SCMPCR
South Asia Centre for Medical Physics and Cancer Research
SCMPCR Newsletter
A sister organization of Alo-BT
July 2025 / Volume 7 / Issue 2
QUALITY EDUCATION AND HEALTH SCIENCE FOR PATIENT BENEFIT

Sequential Sentence Classification in Structured Medical Abstracts: A Performance Comparison of Simple RNN and LSTM

Bushra Intakhab[1] Savera Camran[2]*
[1]Department of Physics, Florida Atlantic University, Boca Raton, FL 33431-0991, USA
[2]Department of Physics, Government Degree College Murad Memon, Karachi, Pakistan
Corresponding Author: bushraintakhab@gmail.com

Abstract

This study investigates the effectiveness of deep learning models SimpleRNN and Long Short-Term Memory (LSTM) for sequential sentence classification in structured medical abstracts. The task involves identifying the functional role of each sentence to facilitate evidence extraction in medical literature. Both models were trained and evaluated on a labeled dataset derived from scientific medical abstracts. Performance metrics indicate that the LSTM model significantly outperforms SimpleRNN, achieving a test accuracy of 75.36% and a macro-averaged F1-score of 0.6867, compared to 35.88% accuracy and 0.1720 F1-score for SimpleRNN. The LSTM model also demonstrated more balanced and accurate predictions across all sentence roles. These findings highlight the importance of using advanced recurrent architectures like LSTM for natural language processing tasks in the medical domain, supporting improved literature mining and automated knowledge extraction.

Introduction

With growing scientific data, the ability of machines to understand and process natural language text has become increasingly valuable especially in domains like medical physics, where practitioners rely heavily on the timely interpretation of scientific literature, clinical trial data, and treatment guidelines. One important task in this domain is sequential sentence classification, where each sentence in a structured document (e.g., a scientific abstract, clinical protocol, or radiotherapy guideline) is classified based on its role in the overall structure such as background, methods, results, or conclusions. In medical physics such automated classification can streamline literature review, identify relevant outcome data, and improve decision making by extracting evidence-based content more efficiently.

Recurrent Neural Networks (RNNs) have traditionally been used for sequence modeling due to their ability to retain context over sequences. However, Long Short-Term Memory (LSTM) networks, an advanced variant of RNNs, have demonstrated superior performance in tasks involving long-range dependencies, making them

more suitable for interpreting structured medical texts.[1] Dernoncourt and Lee introduced the PubMed 200k and 20k RCT datasets, which provide structured biomedical abstracts labeled at the sentence level. Their work demonstrated that sequential models like RNNs and LSTMs significantly outperform non-sequential baselines in identifying sentence roles. This dataset has become a benchmark for evaluating sentence classification in clinical and scientific domains.[2] Mikolov et al. pioneered the use of Recurrent Neural Networks (RNNs) for language modeling, showing their effectiveness in capturing contextual dependencies. Their model marked a foundational shift toward deep learning in NLP tasks. However, the limitations of RNNs, particularly their struggle with long-term dependencies, led to the development of more advanced architectures like LSTM.[3]

This study explores and compares the effectiveness of RNN and LSTM architecture for the task of sequential sentence classification. By evaluating these models on labeled structured abstracts, we aim to determine which architecture better supports the extraction of scientific knowledge, with implications for automated literature mining.

Materials and Methods

The dataset used for this work consists of structured abstracts, where each sentence is labeled according to its role within the document such as background, methods, results, or conclusions. Sentences were first tokenized into individual words, transformed into numerical sequences using a tokenizer, and then padded to a consistent length to prepare them for model input. Two deep learning models were developed and compared: a Recurrent Neural Network (RNN) and a Long Short-Term Memory (LSTM) network. Both models were implemented using Python and Keras and were trained to classify the sentence roles based on word sequences. The training process involved splitting the dataset into training, validation, and test sets. Model performance was optimized using categorical cross-entropy loss and the Adam optimizer.

Model evaluation was conducted using standard classification metrics, including accuracy, F1-score, and a confusion matrix to assess performance across different sentence types. These methods enabled a comprehensive comparison of the two architectures in their ability to understand and structure scientific abstracts.

Results

Two deep learning models SimpleRNN and Long Short-Term Memory (LSTM) were implemented and evaluated for the task of sequential sentence classification on structured medical abstracts. The models were trained and tested using a labeled dataset where each sentence was assigned a role such as Background, Objective, Methods, Results, or Conclusions.

The SimpleRNN model achieved a test accuracy of 35.88% and a macro-averaged F1-score of 0.1720. In contrast, the LSTM model significantly outperformed SimpleRNN, achieving a test accuracy of 75.36% and a macro-average F1-score of 0.6867.

