Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Natural Language Processing and Deep Learning Techniques for Fake Medical News Detection in Digital Healthcare Platforms

Yogesh Dinkar Jadhav[1], Dr. Geetika M. Patel[2], Dr. Prakash Deep[3], Dr. Rakhi Ludam[4], Nainavarapu Radha[5], Leena Deshpande[6], Veerendra Yadav[7], Arivukkodi R[8]

[1]Department of Mechanical Engineering, Sinhgad College of Engineering, Savitribai Phule Pune University Pune, Maharashtra, India, Email: sittpo3@gmail.com, Orcid : 0009-0004-5268-813
[2]Associate Professor, Department of Community Medicine, Parul University, PO Limda, Tal. Waghodia, District Vadodara, Gujarat, India, Email : vicepresident_86@paruluniversity.ac.in , Orcid Id- 0000-0003-3789-184X
[3]professor , MSOPS, Maharishi University of Information Technology, Lucknow, Uttar Pradesh, India, Email: pdpharma@gmail.com, Orcid: https://orcid.org/orcid-search/search?searchQuery=Prakash%20deep
[4]Professor, Department of Respiratory Medicine, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India, Email: rakhiludam@soa.ac.in, Orcid: 0009-0002-7800-6018
[5]Associate Professor, Department of ECE, Aditya University, Surampalem, Andhra Pradesh, 533437, Email: radha.nainavarapu@adityauniversity.in, Orchid id: 0000-0002-5526-1633
[6]Associate Professor, Department of Computer Engineering - Software Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra, 411037 Email: leena.deshpande@vit.edu
[7]Department of Computer Science & Engineering,Noida international University, Greater Noida, Uttar Pradesh 203201, India, Email: veerendra.yadav@niu.edu.in
[8]Computer Science, Assistant Professor, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai, Tamil Nadu, India, Email: arivukodir@maher.ac.in

Abstract

The proliferation of popular digital information and communication platforms, and online information sharing of medical information, have greatly contributed to the dissemination of fake medical information, posing threats to public health, clinical decision making and medical awareness. The linguistic complexity, semantic ambiguity, and the vast amount of medical-related text data created in digital environments have made it difficult to detect misleading medical content. This research presents a Natural Language Processing (NLP) and deep learning-based approach for the detection of fake medical news in digital health care systems. The proposed framework comprises three key stages: text preprocessing, feature extraction, and intelligent classification, which aim to effectively distinguish fake and genuine medical news articles with high accuracy. Experimental evaluation and performance analysis is done using a fake news dataset for healthcare topics that consists of fake and factual medical information. Various deep learning models like Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM), and Bidirectional Encoder Representations from Transformers (BERT) are applied and compared. The experimental results show that the transformer-based models achieve good classification results, demonstrating the effectiveness of transformer-based models in reliable verification of digital healthcare content and detection of misinformation in healthcare.

Keywords: Fake Medical News, NLP, Deep Learning, Healthcare Informatics, BERT, Text Classification

1. Introduction

Digital technologies have revolutionized the way that people communicate and receive information in the healthcare environment. The digital healthcare platforms, such as health websites, social media platforms, telemedicine systems, health blogs and mobile health apps, have emerged as significant sources of medical information for the general public (Allcott & Gentzkow, 2017; Shah et al., 2019). These are the platforms where immediate access to healthcare guidance, disease awareness, treatment and preventive healthcare practices are provided. As digital healthcare ecosystems become more prevalent, access to healthcare knowledge has been enriched and healthcare communications have become more efficient throughout the world. But as healthcare information is so easily available online, it has also come with a flood of fake medical information and misleading content about healthcare in digital spaces (Shu et al., 2017; Ruchansky et al., 2017).

Miscommunication in healthcare has become a significant problem worldwide as false or manipulated information may have detrimental effects on patient behavior, healthcare choices, drug intake, vaccine uptake, and overall public health initiatives (Shah et al., 2019). The swift dissemination of fake medical information via social media and online healthcare forums is often attributed to sensationalist content, emotional rhetoric, and scant verification processes (Allcott & Gentzkow, 2017). Fake medical news detection is a difficult task due to complex medical terminology, semantic ambiguity, context changes and constantly changing information patterns in the healthcare related text data. The traditional rule-based and manual content verification methods are not effective in processing vast amounts of real-time health care data in digital environments (Shu et al., 2017; Kowsari et al., 2019).

With the recent surge in automated misinformation detection systems, especially in the healthcare sector, and intelligent content moderation, it is essential to understand the potential of Natural Language Processing (NLP) and Artificial Intelligence (AI) for these areas. Efficient text specific preprocessing, semantic analysis, feature extraction, and medical content understanding can be achieved using NLP, while deep learning models have the ability to classify fake and genuine healthcare news articles (Kowsari et al., 2019). Recent architectures like Convolutional Neural Networks (CNN) (Graves, 2013); Long Short-Term Memory (LSTM); Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) have been introduced with their performance being promising for complex text classification applications.

