JOIN (Jurnal Online Informatika)
p-ISSN: 2528-1682, e-ISSN: 2527-9165
Volume 10 Number 1 | June 2025: 32-41
DOI: 10.15575/join.v10i1.1486

LLM-Based Information Retrieval for Disease Detection Using Semantic Similarity

Muhammad Adrinta Abdurrazzaq[1], Edwin Lesmana Tjiong[2], Kent Algren Wanady[3]
[1,2,3]Department of Informatics, Kalbis University, Indonesia

Article Info

Article history:

Received October 18, 2024
Revised December 18, 2024
Accepted December 27, 2024
Published April 01, 2025

Keywords:

CRISP-DM Framework
Disease Detection
Information Retrieval System
Large Language Model
Semantic Similarity

ABSTRACT

Information retrieval systems are vital for disease prediction, but traditional methods like TF-IDF struggle with word meanings and produce long, complex vectors. This research uses Large Language Models (LLMs) and follows the CRISP-DM methodology to improve accuracy. Using health forum discussions labeled with specific diseases, we split the data into queries and a corpus. Semantic similarity is used to retrieve the most relevant text from the corpus. After preprocessing, we compare LLMs and TF-IDF, with LLMs achieving an accuracy of 0.911 (Top-K=30), outperforming TF-IDF. LLMs excel by creating shorter, meaningful vectors that preserve context, enabling precise semantic matching. These results demonstrate LLMs' potential to enhance healthcare information retrieval, offering more accurate and context-aware solutions. This research highlights how advanced AI can overcome traditional methods' limitations, opening new possibilities for medical informatics.

Corresponding Author:

Muhammad Adrinta Abdurrazzaq,
Department of Informatics, Kalbis University, Indonesia
Jl. Pulomas Selatan Kav. No.22, RT.4/RW.9, Pulo Gadung, Jakarta Timur.

Email: muhammad.abdurrazzaq@kalbis.ac.id

1. INTRODUCTION

The quality of public health services in Indonesia faces significant challenges, particularly in the availability of health facilities and professionals. This issue is especially acute for low-income populations and those in remote areas, where access to healthcare remains limited [1]. While the National Health Insurance (JKN) program has expanded healthcare access nationwide, the existing resources are insufficient to meet the growing demand for quality services. This inadequacy has placed an excessive burden on health workers, particularly doctors, potentially compromising diagnostic accuracy and patient care quality [2], [3], [4], [5]. Additionally, patients often seek second opinions from multiple doctors when faced with critical diagnoses, highlighting the importance of accurate and collaborative diagnostic decisions, which have been shown to outperform individual diagnoses [6].

Training doctors is a time-consuming and costly process, which worsens these problems. This shows why we need new and effective ways to tackle the lack of healthcare workers. Artificial intelligence (AI) has emerged as a transformative tool in healthcare, with applications ranging from diagnostics to treatment planning. AI models, particularly those based on machine learning and deep learning, can replicate aspects of a doctor's expertise, offering scalable and cost-effective solutions to healthcare challenges.

Existing AI models for disease prediction often rely on specific input features, such as physical characteristics (e.g., age, weight, vital signs), genetic factors, habits (e.g., smoking, alcohol consumption),

and symptoms (e.g., fever, cough). For instance, Tigga and Garg [7] developed a model to predict type 2 diabetes, while Tengnah et al. [8] created a model for hypertension prediction. However, these approaches are inherently limited, as they are designed to predict specific diseases, necessitating the development of numerous models to cover a wide range of conditions. Other studies have used predetermined symptom lists as input [9], [10], [11], [12], but this approach restricts the flexibility of symptom descriptions and may fail to capture the full complexity of a patient's condition. Figure 1 shows the illustration of the disease prediction model based on predetermined symptoms.

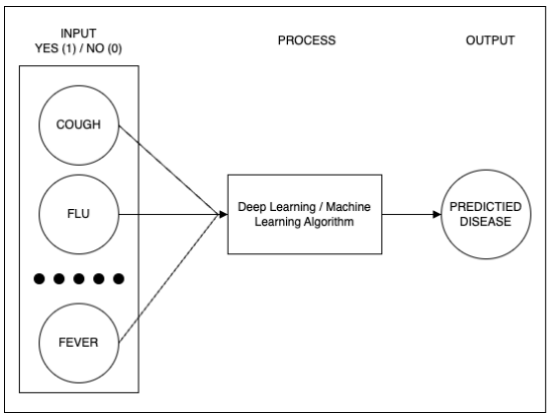


Figure 1. Illustration of a disease prediction model using predetermined symptoms as input [9], [10], [11], [12], [13]

An alternative approach involves using patient-generated text descriptions of symptoms as input. This method allows for more natural and flexible symptom reporting, which is then converted into n-dimensional vectors for similarity measurement against disease symptom vectors stored in the system. Semantic similarity, a method for measuring the similarity of words or sentences [14], is used to identify the most relevant disease match. Aszani et al. [15] developed an information retrieval system using TF-IDF to convert symptom text into vectors. However, TF-IDF has notable limitations, including its inability to capture contextual meaning and its generation of high-dimensional vectors, which can lead to inefficiencies in the system.

