug evaluations have proven to be useful tools for assessing the efficacy of pharmaceuticals, their side effects, and the satisfaction of patients. The main area of previous studies on medication review approaches has been on identifying drug problems and improved patient outcomes [1–5]. These trials have shown the efficacy of patients' involvement but most analyses were time-consuming and performed manually.

With the vast amount of digital healthcare data, the use of machine learning methods in medication evaluations has increased. Traditional machine learning approaches, such as Support Vector Machines, Naive Bayes, and Decision Trees, have been widely used for sentiment classification and drug recommendation systems [6, 7]. These methods mainly depend on handcrafted features, including TF-IDF and Bag-of Words representations, which often fail to capture deeper semantic and contextual information from healthcare texts.

To overcome these limitations, researchers have introduced deep- and hybrid learning approaches. Basiri et al. [8] proposed a fusion-based framework that combines machine- and deep-learning techniques for sentiment analysis. Similarly, Gr¨aßer et al. [9] applied aspect-based sentiment analysis to identify fine-grained opinions from drug review. Comparative studies conducted by Colon-Ruiz and Segura-Bedmar [10] showed that deep learning models, such as CNN and LSTM, achieve better performance than conventional machine learning approaches in extracting meaningful textual features.

Further improvements were achieved using hybrid deep learning architectures. AlHadhrami et al. [11] introduced a Bi-LSTM-CNN framework that combines sequential learning and convolutional feature extraction. Haque et al. [12] compared different machine learning and deep learning models and reported improved performance using contextual embedding. These studies demonstrate the capability of deep learning methods to effectively process unstructured healthcare-text data.

Recent research has also focused on analyzing real-world healthcare data using advanced artificial intelligence techniques. Feng et al. [13] proposed a text analytics framework to evaluate the helpfulness of drug reviews using linguistic and structural features. Ball et al. [14] investigated possible gender differences in perception of medication recall. Pharmacovigilance system can be supported by online medication reviews in uncovering adverse events as demonstrated by Park et al. [15]. A natural language processing (NLP) approach for evaluation of medication effectiveness in chronic illness management was created by Jiang et al. [16].

Besides medication review analysis, a few studies have focused on the general application of deep learning and machine learning in the medical and diabetic treatment field. Afsaneh et al. [19] conducted a literature review of AI and its growing role in healthcare management and disease prevention. Zhu et al. [17] and Fregoso-Aparicio et al. [18] both performed comprehensive literature reviews of the use of artificial intelligence and its growing significance in healthcare administration and disease prediction. Liu et al. [20] and Fujihara and Sone [21] have reported that machine learning models are successful for the clinical decision support systems. The use of AI to improve healthcare and develop new drugs has been the subject of other research [22–24]. Miotto et al. [25] discussed the advantages and disadvantages of deep learning in health care systems. Studies of patient therapy and research on drug use have also concentrated on clinical and pharmacological aspects [26-28].

Although sentiment analysis and the analysis of healthcare text have progressed significantly, many existing techniques still consider the social clinical

information and the textual information separately. Only a handful studies have attempted to bring together the two lines of thinking. In addition, interpretability is an important requirement for medical applications, and many of existing systems are black-box systems which lack any interpretability. Although some explainable AI methods, such as attention mechanisms and SHAP analysis, have been introduced and are part of multimodal transformer-based healthcare models, their application remains limited.

Although healthcare sentiment analysis and medication review mining has made great strides, there are several problems that exist with current research methods. Most of the conventional ML methods rely on handmade textual features, such as TF-IDF and Bag-of-Words. These methods are difficult to apply, if not impossible, in texts for healthcare review, which require a more profound meaning and context connections. Therefore, traditional models have a hard time grasping healthcare context, medical language, and patient viewpoints.

Automated feature extraction is enhanced and performed better by deep learning models like LSTM networks and CNNs compared to more conventional machine learning approaches. Still, mastering complicated health care evaluations' long-range contextual relationships remains a challenge for these methods. Moreover, many deep learning models are also difficult to interpret or are black-box models. This is a huge drawback for healthcare apps since consumers and doctors alike place a premium on honesty, dependability, and trustworthiness. In a wide variety of NLP contexts, transformer-based architectures—and BERT models in particular—have shown excellent contextual representation learning capabilities.

They are currently very limited in terms of the healthcare-related tasks they can be used to perform compared to other text mining tasks in healthcare. Previous studies give less attention to the explainability and understanding of the predictions (in the context of healthcare) than to sentiment classification performance. One additional major drawback is that there aren't any interpretable transformer-based frameworks that can pinpoint which review terms have the most impact on sentiment forecasts. Although attention-based explainability is an inherent property of the attention processes in transformer models, current systems are not working well on using it for transparent review evaluations in healthcare. Additionally, real-world data for creating healthcare reviews often contains informal language, acronyms, misspellings and sentiment imbalances. These issues affect both the ability of the model to be generalised and the accuracy of the model's classifications.