The main goal of this research is to create an NLP and deep learning based approach that will detect fake medical news in digital health platforms. The research focus of the study is on issues concerning contextual understanding, analysis of misinformation in healthcare-related context and intelligent automatic detection mechanisms. The main highlight of this paper is the incorporation of the advanced NLP preprocessing techniques, multiple deep learning models comparative evaluation, extensive performance evaluation, and an efficient framework for the reliable detection of misinformation in healthcare. The rest of this paper will consist of literature review, methodology, experimental analysis, results discussion and conclusion.

2. Literature Review

Digital healthcare communication platforms have been growing in pace, greatly accelerating the delivery of healthcare information via online news sources, social media, blogs, and medical discussion forums. Fake medical news and healthcare misinformation has become a significant problem for the public health agenda, in addition to being a challenge for medical information (Allcott & Gentzkow, 2017; Shu et al., 2017). Most current misinformation detection systems work on the basis of traditional machine learning techniques, content verification techniques, the extraction of linguistic features and statistical text analysis methods. Supervised classification methods like Support Vector Machines (SVM), Naïve Bayes, Decision Trees and Random Forest have been used in several studies to detect fake news (Ruchansky et al., 2017). Identifying medical misinformation is more challenging than identifying fake news, however, due to a number of reasons: Medical information may contain the terms and expressions of a specific domain, the meaning of medical concepts is dependent upon the context, the domain knowledge in healthcare is constantly evolving; etc. In the medical field, misinformation detection is also crucial, due to the potential high-risk consequences of misinformation, as mentioned in (Karimi et al., 2015; Shah et al., 2019).

NLP methods are pivotal for deriving valuable insights from vast amounts of healthcare data in text format. The traditional methods of representation of texts, like the Term Frequency-Inverse Document Frequency (TF-IDF), have proven to be effective in the identification of important keywords and the calculation of weighted feature vectors for classification problems. TF-IDF is efficient for statistical representation of text, unable to give the context's semantics information and word relation analysis. However, to address these challenges, word embedding techniques like Word2Vec, GloVe, and FastText have been developed to capture the semantic similarities and relationship of medical terms (Mikolov et al., 2013; Pennington et al., 2014). In recent years, transformer-based models have shown immense promise for healthcare text analysis, likely due to their ability to capture the context and the meaning in the complex textual information. In the healthcare domain, transformer models like BERT (Devlin et al., 2019) and domain-specific ClinicalBERT models (Alsentzer et al., 2019; Huang et al., 2019; Lee et al., 2020) have demonstrated high accuracy in understanding healthcare-related questions and in classifying medical text.

The content of fake medical news is characterized by hierarchical textual features and contextual representations, which have led to a lot of interest in deep learning techniques for detecting fakes. Healthcare content is commonly analyzed using Convolutional Neural Networks (CNN) to detect significant text patterns and other features. Sequential text analysis and contextual information learning can be done well by Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models (Graves, 2013). Bidirectional Long Short-Term Memory (BiLSTM) models are another advancement that enable processing of textual information in both forward and backward directions, thereby enhancing the contextual understanding. In recent years, the transformer-based architectures like BERT and ClinicalBERT have significantly improved the detection of misinformation in the healthcare domain by enabling the deep representation of the context and context-specific semantic understanding of complex healthcare narratives (Devlin et al., 2019; Huang et al., 2019; Kalyan et al., 2022).

While there has been some progress towards detecting healthcare misinformation, there are still several challenges to be addressed. Previously published research tends to have smaller healthcare-specific fake news dataset, which hinders the robustness and generalizability of the detection models. Many of the old NLP techniques are lacking in semantic contextual understanding, which is essential for proper definition of 'misleading' medical text. Additionally, there are a number of models that have been developed but have less generalizability to other digital healthcare platforms because of the variability of domains and linguistic diversity. In addition, there is a significant concern over high false positive rate as the misclassification of genuine healthcare information as fake may have adverse effects on healthcare communication and public trust (Shu et al., 2017; Ruchansky et al., 2017). The constraints underscore the necessity of sophisticated NLP and deep learning models that can offer accurate, scalable, and context-sensitive identification of fake medical news in the digital healthcare landscape.

3. Proposed NLP and Deep Learning Framework

3.1 Overall System Architecture

The proposed fake medical news detection framework is designed as a sequential NLP and deep learning-based framework to classify healthcare-related news as fake or real. The architecture starts with the data extraction stage, whereby data is gathered from several healthcare news sources, such as news websites, medical blogs, health forums, and social media outlets as shown in Figure 1. These Medical news articles are then passed into the data collection module where textual data related to healthcare is collected and organized for further analysis.

Once data has been gathered, the text preprocessing phase filters out any irrelevant noise in the content that was collected. This stage involves text cleaning, removal of stop words, and lemmatization of the input text to provide a higher quality text. The cleaned text is then tokenized, splitting sentences into individual tokens to be used in computation. The numerical word vectors which are extracted from the embedding layer encode the semantic features from the medical text.

The extracted features then fed into the deep learning classifier, which can be a CNN, LSTM, BiLSTM, or BERT based model. Last, the classification layer classifies the input health care news article as either a fake or real article. Thus, the proposed architecture offers an automated and structured approach for detecting fake medical news in digital healthcare systems in a reliable manner.