To address these limitations, this research proposes the use of Large Language Models (LLMs) as a replacement for TF-IDF. LLMs, such as BERT [16], DistilBERT [17], MPNet [18], and MiniLM [19] excel at generating compact, context-aware vector representations while preserving semantic meaning [20]. These models have been trained on extensive Indonesian language corpora, making them particularly well-suited for this research. By leveraging LLMs, this research aims to develop an information retrieval system that improves upon previous approaches, offering a more accurate and efficient solution for disease prediction. The system will utilize health forum discussions data, where each discussion is labeled with a specific disease, and will be deployed as a web-based application to assist doctors in diagnosis and enable patients to conduct preliminary self-examinations.

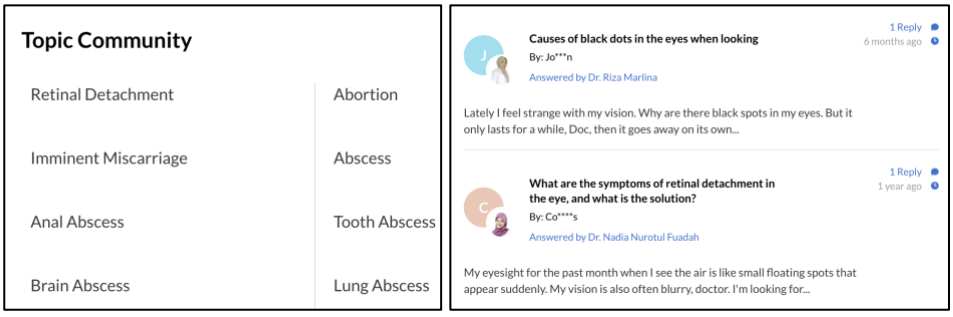
2. METHOD

In developing the proposed information retrieval system, this research adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [21] as its development workflow. The stages of CRISP-DM are as follows:

1. Business Understanding
   At this stage, the research focuses on studying semantic similarity, embedding techniques, and information retrieval systems. The primary objective is to modify the feature extraction method used in previous research by replacing it with Large Language Models (LLMs). Additionally, this research explores how to implement the developed information retrieval system into a user-friendly application.
2. Data Understanding
   Data collection is performed using web scraping techniques, gathering question texts and their corresponding topics from health forums. The question texts represent patient descriptions of

p-ISSN: 2528-1682
e-ISSN: 2527-9165

symptoms, while the topics serve as class labels for specific diseases. Figure 2 provides examples of the collected data, showing question texts and their associated topics.



(a) (b)

Figure 2. (a) Examples of question text topics; (b) Sample question texts with the topic "retinal detachment."

A total of 138,462 question texts were collected, spanning 756 classes. However, the class distribution is highly imbalanced, as shown in Figure 3 and Figure 4, which display the distribution of the 30 most common and 30 least common diseases. For instance, *Infeksi saluran kemih* (urinary tract infection) is the most frequent class, while *Hipogonadisme* (hypogonadism) is the least frequent. Figure 5 provides additional examples of the collected data.

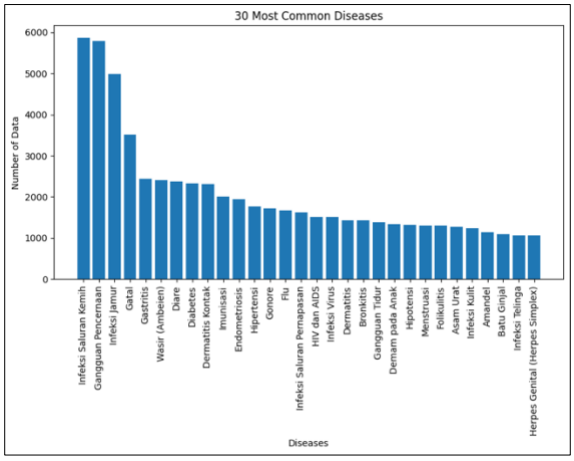


Figure 3. Distribution of the 30 most common diseases in the dataset

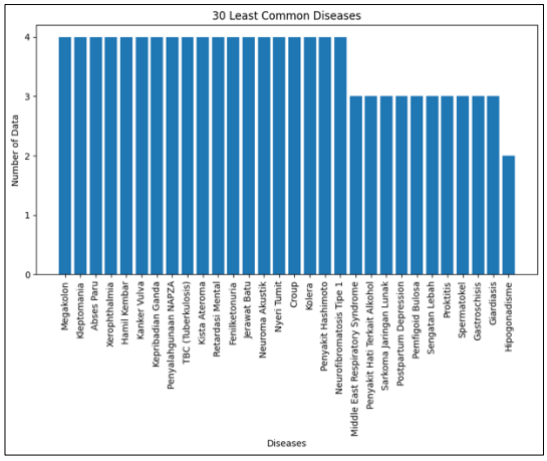


Figure 4. Distribution of the 30 least common diseases in the dataset.