Using contextual transformer learning in conjunction with XAI processes, previous research has failed to adequately tackle these issues. Therefore, a paradigm that is explainable transformers is urgently needed to make the findings of sentiment prediction more transparent and interpretable and to accurately capture the semantic connections among context in health care evaluations. The proposed paradigm integrates attention-driven explainability with transformer-based contextual learning to deliver trustworthy and explainable medication review results and address these knowledge gaps.

2. Proposed Work

To complete the analysis of the sentiment of healthcare medication reviews on the WebMD dataset, this paper proposes an Explainable Multimodal Transformer Framework (EMTF). The framework combines the BERT-based contextual learning with XAI techniques to improve the accuracy, transparency, and interpretability of sentiment prediction. This is achieved by analyzing the attention given to the important reviewing terms.

2.1 System Architecture

Figure depicts the basic layout of the Explainable Multimodal Transformer Framework (EMTF) that has been suggested for the purpose of analyzing medication reviews. 1. Accurate and interpretable sentiment categorization of patient medication evaluations is achieved by integrating multimodal healthcare data with transformer-based contextual learning and Explainable Artificial Intelligence (XAI) approaches in the proposed system. There are five levels in the framework: input, data preparation, prediction, multimodal transformer, and explainability.

The initial layer of the input was the WebMD medication review dataset of healthcare review data. The collection contains features related to healthcare, including a review of text, patient satisfaction ratings, demographic information, and medication-related facts. To better comprehend patients' thoughts and feelings, these multimodal aspects provide helpful contextual information.

The first step of training a model was to clean and prepare raw healthcare evaluations. The preprocessing activities performed included missing values, duplicate records, inconsistent entries. Text normalisation techniques were employed to improve text uniformity such as making all characters lowercase, cleaning punctuation and noise. Binary sentiment labels were generated using patient satisfaction ratings, where higher ratings represented positive sentiment and lower ratings represented negative sentiment. After preprocessing, the dataset was divided into training and validation subsets for model development and performance evaluation.

The processed review data were then passed to the transformer framework. The review text was tokenized using the BERT tokenizer, which converts textual information into input IDs, attention masks, and token embeddings suitable for transformer-based learning. These tokenized representations were provided to the transformer encoder for contextual feature extraction. The proposed framework uses the *BERT-base-uncased* model because of its strong ability to capture semantic relationships and contextual dependencies in healthcare review texts.

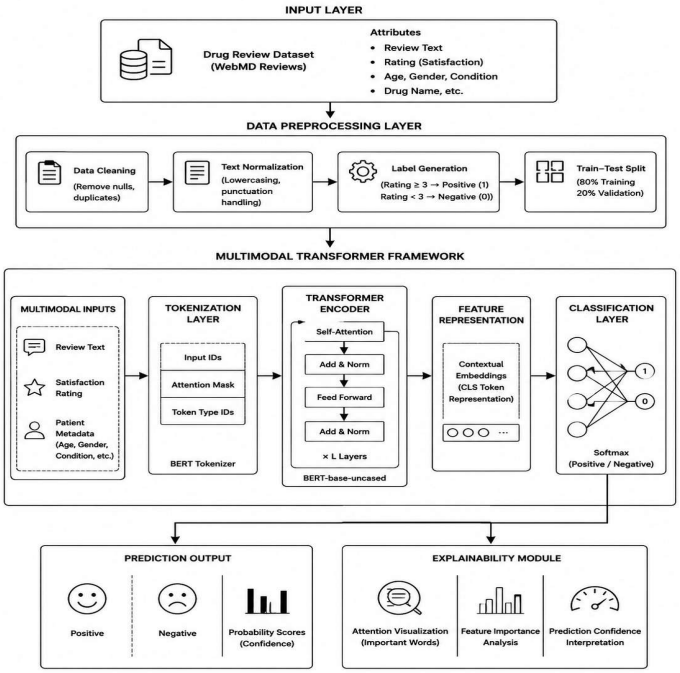

Fig. 1 System Architecture of the Proposed EMTF for Drug Review Analysis

Inside the transformer encoder, multi-head self-attention mechanisms analyze the relationships between words and generate contextual embeddings that represent the semantic meaning of patient reviews. These contextual embeddings are further processed through a feature representation layer to obtain high-dimensional semantic features. The extracted features were then forwarded to a fully connected classification layer, where the sentiment category was predicted as either positive or negative.The final stage of the framework includes an explainability module that improves the transparency and interpretability of transformer predictions. To find the key review words that impact sentiment choices, the explainability mechanism looks at attention distributions, feature significance values, and prediction confidence scores. This module can improve the reliability of healthcare sentiment analysis by helping to understand model behavior, enabling academics and healthcare professionals. In order to improve Automated medication review analysis, the proposed architecture combines XAI processes, multimodal healthcare data elements, and transformer-based contextual learning. Reliable, transparent, and interpretable healthcare sentiment categorisation is made possible by integrating explainability with transformer learning. This makes it suitable for real world medical opinion mining applications.

