

Intelligent preliminary diagnosis system for diseases with similar clinical presentation

Laberiano Andrade-Arenas¹, Cesar Yactayo-Arias²

¹Facultad de Ciencias e Ingeniería, Universidad de Ciencias y Humanidades, Lima, Perú

²Departamento de Estudios Generales, Universidad Continental, Lima, Perú

Article Info

Article history:

Received Jan 16, 2024

Revised Feb 16, 2024

Accepted Feb 20, 2024

Keywords:

COVID-19

Dengue

Diagnostic system

Disease

Healthcare

ABSTRACT

Accurately identifying diseases with similar symptoms, especially in resource-limited medical settings, is a key challenge to diagnostic accuracy. This paper presents a preliminary diagnostic system to address the challenge of diagnosing diseases with similar symptoms. The system has been implemented using the PyQt5 library and employs a unique symptom identification algorithm developed in Python. Furthermore, to carry out the diagnosis, it uses comma-separated values (CSV) and excel files as databases where the diseases and their respective symptoms are stored. The results show that the system has a precision of 95%, a sensitivity of 90%, and a specificity of 93% after evaluating 35 clinical cases covering seven diseases with similar initial symptoms, namely: dengue, zika, chikungunya, COVID-19, influenza, monkey-pox, and the common cold. Furthermore, the positive evaluation of the technical performance of the system by experts supports its practical feasibility and its potential as a valuable tool in medical practice. In conclusion, the system diagnoses diseases by analyzing the symptoms of the information file, highlighting its usefulness in improving the diagnostic accuracy of similar cases and optimizing medical care for the benefit of patients.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Cesar Yactayo-Arias

Departamento de Estudios Generales, Universidad Continental

Lima, Perú

Email: cyactayo@continental.edu.pe

1. INTRODUCTION

Modern medicine is faced with the constant challenge of providing accurate and timely diagnoses to ensure effective medical care and improve patient's quality of life. In addition, people are vulnerable and exposed to various diseases throughout their lives. Neglecting or ignoring some of these diseases because they are considered insignificant can have fatal consequences. Some diseases have a wide range of clinical manifestations, increasing the likelihood that two or more conditions will coexist [1]. For example, diseases caused by dengue, chikungunya, zika, and yellow fever viruses are closely linked to the behavior and development of the aedes aegypti mosquito [2]. Due to similarities in their clinical and epidemiological behavior, these diseases are often misdiagnosed [3], posing a serious threat to public health in tropical and subtropical areas [4]–[6].

In addition, there are others, such as leptospirosis, a disease that shares several symptoms with other diseases such as influenza, dengue, typhoid, rickettsiosis, aseptic meningitis, hepatitis, and malaria. Their similar symptoms make their immediate diagnosis difficult [7]. In this context, the efficient management and analysis of large amounts of clinical information becomes a critical task for informed health decision-making. On the other hand, traditional methods to collect and classify data have resulted in a disorganized

accumulation of information that hinders the effective analysis of diseases and slows down the diagnostic process. Therefore, it is crucial to find a method to prevent the onset of the disease [8]. In addition, it is essential to have automated tools for disease diagnosis.

However, the problem is that many current systems are not designed to efficiently handle the large amount of medical information available, especially in low-resource countries or regions where the medical diagnostic infrastructure is deficient. This can lead to delays in the diagnostic process, a lack of accuracy in results, and difficulties in making informed clinical decisions. Indeed, it is evident that current barriers to disease control are closely related to the lack of adequate diagnostic tools [9]. Therefore, it is important to recognize that accuracy in diagnosing people plays a fundamental role in the development of any solid and efficient health system [10]. For this reason, the solution proposed in this research consists of an innovative health system. Preliminary diagnosis is based on the use of files in comma-separated values (CSV) and Excel format to upload and organize data about diseases and symptoms. This system has been designed to speed up and facilitate the diagnosis process, allowing orderly and rapid management of clinical information.

The importance of this study lies in the need to improve diagnostic accuracy and optimize disease management. The lack of efficient technological solutions has hindered informed clinical decision-making and the use of relevant data, especially in low-resource settings. The implementation of a preliminary diagnostic system based on CSV and Excel files is presented as a practical and accessible solution to address these challenges by allowing greater flexibility in data management for diagnostic purposes. By improving the diagnostic capacity of health professionals, including those at the beginning of their careers, the tool aims to reduce misdiagnosis and provide more accurate and personalized medical care. The rigorous evaluation of the effectiveness and accuracy of the system will demonstrate its validity and reliability in the diagnostic process, highlighting its potential to positively transform medical practice and provide excellent care to patients.