index	symptoms	class
0	Penyeabab bercak putih di batang penis dan penis mengeluarkan cairan berbau Ass dok saya mau nanya nih Penis saya sesekali mengeluarkan cairan putih bening bau seperti air mani dok dan di Batang penis saya ada yg berubah warna menjadi putih tetapi bkn di kepala penisnya dok tapi di Batang penisnya terus saya udah berobat dan di kasih antibiotik untuk infeksi saluran kemih dan salep tetapi warna Batang penis saya masih sama dok dan kadang pada MLM hari badan saya terasa panas tetapi tidak demam kenapa ya dok Apakah saya terkena sefillis atau gonore Mohon di jawab dok	Gonore
1	penyebab bercak merah di lidah dan sakit saat menelan Assalamualaikum donk saya pengen tanya kmrn pagi pas bangun tdr lidah sya aneh dok trs sya cek ada bercak merah gitu dok+kasar telen makanan juga rada sakit dok apa kah itu kanker lidahsariawan dok	Kanker Lidah
2	Sakit kepala setelah operasi pengangkatan tumor telinga Dok kmrin sya operasi pengangkatan tumor jinak di dlm telinga sudah bebberpa hri ini mengapa sya mengalami vertigo ya dok Apakah ini masa penyembuhan atau sperti apa krna mengganggu aktivitas sya	Penyakit Meniere
3	Tangan gemetar pada lansia apa solusi yang tepat Nenek saya usia 75 tahun tangannya suka gemetar nih dok pegang sendok untuk makan saja udah tak sanggup suka tumpah2 makananya dokter Ini sebenernya pengaruh umur ya dok gemeternya apa ada penyakit lain terus gimana caranya agar tangan nenek saya kokoh dan kuat lagi dokter obatnya apa ya	Penyakit Parkinson
4	Penyebab batuk berdarah Salam doktersaya mau tanya harusan saja saya batuk dan berdarah agak bnyak Ada darah yg keluar dari hidung juga Darah yang keluar ada darah yg seperti sudah beku dan mengitam Boleh jelaskan kenapa dokter Saya hanya batuk ketika waktu istirahat tidur siang saja Trima kasih dok	Emboli Paru
5	tekanan darah tinggi di usia 14 tahun assalamualaikum selamat siang doktermohon maaf jika saya merepotkan dokterkenalkan saya rizky usia saya 14walau pun usia saya 14 tahun tapi tekanan darah saya mencapai 13090 apakah itu tanda suatu penyakit yang kronis dok mohon di jawabterima kasih	Hipertensi

Figure 5. Examples of collected data, showcasing question texts and their corresponding disease labels.

3. Data Preparation

The dataset is divided into two parts: corpus data (105,494 entries) and test data (32,968 entries). The question texts from both datasets are then embedded using two methods: LLM and TF-IDF.

a. LLM Embedding: No preprocessing is applied except for the removal of HTML tags.

b. TF-IDF Embedding: The following preprocessing steps are performed to optimize results:

- Convert text to lowercase.
- Remove non-alphabetic characters (except single spaces).
- Replace multiple spaces with a single space.
- Remove leading and trailing spaces.
- Eliminate common stop words.

To ensure consistency, TF-IDF uses 768 terms, matching the longest vector generated by the LLMs in this research. Examples of embedded question texts and their classes are shown in Figure 6.

	symptom (text)	symptom (vector)	class
0	Penyeabab bercak putih di batang penis dan pen...	[0.18721204, 0.09292757, 0.10852494, 0.0475657...	Gonore
1	penyebab bercak merah di lidah dan sakit saat ...	[-0.009280792, -0.073462404, 0.1309922, 0.0173...	Kanker Lidah
2	Sakit kepala setelah operasi pengangkatan tumo...	[-0.019373441, -0.13648549, 0.087172374, -0.08...	Penyakit Meniere
3	Tangan gemetar pada lansia apa solusi yang tep...	[-0.14348951, 0.023910258, 0.32931513, 0.21913...	Penyakit Parkinson
4	Penyebab batuk berdarah Salam doktersaya mau t...	[-0.032741636, 0.074919134, 0.103155546, 0.136...	Emboli Paru
5	tekanan darah tinggi di usia 14 tahun assalamu...	[-0.09482176, 0.26941958, 0.20111617, 0.203372...	Hipertensi

Figure 6. Examples of symptom vectors with their associated texts and disease classes.

4. Modeling

The embedded corpus data is stored in a database, and the vector dimensions generated by the LLMs are summarized in Table 1.

