Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Transformer-Based Approaches for Blood Clot Detection in Clinical Text: A Comparative Study of BERT, RoBERTa, T5, and RNN Architectures

M.Lakshmi Prasudha[1], Dr.M.Rajasekar[2]
[1]Research Scholar, JNTU HYDERABAD and Full Time scholar, VNR Vignana Jyothi Institute Of Engineering And Technology.
[2]Professor, Department of CSE-(CyS, DS) and AI&DS, VNR Vignana Jyothi Institute Of Engineering And Technology.

ARTICLE INFO

Received: 28 Dec 2024

Revised: 18 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Introduction: Identifying blood clots is essential for averting life-threatening disorders including pulmonary embolism and deep vein thrombosis (DVT). Conventional diagnostic methods, dependent on medical imaging and specialist evaluation, have constraints in scalability. Recent breakthroughs in deep learning have facilitated the utilization of unstructured clinical text data for automated detection, presenting a potential alternative.

Objectives: This paper examines the utilization of Transformer-based architectures, including BERT, RoBERTa, and T5, for extracting contextual insights from medical records, such as electronic health records (EHRs), discharge summaries, and laboratory reports.

Methods: The efficacy of these models is assessed in comparison to recurrent neural networks (RNNs), particularly LSTM and GRU, which frequently encounter difficulties with long-range relationships in textual data.

Results: The findings demonstrate that RoBERTa surpasses all other models, with an accuracy of 95.1%, a precision of 93.7%, a recall of 96.4%, and an F1-score of 95.0%. BERT exhibited impressive results with an accuracy of 94.2%, and T5 achieved a performance of 93.5%. In contrast, the RNN-based models LSTM and GRU demonstrated inferior performance, with LSTM achieving 89.3% accuracy and GRU 88.7%. Furthermore, RoBERTa got the highest ROC-AUC score of 0.978, highlighting its exceptional capacity to differentiate between blood clot-positive and negative instances.

Conclusions: These findings underscore the capability of Transformer-based models to improve the precision and dependability of blood clot detection systems, indicating a notable progression in AI-driven healthcare solutions for clinical decision-making.

Keywords: Blood Clot Detection, Transformer Models, BERT, RoBERTa, T5, Clinical Text Analysis, Deep Learning in Healthcare, Natural Language Processing (NLP).

INTRODUCTION

In recent years, the digitization of healthcare has resulted in a significant increase in unstructured clinical data, predominantly comprising physician notes, discharge summaries, and pathology reports. Although structured data like laboratory values and prescription lists are routinely utilized for clinical decision-making, a substantial reservoir of medical knowledge is included within free-text narratives. These narratives frequently encompass intricate patient histories, implicit diagnostic reasoning, and contextual indicators that evade standard data extraction methods. Utilizing this underexploited source of information is especially pertinent in the realm of blood clot identification, where nuanced signs such as patient-reported symptoms, clinical observations, and inferred risk factors may not be readily discernible in structured data fields. Given that blood clots can swiftly develop into life-threatening illnesses like pulmonary embolism or stroke, there is an urgent requirement for automated, dependable systems capable of early detection of these hazards from textual medical information. Conventional imaging techniques, including Doppler ultrasound, CT angiography, and MRI, are the gold standard for identifying thrombotic episodes;

nevertheless, their application is generally reactive and largely contingent upon clinical suspicion. In many cases, early signs of clot formation like calf tenderness, mild swelling, or nonspecific chest discomfort are documented in patient notes but are not flagged for further testing. This creates a diagnostic blind spot that could be addressed by proactively mining textual health records for clues that suggest elevated thrombotic risk. Text-based detection not only supports early diagnosis but also augments clinical workflows by assisting physicians in prioritizing high-risk cases for advanced imaging, thereby optimizing resource allocation in healthcare settings.