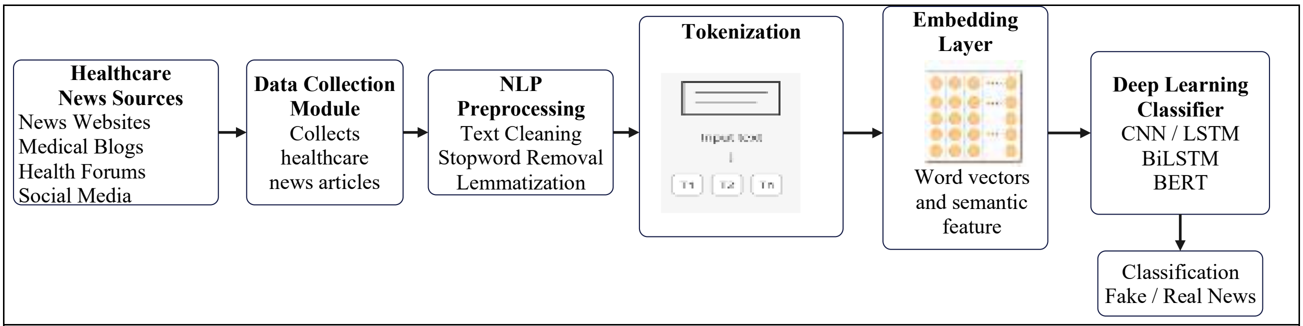

Fig. 1. Overall Architecture of the Proposed Fake Medical News Detection Framework

3.2 Text Preprocessing and NLP Pipeline

The effectiveness of fake medical news detection highly depends on the quality of textual data preprocessing and feature extraction. Proposed system: The selected healthcare news articles are first passed through several of the Natural Language Processing (NLP) operations such as semantic consistency, noise reduction, and creation of semantically meaningful text representation for deep learning based classification. First, stop words are removed, in order to remove commonly occurring words, like conjunctions, articles, auxiliary words and others which do not carry any significant meaning to the semantics. Elimination of words eliminates unnecessary dimensions and increases the speed of calculation.

After stop-word removal, lemmatization is done which changes words into their stem or normal form. This process helps to normalise the medical related textual information, e.g., to remove variations in medical terms and to increase the semantic consistency of the data set. Lemmatization is followed by tokenization, which involves breaking down the text into smaller linguistic tokens to facilitate efficient computational processing and analysis of the context of the text. Named Entity Recognition (NER) is further applied to this tokenized text to extract significant information related to the healthcare domain, like identification of diseases, medication, symptoms, and treatment names from medical news articles. NER helps to get better understanding of a domain-scoped text and boosts classification accuracy.

Medical terminology normalization is also used to ensure semantic uniformity by standardizing abbreviations, clinical terms, and expressions from healthcare used throughout various digital healthcare platforms. The normalization process eliminates ambiguity and makes the text's vocabulary consistent.

The proposed system uses feature weighting method as TF-IDF for feature extraction and vocabulary weighting. TF-IDF representation is based on the concept that considers the frequency of the terms in documents and the inverse frequency of the documents containing such terms to assign greater importance to those that appear rarely in the documents but occur frequently within these documents. The TF-IDF weighting equation is written as shown in Equation (1).

$$TFIDF(t,d) = TF(t,d) \times \log\left(\frac{N}{DF(t)}\right), \qquad (1)$$

Equation (1): TF-IDF Feature Weighting Equation

In Equation (1), $TF(t,d)$ represents the frequency of term t in document (d), (N) denotes the total number of documents in the dataset, and $DF(t,d)$ indicates the number of documents containing the term t. The TF-IDF approach enhances the assessment of the importance of terms, the weighting of the vocabulary, and the representation of the features of a text, by giving the discriminative medical terms more weight, while giving less weight to the words that appear many times, but fortunately, they do not contain discriminative information. This representation enhances the capability of deep learning models to accurately distinguish fake and real healthcare news content.

3.3 Deep Learning Classification Model

It suggests a framework using deep learning with NLP that is used to classify fake medical news automatically on digital health platforms. Once the NLP preprocessing stage, the processed medical text is converted to dense numerical representations via an embedding layer. These are the embedding vectors that represent the semantic and contextual associations between medical words, and transform text into a continuous vector space that can then be analyzed with a deep-learning approach. As shown in Figure 2, the input medical sentence first goes through a tokenization process to break it down into separate words, after which each word is processed by embedding into vectors and capturing the semantics of words.

Then the embedding representations generated are passed into a deep learning architecture, which can be CNN, LSTM, BiLSTM or BERT-based models as per the implementation strategy. Here, a bidirectional LSTM is used as it is able to effectively model context information from both the past and the future. Within the textual content, the BiLSTM layer captures medical terms, symptoms, treatments, and the patterns in the context of healthcare, understanding the relationships between these components. The end-to-end processing capability allows the framework to better grasp the full semantic structure of healthcare news articles as compared to previous sequential learning models.

