International Journal of Applied Operational Research
Vol. 12, No. 3, pp. 25-34, Summer 2024

Journal homepage: ijorlu.liau.ac.ir

Design and development of a customized language model for medicine: PerMed

N. Kheirkhah, M. Faridi Masoule *, A. Bagheri

Received: 28 February 2024; **Accepted:** 25 May 2024

Abstract This paper introduces PerMed, an enhanced version of the large medical language model Meditron, which provides support for the Persian language. Meditron, as a specialized open-source model in the medical field, has demonstrated remarkable performance in English language processing and medical text analysis. However, the lack of Persian language support has been identified as a key limitation of this model, hindering access to accurate medical information for Persian-speaking professionals and patients. In this research, using translation tools and developed scripts, the PerMed model was designed with the capability to translate medical texts from English to Persian, answer questions in Persian, and facilitate full interaction in Persian. PerMed's performance was evaluated on a set of medical tasks, including question answering, text analysis, and translation. The results showed that this model has significantly increased the accuracy and quality of its services in Persian. By expanding access to medical knowledge for Persian speakers, PerMed represents an effective step towards developing multilingual medical models and has high potential for use in multilingual medical environments.

Keyword: Medical Language Model, Persian Language Processing, Medical Text Translation, Artificial Intelligence in Medicine.

1 Introduction

Large language models (LLMs) [1] have recently revolutionized natural language processing (NLP) [2] and its various applications. In the medical field, these models play a crucial role in analysing medical data, answering clinical questions, and providing accurate information to healthcare professionals. One prominent model in this area is Meditron, an open-source large language model designed for English language processing in the medical domain, exhibiting high capability in analyzing specialized texts and providing accurate responses. However, one of Meditron's main limitations is its lack of support for non-English languages, particularly Persian. This has limited access to the model's advanced capabilities for Persian-speaking

* **Corresponding Author. (✉)**
E-mail: m.faridi@ahrar.ac.ir (M. Faridi Masoule)

N. Kheirkhah
Department of Computer Engineering, Ahrar Institute of Technology and Higher Education, Rasht, Iran

M. Faridi Masoule
Department of Computer and Information Technology, Ahrar Institute of Technology and Higher Education, Rasht, Iran

A. Bagheri
Department of Dynamic, Control, and Vibration, Faculty of Mechanical Engineering, University of Guilan, Rasht, Iran

[DOI: 10.71885/ijorlu-2024-1-685]

users, while Persian, as one of the important languages in the medical field, especially in Persian-speaking regions, requires special attention.

1.1 Research contribution and objectives

This research addresses the existing gap in Persian language processing in medical AI models by developing PerMed, an enhanced version of Meditron that supports Persian. The key contributions and objectives of this research are:

- Developing a Persian-enabled medical AI model to enhance accessibility for Persian-speaking users.
- Integrating translation tools to enable bilingual interaction and medical text translation.
- Fine-tuning PerMed to optimize its understanding and generation of Persian medical terminology.
- Evaluating translation accuracy and clinical relevance in Persian language tasks.

1.2 Significance of research

Developing AI-driven medical models that support multiple languages, especially Persian, ensures that medical professionals and patients who rely on Persian-language resources gain access to high-quality medical information. This work has significant implications for improving healthcare services, enabling better access to critical medical knowledge, and fostering multilingual AI applications in the medical field.

1.3 Paper organization

The remainder of this paper is structured as follows: Section 2 presents background information on large language models in the medical field and the significance of multilingual support. Section 3 describes the methodology used in developing PerMed, including model architecture, data sources, and fine-tuning techniques. Section 4 presents the results and evaluation of PerMed's performance in Persian medical tasks. Section 5 provides a detailed discussion on the broader impact of PerMed, while Section 6 addresses its advantages and limitations, along with potential directions for future research and improvements.

2 Background

2.1 Large language models in the medical field

LLMs are recognized as powerful tools for analyzing medical data and providing specialized information. Models such as BioBERT, ClinicalBERT, and PMC-LlaMA, focusing on medical and biomedical data, have performed well in tasks such as medical question answering and specialized text analysis [2, 3].