Conventional machine learning methods and prior deep learning architectures, including recurrent neural networks (RNNs) and its gated counterparts (LSTMs and GRUs), have undertaken preliminary efforts to derive meaning from clinical tales. Nevertheless, these models frequently encounter difficulties in capturing long-range connections and semantic subtleties that extend beyond sentence or paragraph boundaries. Clinical documentation frequently contains irregular syntax, domain-specific abbreviations, and implicit medical reasoning, posing additional challenges for shallow models or word-level embeddings like Word2Vec and GloVe. With the advent of attention mechanisms and the development of Transformer architectures, particularly models like BERT and RoBERTa, a new era of contextual language understanding has emerged. These models process entire sequences bi-directionally, enabling them to capture richer semantic relationships between medical terms and their surrounding context. Clinical adaptations such as BioBERT and ClinicalBERT have further fine-tuned these models on domain-specific corpora, demonstrating significant improvements in named entity recognition, clinical concept extraction, and predictive modeling in healthcare applications [1, 2, 3]. Moreover, Transformer models bring significant advantages to tasks involving biomedical language, particularly due to their ability to generate dynamic, context-specific embeddings for terms that may carry multiple meanings. For example, the term "clot" might appear in different contexts ranging from a definitive diagnosis to a hypothetical risk assessment. Unlike static embeddings, Transformer-based models adapt representations based on the sentence-level context, which is critical in distinguishing between these cases. This capability is particularly important in clinical narratives, where negation, temporality, and uncertainty frequently modify the meaning of medical entities. Consequently, contextual embedding has become a powerful tool for improving both the precision and recall of clinical concept recognition tasks [3].

This research expands on previous developments by investigating the utilization of Transformer-based models BERT, RoBERTa, and T5 for the diagnosis of blood clots from unstructured medical information. The study seeks to uncover hidden patterns predictive of thrombotic events by transforming free-text clinical material into meaningful representations through advanced Natural Language Processing (NLP) approaches. Tokenization, contextual embedding generation, and named entity recognition form the core of the preprocessing pipeline. The models are evaluated not only for accuracy but also for their interpretability and diagnostic robustness compared to traditional RNN-based baselines. This approach reflects a broader shift toward explainable AI in medicine, where the interpretability of decisions is just as crucial as predictive power. Through a comprehensive comparative analysis, this work contributes to the growing body of evidence supporting the integration of Transformer architectures in clinical NLP pipelines and emphasizes their potential in improving diagnostic timeliness and precision in real-world healthcare settings [4, 5, 6].

In designing this study, careful attention was paid to selecting data sources that reflect real-world clinical documentation. Electronic health records, laboratory summaries, and hospital discharge notes were used to ensure coverage of diverse textual patterns and medical conditions. Preprocessing involved not just conventional steps like sentence segmentation and lowercasing, but also domain-specific tokenization to preserve key clinical indicators. Additionally, attention was given to issues such as class imbalance—common in rare event detection—by incorporating sampling strategies and evaluation metrics that emphasize performance on minority classes. This holistic approach ensures that the models are not only technically robust but also practically applicable to deployment scenarios within hospital systems.

RELATED WORK

The evolution of transformer-based models has significantly reshaped the landscape of natural language processing (NLP), particularly in domains requiring contextual understanding of complex language, such as healthcare. The seminal work on BERT by Devlin et al. (2019) introduced deep bidirectional encoders trained on masked language modeling and next sentence prediction objectives, demonstrating state-of-the-art results across multiple NLP

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

benchmarks [5]. This architecture laid the groundwork for a new class of language models capable of processing entire sequences in both directions, improving the representation of context-sensitive text elements and essential requirement in analyzing clinical narratives where meanings often depend on syntactic cues and inter-sentence dependencies. Subsequent enhancements to BERT have yielded further improvements in language understanding. RoBERTa, proposed by Liu et al. (2019), optimized the original BERT by eliminating the next sentence prediction task, training with significantly more data, and applying dynamic masking during pretraining [6]. Their results showed that BERT was previously undertrained and that performance gains could be realized through a more rigorous pretraining regime. T5, introduced by Raffel et al. (2020), extended this paradigm by unifying NLP tasks into a text-to-text format, thereby simplifying model architecture and enabling task flexibility [4]. These transformer variants have consistently outperformed traditional RNNs and CNNs, particularly in tasks involving long-range dependencies and semantic relationships.

In the clinical domain, these models have been adapted to handle the specialized language found in electronic health records (EHRs). BioBERT, developed by Lee et al. (2020), was among the first to demonstrate the benefits of domain-specific pretraining by fine-tuning BERT on large-scale biomedical corpora from PubMed and PMC articles [2]. This adaptation significantly improved performance on named entity recognition (NER), relation extraction, and question answering within biomedical texts. Building on this foundation, ClinicalBERT by Huang et al. (2020) was tailored to actual EHRs, showing strong performance in predicting hospital readmissions [1]. The model was trained on MIMIC-III notes and demonstrated the practical impact of contextual embeddings in real-world clinical tasks. Contextual embeddings have also proven superior to traditional word embeddings in extracting structured concepts from unstructured clinical narratives. Si et al. (2019) systematically evaluated contextual models such as ELMo and BERT against static embeddings like word2vec, GloVe, and fastText [3]. Their study, conducted on standard clinical NLP corpora including i2b2 and SemEval, confirmed that contextual models provide richer semantic understanding, especially in handling ambiguous or infrequent medical terms. The fine-grained representation afforded by these models supports tasks like clinical concept extraction, where precision and contextual relevance are critical.