There is an additional BiLSTM feature extraction layer, followed by an attention mechanism, to further improve the feature learning. In the attention layer, the important medical words and health contextually meaningful entities are given more weight, enabling the model to focus on highly informative text patterns when it comes to the classification. Next, the weighted contextual features are passed to a fully connected dense layer, which is used to learn high level abstract features and transform the dimension of the features. Lastly, the classification layer classifies the healthcare news article as fake or real. The proposed overall architecture for the NLP classification model of deep learning is shown in Figure 2.

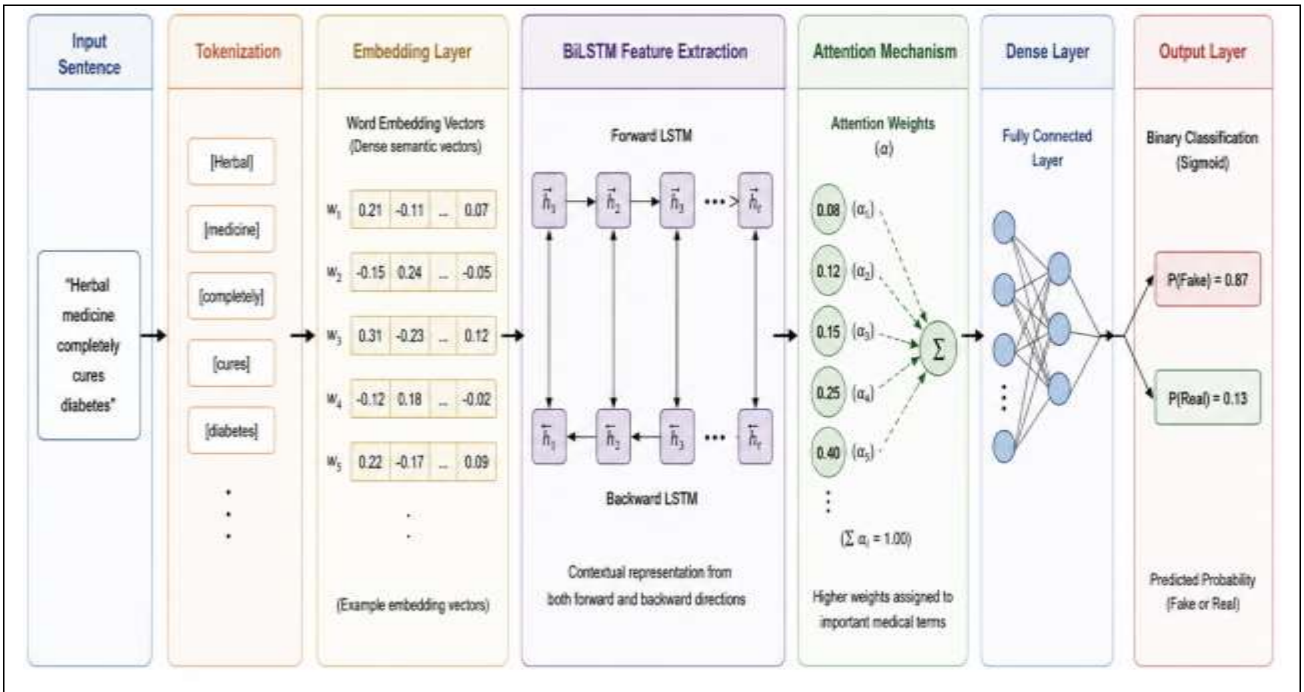

Fig. 2. Deep Learning-Based NLP Classification Model for Fake Medical News Detection

The proposed framework uses the Binary Cross-Entropy Loss Function for optimizing the model for classification during the training phase. The loss function evaluates the discrepancy between the predicted probabilities and the true class labels in order to make predictions as accurate as possible and minimize prediction errors. In equation (2) the Binary Cross-Entropy Loss Function is defined.

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)], \quad (2)$$

Equation (2): Binary Cross-Entropy Loss Function

In Equation (2), y_i represents the actual class label, $\hat{y}_i$ denotes the predicted probability generated by the classification model, and N indicates the total number of training samples. The Binary Cross-Entropy Loss Function is a very popular loss function in binary classification because it helps optimise the model, enhance the classification learning objective and reduce prediction errors from deep learning training. The proposed framework learns the discriminative semantic patterns and increases the accuracy of the fake medical news classification in healthcare websites by minimizing this loss function.

4. Experimental Setup and Dataset Description

4.1 Dataset Description

The proposed NLP and deep learning framework is tested through experiments with a fake news dataset on healthcare collected from publicly available digital healthcare platforms, online medical news portals, healthcare discussion forums, and social media platforms. The dataset includes fake and authentic healthcare-related news articles covering a range of topics such as medical treatments, disease information, vaccination awareness, herbal remedies, pharmaceutical claims, public health advisories, and healthcare recommendations. The process of collecting and pre-processing of text data from websites and blogs is very careful in order to improve the reliability of the classification results and to obtain a balanced number of fake and real medical news categories.