2. LITERATURE REVIEW

The purpose of this section is to enrich research by collecting relevant information related to previous studies on disease diagnosis. This review is intended to provide a deeper understanding of the proposed solutions and technologies used in the field of disease diagnosis. In this way, a broader panorama will be obtained for the investigation.

Yu *et al.* [11] used conventional microscopy to diagnose malaria in endemic areas but noted the lack of trained microscopists and the propensity for errors. In response, they evaluated Malaria Screener, a smartphone application, in a study in rural Sudan. The application achieved a diagnostic accuracy of 74.1%. On the other hand, Yazdani *et al.* [12] focused on addressing the high prevalence of upper respiratory diseases in children. They used a novel approach to design an expert system, integrating logical criteria and variable evocative strength to improve diagnostic accuracy. In their retrospective study, they reviewed the electronic medical records of children with upper respiratory diseases at a university health system in the United States. The average accuracy of their tool for the initial diagnosis was 75%, and it reached 84% for matching one of the two major differential diagnoses. They concluded that the integration of logical diagnostic criteria and variable evocativeness into expert systems can significantly improve accuracy. Similarly, Dansy *et al.* [13] addressed the broad global impact of COVID-19 disease, recognizing that the similarity of its symptoms to other conditions, such as influenza and the common flu, complicates its early detection. To address these challenges, she proposed an expert system for the early detection of corona. Her approach included interviews with medical professionals to gather data on the symptoms of these diseases, along with a review of peer-reviewed articles. The certainty factor method was used during the research.

Similarly, Melendez-Acosta *et al.* [14] presented a fuzzy system to assess the level of infection of viral diseases such as dengue, zika, chikungunya, and yellow fever. Using six relevant symptoms and fuzzy sets implemented in a matrix laboratory (MATLAB), the system achieved an accuracy of 86.6% in tests with 15 cases of viral diseases, highlighting its usefulness in early diagnosis. In a similar approach, Sultana *et al.* [15] addressed the early detection of the chikungunya virus (CHIKV) using a belief rule-based expert system (BRBES). She applied BRBES to real-world data and compared its performance with deep learning and machine learning models, demonstrating that BRBES is an efficient and fast option for assessing CHIKV in the early stages of infection.

Likewise, de Araújo *et al.* [16] contributed to the development of an expert system in the form of an Android application for the diagnosis of dengue, zika, and chikungunya. This expert system associated symptoms with a probability score of suffering from these diseases, with a success rate of 96.88% in tests with 96 cases, exceeding the success rate of resident physicians. In the area of disease surveillance, Kim *et al.* [17] evaluated the use of parent-reported data through the fever coach app for real-time detection of influenza activity. The app predicted influenza activity with a significant correlation ($\rho = 0.878$) compared

to data from the Korea center for disease control and prevention, detecting the epidemic 10 days before the official alert. On the other hand, Karsi *et al.* [18] proposed an approach to improve classification in automated health surveillance systems. Using deep contextualized word embeddings from embeddings from language models (ELMo) to enrich training samples with semantically similar terms and a weighting scheme, their improved model, evaluated with support vector machine (SVM) on the i2b2 dataset, showed a significant improvement in classification, achieving an F-measure of 86.54%. These investigations provide different perspectives and approaches for improving disease diagnosis and monitoring.

In conclusion, various researchers have presented solutions for the diagnosis of diseases such as chikungunya, COVID-19, malaria, influenza, and others, using approaches ranging from rule-based expert systems to the use of certainty factors and machine learning. These investigations have yielded positive results, although they have mostly focused on the diagnosis of one or, in some cases, two specific diseases. Therefore, this research develops a diagnostic system capable of addressing any disease with similar symptoms, which represents a significant advance in the field.

3. METHOD

3.1. System development method

The development of the diagnostic system for diseases with similar symptoms has been divided into four essential stages, as shown in Figure 1, based on the stages of the rational unified process (RUP) methodology, a method that provides a structured way for the development of the software [19]. These stages collectively formulate a comprehensive strategy for the successful establishment of the mentioned diagnostic system. In the development of the system, several fundamental stages were carried out, which are detailed below.