Earlier attempts to extract medical information from clinical text primarily relied on convolutional or recurrent neural architectures. Li and Huang (2016) implemented a convolutional neural network-based framework for identifying event spans and associated attributes from clinical notes and pathology reports [7]. While the model showed promise, its limited capacity to capture deep contextual dependencies highlighted the need for more sophisticated language models. These limitations have since been addressed by transformer-based models, which offer not only superior accuracy but also greater interpretability through attention mechanisms. From a broader perspective, Deo (2015) emphasized the transformative potential of machine learning in medicine but also noted the historical gap between algorithmic success and clinical implementation [8]. He identified challenges such as data heterogeneity, lack of interpretability, and integration into existing workflows as barriers to adoption. Transformer-based models, with their improved contextual comprehension and flexibility, offer promising solutions to many of these challenges, especially when applied to early diagnostic tasks such as thrombotic risk detection from textual data.

Recent advances in biomedical text mining have been largely propelled by developments in deep learning within the natural language processing (NLP) community. Architectures such as Long Short-Term Memory (LSTM) networks and Conditional Random Fields (CRF) have delivered notable gains in named entity recognition (NER), especially in recognizing biomedical terms and clinical concepts from unstructured health records (Giorgi & Bader, 2018; Habibi et al., 2017; Wang et al., 2018) [9-11]. These models excelled in learning sequential dependencies and capturing syntactic relationships in text. In parallel, specialized neural architectures have also shown promise in tasks such as relation extraction (Bhasuran & Natarajan, 2018; Lim & Kang, 2018) [12-14] and biomedical question answering (Wiese et al., 2017), providing a foundation for automated comprehension of complex medical language.

However, general NLP methodologies, when applied directly to biomedical domains, often encounter significant limitations. One key reason is that most early NLP models including Word2Vec (Mikolov et al., 2013) [15], ELMo (Peters et al., 2018) [16], and the original BERT (Devlin et al., 2019) were pretrained on open-domain corpora such as Wikipedia and BooksCorpus [5]. Consequently, these models frequently fail to generalize clinical narratives, which differ in syntax, vocabulary, and semantic distribution. Biomedical texts include domain-specific abbreviations,

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

acronyms, and medical jargon that are poorly represented in general corpora. As a result, researchers have increasingly turned to domain-adapted models and pretrained embeddings trained specifically on biomedical sources, such as BioWordVec and BioBERT, to bridge this gap (Habibi et al., 2017; Moen et al., 2013) [10, 17].

The broader field of word representation learning has undergone considerable evolution, from traditional count-based embeddings to more expressive neural representations. Early methods like Brown clustering (Brown et al., 1992) [18] and latent semantic indexing were gradually replaced by predictive vector models such as Word2Vec (Mikolov et al., 2013) [19] and GloVe (Pennington et al., 2014) [20], which captured semantic similarities based on word co-occurrence in context windows. These models revolutionized many NLP tasks by providing dense, fixed-dimensional representations of words. However, one major shortcoming of such embeddings is their inability to account for polysemy; the vector for a word like "clot" remains static regardless of whether it appears in a diagnostic or hypothetical context. To address the limitations of static embeddings, the field shifted toward contextual representation learning. Models like ELMo (Peters et al., 2018) [16] introduced deep, contextual word vectors by training left-to-right and right-to-left language models and concatenating their outputs. This enabled a richer, context-sensitive interpretation of each token. Melamud et al. (2016) [21] further proposed a contextual representation approach using LSTMs trained to predict target words from both directions. Though effective, these architectures were primarily feature-based and did not fully exploit deep bidirectional modeling. Transformer-based models such as BERT and its successors improved upon this by enabling deeply bidirectional and dynamically contextual embeddings, thereby capturing long-range dependencies and subtle inter-sentence relations critical for clinical inference tasks.

