

Medical Specialty Classification: An Interactive Application with Iterative Improvement for Patient Triage [†]

Anas Chahid ^{1,*}, Ismail Chahid ², Mohamed Emharraf ¹ and Mohammed Ghaouth Belkasmi ¹

¹ National School of Applied Sciences, Mohamed First University, Oujda 60000, Morocco; m.emharraf@ump.ac.ma (M.E.); m.belkasmi@ump.ac.ma (M.G.B.)

² Faculty of Science, Mohamed First University, Oujda 60000, Morocco; ismail.chahid@ump.ac.ma

* Correspondence: anas.chahid.d23@ump.ac.ma

[†] Presented at the 7th edition of the International Conference on Advanced Technologies for Humanity (ICATH 2025), Kenitra, Morocco, 9–11 July 2025.

Abstract

The challenge of accurately identifying the appropriate medical specialty based on patient symptoms leads to delays in diagnosis and treatment. This paper presents an AI model developed to classify medical specialties from symptom descriptions. The model, implemented with BERT, hosted via a Python-based Flask API v3, and integrated with an interactive frontend application, allows users to input symptoms textually or interactively select affected body parts and answer multiple choice questions. Following deployment, feedback data from doctors and residents was collected and utilized to enhance the model performance, supplemented by additional data from online medical forums. This study demonstrates significant improvements in finding the correct medical specialty, contributing to more efficient patient triage, reducing the time to diagnose and treat patients, and eliminating the presence of doctors in the initial process as they are often busy in emergency departments. The use of generative AI and large language models, notably BERT, is highlighted as a key factor in the model's success.

Keywords: medical specialty classification; symptom description; BERT; patient triage; emergency room; diagnosis; treatment efficiency



Academic Editors: Brahim El Bhiri
and Ayman S. Mosallam

Published: 4 November 2025

Citation: Chahid, A.; Chahid, I.; Emharraf, M.; Belkasmi, M.G. Medical Specialty Classification: An Interactive Application with Iterative Improvement for Patient Triage. *Eng. Proc.* **2025**, *112*, 64. <https://doi.org/10.3390/engproc2025112064>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Medical shortages and the growing number of patients who need specialized care require new solutions for efficient patient guidance. Traditional methods for determining the correct medical specialty are often slow and require medical knowledge that most patients do not have, leading to delays in diagnosis and treatment.

Our research project uses AI to make this process smoother, ensuring patients receive timely and accurate medical specialty referrals. Using a dataset of over 16,000 medical entries and testing multiple models, we achieved a high accuracy with the BERT model. This model was put into a web-based application, showing its potential to improve patient triage in healthcare.

The interactive website allows users to input symptoms through text or select affected body parts and answer some questions. This approach improves user engagement and symptom understanding, leading to more accurate classifications. Feedback from doctors and residents and data from online medical forums were used to improve the model's performance.

Our AI model addresses medical shortages and improves healthcare efficiency. This study highlights the significant potential of generative AI and large language models (LLMs) in modernizing patient triage and ensuring specialized care.

The remainder of this paper is organized as follows: Section 2 reviews related studies on medical text classification and symptom-based specialty prediction systems. Section 3 details our model architecture and BERT implementation. Section 4 describes our iterative model enhancement process, including post-deployment feedback collection, data augmentation from medical forums, and model retraining methodology. Section 5 presents our interactive web application, covering both technical implementation details and user interface design. Finally, Section 6 concludes with a discussion of our results and future research directions, highlighting the potential for voice recognition integration and advanced natural language processing techniques to further improve patient triage efficiency.