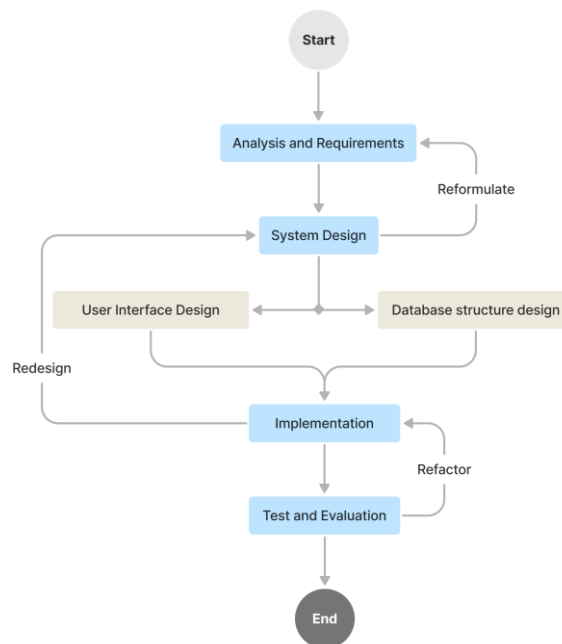


Figure 1. Development phases of the proposed system

In the initial analysis and requirements phase, a thorough analysis of the system requirements was carried out, identifying the essential functionalities for the loading and processing of data related to diseases and symptoms from CSV and Excel files. Likewise, the objectives, scope, use cases, and workflows of the system were established. In the subsequent stage of system design, detailed planning was carried out that included creating a user interface and specifying the layout of elements such as windows, buttons, and input fields. The structure of the database responsible for storing relevant information on diseases and symptoms was also designed. The third phase, implementation, focused on coding the system according to the previously conceived design, developing the functionality for loading and viewing data from CSV and Excel files, and incorporating the preliminary diagnosis logic that enables the identification of diseases from the symptoms provided by users. Finally, in the test and evaluation stage, extensive tests were carried out to ensure the correct functioning of the system. The accuracy of data loading from CSV and Excel files was

validated, the accuracy of the preliminary diagnosis was evaluated, and possible errors identified were addressed and corrected to ensure the integrity and efficiency of the system.

3.2. Diagnosis inference method

The algorithm developed for the disease diagnosis system is based on a unique symptom identification algorithm. The algorithm first accesses the extensive database that associates specific symptoms with various diseases, creating a set of stored symptom relationships. Then, after receiving the user's symptoms, the algorithm meticulously compares each symptom entered with the symptom profiles associated with each disease in the database. The weighting of the match is based on the relevance and frequency of the symptoms in each case. Through this process, the algorithm identifies and selects the disease whose symptoms best match those provided by the user, thus providing an accurate diagnostic result. This approach ensures a thorough and personalized evaluation, highlighting the most likely disease based on the patient's symptom presentation.

3.3. Evaluation

3.3.1. Confusion matrix

The confusion matrix was used to evaluate the effectiveness of the system. The matrix is organized in rows and columns [20] and provides a detailed analysis of the results, classifying them into four categories [21]–[23]. True positives (TP), are correct predictions of positive events; false positives (FP), are incorrect predictions of positive events when they are negative; true negatives (TN), are accurate predictions of negative events; and false negatives (FN), are incorrect predictions of negative events when, they are positive.

3.3.2. Performance metrics

Performance metrics such as precision, sensitivity, and specificity were used to analyze its performance and ability to ensure reliable and accurate results. This methodology allowed us to evaluate the effectiveness of the system in accurately detecting positive cases and accurately rejecting negative cases, thus providing a comprehensive evaluation of its diagnostic capacity. Each metric is described below:

Precision: Evaluate the proportion of correct diagnoses concerning the total number of diagnoses generated by the system. In simple terms, it represents the proportion of values identified as positive that were correctly predicted [24], [25]. This indicator can be calculated using (1) and provides an essential quantitative measure for the overall evaluation of the performance of the diagnostic system.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

Sensitivity: evaluates the ability of the system to positively detect true disease among all disease cases. The higher the sensitivity, the fewer false negatives [26]. As sensitivity increases, the likelihood of false negatives decreases. This parameter can be calculated using (2) and provides an essential quantitative measure for evaluating the performance of the diagnostic system.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (2)$$

Specificity: measures the system's ability to avoid incorrect positive diagnoses in individuals who do not suffer from the disease; the higher the specificity, the fewer false positives [27]. The higher the specificity, the lower the probability of false results. This parameter can be calculated using (3) and provides an essential quantitative metric for evaluating the performance of a diagnostic system.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (3)$$

3.3.3. Expert validation

To evaluate the technical aspects of the system, a survey was implemented consisting of an online questionnaire with a total of ten questions, as detailed in Table 1. These questions were designed using a Likert scale, which allowed the experts to express their level of agreement or disagreement with the different functionalities offered by the system. After obtaining the results, they were analyzed using descriptive statistical techniques. This included calculating the average of the answers provided by the experts to each question, as well as determining the standard deviation (SD) corresponding to each of the questions. These calculations allow a precise and detailed assessment of the level of efficiency of the system based on the opinions and perceptions of the experts involved in the evaluation.