Meditron, as an open-source model, represents a significant advancement in this field. Using validated medical data and leveraging advanced natural language processing methods, this model has achieved better performance in tests such as MedQA and PubMedQA

[DOI: 10.71885/ijorlu-2024-1-685]

compared to previous models. However, its English language limitation restricted its capabilities to non-English users [4, 5].

2.2 Translation and support for non-english languages

The use of multilingual models and machine translation techniques has been proposed as a solution to increase access to knowledge in various fields. Models such as TOWER and Sailor have been developed to support non-English languages, demonstrating that combining parallel and monolingual data can improve translation quality [5, 6]. TOWER uses a multi-layer transformer-based architecture and is designed for analyzing complex specialized data. This model uses multi-stage learning, which allows it to improve processing and translation accuracy using specialized data. In contrast, Sailor, focusing on low-resource languages and using advanced machine translation techniques, not only accurately preserves the meaning of sentences but is also very efficient for improving translation in multilingual environments. Inspired by these two models and using advanced translation tools such as the Google Translate API and prompt engineering techniques [4, 7, 8], PerMed combines their key features and, using parallel Persian and English data, has significantly increased the accuracy and quality of Persian medical text translation [6, 9].

2.3 Challenges

Despite advancements, the development of multilingual models in the medical field faces challenges such as maintaining scientific accuracy, reducing language bias, and improving translation quality in complex texts. PerMed is designed to mitigate these challenges by combining English medical data and Persian translations, offering new capabilities to users [5, 10].

3 Methodology

3.1 Meditron’s overall structure

Meditron, as a large medical language model with a Llama-2 base architecture, uses specialized medical data such as PubMed Central and medical guidelines for pre-training. This model has high capabilities in medical question answering, specialized text analysis, and chain-of-thought reasoning, but its English language limitation has restricted its use for non-English users [4, 5]. Llama-2 is an advanced transformer-based architecture designed for developing large language models. Developed by Meta, this architecture offers significant improvements over previous versions. Llama-2 uses a CLM [5] structure, which is highly optimized for predicting the next word in the text. Using a large number of parameters (from 7 billion to 70 billion) and diverse data, this model has achieved high accuracy and capability in natural language processing. One of Llama-2’s key features is its ability to process long text data and maintain semantic coherence throughout the text, making it highly efficient for applications such as specialized text analysis, question answering, and content generation. Using this architecture in Meditron has increased its accuracy and ability in medical reasoning [4, 5].

[DOI: 10.71885/ijorlu-2024-1-685]

3.2 Persianizing meditron: PerMed development

To address Meditron's limitations, the PerMed project was designed to add Persian language support. The main development steps include:

- Using Translation Tools: Google Translate API and other advanced translation tools were used to translate English medical texts into Persian. These translations were used as training data for PerMed development [7].
- Prompt Engineering: Using chain-of-thought prompting [7] and self-consistency [8] techniques, Persian question answering capability was added to PerMed. These methods allow the model to perform the answering process step-by-step and increase final accuracy [8].
- Integration with the Base Model: PerMed was designed to retain all of Meditron's capabilities while also providing Persian language support. This integration was done using automatic translation techniques and fine-tuning of the model [6].

4 Modeling

4.1 Model performance comparison

PerMed demonstrated similar performance to Meditron (the base model) in answering English questions on standard medical benchmarks such as MedQA and PubMedQA. This indicates that adding Persian capabilities to PerMed did not negatively impact its English language performance.

The key finding is PerMed's superiority in Persian language-related tasks. This superiority was observed in the accuracy and quality of answering Persian questions, as well as in translating medical texts from English to Persian. This demonstrates the success in localizing medical knowledge for Persian-speaking users [5, 10].