2. Related Work

Previous studies have explored various machine learning techniques for medical text classification. Kim et al. [1] fine-tuned BERT for predicting medical specialties from English texts; their analysis focused on predicting 1 of 27 medical specialties based on medical question texts. The dataset, collected from a medical question answering platform, is used for model training and evaluation. The results highlight the proposed model's superior performance compared with four competitive deep learning-based NLP models, demonstrating a higher accuracy, precision, recall, and F1 score. The algorithms compared include BERT, KR-BERT, M-BERT, CNN, and LSTM. The study also acknowledges challenges related to data quality and the complexity of predicting certain medical specialties, proving the importance of the careful interpretation of the prediction results. Hossam et al. [2] developed a system aimed at automatically classifying medical questions from patients into 15 medical specialties, addressing the complexities of the Arabic language, especially its common dialectical variations on the internet. The solution leverages swarm intelligence and an ensemble of support vector machines (SVMs), utilizing particle swarm optimization (PSO) for feature selection and hyperparameter tuning. The integration of PSO with binarization techniques significantly enhances performance. Their approach, which combines PSO, SVM, TF-IDF, and tSNE, achieves 85% accuracy with an impressive 95.9% feature reduction rate. Comparative analysis shows that the method outperforms other machine learning classifiers. However, the study is limited by focusing on a subset of medical specialties and questions, which may reduce its overall applicability. Chao Mao et al. [3] introduced a novel model, HyM, designed to automatically direct patients to one of eight appropriate medical specialties based on their symptom descriptions in Chinese. HyM integrates feature extraction from LSTM, TEXT-CNN, BERT, and TF-IDF to capture both local and global semantic information from the symptom texts. The model is evaluated on a dataset containing over 40,000 records from both online and offline sources, and its performance is compared with models such as TEXT-CNN, LSTM, and BERT. The results demonstrate that HyM achieves the highest accuracy of 93.5%, outperforming the other models in terms of classification precision and stability; however, the number of medical specialties was significantly less than the other studies. Weng, W. et al. [4] aimed to develop a machine learning-based NLP pipeline for classifying medical subdomains in clinical notes, utilizing various data representation techniques and supervised learning algorithms. The study evaluated classifier performance using 431 clinical notes from iDASH and 91,237 from MGH. The linear SVM with TF-IDF weighting and a hybrid feature representation (combining bag-of-words and UMLS concepts) emerged as the best-performing classifier across both datasets. While deep learning classifiers using neural embeddings showed a higher AUC, they yielded lower F1 scores compared with shallow learning

classifiers with interpretable clinical features. The study employed seven shallow learning algorithms (including Naive Bayes, logistic regression, SVM, SGD, Random Forest, and adaptive boosting) and two deep learning models (CNN and CRNN). Metrics such as F1 score, precision, recall, balanced accuracy, and AUC were reported, with the best classifier for the iDASH dataset achieving an F1 score of 93.2% and an AUC of 95.7%. For the MGH dataset, the top-performing classifier attained an F1 score of 93.4% and an AUC of 96.4%. However, limitations include the reliance on a single clinical NLP system (cTAKES), the analysis of only two datasets, an imperfect de-identification process, disregard for concept relationships in UMLS, and the absence of external clinical validation.

Additionally, studies have discussed online-based medical diagnosis applications. Mosa et al. [5] conducted a systematic review of healthcare applications for smartphones, grouping them by targeted users (clinicians, students, patients) and discussing their functionalities. S. Madanian et al. [6] reviewed patients' perspectives on digital health tools, identifying facilitators like empowerment and self-management, as well as barriers such as digital literacy and privacy concerns. H. L. Semigran et al. [7] evaluated online symptom checkers to determine their diagnostic and triage accuracy, Tao et al. [8] presented their medical diagnosis LLM (AMIE) capable of dialog. Finally, A. Jutel et al. [9] sought to describe and catalog available diagnosis apps and explore their impact on the diagnostic process.

3. Model Architecture and Implementation

Our implementation utilizes the BERT (Bidirectional Encoder Representations from Transformers) model, specifically the bert-base-uncased variant with 12 transformer blocks, 12 attention heads, and 110 million parameters. This pre-trained model was fine-tuned on our medical specialty dataset to optimize the performance for our specific classification task.