Table 1. Technical aspect

Id	Technical aspect	Id	Technical aspect
1	Data import capability	6	Error handling
2	System performance	7	Data loading validation
3	User experience optimisation	8	Result export functionality
4	Diagnosis accuracy and clarity	9	Symptom input validation
5	Symptom diagnosis algorithm	10	System scalability

3.4. Development tools

For the development of the system, Python was chosen, a widely recognized programming language [28] and used in the field of medical informatics. In addition, PyQt5 was used as the main library for creating the graphical interface. PyQt5, a Python interface to the Qt library, can be used to develop software that takes up less disk space and runs faster, making it one of the most used modules for creating graphical user interfaces (GUIs) in Python due to its simplicity [29], [30], which allowed an intuitive presentation of the diagnostic results. Python will be used to implement medical data analysis and processing algorithms, taking advantage of its ability to manipulate information and the wide availability of specialized libraries in the health sciences. The choice of these tools is justified by their efficiency, versatility, and ability to facilitate the development of an accurate and accessible system for diagnosing diseases in a variety of clinical contexts.

4. SYSTEM DEVELOPMENT

This section presents a detailed description of the crucial stages in the development of the preliminary diagnostic system. The detailed exploration facilitates a more accurate and complete understanding of the design and construction of the innovative tool designed to assist in the diagnosis of various diseases. The presentation aims to provide a deeper insight into the approaches and foundations of the advanced medical technology in question.

4.1. Analysis and requirements

4.1.1. Use case

The use case diagram of the disease diagnosis system covers several essential use cases, as shown in Figure 2. The most prominent of these is the upload CSV/Excel File use case, which allows users to import disease and symptom records for diagnosis. The open diagnostic window use case allows the user to open a new interface for performing diagnostics. The perform diagnosis use case, in turn, allows users to enter symptoms and obtain diagnoses based on this information, even offering the possibility to perform multiple diagnoses according to their individual needs. Furthermore, the export result use case provides the option to save the diagnostic results in PDF format for later review. The system incorporates error detection through the validate symptom use case, which alerts the user if they fail to enter symptoms before proceeding. Similarly, show result displays the corresponding disease diagnosis based on the symptoms provided. Finally, cancel diagnosis allows the user to cancel the diagnostic process and exit the window.

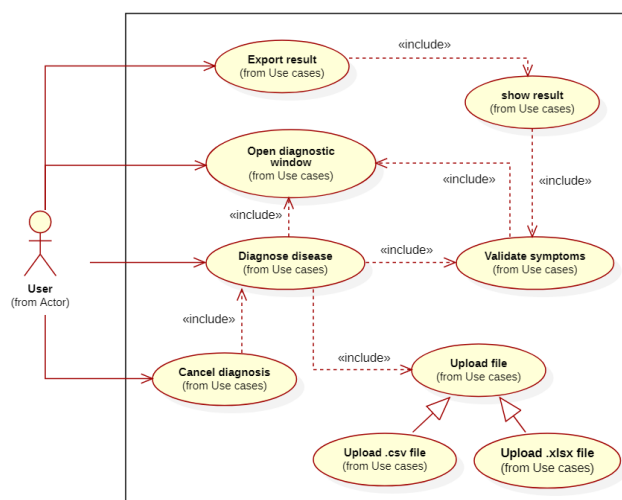


Figure 2. System use case diagram

4.1.2. Diagnostic process flowchart

Figure 3 shows a visual diagram that illustrates the workflow of the disease diagnosis process. This diagram not only serves to clarify the process itself but also adds significant value to the system by guiding the user through several critical stages of diagnosis. In addition to its usefulness to the system, it provides a clear and consistent structure that optimizes the user experience during the diagnostic process. This flow provides the user with intuitive guidance that ranges from loading data from CSV and Excel files to entering the patient's symptoms and presenting the corresponding results. This ensures a smooth and well-structured diagnostic process.