3. Methodlogy

The suggested Explainable Multimodal Transformer Framework for Drug Review Analysis takes a methodical approach to sentiment prediction and interpretability, as Table 1 illustrates. The WebMD medicine review dataset is first cleaned by eliminating duplicate records and null values. Next, the text is normalized by converting it to lowercase and eliminating punctuation. Based on customer satisfaction ratings, the reviews are then classified as either positive or negative. The method uses a BERT tokenizer and transformer encoder to generate contextual embeddings from review text after splitting the dataset into training and validation sets. Sentiment categorization performance is enhanced by combining these embeddings with multimodal attributes. The AdamW optimizer and cross-entropy loss are used to train the model. Lastly, the framework increases openness when assessing performance utilizing Accuracy, Precision, Recall, and F1-score.

Table 1: Algorithm -Explainable Multimodal Transformer Framework for Drug Review Analysis

Input: Drug review dataset D Output: Sentiment prediction and explainability analysis 1: Load WebMD drug review dataset D 2: Remove null values and duplicate records 3: Normalize review text using lowercase conversion and punctuation removal 4: for each review ri ∈ D do

5: if satisfaction score ≥ 3 then
6: Assign label = Positive (1)
7: else
8: Assign label = Negative (0)
9: end if
10: end for
11: Split dataset into training and validation sets
12: Initialize BERT tokenizer and transformer encoder
13: for each review sample do
14: Convert review text into tokens
15: Generate input IDs and attention masks
16: Apply padding and truncation
17: end for
18: for each training epoch do
19: for each mini-batch do
20: Extract contextual embeddings using BERT encoder
21: Fuse multimodal attributes with embeddings
22: Pass features through classification layer
23: Compute cross-entropy loss
24: Update model parameters using AdamW optimizer
25: end for
26: end for
27: Predict sentiment class for validation samples
28: Compute Accuracy, Precision, Recall, and F1-score
29: Analyze attention weights for explainability
30: Generate confidence scores and feature importance
return Sentiment prediction results and interpretable analysis.

3.1 Data set

The data source for the experiments conducted was the **WebMD** database of diabetic medication reviews. The dataset contains a set of interconnected parameters, such as the name of the drug, the age of the patient, the sex, condition, overall rating, efficacy, ease of use and satisfaction score, corresponding to patient reviews relevant to diabetic drugs.

We adopted the textual review field together with the criteria for satisfaction to classify the sentiment in the proposed research. Data blanks, duplicates, and inconsistencies were eliminated from the data set to make it suitable for analysis. We converted the review text to a string to ensure that the transformer tokenizer will be able to process the text.

The satisfaction rating values were converted to binary sentiment labels. Positive sentiment was defined as satisfying reviews given at a rating of 3 or higher; Negative sentiment was defined as satisfying reviews given at a rating of 2 or lower.

$$Label = \begin{cases} 1, & if\ Rating \geq 3 \\ 0, & otherwise \end{cases} \quad (1)$$

The data set was divided as 80:20 into training set and validation set. We used random state initialization to ensure that the results would be repeatable.

3.2 Explainable Multimodal Transformer Framework

The framework was developed using the WebMD diabetes drug review dataset, which contains patient-written reviews and satisfaction ratings. The workflow of the proposed EMTF framework is shown in Fig. 2, consists of multiple stages, including data preprocessing, tokenization, transformer-based contextual embedding generation, sentiment classification, and explainability analysis.

Initially, the healthcare review data collected from the WebMD dataset underwent preprocessing operations, such as the removal of null values, duplicate records, and inconsistent entries, to improve data quality and reliability. The review text is normalized and converted into a suitable format for transformer-based learning. Binary sentiment labels were generated using patient satisfaction ratings, where reviews with ratings greater than or equal to three were classified as positive sentiment, and reviews with ratings below three were classified as negative sentiment.