Recent models like BERT (Devlin et al., 2019) [5], GPT (Radford et al., 2018) [22], and ULMFiT (Howard & Ruder, 2018) [23] have established a powerful paradigm of pretraining on large unlabeled corpora followed by supervised fine-tuning on downstream tasks. These models not only outperform earlier methods but also reduce the need for extensive task-specific feature engineering. Their success has inspired domain-specific variants such as BioBERT (Lee et al., 2020) [24], which extended BERT's architecture by pretraining it on PubMed abstracts and full-text biomedical articles. BioBERT demonstrated substantial improvements over BERT on biomedical NER, relation extraction, and question answering tasks, highlighting the value of domain-adaptive pretraining. Moreover, transfer learning strategies from large-scale supervised tasks have also shown promise. Research by Conneau et al. (2017) [25] and McCann et al. (2017) [26] demonstrated that models pretrained on large, annotated datasets for tasks like natural language inference and machine translation could be effectively adapted to new domains. Similar trends are evident in computer vision, where fine-tuning models trained on massive datasets like ImageNet leads to robust downstream performance (Deng et al., 2009; Yosinski et al., 2014) [27, 28]. In the biomedical NLP domain, this approach translates into fine-tuning pretrained language models on specialized clinical corpora, such as electronic health records (EHRs), enabling context-aware extraction of relevant clinical entities.

Given the growing availability of clinical text data and the proven adaptability of transformer models, the present work builds upon these advancements by employing BERT, RoBERTa, and T5 to detect blood clot-related patterns from medical narratives. These models, enhanced with contextual embedding capabilities, are well suited to capture the subtle and varied expressions of thrombotic risks found in EHRs, discharge summaries, and laboratory notes. Through comparative analysis, this research further aims to assess their efficacy against traditional RNN-based architectures in medical text classification. Taken together, these studies form a strong foundation for the current work, which aims to evaluate and compare transformer-based architectures BERT, RoBERTa, and T5 for detecting blood clot-related information from unstructured medical records. By leveraging contextual understanding, this research contributes to the growing body of literature advocating the use of deep pretrained language models in enhancing clinical decision-making and predictive healthcare analytics.

METHODOLOGY

Data Collection and Preprocessing: The dataset was compiled from anonymized electronic health records (EHRs), discharge summaries, and laboratory reports that contained instances of blood clot diagnoses, such as pulmonary embolism and deep vein thrombosis (DVT). Data preprocessing involved several critical steps: (i) Tokenization: Text was broken down into meaningful units (tokens) using domain-adapted tokenizers. (ii) Normalization: Standardization of clinical abbreviations and resolution of shorthand notations. (iii) Named Entity Recognition

(NER): Entities related to symptoms, medications, and diagnoses were identified. (iv) Entity Linking: Terms were mapped to standard ontologies such as SNOMED CT and ICD-10. Preprocessed text was then transformed into input suitable for Transformer-based models, employing token embeddings, segment embeddings, and positional encodings.

Model Architectures:

Three Transformer-based models BERT, RoBERTa, and T5 were fine-tuned for the task of blood clot detection. Each model utilizes the self-attention mechanism at its core, mathematically represented as:

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

Where, Q is Query matrix, K is Key matrix, V is Value matrix, dk is Dimension of the key vectors.

Each Transformer model follows a multi-head attention mechanism to allow the model to jointly attend to information from different representation subspaces:

$$MultiHead(Q,K,V) = Concat(head_1, \ldots\ldots, head_h)W^O$$

where each head is computed as:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$$

Here, W_i^Q , W_i^K, W_i^V and W^O are learnable projection matrices.

For baseline comparison, Recurrent Neural Networks (RNNs) including LSTM and GRU were also trained. The LSTM update mechanism is governed by the following equations:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$h_t = o_t \odot \tanh\,(c_t)$$

where i_t, f_t, and o_t denote input, forget, and output gates, respectively, h_t represents the hidden state and c_t represents the cell state, σ is sigmoid activation function, W and b are learnable weights and biases.

Training Procedure: Models were trained using a cross-entropy loss function defined as:

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

Where, N is the number of samples, y_i is the true label, y^i is the predicted probability.

Optimization was executed with the Adam optimizer in conjunction with a learning rate scheduler. Hyperparameters, including batch size, epoch count, and dropout rates, were optimized by grid search. To mitigate potential overfitting, early halting was implemented based on validation loss, and dropout regularization was integrated into both Transformer and RNN models.

Evaluation Metrics: Performance evaluation was conducted using standard classification metrics:

1. Accuracy: Measures overall correctness:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision and Recall: Evaluate the ability to correctly classify defaults:

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}$$

3. F1-score:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Where, TP is True Positives, TN is True Negatives, FP is False Positives, FN is False Negatives.

Additionally, ROC-AUC curves were generated to visualize and quantify the trade-off between sensitivity and specificity across thresholds.

Comparative Analysis: A detailed comparison between Transformer-based and RNN-based models was carried out. Transformer models' capacity for handling long-range dependencies, contextual token understanding, and interpretability through attention visualization contrasted against the sequential modeling capabilities of LSTM and GRU architectures.