The dataset preparation process consists of duplicate removal, text normalization, token filtering, and verification of the medical terms to ensure semantic consistency of medical content. The articles in the healthcare section are divided into several domains that relate to healthcare including infectious diseases, chronic illness management, nutrition and wellness, pharmaceutical information, and preventive healthcare awareness. The proposed deep learning-based classification framework is evaluated through diverse healthcare categories to enhance the utility of the classification method in various medical information contexts and to demonstrate its utility in making stronger classifications.

In order to ensure consistency of classification and to ensure unbiased performance analysis, the data is split into a training set consisting of 80% of the data and a test set consisting of 20% of the data. The training dataset is used for optimizing the deep learning model and learning parameters, while the testing dataset is used to test the generalization ability and the prediction ability of the deep learning model on health-related news articles that have never been encountered before. The description of the data set and the class distribution of the proposed framework are shown in Table 1.

Table 1. Dataset Characteristics and Class Distribution

Dataset Parameter	**Value**
Total News Articles	12,500
Fake News Samples	6,200
Real News Samples	6,300
Training Split	80%
Testing Split	20%
Average Text Length	245 words

As seen in Table 1, the dataset is balanced with an equal number of fake and real healthcare news samples, helping to minimize classification bias when training the model. The average length of texts implies that there are messages of decent length and healthcare related facts and information in the texts for the semantic learning and extraction of textual features. Along with the reliability of its experimental foundation for the evaluation of performance in subsequent result analysis sections, the prepared dataset is used for the embedding generation, BiLSTM feature extraction, attention-based learning and binary classification stages mentioned in the previous section.

4.2 Training Configuration

High Performance Computing environment is setup for implementing the proposed deep learning-based classification framework in NLP and its experimental analysis is reliable and efficient. The training process is performed on a workstation that is equipped with the NVIDIA RTX T20 chipset (GPU generations 7, 11, and 21), an Intel i7 processor, and 32 GB of RAM to enable faster processing of deep learning calculations and embedding operations. The implementation is conducted by creating models, optimizing them, and testing their performance in deep learning libraries using Python programming language, such as TensorFlow and Keras.

The classification model is trained using the best efficient learning method – Adam optimizer, which can adjust itself according to the learned model and have a superior gradient based optimization ability. The initial learning rate will be fixed at 0.001 to ensure stability of training while also providing a reasonable speed of convergence during the process of parameter optimization. The batch size is set to 32, which is good for memory and gradient updating when training the model. This deep learning model is trained for 50 epochs, providing adequate learning of semantic and contextual healthcare patterns from the dataset.

The embedding layer is based on dense semantic vector representations with an embedding dimension of 100, which allow for a good understanding of the semantic context of textual information in the field of healthcare. These embedding vectors enable the model to learn semantic similarity, while enhancing the ability of the BiLSTM architecture to understand the context between medical terms and healthcare entities. The dropout regularization is also used in the deep learning model to minimize overfitting and to boost the model performance on unseen healthcare news samples.

The optimization of the embedding layer, BiLSTM feature extraction module, attention mechanism and binary classification layer explained in previous sections are supported efficiently with the selected training configuration. Moreover, these experimental conditions can serve as a sound computational base for the comparison of the experimental performance and evaluation classifications to be presented in the following results/discussions.

5. Results and Performance Analysis

5.1 Classification Performance Comparison

To test the feasibility of the proposed fake medical news detection framework, various classification models derived from machine learning and deep learning are tested for their performance in analyzing the effectiveness for detecting fake news and real news in the medical field. Along with the aforementioned traditional machine learning techniques and advanced deep learning architectures discussed above, support vector machine (SVM), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM) and Bidirectional Encoder Representations from Transformers (BERT) are also included in the comparative analysis. The evaluation is conducted by applying the commonly used classification metrics, including accuracy, precision, recall, F1-score and Area Under the Curve (AUC), which thoroughly covers the reliability of classification and contextual learning capability.

The experimental results show that the deep learning based methods are more successful than traditional machine learning methods due to their more powerful semantic feature extraction and understanding of healthcare related information in text content. Transformer-based models give better classification tasks because they are capable of modeling the long-range contextual dependency and semantic relationships among the content in medical news. A comparative performance analysis of the various machine learning and deep learning models is given in the following Table 2.

Table 2. Performance Comparison of Machine Learning and Deep Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC
SVM	88.6	87.9	88.1	88.0	0.90
CNN	92.8	92.1	91.7	91.9	0.94
LSTM	94.1	93.5	93.8	93.6	0.96
BiLSTM	96.2	95.8	96.0	95.9	0.98
BERT	97.5	97.1	97.3	97.2	0.99

As indicated in Table 2, the SVM model performs well at the basic level of classification but has its drawbacks in terms of learning the hidden meaning of text for complex healthcare-related information. CNN-based classification enhances the feature extraction performance by conducting local semantic pattern analysis, and the LSTM models delivered better sequential contextual understanding for healthcare news classification. The BiLSTM network is used to improve prediction capability by inputting the word's information in a forward and backward way to obtain better semantic representation and find less classification error.