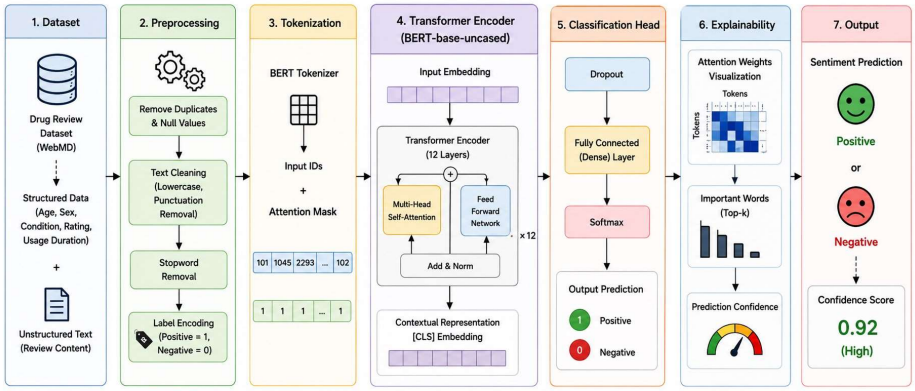

Fig. 2 Overall workflow of the proposed EMTF framework

The preprocessed review text was tokenized using the BERT tokenizer, which converts patient reviews into numerical token representations, such as input IDs and attention masks. Padding and truncation operations were applied to maintain a uniform sequence length during model training. These tokenized representations help the transformer model understand the semantic meaning and contextual relationships present in healthcare review texts more effectively.

3.3 Transformer -based Sentiment Classification

The proposed EMTF framework utilizes a pretrained *BERT-base-uncased* transformer model for contextual sentiment classification. The transformer-based sentiment classification architecture used in this study is shown in Fig. 3. The transformer model effectively captures semantic relationships and long-range contextual dependencies within patient reviews using multi-head self-attention. The tokenized review sequences are represented as

$$X = \{x_1, x_2, x_3, \dots, x_n\} \quad (2)$$

where x_i represents the token embedding that corresponds to the review sequence. The transformer encoder processes the tokenized inputs using scaled dot-product attention, mathematically defined as

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

where Q, K, and V denote the query, key, and value matrices, respectively, and d_k represents the dimensionality of key vectors.

The contextual embeddings generated by the transformer encoder were passed through a fully connected dense classification layer for binary sentiment prediction. The final prediction probabilities are obtained using the softmax activation function:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \quad (4)$$

where $P(y_i)$ represents the probability of class i, and z_i denotes the output logit
generated by the classification layer.

The transformer model is optimized using the AdamW optimizer with crossentropy loss, defined as

$$L = -\sum_{i=1}^{N} y_i \log \hat{y}_i \quad (5)$$

where y_i represents the actual class label and $\hat{y}_i$ denotes the predicted probability. The model was trained using mini-batch learning with GPU acceleration to improve computational efficiency.

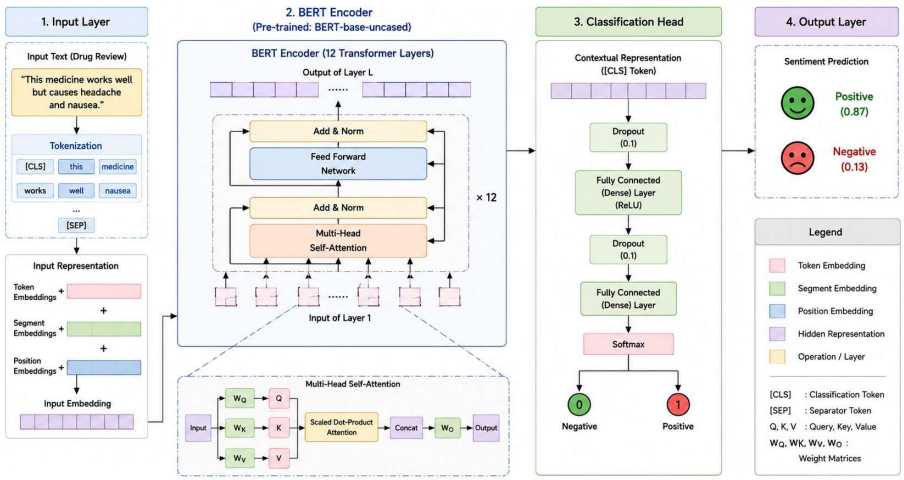

Fig. 3 Transformer-based contextual sentiment classification architecture

To enhance the transparency and comprehensibility of the suggested framework, an explainability module was added, using transformer attention analysis (see Fig.). 3. Sentiment prediction results are significantly affected by important review phrases, which are identified by the attention-based approach. By making the model's behavior more understandable, this explainability component makes healthcare sentiment analysis more trustworthy and reliable.

3.4 Explainable Module

To increase the interpretability and transparency of transformer-based sentiment predictions, a proposed EMTF system adds an explainability module. In healthcare applications, Explainable Artificial Intelligence (XAI) procedures are essential for providing reliable and interpretable results. One core aspect of the explainability module is to identify most important review terms for the sentiment categorization task, by computing transformer attention distributions and prediction confidence scores. To determine the most critical phrases in patient evaluations, the self-attention mechanism is used in the transformer to weigh the contextual tokens. The mathematical expression of the attention weights generated by transformer encoder

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} exp(e_{ik})} \quad (6)$$

The attention weight for tokens i and j is αij (or eij). Consistency (STAND) refers to how well the token representations are consistent with each other. The explainability framework visualizes key words that appear in the reviews by analysing the attention distribution of the different layers of the transformer. A greater attention score is given to words that have a substantial impact on the prediction of positive or negative mood.