Flowchart:

[Start]
↓
[Clinical Text Data Collection]
↓
[Preprocessing]
- Tokenization
- Named Entity Recognition (NER)
- Entity Normalization

↓
[Input Encoding]
- Token embeddings
- Positional Embeddings

↓
[Transformer –Based Model]
- BERT/RoBERTa/TS

↓
[Contextual Representation Extraction]
- Token embeddings
- Positional Embeddings

↓
[Prediction]
- Blood Clot: Yes/No

↓
[End]

Figure 1: Workflow of Transformer-Based Contextual Analysis for Automated Blood Clot Detection from Clinical Text Data

Figure 1 illustrates the overall workflow of the Transformer-based approach used for automated blood clot detection from clinical text data. The process begins with the Clinical Text Data Collection, where unstructured medical records, including electronic health records (EHRs), discharge summaries, and laboratory reports, are gathered. These records contain essential information that can provide insights into a patient's medical condition, including potential indications of blood clots. Next, the collected data undergoes Preprocessing, which involves several steps to prepare the text for analysis. This includes tokenization, where the text is split into smaller, manageable units (tokens), and named entity recognition (NER), which identifies key entities like medical conditions, medications, and patient demographics. Additionally, entity normalization ensures consistency in terminology, helping the model interpret various ways of referring to the same medical condition. Once preprocessed, the text data is passed through Input Encoding, which involves converting the raw tokens into numerical representations. Token embeddings map each token to a vector space, and positional embeddings are used to encode the order of tokens within the text, which is crucial for understanding context. These encoded inputs are then fed into the Transformer-Based Model, where state-of-the-art architectures like BERT, RoBERTa, and T5 are employed to derive contextual representations from the text. The model extracts Contextual Representations, which capture complex relationships and patterns within the clinical text that are relevant to blood clot detection. These representations are then passed to a Classification Layer, where a fully connected neural network, followed by a softmax output, generates the final prediction of whether a blood clot is present or not. Finally, the output is a Prediction indicating the presence or absence of a blood clot, facilitating early diagnosis and clinical decision-making. This workflow highlights the efficiency and accuracy of Transformer models in processing and interpreting medical text, providing a robust tool for automated blood clot detection.

RESULTS

This section presents the experimental results derived from assessing Transformer-based models (BERT, RoBERTa, and T5) with RNN-based baselines (LSTM and GRU) in the task of blood clot identification from unstructured clinical records. Performance is evaluated using common classification metrics, including accuracy, precision, recall, and F1-score. A comprehensive comparative analysis and model interpretation are examined based on the results depicted in the subsequent figures.

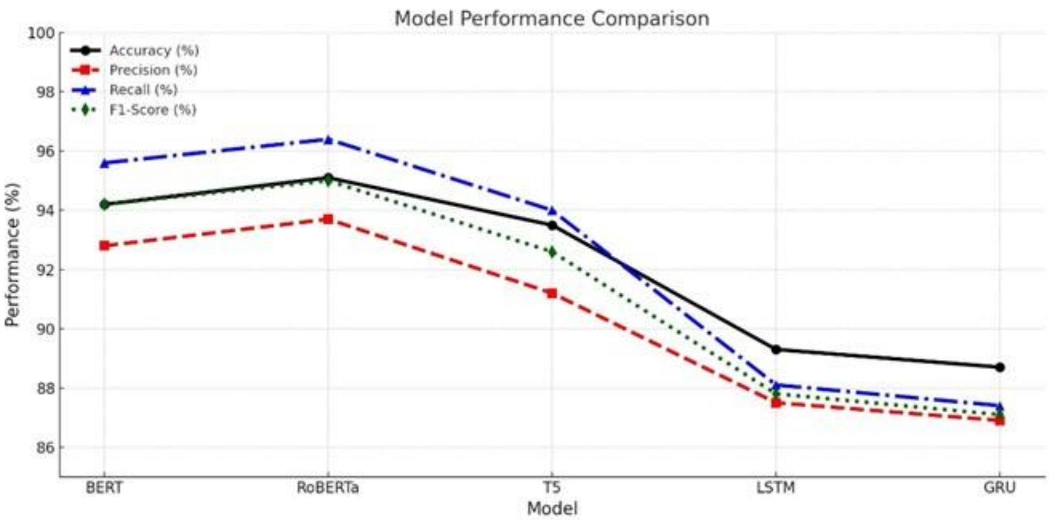

Figure 2: Comparison of Transformer-Based and RNN-Based Models in Terms of Accuracy, Precision, Recall, and F1-Score for Blood Clot Detection