MedQA and PubMedQA are two reputable standard benchmarks for evaluating the capabilities of language models in the medical domain. MedQA comprises a set of multiple-choice questions based on standard medical examinations such as the USMLE. It assesses the models' abilities in reasoning, analyzing specialized texts, and providing accurate answers to complex questions. MedQA questions typically cover concepts like physiology, pharmacology, and clinical diagnosis, requiring a deep understanding of medical knowledge from the model. PubMedQA, on the other hand, is designed based on research questions extracted from PubMed scientific articles. This benchmark evaluates the model's ability to extract accurate answers from research texts and specifically focuses on evidence-based answers. In this comparison, PerMed showed similar performance to Meditron on English questions, but in Persian tasks such as translation and answering specialized questions, it achieved superior performance with high accuracy and semantic consistency. This demonstrates PerMed's ability to localize medical knowledge and expand the model's capabilities for the Persian language [8, 10].

In this section, the performance evaluation of the PerMed model in answering medical questions from the perspective of various specialists is reviewed. The table below presents a comparison of accuracy, content comprehensiveness, statistical accuracy, clinical recommendations, and response time across multiple medical specialties. The evaluations were conducted by a group of physicians specializing in different fields, and the results

[DOI: 10.71885/ijorlu-2024-1-685]

indicate that PerMed has demonstrated satisfactory performance across most of the criteria. Following these results, a more detailed analysis of each specialty will be conducted.

Table 1 Evaluation of Model Accuracy and Performance in Various Medical Field

Criterion	General	Orthopedic Physician	Internist	Neurology	Obstetrics and Gynecology	Ophthalmology/ Vision
Accuracy of Content	90%	90%	90%	85%	85%	90%
Accuracy of Statistics	85-90%	85%	85%	80%	80%	85%
Comprehensiveness of Content	75%	80%	80%	70%	75%	75%
Clinical Recommendations	80%	85%	85%	85%	85%	80%
Response Time	8-6/5-3 seconds	8-6/5-3 seconds	8-6/5-3 Seconds	8-5seconds	3-8 seconds	5-8 seconds

Table 2, PerMed was evaluated on a set of tasks including medical question answering, medical text translation, and interaction in Persian. The criteria used to compare its performance with the original Meditron version included accuracy, translation quality, and response time. As shown in the table above, PerMed demonstrated high accuracy in content and clinical recommendations, as well as solid performance across various specialties, particularly in Persian language tasks.

The evaluations were conducted by a group of medical experts, including general practitioners and clinical specialists. After thoroughly reviewing the model's answers, the quality of medical text translation, and the model's ability to interact in Persian, the physicians confirmed PerMed's performance. These experts highlighted that PerMed's answers not only showed high accuracy but also maintained consistency with specialized medical knowledge. Moreover, the model's translation capability effectively conveyed complex medical terms and concepts, making it especially valuable for educational and research purposes.

In summary, the physicians emphasized that PerMed's ability to interact in Persian is an essential step towards improving access to medical knowledge for Persian-speaking users. This could significantly enhance digital medical services in the future, especially in regions with Persian-speaking populations.

The table below presents a comparison of the performance of different models in medical question answering, medical text translation, and Persian language support.

Based on the results shown in Table 2, PerMed outperforms the other models in terms of both medical question answering and medical text translation, with particular strength in Persian language support. While models like MEDITRON, PMC-LlaMA, and TOWER performed well in English, PerMed excels in providing high-quality translation and accuracy for Persian language tasks. This highlights the successful incorporation of Persian language capabilities into PerMed, making it a more effective tool for Persian-speaking users in the medical domain.

[DOI: 10.71885/ijorlu-2024-1-685]

Table 2 Comparison of Model Performance in Medical Tasks

Model	Question Answering (MedQA)	Medical Text Translation (Human Evaluation)	Persian Language Support	Description
MEDITRON	85%	85%	No	Strong performance in English.
PMC-LlaMA	80%	78%	No	Good performance in English and medical domain.
TOWER	75%	82%	No	Focus on multilingual translation (excluding Persian).
PERMED	87%	90%	Yes	Support for Persian and high-quality translation.