Metric	**SimpleRNN**	**LSTM**
Accuracy	35.88%	**75.36%**
F1-Score (Macro)	0.1720	**0.6867**

Table 1. Performance comparison between SimpleRNN and LSTM models on the test dataset.

The classification report also showed notable differences in class-wise performance:

Sentence Role	SimpleRNN F1	LSTM F1
BACKGROUND	0.00	0.51
OBJECTIVE	0.03	0.59
CONCLUSIONS	0.01	0.67
RESULTS	0.27	0.81
METHODS	0.55	0.85

Table 2:Class-wise F1-score comparison between SimpleRNN and LSTM models across different sentence roles in structured medical abstracts.

The confusion matrices for both models further illustrate their classification behavior (Figure 1). The SimpleRNN model exhibits significant off-diagonal errors, particularly for the Background, Objective, and Conclusions classes, which are often misclassified as Methods or Results. In contrast, the LSTM model produces a confusion matrix with strong diagonal elements, indicating a higher degree of alignment between predicted and actual sentence roles.

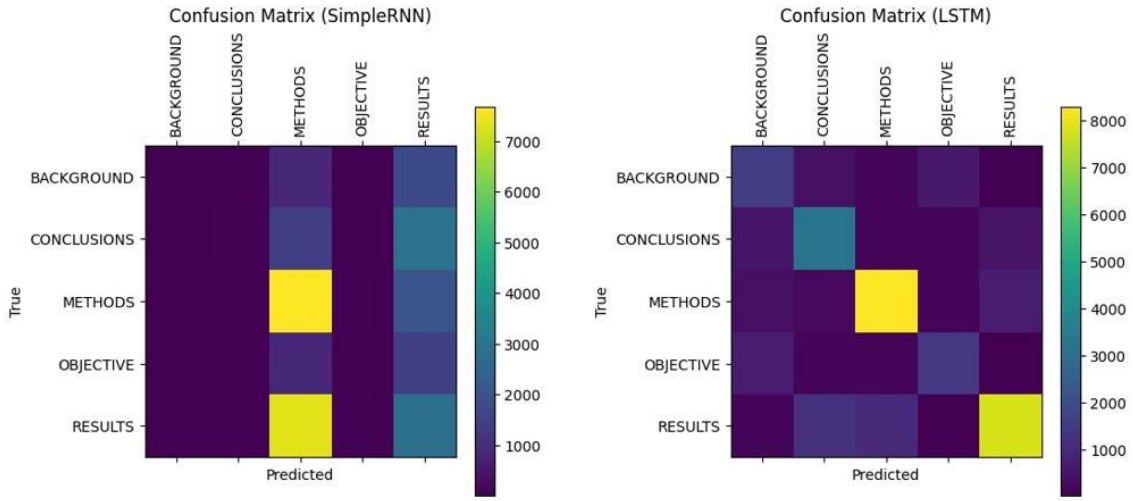

Figure 1:Confusion matrices for SimpleRNN (left) and LSTM (right) showing model predictions across five sentence roles in structured medical abstracts. LSTM exhibits stronger diagonal dominance, indicating more accurate classification.

Discussion

The results clearly demonstrate that the LSTM model is more effective than SimpleRNN for classifying sentence roles in structured medical abstracts. This can be attributed to LSTM's architectural advantage in capturing long-range dependencies and retaining contextual information across sequences a critical requirement in the interpretation of medical literature.

Simple RNN's low performance, particularly in identifying BACKGROUND, CONCLUSIONS, and OBJECTIVE sentences, suggests that it struggles to distinguish among sentence roles when context from surrounding sentences is required. LSTM mitigates this limitation through its memory cells and gating mechanisms, which allow it to model complex patterns over longer textual sequences.

The substantial improvements observed in macro F1-score and overall accuracy suggest that LSTM not only classifies dominant classes like METHODS and RESULTS well but also maintains robustness in handling minority classes. This is particularly important in medical domains, where extracting conclusions or objectives from text can directly impact clinical decision-making and literature synthesis.

Overall, the findings confirm that for tasks involving sequential sentence classification in structured scientific texts, more advanced recurrent architectures like LSTM should be preferred over simpler RNNs. These insights contribute to ongoing efforts in automating evidence extraction and enhancing information retrieval in medical physics and related fields.

Conclusion

Overall, the LSTM model demonstrates a robust capacity for sequential sentence classification in medical physics literature, which can facilitate automated extraction of evidence-based content and support efficient literature reviews. These findings highlight the importance of choosing advanced sequential models for downstream applications in clinical natural language processing.

References:

1. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780
2. Dernoncourt, F., & Lee, J. Y. (2017). PubMed 200k RCT: a dataset for sequential sentence classification in medical abstracts. https://arxiv.org/abs/1710.06071
3. Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. (2010). Recurrent neural network-based language model. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH)* (pp. 1045–1048).