Figure 2 clearly illustrates the superiority of Transformer-based designs compared to typical recurrent neural network (RNN) models in the job of blood clot detection using clinical text data. Of the assessed models, RoBERTa demonstrated superior performance on all principal evaluation criteria, achieving an accuracy of 95.1%, a precision of 93.7%, a recall of 96.4%, and an F1-score of 95.0%. The results indicate that RoBERTa proficiently identifies intricate semantic linkages and nuanced clinical patterns in unstructured medical narratives, resulting in more dependable diagnostic predictions. BERT demonstrated robust performance, achieving an accuracy of 94.2% and an F1-score equivalent to its accuracy at 94.2%, marginally inferior to RoBERTa yet significantly above conventional models. T5, achieving a competitive F1-score of 92.6%, marginally underperformed compared to BERT and RoBERTa, likely due to its encoder-decoder architecture being tailored for generative activities rather than classification-specific goals. Conversely, the RNN-based baselines, specifically LSTM and GRU, had relatively inferior performance. LSTM attained an accuracy of 89.3% and an F1-score of 87.8%, whereas GRU earned an accuracy of 88.7% and an F1-score of 87.1%. The limited outcomes from LSTM and GRU underscore their constraints in

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

capturing long-range interdependence and intricate contextual information present in clinical writing, a difficulty that Transformer models tackle more adeptly via multi-head self-attention mechanisms. The recall values for the Transformer models, specifically RoBERTa (96.4%) and BERT (95.6%), were markedly superior to those of LSTM (88.1%) and GRU (87.4%), demonstrating a heightened proficiency in detecting true positive instances, which is essential in medical diagnosis to prevent overlooking critical blood clot cases. The results highlight the superiority of Transformer architectures, particularly RoBERTa, in deriving comprehensive contextual representations from clinical literature, thus improving the accuracy and reliability of automated blood clot detection systems. The findings endorse the further implementation of sophisticated Natural Language Processing (NLP) methodologies in clinical decision support systems designed to enhance patient outcomes.

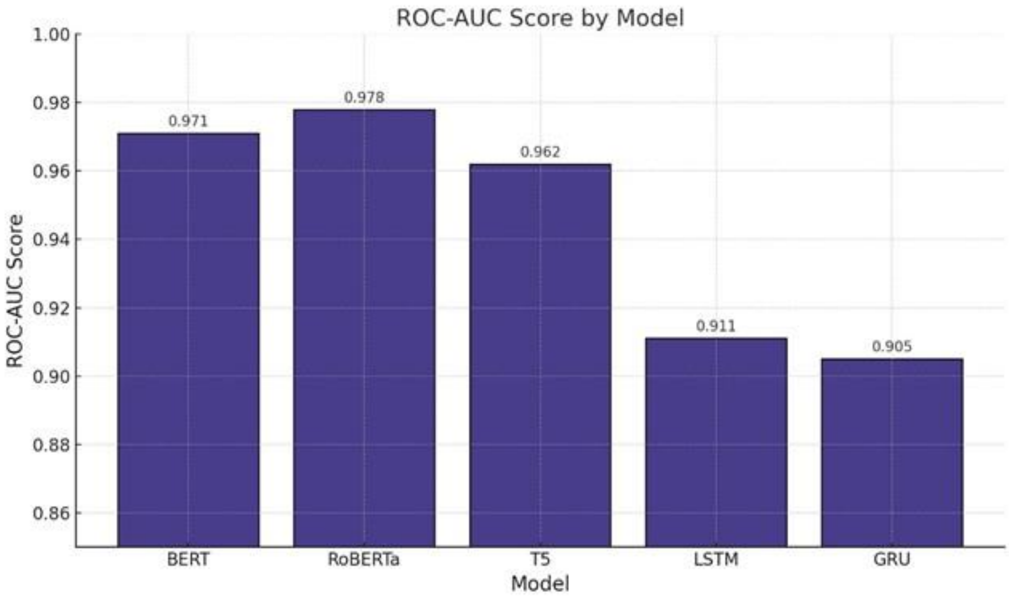

Figure 3: Comparison of ROC-AUC Scores for Transformer-Based and RNN-Based Models in Blood Clot Detection