Table 1. Dimensions of vectors generated using the LLMs employed in this research

Feature Extraction Model	Feature Vector Dimension
BERT	768
DistilBERT	512
MiniLMv2	384
MPNet	768

To measure the similarity between the embedded query (test data) and the embedded corpus, two schemes are employed:

a. Cosine Similarity: This scheme calculates the similarity between the embedded query and all embedded corpus data using the cosine similarity formula (Formula 1). The results are sorted in descending order.

$$\text{Cosine Similariy (A, B)} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (1)$$

Here, A and B are the vectors to be compared, i is the index of the vector values, and n*n* is the vector dimension.

b. K-Nearest Neighbor (KNN): This scheme uses an additional classifier, where the embedded corpus and their corresponding classes serve as training data. The KNN model calculates the probability score, which represents the similarity between the corpus data and the embedded query. The KNN configuration uses cosine distance as the metric and tests various numbers of

p-ISSN: 2528-1682
e-ISSN: 2527-9165

neighbors (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30). This scheme is applied only to the best-performing model from the first scheme.

The corpus data with the highest similarity score to the embedded query is used to classify the query. Since this research focuses on information retrieval, classification is performed using a Top-K scenario, where a query is assigned to K classes. The values of K tested are 1, 3, 5, 10, 20, and 30.

5. Evaluation
System performance is measured using accuracy and balanced accuracy metrics to account for class imbalance. The formulas for these metrics are as follows:
   a. Accuracy (Formula 2): Measures the proportion of correctly classified queries.

$$\text{Accuracy} = \frac{True\ Predictions}{Total\ Predictions} \tag{2}$$

   b. Balanced Accuracy (Formula 3): Calculates the average accuracy across all classes, ensuring fairer performance evaluation for imbalanced datasets [22].

$$\text{Balanced Accuracy} = \frac{\sum_{i=0}^{n} Accuracy_i}{Total\ Label} \tag{3}$$

In both metrics, a prediction is considered correct if the actual class of the query is among the Top-K predicted classes.

6. Deployment
The best-performing model is deployed on a cloud computing server. For ease of use, a web application is developed as the user interface (UI). The application connects to the model via FastAPI, a web framework for building APIs. The front end is developed using HTML, CSS, and JavaScript, while the model is implemented using the HuggingFace Transformers library [23] in Python. Figure 7 illustrates the data flow between the web application and the semantic similarity model.

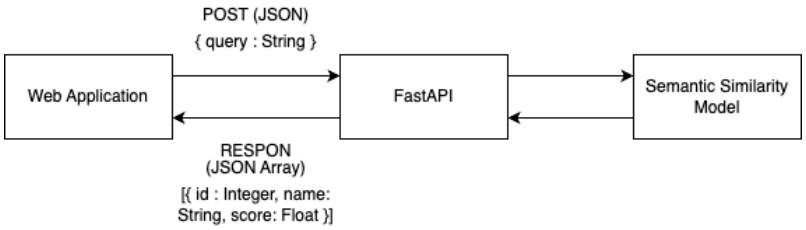


Figure 7. Data flow between the web application and the semantic similarity model

Figure 8 demonstrates the prediction process, where the model calculates semantic similarity scores between the embedded query and corpus data, returning the results in descending order.

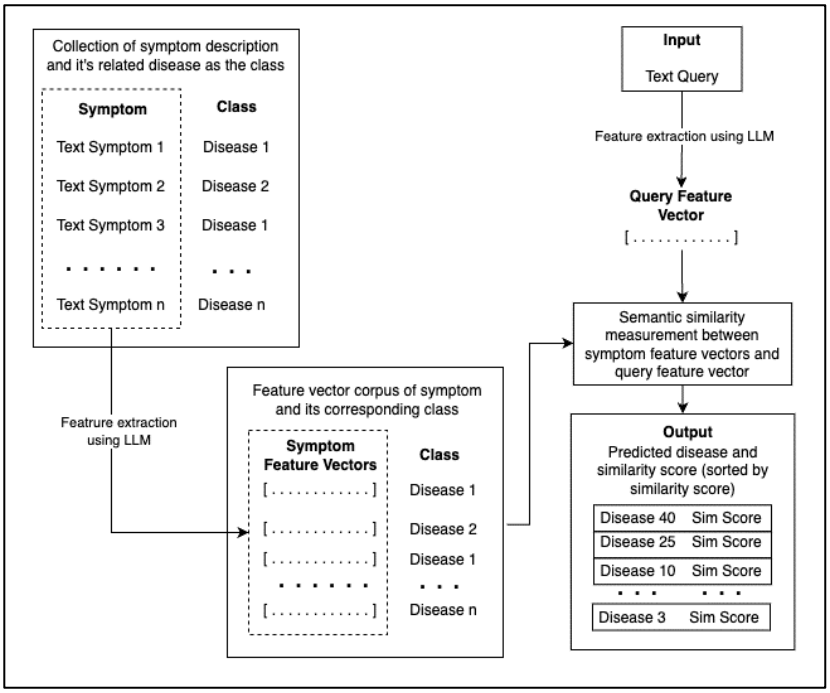


Figure 8. Process of the semantic similarity model for disease prediction

Figure 9 shows the application mockup, displaying semantic similarity scores for each class, sorted from highest to lowest.