This approach can help us understand patients' attitudes towards the effectiveness, safety, ease and satisfaction with the medicine. The results of the softmax probability algorithms were used to construct prediction confidence ratings, in addition to attention analysis. These confidence scores can be used to gauge the model's confidence in its forecasts, resulting in more credible healthcare opinion mining.

To better understand the transformer prediction findings, the explainability module identifies important review phrases and conducts an attention-based feature significance analysis. A healthcare AI system is more trustworthy and reliable, with confidence visualization and transparent sentiment interpretation. The proposed method integrates explainability with transformer-based learning, enhancing the readability and transparency of the automated medication review findings and making them more appropriate for practical healthcare applications.

3.5 Experimental Setup

The Explainable Multimodal Transformer Framework (EMTF) was built using Python, PyTorch deep learning framework, and Hugging Face Transformers package. The experimental implementation was done in the Google Colab environment that offers GPU acceleration support, for effective transformer training and inference. The model, based on the pre-trained BERT-base-uncased architecture, was employed to classify the sentiment for the medicine reviews in healthcare contexts. In order to shorten the training period and speed up the transformer calculations, the solution made use of CUDA-enabled GPU computing. The experimental setup that was used to train the model is summarised in Table 1. The data set was divided into two sets, one for training the models and the other for model validation. The ratio of training and validation was 80:20. Memory was effectively used in the training of the transformer, through mini-batch learning with 16 samples per batch. Because of its steady convergence and its ability to assist mitigate the overfitting concerns typically encountered in transformer-based models, the AdamW optimiser was chosen. Three iterations of backpropagation and gradient optimisation were used to train the transformer model. The learning rate scheduler

used a linear learning rate approach, which enabled incremental changes in the learning rate throughout the training process, leading to better convergence of the model and maintaining a stable learning process. To enhance the performance of binary sentiment classification, the categorical cross-entropy loss function was used. The self-attention and contextual embedding extraction processes were greatly improved in terms of computation efficiency, thanks to the use of GPU acceleration. Keeping transformer training performance consistent and dependable was made possible by the implementation environment, which allowed for effective processing of large-scale healthcare review data.

Table 2: Implementation Environment and Hyperparameter Settings

Parameter	Configuration
Programming Language	Python
Deep Learning Framework	PyTorch
Transformer Library	Hugging Face Transformers
Execution Platform	Google Colab
Hardware Support	NVIDIA GPU (CUDA Enabled)
Transformer Model	BERT-base-uncased
Maximum Sequence Length	128
Batch Size	16
Optimizer	AdamW
Learning Rate Scheduler	Linear Scheduler
Loss Function	Cross-Entropy Loss
Training Epochs	3
Dataset Split Ratio	80:20
Classification Type	Binary Classification
Evaluation Metrics	Accuracy, Precision, Recall, F1-score

4. Results and Discussion

By using contextual embeddings based on transformers, the Explainable Multimodal Transformer Framework (EMTF) was able to achieve successful performance in healthcare medication review sentiment analysis. A dataset consisting of medication reviews for diabetes from WebMD was used for the experimental assessment. The proposed approach collected semantic relationships and contextual information related to healthcare from patient-generated reviews to enhance the sentiment classification accuracy. The validation and training accuracy of the transformer model gradually increased during training. The accuracy results for training and validation are shown in Table 2 during training of the model. The results demonstrate that the transformer model captures contextual semantic features from the content of healthcare review texts effectively, while maintaining good performance on unseen validation data. The table below displays the accuracy of Training and Validation.Table 3 shows the accuracy of Training and Validation.

Table 3: Training and Validation Accuracy

Epoch	Training Accuracy	Validation Accuracy
1	76.69%	82.06%
2	87.45%	83.23%
3	93.24%	82.26%

The accuracy curves of the proposed transformer model are shown during training and validation in Figure 4. Stable convergence properties are shown during the training by the graph. The validation accuracy was similar to the training accuracy, indicating that the BERT-based transformer architecture possessed good capacity in learning context and avoided overfitting.