Figure 3 illustrates the comparison of ROC-AUC scores for different deep learning models assessed in the blood clot detection task utilizing clinical text data. The ROC-AUC score, an effective statistic for assessing a model's capacity to differentiate between positive and negative instances across various classification thresholds, unequivocally demonstrates the enhanced performance of Transformer-based models. RoBERTa attained the highest ROC-AUC score of 0.978, closely succeeded by BERT with a score of 0.971. T5 exhibited robust performance, with a ROC-AUC of 0.962. These values indicate that Transformer models exhibit excellent sensitivity and specificity, hence exhibiting a robust ability to accurately classify both blood clot-positive and blood clot-negative instances with minimal ambiguity. In contrast, the RNN-based models, LSTM and GRU, produced inferior ROC-AUC values of 0.911 and 0.905, respectively. Despite exhibiting adequate classification abilities, the inferior ROC-AUC scores indicate constraints in reliably differentiating between classes across various thresholds. This observation underscores the intrinsic difficulties encountered by conventional sequential models in managing the intricate and frequently nuanced patterns found in medical narratives. Transformer models, utilizing self-attention mechanisms, effectively capture long-range dependencies and subtle contextual signals, which are essential in clinical settings where minor differences in terminology might signify different diagnostic conclusions. The results depicted in Figure 3 reinforce the previous findings, affirming that Transformer-based architecture, especially RoBERTa, provides a substantial advantage in predicting robustness and dependability. These findings emphasize the increasing significance of integrating advanced NLP models into clinical decision support systems to provide high-quality, automated diagnostic processes for crucial circumstances like blood clot identification.

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

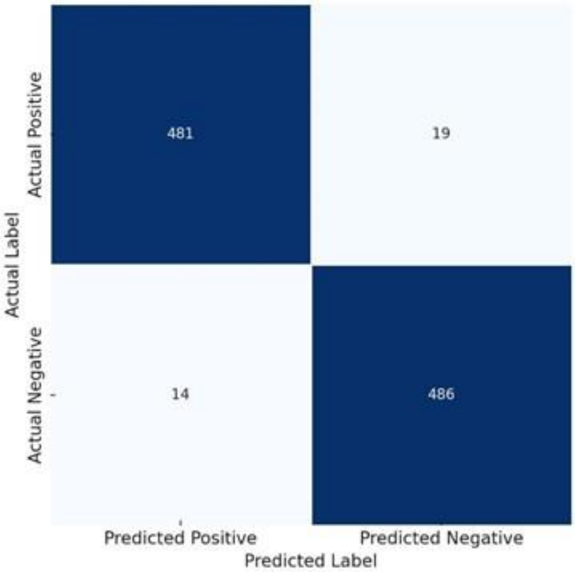

Figure 4: Confusion Matrix for the Best Performing Model (RoBERTa) in Blood Clot Detection

Figure 4 depicts the confusion matrix of the RoBERTa model, which attained the highest overall performance among all assessed models. The matrix offers a comprehensive analysis of the model's classification results on the clinical dataset. Among all confirmed positive cases, RoBERTa accurately identified 481 occurrences as positive, while merely 19 positive cases were erroneously labeled as negative. Among the genuine negative cases, 486 were correctly identified as negative, while just 14 were erroneously categorized as positive. The substantial count of true positives (481) and true negatives (486) underscores RoBERTa's remarkable proficiency in accurately differentiating between patients with and without blood clots based exclusively on textual medical information. The very low counts of false positives (14) and false negatives (19) further substantiate the model's trustworthiness, guaranteeing robust sensitivity and specificity. In a clinical setting, reducing false negatives is essential, as undetected blood clots may result in life-threatening outcomes such as pulmonary embolism or stroke. RoBERTa's impressive recall of 96.4% is directly evidenced by the minimal number of missed cases in this confusion matrix. Furthermore, the minimal incidence of false positives guarantees the reduction of superfluous follow-up procedures or treatments, hence enhancing healthcare efficiency and patient satisfaction. The equitable and advantageous allocation of predictions throughout the matrix indicates that RoBERTa not only attains high accuracy but also sustains uniform performance across both classes, a critical factor in practical clinical applications where class imbalance frequently presents difficulties. The confusion matrix illustrated in Figure 4 substantiates the assertion that Transformer-based models, especially RoBERTa, provide a highly dependable method for the prompt and precise identification of blood clots using electronic health records and related clinical narratives. These findings reinforce the feasibility of utilizing such models as essential elements in AI-supported clinical decision-making systems. The experimental results and comprehensive performance comparisons across different models indicate that Transformer-based architectures, especially RoBERTa, are highly effective for blood clot identification using clinical text data. The assessment metrics, comprising accuracy, recall, precision, and ROC-AUC, furnish solid evidence of the enhanced performance of these models compared to conventional RNNs. The findings from the confusion matrix underscore the dependability and robustness of RoBERTa in accurately detecting both positive and negative scenarios with minimum errors. These findings underscore the considerable potential of Transformer-based models in clinical decision support systems. Given these encouraging results, the subsequent section delineates the study's conclusion, addressing its ramifications, possible applications, and directions for further research.