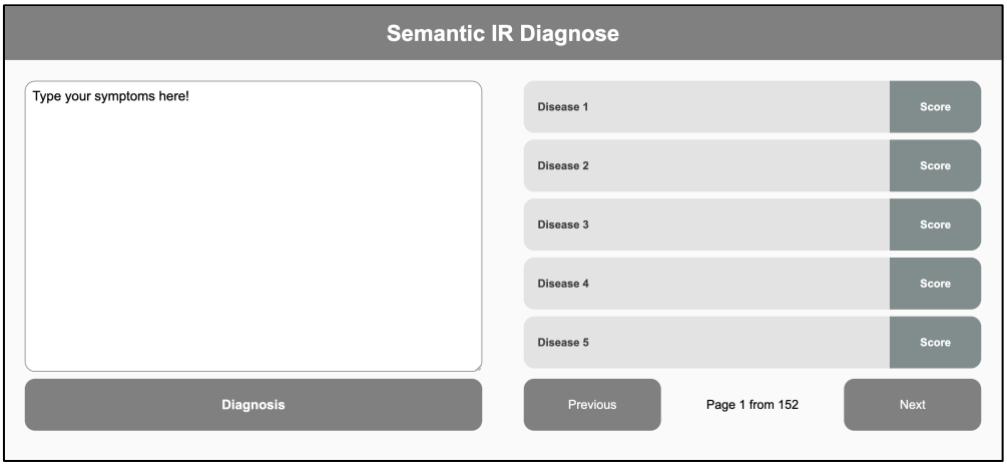


Figure 9. Mockup of the web application interface for the disease prediction system

3. RESULT AND DISCUSSION

The retrieval system using MPNet for feature extraction achieved the highest performance, with an accuracy score of 0.911 and a balanced accuracy score of 0.687 at Top-K=30 when using cosine similarity. In comparison, MiniLMv2 outperformed TF-IDF, while BERT and DistilBERT showed better balanced accuracy scores at smaller k*k* values (1 and 3). However, TF-IDF surpassed both BERT and DistilBERT at larger k*k* values. A comprehensive comparison of scores is provided in Table 1.

Table 1. Performance comparison of models using cosine similarity as the classifier

K	Accuracy					Balanced Accuracy				
	TF-IDF	BERT	DistilBERT	MiniLMv2	MPNet	TF-IDF	BERT	DistilBERT	MiniLMv2	MPNet
1	0.290	0.276	0.262	0.307	**0.343**	0.116	0.137	0.123	0.170	**0.193**
3	0.497	0.477	0.465	0.523	**0.570**	0.226	0.236	0.227	0.298	**0.334**

K	Accuracy					Balanced Accuracy				
	TF-IDF	BERT	DistilBERT	MiniLMv2	MPNet	TF-IDF	BERT	DistilBERT	MiniLMv2	MPNet
5	0.597	0.582	0.562	0.626	**0.672**	0.289	0.299	0.285	0.362	**0.413**
10	0.714	0.712	0.691	0.747	**0.792**	0.388	0.393	0.385	0.461	**0.520**
20	0.810	0.815	0.799	0.842	**0.875**	0.497	0.495	0.488	0.572	**0.623**
30	0.855	0.863	0.849	0.884	**0.911**	0.564	0.556	0.552	0.634	**0.687**

When KNN was used as the similarity measurement method, none of the tested n values (number of neighbors) outperformed cosine similarity. This is attributed to the imbalanced class distribution in the dataset, where some classes contain fewer than five samples. Since KNN relies on a voting mechanism, predictions tend to favor majority classes, resulting in poor performance for minority classes. The detailed scores for this approach are presented in Table 2.

Table 2. Performance of the MPNet model using K-Nearest Neighbor (KNN) as the classifier across various n values

n	Accuracy						Balanced Accuracy					
	K=1	K=3	K=5	K=10	K=20	K=30	K=1	K=3	K=5	K=10	K=20	K=30
1	0.343	0.344	0.346	0.351	0.354	0.358	0.193	0.196	0.199	0.204	0.216	0.227
2	0.329	0.463	0.465	0.470	0.473	0.476	0.181	0.261	0.264	0.269	0.282	0.291
3	0.361	0.531	0.532	0.536	0.539	0.543	0.189	0.306	0.308	0.313	0.325	0.335
4	0.383	0.544	0.577	0.581	0.584	0.587	0.190	0.311	0.335	0.339	0.349	0.362
5	0.398	0.555	0.613	0.617	0.620	0.623	**0.197**	0.317	0.362	0.367	0.376	0.387
6	0.405	0.564	0.632	0.644	0.647	0.650	0.196	0.315	0.375	0.386	0.395	0.407
7	0.415	0.575	0.642	0.665	0.669	0.672	0.191	0.316	0.381	0.405	0.414	0.424
8	0.419	0.583	0.650	0.683	0.686	0.689	0.189	0.319	0.384	0.417	0.426	0.435
9	0.423	0.589	0.655	0.699	0.703	0.705	0.191	0.320	0.385	0.431	0.438	0.449
10	0.427	0.596	0.659	0.712	0.715	0.718	0.189	**0.321**	0.386	0.441	0.447	0.458
20	0.433	0.630	0.697	0.771	0.797	0.800	0.166	0.319	0.392	0.489	0.523	0.531
30	**0.435**	**0.639**	**0.714**	**0.789**	**0.831**	**0.835**	0.156	0.309	**0.393**	**0.490**	**0.561**	**0.573**

Misclassifications or unsuccessful retrievals (where the correct class is not within the Top-K=30 results) often occur due to the overlapping symptoms of many diseases. Additionally, the presence of identical words in the text, even if unrelated to symptoms, can artificially inflate similarity scores, leading to incorrect matches.