The highest classification performance are model based on BERT which has an accuracy of 97.5%, precision of 97.1%, recall of 97.3%, F1-score of 97.2%, and AUC value of 0.99 among all evaluated model. The statistics show that, thanks to their strong semantic understanding and ability to learn contextual features, Transformer-based contextual language models are highly effective for fake medical news detection. As mentioned in the previous subsections, the performance of the embedding layer, attention mechanism and deep contextual learning architecture presented in the methodologies sections are also confirmed by these performance trends that have been observed in Table 2, and this provides a solid basis for the graphical performance analysis in the following subsections.

5.2 Accuracy and Loss Analysis

Training and validation accuracy and/or loss curves are used to analyze the learning behavior and the performance for optimization of the proposed deep learning-based fake medical news detection framework. Due to the fact that the model exhibits stable convergence behavior throughout the training process as shown in Figure 3, it implies that the model is able to learn the semantics properly and achieve the reliable capability of classification for healthcare-related textual information.

The training and validation accuracy curves for varying training epochs are shown in figure 3(a). The accuracy on the training dataset increases gradually from 70.2% in the first epoch to 98.1% in the last epoch and the accuracy on the validation dataset increases gradually from 68.5% in the first epoch to 97.4% in the last epoch. The proposed BiLSTM-attention-based architecture is well-trained and generalized, as evidenced by the high similarity between the accuracy of both the training and validation sets. Another smooth increments in the accuracy further validates the effectiveness of the embedding layer, the contexts feature extraction mechanism, and the attention based semantic learning described in the former sections.

Figure 3(b) shows the loss curves for training and validation while optimizing the model. The number of epochs increases and the training loss decreases to 0.04 and the validation loss decreases to 0.07. The continuous decreases in the loss values indicate that the optimization of the Binary Cross-Entropy Loss Function is effective and the gradient is converging during training. Moreover, the high similarity of the average values for training and validation losses suggests that the models are better able to predict the unseen healthcare news samples, which means that the models are stable.

The performance trends reported in Fig. 3 also confirm the classification results provided in Tab. 2 as well as the successful training of the selected configuration, such as the Adam optimizer, the learning rate, the embedding size and the BiLSTM-attention structure. The proposed framework's growing accuracy and fewer loss patterns validate it for learning discriminative semantic features from healthcare textual data and a reliable performance of fake medical news detection on digital healthcare platforms.

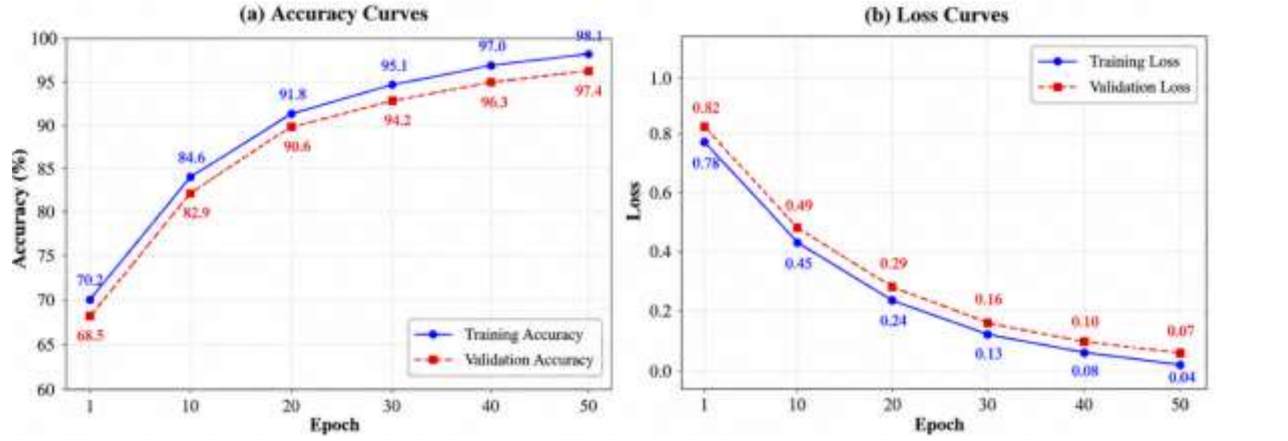

Fig. 3. Training and Validation Accuracy/Loss Curves of the Proposed Deep Learning Model

5.3 Comparative Performance Evaluation

Comparative performance evaluation of different machine learning and deep learning models is done to analyze their effectiveness to detect fake medical news in digital healthcare platforms. To assess the reliability of each model, its semantic-learning ability and contextual classification performance, four key classification metrics are used: accuracy, precision, recall, and F1-score. As shown in Figure 4, the proposed deep learning based techniques outperform traditional machine learning approaches in terms of the contextual understanding and extracting semantic feature.

The comparative performance analysis of the SVM, CNN, LSTM, BiLSTM and BERT models is carried out on the fake news data set on health-related topics discussed in the previous section, as shown in Figure 4. Compared to the other models, SVM has the lowest classification accuracy due to the lack of the ability to learn complex semantic relationships and contextual healthcare information with traditional machine learning approaches. CNN-based classification yields better results by extracting local features and learning semantic patterns and LSTM models are used to provide more significant understanding of the sequence for textual analysis in healthcare.