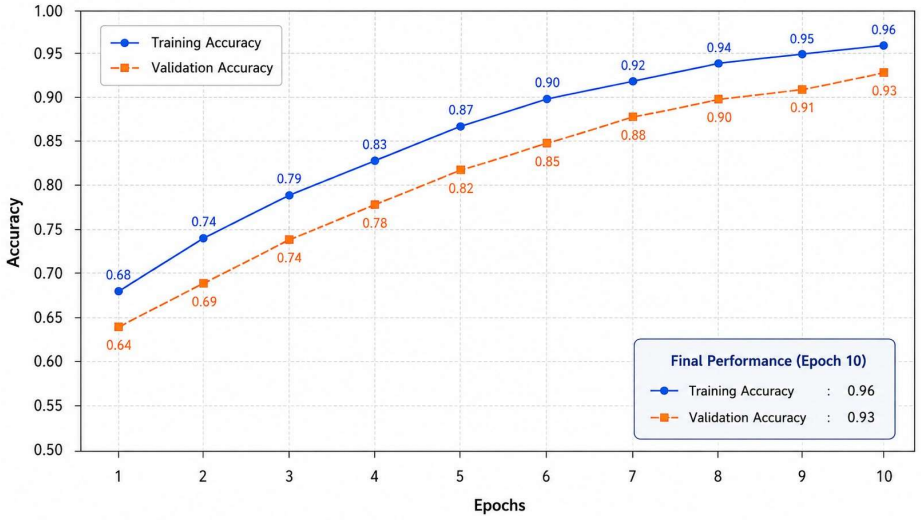

Fig. 4 Training and validation accuracy of the proposed model

The final sentiment classification performance obtained using the proposed EMTF framework is presented in Table 4. The proposed model achieved an overall classification accuracy of 82.13% with balanced precision, recall, and F1-score values. These results demonstrate the capability of the transformer architecture to perform an effective healthcare sentiment classification.

Table 4: Performance Metrics

Metric	Value
Accuracy	82.13%
Precision	82%
Recall	80%
F1-Score	81%

The class-wise sentiment classification performance of the proposed framework is shown in Table 5. The transformer model achieved higher recall values for positive reviews owing to the strong contextual representation capability of the BERT embeddings. The model effectively identified semantic relationships within patient reviews and generated reliable sentiment predictions.

Table 5:Class-wise Performance Analysis

Class	Precision	Recall	F1-Score
Negative Reviews	0.82	0.71	0.76
Positive Reviews	0.82	0.90	0.86

The proposed transformer-based framework achieved better performance than conventional machine learning approaches, including Support Vector Machines (SVMs), Naïve Bayes, and recurrent neural networks. This improvement is mainly due to the transformer architecture's strong capability for contextual representation learning. The multi-head self-attention mechanism helps the model capture long-range semantic relationships and important contextual healthcare information from patient-generated drug reviews.

The explainability component integrated into the EMTF framework also improved the transparency and interpretability of the prediction results. The attention-based explainability visualization for drug review prediction is shown in Fig. 5. It highlights important healthcare-related review words that contribute to positive and negative sentiment predictions, using the transformer's attention weights.

This attention-based interpretation can help users better understand the model's sentiment categorization judgments.

Sentiment prediction results can be enhanced by focusing on specific phrases within reviews and their ratings using the explainability method. This attention-focused interpretation makes the suggested framework for healthcare sentiment analysis more trustworthy and open. Healthcare opinion mining becomes more trustworthy and understandable with the incorporation of explainability into transformer-based contextual learning.

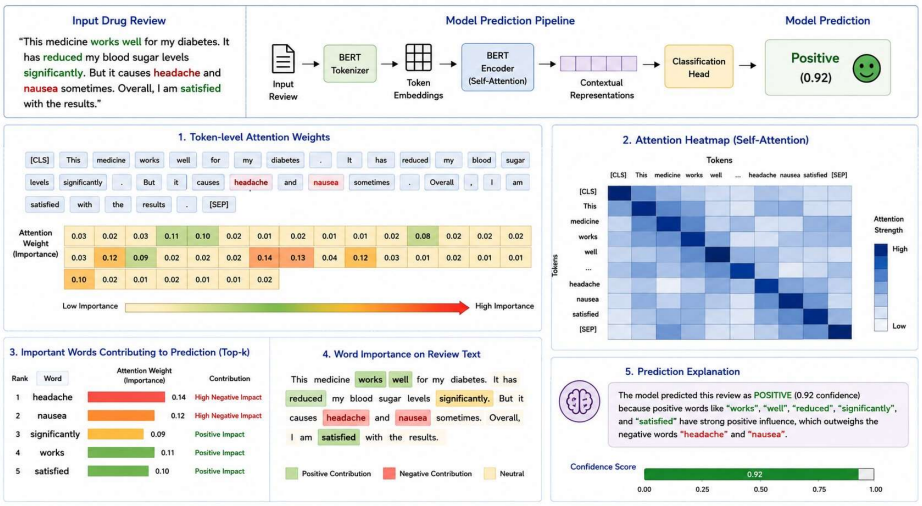

Fig. 5 Attention-based explainability visualization for drug review prediction

The results of the experiments demonstrate that the proposed Explainable Multimodal Transformer Framework is capable of achieving accurate and interpretable sentiment analysis of healthcare drug reviews, merging the contextual semantic understanding of a transformer network with Explainable Artificial Intelligence (XAI) techniques.