4.2 Translation quality analysis

The translation of medical texts from English to Persian using PerMed was of high quality and received favorable scores in terms of semantic accuracy and fluency. These translations were evaluated by medical experts and were deemed reliable in over 85% of cases [3, 7].

4.3 User experience in Persian interaction

PerMed was able to fully answer Persian questions and manage multi-turn dialogues with high accuracy. This feature improved the use of the model in Persian-speaking environments and demonstrated success in Persian language integration [6, 8].

4.4 Advantages of PerMed

PerMed, as the Persianized version of Meditron, offers significant advantages. By adding Persian language support, Persian-speaking users, including physicians, researchers, and patients, can benefit from the advanced capabilities of this model [3, 7]. This model has managed to maintain the accuracy and quality of Meditron's responses and extend this accuracy to Persian interactions. [4, 5]. Furthermore, by reducing language barriers, PerMed has filled the gap in providing language processing services for the Persian language and added multilingual capabilities to the model [6, 9].

4.5 Limitations of PerMed

Despite its successes, PerMed still faces limitations. In some highly specialized texts, the translation quality has decreased due to the complexity of scientific Persian [7]. Additionally, adding the translation process may increase the model's response time, which needs improvement in clinical settings [6, 8]. The lack of sufficient specialized Persian medical data is also a barrier to further model improvement [10].

[DOI: 10.71885/ijorlu-2024-1-685]

5 Future opportunities

With the successful Persianization of Meditron, the approach used for creating PerMed can be applied to other languages, particularly those in regions with limited access to advanced language models [6, 9]. By collecting and using specialized Persian data, the model's ability to translate and analyze medical texts can be significantly improved, leading to higher accuracy and more relevant insights in Persian [3]. Additionally, the combination of PerMed with newer, state-of-the-art translation models holds the potential to further enhance the quality and precision of both translations and medical question answering systems [7, 8].

One of the most promising future opportunities for PerMed lies in extending the approach of Persianization to other underrepresented languages. Many regions across the globe, particularly in the Middle East, Africa, and Asia, lack high-quality language models for medical applications. As a result, these regions experience significant barriers to accessing healthcare information and medical services. By applying the lessons learned from the Persianization of Meditron to languages such as Arabic, Urdu, and various African and Asian languages, these barriers can be reduced. These languages often lack comprehensive medical datasets, which limits the ability of AI models to provide accurate medical information. By creating and utilizing specialized datasets in these languages, PerMed can help make AI-driven healthcare tools more accessible and effective, ultimately improving healthcare delivery in these regions.

Moreover, the increasing digitalization of the healthcare sector, particularly in non-English-speaking countries, creates a growing demand for AI-powered tools capable of understanding and processing medical information in various languages. Extending PerMed's capabilities to these languages would play a pivotal role in improving the global reach of medical AI models. This expansion could significantly enhance access to reliable medical information and advice in regions where language barriers currently pose a significant challenge. As a result, healthcare professionals and patients alike would benefit from better digital healthcare tools, and this can ultimately lead to improved health outcomes worldwide.

In addition to expanding language support, another avenue for enhancing PerMed's capabilities involves incorporating specialized medical data in Persian. The existing version of PerMed has shown commendable performance in medical text translation and answering medical questions, but it could still be improved with more diverse and specific medical datasets. By collaborating with medical institutions and universities to create and curate high-quality, annotated Persian-language medical corpora, PerMed's understanding of complex medical concepts could be further refined. For example, incorporating more detailed datasets related to specific medical specialties, such as cardiology or oncology, would allow PerMed to better handle specialized medical terminologies and offer more accurate, nuanced responses. Such datasets would improve not only medical translations but also the overall performance of the model in answering specialized questions across a wide range of medical fields.