Table 3. Examples of misclassified data, including predicted classes and the most similar question texts.

Query (Test Data)	Most Similar Text	Similarity Score	True Label	Predicted Label	Analysis
Penyebab indra perasa dan penciuman terasa lebih sensitif setelah sembuh dari demam Dok saya wanita umur 25th Saya mau tanya indra perasa dan indra penciuman saya sejak saya sakit demam dari hari minggu hingga hari ini jumat jadi sangat sensitif sekali Kira2 kenapa yaa dok Selama saya sakit saya hanya minum obat2 penurun demam seperti biasa diapotek	Penyebab gangguan pada indera perasa dan penciuman Dok saya kemarin habis berlibur kepuncak setelah berlibur saya pulang kerumah sakit dan mengalami panas badan tidak enak tapi setelah berobat itu sembuh dok Nah masalahnya knapa ya Indara perasa saya sama penciuman tidak normal dok mohon jawabannya	82.5%	Sindrom Cushing	Flu	Both questions focus on changes in the senses of taste and smell after recovering from a fever or illness.
Penyebab mual disertai lemas mata nyeri panas naik turun pegalpegal dan tenggorokan kering Saya memiliki gejala mual lemes mata sakit panas naik turun pegal pegal Tenggorokan kering Menurut ciri ciri diatas saya terkena penyakit apa ya dok Terimakasih	Penyebab demam disertai pusing mata perih dan bab cair Saya mau bertanya dok Tadi pagi tiba2 saya sakit padahal habis makan panas tinggi terasa terbakar kepala depan pusing mata terasa perih kalo berdiri gak kuat Buang air besar mencret badan terasa linu Saya sudah berobat Belum ada reaksi	84.6%	Tifus	Chikungunya	Both questions involve eye discomfort, body aches, and gastrointestinal issues. Tifus and Chikungunya often share same symptoms [24]

Query (Test Data)	Most Similar Text	Similarity Score	True Label	Predicted Label	Analysis
	sembuh Mohon bantuannya Makasih				
Bercak putih pada gusi Slmt mlm dok saya mau tanya Bbrp hari lalu gusi sblh kiri bawah saya ada bercak putih dan sekitar bercak itu sedikit memerah kira itu kenapa ya dok Tp saya tdk mengetahui pasti sudah bbrp lama bercak itu ada yg saya rasakan saat terkena mknan seperti sariawan tp mlm harinya bercak itu sudah hilang Di gusi kanan bawah saya jg ada kyk daging tumbuh dan itu sudah lama dok dan tdk ada keluhan apapun Kira itu kenapa ya dok Berbahaya atau tidak ya dok	Penyebab muncul benjolan putih di bawah lidah yang berulang Mau tanya dok Beberapa bulan lalu muncul benjolan kecil di bawah lidah Benjolannya itu berwarna putih Karena sedikit mengganggu akhirnya saya congkel dengan kuku dan hilang Beberapa hari kemudian muncul lagi dan lebih besar seperti ada cairan putih macam jerawat Ini kenapa ya dok	88.7%	Infeksi Gusi	Mucocele	Both questions revolve around oral health concerns involving white lesions. The similarity lies in the concern about recurring white lesions in the mouth and the desire to understand whether they are harmful
sakit pada telapak kaki bagian tengah dok saya mau tanya akhirakhir ini mama saya mengeluh sakit pada telapak kaki bagian tengahbiasanya terasa pada saat bangun tidur menurut dokter mama saya menderita penyakit apa	Bengkak di kaki yang terasa nyeri selamat pagi dok mau tanya dok ibu saya itu mengalami pembengkakkan di kakinya dan kadang itu terasa nyeri tapi kadang bengkak itu hilang tapi kadang di lain hari muncul lagi bengkaknya itu apa penyakit asam urat ya dok	90.7%	Kapalan	Asam Urat	Both questions involve foot-related discomfort and recurring symptoms. The similarity lies in the concern about foot pain and the desire to understand the cause
penyebab jari kaki sebelah kiri kebas setelah berjalan jauh Halo Dok Saya ingin bertanya Selepas 2 hari lalu saya melakukan traveling dan selalu berjalan kaki jauh Hal ini menimbulkan kebas pada jari kaki kiri saya yang timbul sejak 2 hari lalu dan sampai saat ini kebas masih terasa dan belum kunjung hilang Mohon info nya mengenai masalah ini Terima Kasih	Nyeri jari kaki saat dibuat berjalan Dok beberapa hari lalu sya mengalami sakit pada telunjuk kaki kiri rasanya nyeri tapi sakitnya hanya muncul saat saya berjalan itu saja dok pertanyaan saya mohon penjelasanya	88.3%	Neuropati Diabetik	Asam Urat	Both questions involve discomfort in the toes of the left foot that is related to walking. The similarity lies in the concern about toe-related symptoms and their association with walking