Model Architecture

- Input Layer: Processes tokenized text with a maximum length of 128 tokens.
- Embedding Layer: Combines token, segment, and position embeddings.
- Transformer Encoder Layers: Comprises 12 layers of self-attention and feed-forward networks.
- Classification Layer: Custom layer added during fine-tuning to map BERT outputs to specialty labels.

Figure 1 shows the general model architecture.

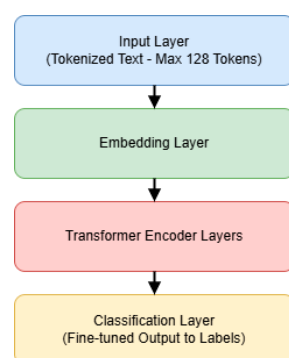


Figure 1. BERT model architecture.

The bidirectional nature of BERT allows it to consider the full context of words in symptom descriptions, which is crucial for understanding medical terminology and symptom relationships.

4. Model Enhancement

4.1. Data Collection and Initial Model

Our initial model, as discussed in our previous research paper [10], was trained using a dataset of over 16,000 medical questions categorized into 41 specialties. In this study, we tested four models, SVM, Random Forest, BERT, and FastText, with BERT achieving the highest accuracy of 94%. The diversity of the dataset allowed the model to learn the relationships between symptoms and specialties effectively.

4.2. Incorporating User Feedback

Post deployment, we collected feedback from doctors and residents using a form integrated into the application which helped collect 543 input data points with 88% positive accuracy. This form allowed users to indicate whether the predicted specialty was correct, providing valuable data for model refinement. We gathered significant feedback, identifying and correcting biases in certain specialties especially “Cardiology” where there was a noticeable bias toward “Gastroenterology” when selecting the upper parts of the abdomen.

We summarize the whole architecture in Figure 2.

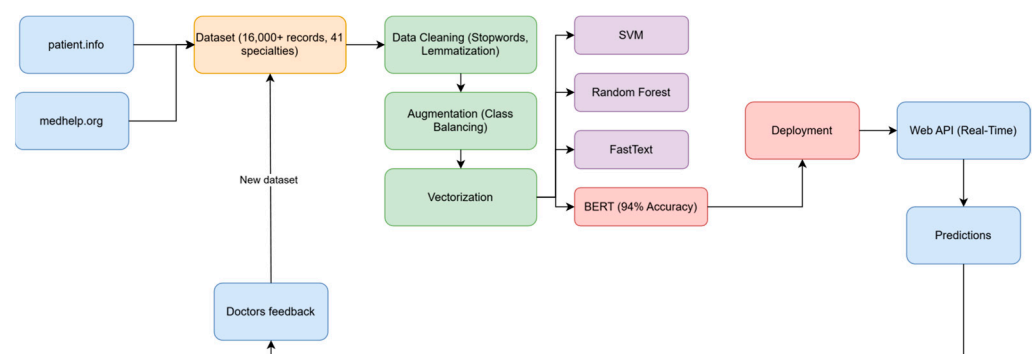


Figure 2. Data feedback and retraining architecture.

4.3. Data Augmentation from Medical Forums

To further improve the model, we augmented our dataset to reach more than 18,000 entries with more data from the popular medical QA forums MedHelp [11] and PatientInfo [12] and the data collected from after the deployment of the application with a focus on data about the specialties affected by the bias. This additional data provided a broader range of symptom descriptions and patient questions, enhancing the model’s ability to generalize across different contexts and reducing bias.

4.4. Model Retraining and Evaluation

The enhanced dataset was used to retrain the previously trained BERT model, resulting in improved accuracy and robustness. The model’s performance was re-evaluated, showing a minor improvement in precision and recall for previously biased specialties which led to a small increase in the model performance from 94% to 94.7% accuracy. This iterative enhancement process demonstrates the importance of continuous data collection and model refinement in developing AI solutions for healthcare.

4.5. Challenges and Solutions in Model Development

During development, we encountered several challenges that required some solutions: **Class Imbalance:** Some specialties were underrepresented in our initial dataset. We addressed this by implementing weighted loss functions and oversampling minority classes.