The BiLSTM model further enhances classification accuracy by reading the healthcare text bi-directionally, which gives it more depth to represent the text context and allow more semantic understanding about the medical news articles. The transformer models, particularly BERT, are the best-performing models on every evaluation metric: accuracy, precision, recall, and F1 score, among all the evaluated models. The remarkable performance shows how well the contextual language model, learned as a transformer, can be used for healthcare misinformation detection and semantic feature learning.

The advantage of the embedding layer, the BiLSTM feature extraction method, attention-based contextual learning, and the binary classification shown in the preceding methodology parts are validated by the progressive improvement that has been seen on the SVM to BERT in Figure 4. In addition, the comparative outcomes demonstrate that sophisticated deep learning architectures ensure better generalization ability, lower classification error, and better semantic understanding in the context of spotting fraudulent medical information in the digital healthcare area.

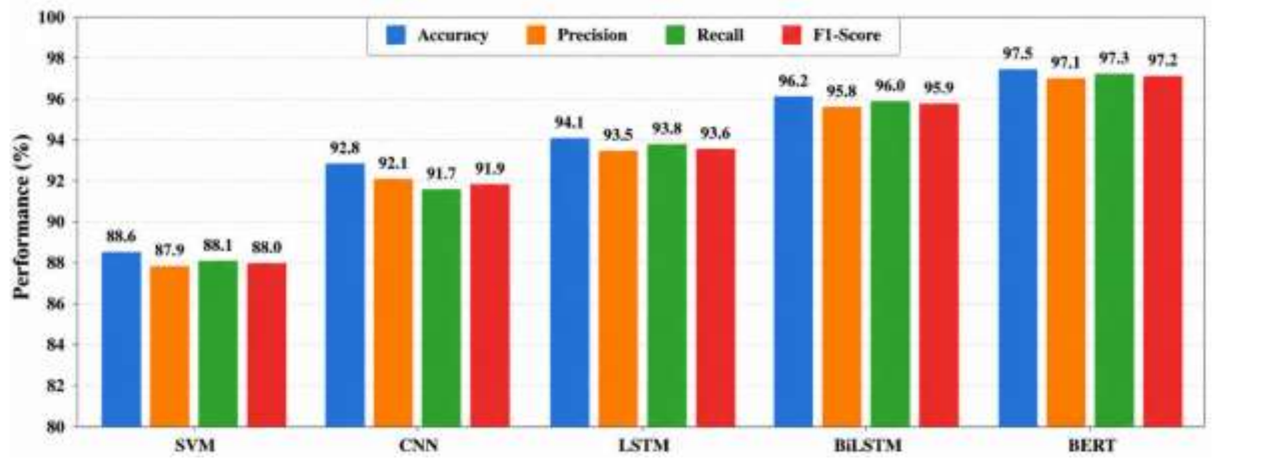

Fig. 4. Comparative Performance Analysis of Different NLP and Deep Learning Models

6. Discussion

Finally, the experimental findings on the proposed NLP and deep learning approach show that using sophisticated contextual language modeling techniques for fake medical news detection in digital healthcare platforms is effective. As discussed in previous sections, architectures of deep learning outperform traditional machine learning methods because they have high performance across two areas. First, they excel in semantic feature extraction from the text and second, in understanding the information within a text context that is relevant to healthcare. As the results of the four classification models (SVM, CNN, LSTM, BiLSTM, BERT) demonstrate, the contextual semantic learning method is essential for healthcare misinformation detection tasks, as the classification accuracy has improved with each of the models.

Transformer-based models like BERT perform best in terms of the classification accuracy for all the evaluation metrics (accuracy, precision, recall, F1 score, AUC values) among the tested models. The superiority of transformer models is mainly due to their bi-directional contextual learning and self-attention mechanisms, which enable deep semantic understandings of the healthcare-related textual data. Transformer-based models are able to learn contextual dependencies, semantic relationships, and long-range linguistic patterns in healthcare news articles, without requiring extensive manual feature engineering, unlike traditional machine learning models. In the field of medical misinformation detection, the ability to handle such text is especially crucial, as medical data is frequently loaded with specialized terminology and expressions, context-dependent ambiguity, and multiple layers of abstraction.

BiLSTM-attention's bidirectional processing of textual information also yields good classification performance. The integrated attention mechanism enhances semantic learning by boosting the relevance of specific terms related to healthcare, like diseases, treatments, symptoms, and pharmaceutical terms, among others. As shown in the accuracy/loss curves in Figure 3, training and validation accuracy and loss are stable in the training process, and the convergence and optimization ability are good, which verifies that the proposed method does not overfit the training process, and a discriminative semantic representation is learned from all the textual data in the healthcare field.

The findings also confirm high model robustness on various categories of health care text and structures. The slight difference between training and validation accuracy suggests that the model has good generalization ability for recognizing news articles in unseen healthcare fields. But, the problems of false positives and false negatives remain significant issues for misinformation detection algorithms in healthcare. Predictions can be wrong, leading to a misclassification of genuine healthcare information as fake, and influencing the reliability of healthcare communication and trust in recipients. Likewise, false negative forecasts can lead to the dissemination of detrimental healthcare-related misinformation on digital healthcare platforms, posing threats to healthcare awareness and patient safety.