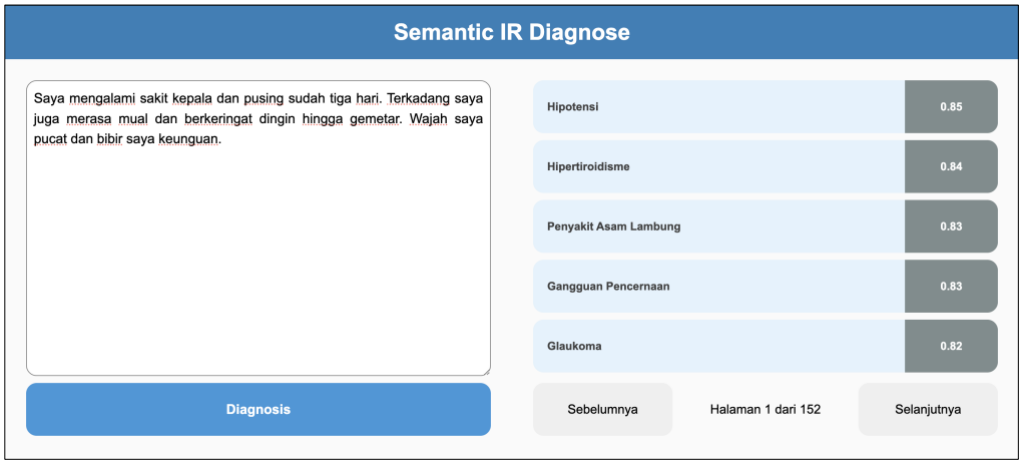


Figure 10. Screenshot of the deployed system in the web application

Table 3 presents five randomly selected examples of misclassified data, highlighting cases where the query text and the most similar corpus text share overlapping words and contextual meanings. The primary cause of misclassification is the high similarity of symptoms across different diseases. To improve accuracy, future queries should be more specific, ensuring clearer distinctions between closely related conditions.

At the deployment stage, the model was successfully integrated into a web application using the REST API architecture, built with the FastAPI framework. The system calculates semantic similarity between user queries and the entire corpus, categorizing results by class. The final user interface of the application is shown in Figure 10.

4. CONCLUSION

This research developed an information retrieval system using Large Language Models (LLMs) to address the critical challenges of diagnostic accuracy and healthcare access in Indonesia. By replacing traditional methods like TF-IDF with advanced LLMs such as MPNet, the system achieved an accuracy score of 0.911 and a balanced accuracy score of 0.687 at Top-K=30, significantly outperforming existing approaches. The system's ability to capture contextual meaning and generate compact, context-aware vectors directly addresses the limitations of traditional methods highlighted in the research background.

Deployed as a user-friendly web application, the system provides a scalable and accessible tool for disease prediction, offering a practical solution to the shortage of healthcare professionals and the need for accurate, collaborative diagnostics. While challenges such as overlapping symptoms and class imbalance remain, this research demonstrates the potential of LLMs to revolutionize healthcare information retrieval, paving the way for more efficient and equitable healthcare services.

ACKNOWLEDGEMENTS

This research was supported by DRTPM Ministry of Education, Culture, Research, and Technology of Indonesia through Penelitian Dosen Pemula Grant scheme in 2024, under contract number 105/E5/PG.02.00.PL/2024; 802/LL3/AL.04/2024; 014/LPPM-SRT/UK/VI/2024. Additionally, the authors acknowledges the invaluable contributions of all other parties who have directly or indirectly supported the successful completion of this research.

REFERENCES

[1] N. Ghaffar Nia, E. Kaplanoglu, and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," *Discover Artificial Intelligence*, vol. 3, no. 1, p. 5, Jan. 2023, doi: 10.1007/s44163-023-00049-5.

[2] T. A. Sugiyatmi, U. Hadi, D. Chalidyanto, F. Hafidz, and M. Miftahussurur, "Does the implementation of national health insurance affect the workload of a doctor and have an impact on service quality? A systematic literature review," *J Public Health Afr*, Oct. 2019, doi: 10.4081/jphia.2019.1198.

[3] F. M. Ekawati and M. Claramita, "Indonesian General Practitioners' Experience of Practicing in Primary Care under the Implementation of Universal Health Coverage Scheme (JKN)," *J Prim Care Community Health*, vol. 12, p. 215013272110237, Jan. 2021, doi: 10.1177/21501327211023707.