The integration of PerMed with newer and more advanced machine translation (MT) models represents another exciting opportunity. Modern advancements in neural machine translation (NMT) and transformer-based models have significantly improved translation accuracy. By combining PerMed with the latest advancements in translation models, such as multilingual transformers or zero-shot models, the quality of both medical text translations and question answering can be further enhanced. Hybridizing PerMed with these models could help the system better adapt to changes in medical terminology, new diseases, and emerging treatments, while ensuring high-quality translations in real-time. Furthermore, the

[DOI: 10.71885/ijorlu-2024-1-685]

integration of PerMed with domain-specific translation models can address specific challenges in medical language translation, such as the translation of highly technical terms, colloquial phrases, or region-specific medical concepts.

Furthermore, PerMed could be integrated into Clinical Decision Support Systems (CDSS). CDSS are increasingly being used in healthcare to assist medical professionals in making data-driven decisions by analyzing patient data, offering suggestions, and providing medical insights. By embedding PerMed into CDSS, healthcare providers who speak Persian would be able to receive AI-assisted support in their native language, making the decision-making process more efficient and accurate. PerMed could assist healthcare professionals by processing medical records, suggesting treatment options, and even identifying potential diagnoses. The ability to interact with PerMed in Persian would also reduce the likelihood of miscommunication between healthcare providers and their patients, as language barriers could be significantly mitigated. This integration could help improve patient care, especially in under-resourced areas where specialized expertise may not always be available.

As PerMed continues to evolve, collaboration with medical experts, linguists, and researchers will remain critical for its ongoing improvements. Regular evaluations by professionals in various medical fields will ensure that PerMed stays up-to-date with the latest medical knowledge, terminologies, and practices. Collaboration with these experts will also help to identify potential areas of weakness in the model, such as challenges with emerging medical terms or newly developed treatments, and address them in future iterations. Such continuous expert involvement will guarantee that PerMed remains a reliable, accurate, and trusted resource for medical professionals and researchers.

Another exciting avenue for future development lies in the creation of personalized healthcare assistants using PerMed's capabilities. Leveraging PerMed's language understanding and medical knowledge, a personalized healthcare assistant could offer tailored advice and recommendations based on an individual's health profile, medical history, and preferences. For instance, by analyzing a user's medical records and other relevant data, PerMed could provide personalized health advice, chronic disease management suggestions, or even mental health support. This would be particularly beneficial in preventive healthcare, where personalized guidance could help people make better lifestyle choices and manage health risks more effectively. The ability to provide personalized healthcare recommendations would also empower patients to take an active role in their health management, improving both engagement and overall health outcomes.

PerMed also presents significant opportunities for enhancing telemedicine platforms. As telemedicine becomes more common worldwide, the ability to bridge language gaps between patients and healthcare providers will be essential. PerMed's ability to provide accurate, real-time translations of medical questions and answers could be integrated into telemedicine platforms, allowing doctors and patients to communicate effectively, regardless of language differences. This would be particularly valuable in regions where access to healthcare providers is limited or where 32ommunice barriers prevent effective 32ommunicateon. With real-time translation, healthcare providers could offer timely and accurate advice to patients, while patients would feel more comfortable sharing their symptoms and concerns, leading to more effective consultations.

Lastly, scaling PerMed to meet the increasing demand for multilingual medical models will require significant investment in data collection, model training, and collaboration with stakeholders in both the medical and linguistic fields. Large-scale field tests and collaborations with healthcare organizations, research institutions, and government agencies will be crucial for ensuring that PerMed is both effective and scalable across diverse regions.

[DOI: 10.71885/ijorlu-2024-1-685]

Additionally, as more languages are incorporated, the model's adaptability and ability to generalize across various medical fields will need continuous refinement to ensure that it remains a powerful tool for healthcare professionals and patients worldwide.

In conclusion, the successful Persianization of Meditron through PerMed is just the beginning. With opportunities for expanding to other languages, improving model performance through specialized data, integrating with advanced translation models, and embedding PerMed into real-world healthcare systems, the future holds tremendous promise. By continuously adapting to the evolving needs of the healthcare industry, PerMed can significantly enhance access to healthcare information, reduce language barriers, and improve patient care on a global scale.