Ambiguous Symptoms: Many symptoms can be associated with multiple specialties. We enhanced disambiguation by adding symptom severity, duration, and location in our feature engineering process.

Medical Terminology Variation: Patient descriptions often use general known terms rather than medical terminology.

5. The Prediction Application

Our application features a user-friendly interface, allowing patients to input symptoms either through text via a text field or interactively. The interactive option includes selecting body parts and specifying pain characteristics such as pain type, severity, and duration. The backend, powered by our Flask API and hosted in Azure, processes these inputs to predict the appropriate medical specialty.

Our application has five pages, each containing specific input fields about each step in identifying the correct medical specialty: a location page, an area of pain page, a pain description page, a severity page, and a result page.

We combined all the fields' information on each page provided by the user to generate a pre-defined phrase that can help give an understandable context for our BERT model.

An example of a phrase provided to the model, which gave the "Cardiology" specialty as a result, is as follows: "I've been experiencing pain of 5 from a 0 to 10 scale in my chest. It feels Aching and it has been going on for 1–3 Weeks".

5.1. Technical Implementation Details

The application architecture consists of three main components:

- Frontend: Developed using React.js for responsive design, ensuring compatibility across devices including mobile phones and tablets. The frontend communicates with the backend via RESTful API calls.
- Backend API: Built with Flask, the API handles request processing, text normalization, and model inference. Key features include:
 - CPU-based model inference to lower the running cost.
 - User feedback collection and storage.
 - Authentication for medical professionals.
- Deployment Infrastructure: The application is containerized using Docker and deployed on the Azure App Service with autoscaling capabilities to handle varying loads. This ensures a high availability and consistent performance.

5.2. Performance Metrics Pages

Load testing was conducted to ensure that the application could handle multiple simultaneous users. The results showed an average result response time of 6.3 s which can be improved by using a GPU-based inference.

These metrics confirm the application's readiness for real-world deployment in clinical settings.

The main pages of our application are the home page and the result page, as presented in the following sections.

5.2.1. Home Page

The home page serves as the application's entry point, providing the first step in identifying the medical specialty and it offers the user the choice between using a text field to describe the issue and symptoms or an interactive body parts selector followed by several steps to better describe the issue, as shown in Figure 3.

1 First step Location — 2 Second step Area of pain — 3 Third step Pain Description — 4 Fourth step Severity — 5 Final step Duration

You can use this text area to describe your issue and symptoms better!

Describe your issue in a detailed way

Your detailed symptoms here...

Get the medical Specialty

OR

Select the body parts you have issue in:

Reset Selected Parts

Next step

Figure 3. First page of the application.

5.2.2. Result Page

This page gives the user the prediction of the medical specialty, based on how the model takes the user input data and uses it to predict the correct medical specialty. It also contains a form to help doctors and residents easily give feedback on the prediction correctness, as shown in Figure 4.

✓ First step Location — ✓ Second step Area of pain — ✓ Third step Pain Description — ✓ Fourth step Severity — ✓ Final step Duration

You should go to the **Cardiology Medical Specialty.**

Was It the correct specialty?

Yes, That's Correct !

Was it the wrong specialty?

Write the Correct Specialty *

Submit

Try Again

Figure 4. Result page.

6. Conclusions

Our AI-based solution for medical specialty classification demonstrates significant potential in improving patient triage efficiency. By integrating user feedback and augmenting our dataset with forum data, we enhanced the model's accuracy and reduced bias by increasing the model accuracy by 0.7%, reaching 94.7%. Future work will focus on expanding the dataset further, incorporating more diverse medical queries, and exploring additional NLP techniques to refine the model, such as fine-tuning LLMs and comparing their accuracy with the BERT-based solution. This approach highlights the critical role of continuous data collection and model enhancement in developing effective AI solutions for healthcare. We are also planning on using voice recognition to input the symptom

description text and then using this as part of a chatbot to enhance the user experience and therefore make the triage process faster.