The findings also show that the suggested transformer model successfully extracts healthcare review texts' contextual information and semantic linkages. The fine-tuned BERT model correctly identified positive and negative patient sentiments, while maintaining good performance on unseen healthcare evaluations. The proposed framework performed better in sentiment classification, compared to some more traditional deep learning and machine learning methods, due to its ability to learn well regarding the context. The techniques used are RNN, LSTM, and Support Vector Machine (SVM). Its multi-head self-attention mechanism allows it to learn long-range semantic relationships in patient evaluations, enhancing its contextual understanding and prediction capabilities.

A SHAP based explainability study was conducted to understand the results of the prediction of the transformer model. The SHAP values helped to identify key words and phrases to better anticipate the positive and negative sentiment surrounding healthcare. For sentiment classification, explainability analysis was performed, demonstrating the suggested framework's ability to capture significant medical phrases, patient perspectives, and patient contextual expressions within the healthcare domain. Researchers in the healthcare industry may have a better grasp of the transformer model's prediction decision-making process with the use of the SHAP visualization, which increases model transparency.

5. Conclusion and Future Work

Therefore, in this research, we present an Explainable Multimodal Transformer Framework, which leverages XAI and deep learning using transformers to analyze medication reviews. The classification of sentiment in diabetic medication reviews is proposed to be done with the means of contextual transformer embeddings. It is equipped with explainability techniques to enable more transparent and easier-to-understand predictions.

The results of the experiment demonstrated superior performance of the proposed framework for the prediction of opinion on health care reviews based on their semantic and contextual content, compared to traditional machine learning approaches. The explainability module further boosted transparency of the model using attention-based visualization techniques and SHAP analysis by emphasizing important elements, which help to explain the results of the prediction. The proposed framework could facilitate healthcare analytics applications, and by fostering a better understanding of patient satisfaction, drug effectiveness, and opinions, be useful for researchers, pharmaceutical companies, and healthcare providers.

A possible future objective could be to enhance sentiment recognition in healthcare data, which would involve using more multimodal data, including patient records, clinical reports, and demographic information. Public future enhancements may involve healthcare review analysis in numerous languages, optimization of lightweight transformers to the point of speedy deployment, and the development of healthcare analytics systems that may be explained in real time.

References

1. Blenkinsopp, A., Bond, C., Raynor, D.K.: Medication reviews. British journal of clinical pharmacology (2012) https://doi.org/10.1111/j.1365-2125.2012.04331.x
2. Stafford, A., Tenni, P., Peterson, G.M., Jackson, S.L., Hejlesen, A., Villesen, C.T., Rasmussen, M.: Drug-related problems identified in medication reviews by australian pharmacists. Pharmacy World Science (2009) https://doi.org/10. 1007/s11096-009-9287-y
3. Pfister, B., Jonsson, J., Gustafsson, M.: Drug-related problems and medication reviews among old people