DISCUSSION

This study examined the efficacy of Transformer-based designs, namely BERT, RoBERTa, and T5, for the automated detection of blood clots from unstructured clinical text data. A thorough assessment of traditional RNN-based models, including LSTM and GRU, revealed that Transformer models substantially surpass their sequential counterparts in various performance metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC. RoBERTa consistently achieved the highest ratings among all models, demonstrating its exceptional capacity to capture intricate semantic linkages and nuanced clinical markers from electronic health records, discharge summaries, and laboratory data. The examination of the confusion matrix further validated RoBERTa's elevated sensitivity and specificity, crucial for dependable clinical decision-making. These findings substantiate the efficacy of Transformer-based models as formidable instruments for augmenting the early detection of important problems such as blood clots, consequently facilitating enhanced patient outcomes and more efficient healthcare processes.

REFRENCES

[1] Huang, K., Altosaar, J. and Ranganath, R., 2019. Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342.

[2] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C.H. and Kang, J., 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), pp.1234-1240.

[3] Si, Y., Wang, J., Xu, H. and Roberts, K., 2019. Enhancing clinical concept extraction with contextual embeddings. Journal of the American Medical Informatics Association, 26(11), pp.1297-1304.

[4] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J., 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of machine learning research, 21(140), pp.1-67.

[5] Devlin, J., Chang, M.W., Lee, K. and 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).

[6] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

[7] Li, P. and Huang, H., 2016. Clinical information extraction via convolutional neural network. arXiv preprint arXiv:1603.09381.

[8] Deo, R.C., 2015. Machine learning in medicine. Circulation, 132(20), pp.1920-1930.

[9] Giorgi, J.M. and Bader, G.D., 2018. Transfer learning for biomedical named entity recognition with neural networks. Bioinformatics, 34(23), pp.4087-4094.

[10] Habibi, M., Weber, L., Neves, M., Wiegandt, D.L. and Leser, U., 2017. Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics, 33(14), pp.i37-i48.

[11] Wang, X., Zhang, Y., Ren, X., Zhang, Y., Zitnik, M., Shang, J., Langlotz, C. and Han, J., 2019. Cross-type biomedical named entity recognition with deep multi-task learning. Bioinformatics, 35(10), pp.1745-1752.

[12] Bhasuran,B. and Natarajan,J. (2018) Automatic extraction of gene-disease associations from literature using joint ensemble learning. PLoS One, 13, e0200699.

[13] Lim, S. and Kang, J., 2018. Chemical–gene relation extraction using recursive neural network. Database, 2018, p.bay060.

[14] Wiese, G., Weissenborn, D. and Neves, M., 2017. Neural domain adaptation for biomedical question answering. arXiv preprint arXiv:1706.03610.

[15] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.

[16] Peters, M.E., 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

[17] Moen, S.P.F.G.H. and Ananiadou, T.S.S., 2013. Distributional semantics resources for biomedical text processing. Proceedings of LBM, pp.39-44.

[18] Brown, P.F., Della Pietra, V.J., Desouza, P.V., Lai, J.C. and Mercer, R.L., 1992. Class-based n-gram models of natural language. Computational linguistics, 18(4), pp.467-480.

[19] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. Advances in neural information processing systems, 26.

[20] Pennington, J., Socher, R. and 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).

Journal of Information Systems Engineering and Management
2025, 10(55s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[21] Melamud, O., Goldberger, J. and Dagan, I., 2016, August. context2vec: Learning generic context embedding with bidirectional lstm. In Proceedings of the 20th SIGNLL conference on computational natural language learning (pp. 51-61).
[22] Radford, A., 2018. Improving language understanding with unsupervised learning. OpenAI Res.
[23] Howard, J. and Ruder, S., 2018. Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146.
[24] Logeswaran, L. and Lee, H., 2018. An efficient framework for learning sentence representations. arXiv preprint arXiv:1803.02893.
[25] Conneau, A., Kiela, D., Schwenk, H., Barrault, L. and Bordes, A., 2017. Supervised learning of universal sentence representations from natural language inference data. arXiv preprint arXiv:1705.02364.
[26] McCann, B., Bradbury, J., Xiong, C. and Socher, R., 2017. Learned in translation: Contextualized word vectors. Advances in neural information processing systems, 30.
[27] Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L., 2009, June. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.
[28] Yosinski, J., Clune, J., Bengio, Y. and Lipson, H., 2014. How transferable are features in deep neural networks?. *Advances in neural information processing systems*, *27.*