Author Contributions: Conceptualization, A.C., M.E., M.G.B., and I.C.; methodology, A.C., M.E., M.G.B., and I.C.; software, A.C.; validation, M.E., M.G.B., and I.C.; formal analysis, M.E.; investigation, A.C., M.E., M.G.B., and I.C.; resources, A.C.; data curation, A.C.; writing—original draft preparation, A.C.; writing—review and editing, A.C., M.E., M.G.B., and I.C.; visualization, A.C., M.E., M.G.B., and I.C.; supervision, M.E.; project administration, M.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Formal ethics approval from an Institutional Review Board (IRB) was not required for this study as it was conducted using publicly available, anonymized data collected from online health forums. No personally identifiable information (PII), such as usernames, locations, or specific dates, was collected or used in this research, ensuring the complete anonymity of the original posters. As the data was publicly accessible and fully anonymized, the requirement for informed consent was waived. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.

Informed Consent Statement: Not applicable.

Data Availability Statement: The application is not yet published for public usage as it is still not perfectly ready for usage by patients.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kim, Y.; Kim, J.-H.; Kim, Y.-M.; Song, S.; Joo, H.J. Predicting medical specialty from text based on a domain-specific pre-trained BERT. *Int. J. Med. Inform.* **2022**, *170*, 104956. [CrossRef] [PubMed]
2. Faris, H.; Habib, M.; Faris, M.; Alomari, M.; Alomari, A. Medical speciality classification system based on binary particle swarms and ensemble of one vs. rest support vector machines. *J. Biomed. Inform.* **2020**, *109*, 103525. [CrossRef] [PubMed]
3. Mao, C.; Zhu, Q.; Chen, R.; Su, W. Automatic medical specialty classification based on patients' description of their symp-toms. *BMC Med. Inform. Decis. Mak.* **2023**, *23*, 15.
4. Weng, W.-H.; Wagholikar, K.B.; McCray, A.T.; Szolovits, P.; Chueh, H.C. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. *BMC Med. Inform. Decis. Mak.* **2017**, *17*, 1–13. [CrossRef] [PubMed]
5. Mosa, A.S.M.; Yoo, I.; Sheets, L. A Systematic Review of Healthcare Applications for Smartphones. *BMC Med. Inform. Decis. Mak.* **2012**, *12*, 67. [CrossRef] [PubMed]
6. Madanian, S.; Nakarada-Kordic, I.; Reay, S.; Chetty, T. Patients' perspectives on digital health tools. *PEC Innov.* **2023**, *2*, 100171. [CrossRef] [PubMed]
7. Semigran, H.L.; Linder, J.A.; Gidengil, C.; Mehrotra, A. Evaluation of symptom checkers for self diagnosis and triage: Audit study. *BMJ* **2015**, *351*, h3480. [CrossRef] [PubMed]
8. Tu, T.; Palepu, A.; Schaekermann, M.; Saab, K.; Freyberg, J.; Tanno, R.; Wang, A.; Li, B.; Amin, M.; Tomasev, N.; et al. Towards Conversational Diagnostic AI. *arXiv* **2024**, arXiv:2401.05654. [CrossRef]
9. Jutel, A.; Lupton, D. Digitizing diagnosis: A review of mobile applications in the diagnostic process. *Diagnosis* **2015**, *2*, 89–96. [CrossRef] [PubMed]
10. Anas, C.; Mohamed, E.; Ghaouth, B.M.; Wafae, M.; Yassine, C. Medical Specialty Triage Using Patients' Questions and Symptoms. *SSN* **2023**. [CrossRef]
11. MedHelp-Health Community, Health Information, Medical Questions, and Medical Apps. Available online: <https://www.medhelp.org/> (accessed on 11 November 2023).
12. Symptom Checker, Health Information and Medicines Guide, Patient. Available online: <https://patient.info/> (accessed on 11 November 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.