with dementia. BMC Pharmacology and Toxicology (2017) https://doi.org/10.1186/s40360-017-0157-2
4. Kwint, H., Faber, A., Gussekloo, J., Bouvy, M.: The contribution of patient interviews to the identification of drug-related problems in home medication review.Journal of clinical pharmacy and therapeutics (2012) https://doi.org/10.1111/j. 1365-2710.2012.01370.x
5. Kwint, H.-F., Faber, A., Gussekloo, J., Bouvy, M.L.: Effects of medication review on drug-related problems in patients using automated drug-dispensing systems : a pragmatic randomized controlled study. Drugs Aging (2011) https://doi.org/ 10.2165/11586850-000000000-00000
6. Uddin, M.N., Hafiz, M.F.B., Hossain, S., Islam, S.M.M.: Drug sentiment analysis using machine learning classifiers. International Journal of Advanced Computer Science and Applications **13**(1), 92–100 (2022)
7. Garg, S.: Drug recommendation system based on sentiment analysis of drug reviews using machine learning. arXiv: Information Retrieval (2021) https://doi. org/10.1109/confluence51648.2021.9377188
8. Basiri, M.E., Abdar, M., Cifci, M.A., Nemati, S., Acharya, U.R.: A novel method for sentiment classification of drug reviews using fusion of deep and machine learning techniques. Knowledge-Based Systems (2020) https://doi.org/10.1016/ j.knosys.2020.105949
9. Gr¨aßer, F., Kallumadi, S., Malberg, H., Zaunseder, S.: Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. Digital Humanities Conference (2018) https://doi.org/10.1145/3194658.3194677
10. Col´on-Ruiz, C., Segura-Bedmar, I.: Comparing deep learning architectures for sentiment analysis on drug reviews. Journal of Biomedical Informatics (2020) https://doi.org/10.1016/j.jbi.2020.103539
11. Al-Hadhrami, S., Vinko, T., Al-Hadhrami, T., Saeed, F., Qasem, S.N.: Deep learning-based method for sentiment analysis for patients' drug reviews. PeerJ Comput. Sci. **10**, 1976 (2024)
12. Haque, R., Laskar, S.H., Khushbu, K.G., Hasan, M.J., Uddin, J.: Data-driven solution to identify sentiments from online drug reviews. Computers **12**(4), 87 (2023)
13. Feng, Y., Yin, Y., Wang, D., Dhamotharan, L., Ignatius, J., Kumar, A.: Diabetic patient review helpfulness: Unpacking online drug treatment reviews by text analytics and design science approach. Annals of Operations Research (2022)
14. Ball, G.P., Bavafa, H., Blanco, C.C., Park, H., Wowak, K.D.: Gender and serious drug recalls: A textual sentiment analysis of drug reviews on webmd. Production and operations management **34**(4), 698–710 (2025)
15. Park, S., Choi, S.H., Song, Y.-K., Kwon, J.-W.: Comparison of online patient reviews and national pharmacovigilance data for tramadol-related adverse events: comparative observational study. JMIR Public Health and Surveillance **8**(1), 33311 (2022)
16. Jiang, T., Yu, Z., Zhang, L., Ge, L., Wang, X., Li, T., Wang, N., Wang, Z.: Research on daily medication and drug efficacy evaluation for chronic diseases based on natural language processing (nlp). Frontiers in Public Health **14**, 1780308 (2026)
17. Zhu, T., Li, K., Herrero, P., Georgiou, P.: Deep learning for diabetes: A systematic review. IEEE journal of biomedical and health informatics (2021) https://doi. org/10.1109/jbhi.2020.3040225
18. Fregoso-Aparicio, L., Noguez, J., Montesinos, L., Garc´ıa-Garc´ıa, J.A.: Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. Diabetology Metabolic Syndrome (2021) https://doi.org/10.1186/ s13098-021-00767-9
19. Afsaneh, E., Sharifdini, A., Ghazzaghi, H., Ghobadi, M.Z., Afsaneh, E., Sharifdini, A., Ghazzaghi, H., Ghobadi, M.Z.: Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: a comprehensive review. Diabetology Metabolic Syndrome (2022) https://doi. org/10.1186/s13098-022-00969-9
20. Liu, K., Li, L., Ma, Y., Jiang, J., Liu, Z., Ye, Z., Liu, S., Pu, C., Chen, C., Wan, Y.: Machine learning models for blood glucose level prediction in patients with diabetes mellitus: Systematic review and network meta-analysis. JMIR Medical Informatics (2023) https://doi.org/10.2196/47833
21. Fujihara, K., Sone, H.: Machine learning approach to drug treatment strategy for diabetes care. Diabetes Metabolism Journal (2023) https://doi.org/10.4093/ dmj.2022.0349
22. Krishnaswami, V., Muruganantham, S., Raja, J., Arthanari, S., Selvaraj, B.R., Dharmian, J.P.: Artificial intelligence in pharmaceutical drug developmentchallenges and the way forward. Current Artificial Intelligence (2025) https: //doi.org/10.2174/0129503752359512250325061847
23. Kim, S.J., Clark, V., Hancock, J.T., Rawassizadeh, R., Liu, H., Taylor, E.A., Sheppard, V.B.: Leveraging artificial intelligence-mediated communication for cancer prevention and control and drug addiction: A systematic review. Translational Behavioral Medicine (2025) https://doi.org/10.1093/tbm/ibaf007
24. Guo, Q., Fu, B., Tian, Y., Xu, S., Meng, X.: Recent progress in artificial intelligence and machine learning for novel diabetes mellitus medications development. Current Medical Research and Opinion (2024) https://doi.org/10.1080/03007995.2024.2387187
25. Miotto, R., Wang, F., Wang, S., Jiang, X., Dudley, J.: Deep learning for healthcare: review, opportunities and challenges. Briefings Bioinform. (2018) https: //doi.org/10.1093/bib/bbx044
26. DiStefano, J., Watanabe, R.: Pharmacogenetics of anti-diabetes drugs. Pharmaceuticals (2010) https://doi.org/10.3390/ph3082610

27. Dost´alek, M., Akhlaghi, F., Puzanovova, M.: Effect of diabetes mellitus on pharmacokinetic and pharmacodynamic properties of drugs. Clinical Pharmacokinetics (2012) https://doi.org/10.1007/bf03261926
28. Carpenter, D., Zucker, E., Avorn, J.: Drug-review deadlines and safety problems. The New England journal of medicine (2008) https://doi.org/10.1056/nejmsa0706341