On a practical level, the proposed framework shows great promise for implementation in real-time digital healthcare moderation systems, online medical news verification systems, healthcare social media monitoring systems, and automated misinformation filtering applications. Overall, this NLP preprocessing, generation of contextual embedding, feature extraction using BiLSTM, attention mechanisms, and transformer-based learning will make the misinformation detection system scalable and automated enough to handle larger-scale digital healthcare environments.

Although the study provided positive findings, it had some limitations. However, the experimental analysis relies on the use of fake news data sets related to healthcare and their availability and quality, which are still scarce compared to those available for other types of news. The outlined architecture is further centered around English-language healthcare text data and might want to be multilingualized for wider deployment to healthcare communication platforms all over the world. Furthermore, transformer models can be large and complex, making them less practical for deployment in resource-constrained settings where computational resources are limited. Potential limitations can be tackled in future research by integrating multi-lingual healthcare datasets, explainable AI techniques, lightweight transformer architectures, and federated learning privacy preserving framework for detecting misinformation in healthcare.

7. Conclusion and Future Work

The study proposed a framework for detecting fake medical news in digital healthcare platforms, which combines the capabilities of NLP and deep learning. The proposed framework combined the text preprocessing, embedding generation, contextual feature extraction, attention based semantic learning, and binary classification mechanisms to effectively detect fake and real healthcare news articles. Through experimental analysis, it was observed that deep learning architecture significantly outperforms the traditional machine learning techniques as the former has a better capacity for semantic understanding and finding out the context-based features. Transformer-based models like BERT outperformed other models evaluated in terms of accuracy, precision, recall, F1-score and AUC, establishing the success of contextual language modeling in healthcare misinformation detection.

The strength of this research is introducing an intelligent and scalable framework for identifying fake medical news, which integrates the NLP preprocessing methods, contextual learning through BiLSTM, attention mechanisms, and AI-based semantic representations for medical text analysis using a transformer architecture. The proposed framework is found to be suitable for capturing information about the semantic relationships while enhancing the classification accuracy of the digital environment of healthcare communication. In addition, embedding layers and contextual feature extraction with Binary Cross-Entropy optimization exhibited robust convergence and enhanced model generalization capability over a wide range of textual datasets from healthcare.

The results also highlight several advantages of deep learning techniques in healthcare misinformation detection. Advanced architectures are outperforming the traditional machine learning models in terms of semantic understanding, learning iteration in the context, automatic feature extraction and better robustness for classification. The attention-based semantic learning mechanism also helps in the recognition of more key entities, symptoms, treatments, and other pharmaceutical information related to healthcare from medical news articles. The capabilities mentioned above make deep learning models extremely applicable to healthcare misinformation monitoring systems based on automation and intelligent healthcare content moderation systems at a large scale.

The proposed framework gave promising results, however there are several directions for future research that would give future improvement. Enhancements in the future could also feature multilingual fake news detection, for healthcare communication in multiple languages and healthcare platforms globally. Deep learning predictions can also be enhanced by incorporating Explainable Artificial Intelligence (XAI) techniques to make the models more interpretable and transparent for healthcare professionals and end users. Furthermore, deployment optimisation in real-time can be explored for large scale healthcare social media monitoring and automated filtering misinformation applications. Future research could also involve federated learning-based architectures to support privacy-preserving collaborative healthcare misinformation detection in distributed healthcare settings, ensuring data confidentiality and security.

References

1. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236.
2. Alsentzer, E., Murphy, J., Boag, W., Weng, W. H., Jindi, D., Naumann, T., & McDermott, M. (2019, June). Publicly available clinical BERT embeddings. In Proceedings of the 2nd Clinical Natural Language Processing Workshop (pp. 72–78).
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171–4186).
4. Graves, A. (2013). Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.
5. Huang, K., Altosaar, J., & Ranganath, R. (2019). ClinicalBERT: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342.
6. Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2022). AMMU: A survey of transformer-based biomedical pretrained language models. Journal of Biomedical Informatics, 126, 103982.
7. Karimi, S., Metke-Jimenez, A., Kemp, M., & Wang, C. (2015). CADEC: A corpus of adverse drug event annotations. Journal of Biomedical Informatics, 55, 73–81.
8. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
9. Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). Text classification algorithms: A survey. Information, 10(4), 150.
10. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240.
11. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
12. Pennington, J., Socher, R., & Manning, C. D. (2014, October). GloVe: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) (pp. 1532–1543).
13. Ruchansky, N., Seo, S., & Liu, Y. (2017, November). CSI: A hybrid deep model for fake news detection. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (pp. 797–806).
14. Shah, Z., Surian, D., Dyda, A., Coiera, E., Mandl, K. D., & Dunn, A. G. (2019). Automatically appraising the credibility of vaccine-related web pages shared on social media: A Twitter surveillance study. Journal of Medical Internet Research, 21(11), e14007.
15. Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.