[4] C. Maharani, S. R. Rahayu, M. Marx, and S. Loukanova, "The National Health Insurance System of Indonesia and primary care physicians' job satisfaction: a prospective qualitative study," *Fam Pract*, vol. 39, no. 1, pp. 112–124, Jan. 2022, doi: 10.1093/fampra/cmab067.

[5] R. Pratama and A. Yufika, "Physicians' Workload and Quality Healthcare in Indonesia," *Trends in Infection and Global Health*, vol. 3, no. 1, pp. 43–55, Jun. 2023, doi: 10.24815/tigh.v3i1.32363.

[6] M. L. Barnett, D. Boddupalli, S. Nundy, and D. W. Bates, "Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians," *JAMA Netw Open*, vol. 2, no. 3, p. e190096, Mar. 2019, doi: 10.1001/jamanetworkopen.2019.0096.

[7] N. P. Tigga and S. Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods," *Procedia Comput Sci*, vol. 167, pp. 706–716, 2020, doi: 10.1016/j.procs.2020.03.336.

[8] M. A. J. Tengnah, R. Sooklall, and S. D. Nagowah, "A Predictive Model for Hypertension Diagnosis Using Machine Learning Techniques," in *Telemedicine Technologies*, Elsevier, 2019, pp. 139–152. doi: 10.1016/B978-0-12-816948-3.00009-X.

[9] S. Grampurohit and C. Sagarnal, "Disease Prediction using Machine Learning Algorithms," in *2020 International Conference for Emerging Technology (INCET)*, IEEE, Jun. 2020, pp. 1–7. doi: 10.1109/INCET49848.2020.9154130.

[10] P. Hamsagayathri and S. Vigneshwaran, "Symptoms Based Disease Prediction Using Machine Learning Techniques," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, IEEE, Feb. 2021, pp. 747–752. doi: 10.1109/ICICV50876.2021.9388603.

[11] P. Hema, N. Sunny, R. Venkata Naganjani, and A. Darbha, "Disease Prediction using Symptoms based on Machine Learning Algorithms," in *2022 International Conference on Breakthrough in Heuristics And Reciprocation of Advanced Technologies (BHARAT)*, IEEE, Apr. 2022, pp. 49–54. doi: 10.1109/BHARAT53139.2022.00021.

[12] A. Divya, B. Deepika, C. H. Durga Akhila, A. Tonika Devi, B. Lavanya, and E. Sravya Teja, "Disease Prediction Based on Symptoms Given by User Using Machine Learning," *SN Comput Sci*, vol. 3, no. 6, p. 504, Oct. 2022, doi: 10.1007/s42979-022-01399-0.

[13] J. H. Kamdar, J. Jeba Praba, and J. J. Georrge, "Artificial Intelligence in Medical Diagnosis: Methods, Algorithms and Applications," 2020, pp. 27–37. doi: 10.1007/978-3-030-40850-3_2.

[14] S. Haque, Z. Eberhart, A. Bansal, and C. McMillan, "Semantic similarity metrics for evaluating source code summarization," in *Proceedings of the 30th IEEE/ACM International Conference on Program Comprehension*, New York, NY, USA: ACM, May 2022, pp. 36–47. doi: 10.1145/3524610.3527909.

[15] A. Aszani, H. I. Wicaksono, U. Nadzima, and L. Heryawan, "Information Retrieval for Early Detection of Disease Using Semantic Similarity," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 1, p. 45, Feb. 2023, doi: 10.22146/ijccs.80077.

[16] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," Aug. 2019.

[17] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter," Oct. 2019.

[18] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, "MPNet: Masked and Permuted Pre-training for Language Understanding," Apr. 2020.

[19] W. Wang, H. Bao, S. Huang, L. Dong, and F. Wei, "MiniLMv2: Multi-Head Self-Attention Relation Distillation for Compressing Pretrained Transformers," Dec. 2020.

[20] E. De Santis, A. Martino, F. Ronci, and A. Rizzi, "From Bag-of-Words to Transformers: A Comparative Study for Text Classification in Healthcare Discussions in Social Media," *IEEE Trans Emerg Top Comput Intell*, pp. 1–15, 2024, doi: 10.1109/TETCI.2024.3423444.

[21] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining."

[22] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The Balanced Accuracy and Its Posterior Distribution," in *2010 20th International Conference on Pattern Recognition*, IEEE, Aug. 2010, pp. 3121–3124. doi: 10.1109/ICPR.2010.764.

[23] T. Wolf *et al.*, "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2020, pp. 38–45. doi: 10.18653/v1/2020.emnlp-demos.6.

[24] S. Pathak, N. Chaudhary, P. Dhakal, S. R. Yadav, B. K. Gupta, and O. P. Kurmi, "Comparative Study of Chikungunya Only and Chikungunya-Scrub Typhus Coinfection in Children: Findings from a Hospital-Based Observational Study from Central Nepal," *Int J Pediatr*, vol. 2021, pp. 1–6, Apr. 2021, doi: 10.1155/2021/6613564.