6 Conclusion

In this paper, PerMed was introduced as the Persianized version of Meditron, which added Persian language support to Meditron's advanced capabilities. Using translation tools, prompt engineering techniques, and careful integration, this model was able to successfully answer Persian questions, translate English texts into Persian, and manage multilingual interactions.

PerMed, by increasing access to medical information for Persian speakers, is an effective step in developing multilingual large language models in the medical field. However, improving translation quality and reducing response time are issues that need to be addressed in the future. Developing specialized Persian data and using more advanced techniques can further enhance PerMed's performance.

While PerMed has made significant achievements in supporting Persian in the medical domain, there are still challenges related to scalability and optimizing the model's performance. To extend the applicability of this model, the use of more advanced techniques in natural language processing (NLP) is necessary. In this regard, focusing on training the model with larger and more diverse datasets, especially specialized Persian medical data, can improve the accuracy and quality of the model's responses.

Furthermore, given the widespread use of language models in healthcare and research centers, models like PerMed can contribute to improving disease diagnosis systems, offering treatment recommendations, and enhancing the quality of medical care in Persian-speaking countries. This will not only increase the accuracy of clinical decision-making but also facilitate access to medical knowledge for physicians and patients in regions with limited resources.

Finally, considering the unique features of PerMed, expanding it to other languages is also noteworthy. Turning PerMed into a multilingual model for other languages could play a significant role in the future, not only in the medical field but also in other scientific domains. This process could ultimately lead to more widespread access to reliable and credible information for individuals worldwide, especially in regions facing language barriers.

References

1. Liévin, V., Hother, C. E., Motzfeldt, A. G., & Winther, O. (2024). Can large language models reason about medical questions?. *Patterns*, *5*(3).
2. Wu, C., Lin, W., Zhang, X., Zhang, Y., Xie, W., & Wang, Y. (2024). PMC-LlaMA: toward building open-source language models for medicine. *Journal of the American Medical Informatics Association*, ocae045.

[DOI: 10.71885/ijorlu-2024-1-685]

3. Dou, L., Liu, Q., Zeng, G., Guo, J., Zhou, J., Lu, W., & Lin, M. (2024). Sailor: Open Language Models for South-East Asia. *arXiv preprint arXiv:2404.03608*.
4. Nazi, Z. A., & Peng, W. (2024, August). Large language models in healthcare and medical domain: A review. In *Informatics* (Vol. 11, No. 3, p. 57). MDPI.
5. Chen, Z., Cano, A. H., Romanou, A., Bonnet, A., Matoba, K., Salvi, F., & Bosselut, A. (2023). Meditron-70b: Scaling medical pretraining for large language models. *arXiv preprint arXiv:2311.16079*.
6. Alves, D. M., Pombal, J., Guerreiro, N. M., Martins, P. H., Alves, J., Farajian, A., … & Martins, A. F. (2024). Tower: An open multilingual large language model for translation-related tasks. *arXiv preprint arXiv:2402.17733*.
7. Maharjan, J., Garikipati, A., Singh, N. P., Cyrus, L., Sharma, M., Ciobanu, M., … & Das, R. (2024). OpenMedLM: prompt engineering can out-perform fine-tuning in medical question-answering with open-source large language models. *Scientific Reports*, *14*(1), 14156.
8. Wang, X., Chen, Y., Yuan, L., Zhang, Y., Li, Y., Peng, H., & Ji, H. (2024). Executable code actions elicit better llm agents. *arXiv preprint arXiv:2402.01030*.
9. Zhang, G., Jin, Q., Zhou, Y., Wang, S., Idnay, B., Luo, Y., … & Peng, Y. (2024). Closing the gap between open source and commercial large language models for medical evidence summarization. *Npj Digital Medicine*, *7*(1), 239.
10. Yang, R., Tan, T. F., Lu, W., Thirunavukarasu, A. J., Ting, D. S. W., & Liu, N. (2023). Large language models in health care: Development, applications, and challenges. *Health Care Science*, *2*(4), 255-263.

[DOI: 10.71885/ijorlu-2024-1-685]