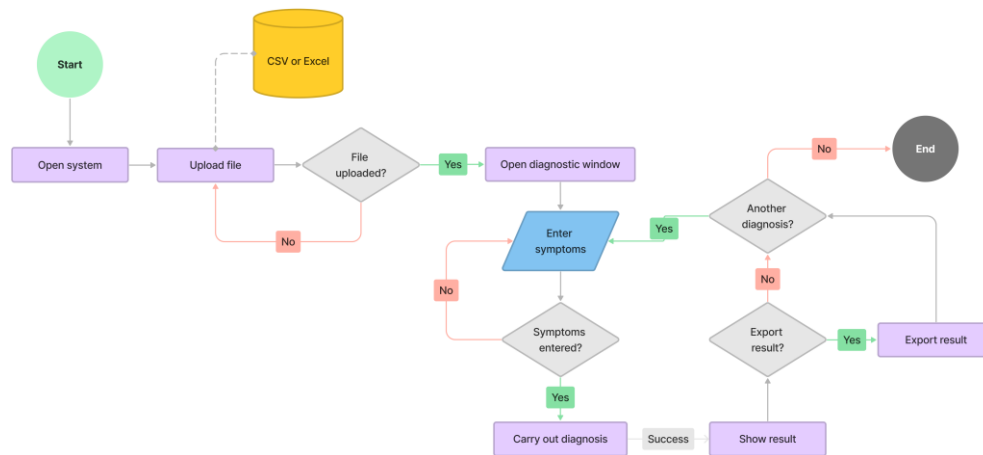


Figure 3. Comprehensive diagnostic process flowchart illustrating the sequential steps and decision points in the diagnostic procedure

4.2. Design

4.2.1. User interface design

The user interface, shown in Figure 4, has been designed with key elements such as a menu bar consisting of two main sections: file and diagnostics. The file section offers three options, namely: i) upload from CSV, which facilitates uploading files in CSV format; ii) upload from Excel, which allows uploading files in Excel format; and iii) exit, which allows closing the system. Once the file is loaded, it is displayed in the main view in the form of a table, organized in rows and columns, with a title and, at the bottom, the number of rows and columns present. The diagnostics section contains the perform diagnostics option, which, when selected, opens a new interface with a text box and a button. Here, the user can enter symptoms and proceed with the diagnosis. The diagnostic result is displayed in a confirmation pop-up window, with the option to save the result.

Disease and its symptoms	
Disease	Symptoms
1 Zika	mild fever
2 Zika	rash
3 Zika	conjunctivitis
4 Zika	muscle pain
5 Zika	joint pain
6 Zika	headache
7 Zika	general malaise
8 Dengue	high fever
9 Dengue	moderate fever
10 Dengue	headache
11 Dengue	pain behind the ...
12 Dengue	muscle pain
13 Dengue	joint pain

Row: 93 Column: 2

Figure 4. Intuitive user interface design for enhanced user experience

4.2.2. Data structure

In the context of data structure, the system relies on files in CSV or Excel format as the main source of information. These files store detailed information about diseases and their symptoms, following an organization of rows and columns, as shown in Figure 5 for both file formats. Figure 5(a) illustrates the structure in CSV format, while Figure 5(b) presents the corresponding provision in Excel format. Regarding the columns, the first contains the names of the diseases, while the second contains the list of symptoms associated with each disease. As for the rows, the first row contains unique titles: disease and symptoms. In the second line, each disease is listed along with its symptoms. Note that the name of the disease is repeated depending on the number of symptoms associated with it.

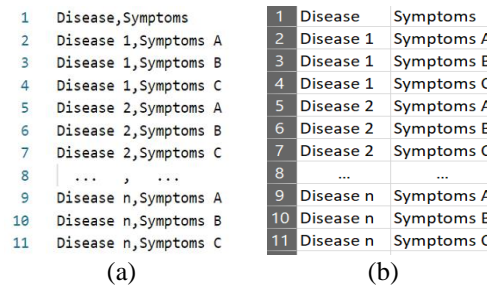


Figure 5. Illustration of information storage structure in (a) CSV and (b) Excel formats, highlighting data organization and storage schema

4.3. Implementation of system diagnostic logic

The diagnostic logic implemented in the system is based on a unique symptom identification algorithm, which follows a structure composed of five fundamental stages, detailed in Figure 6. Each of these stages is detailed in the subsequent sections, providing a thorough understanding of the diagnostic process and strengthening the reliability of the system for the accurate detection of medical conditions.

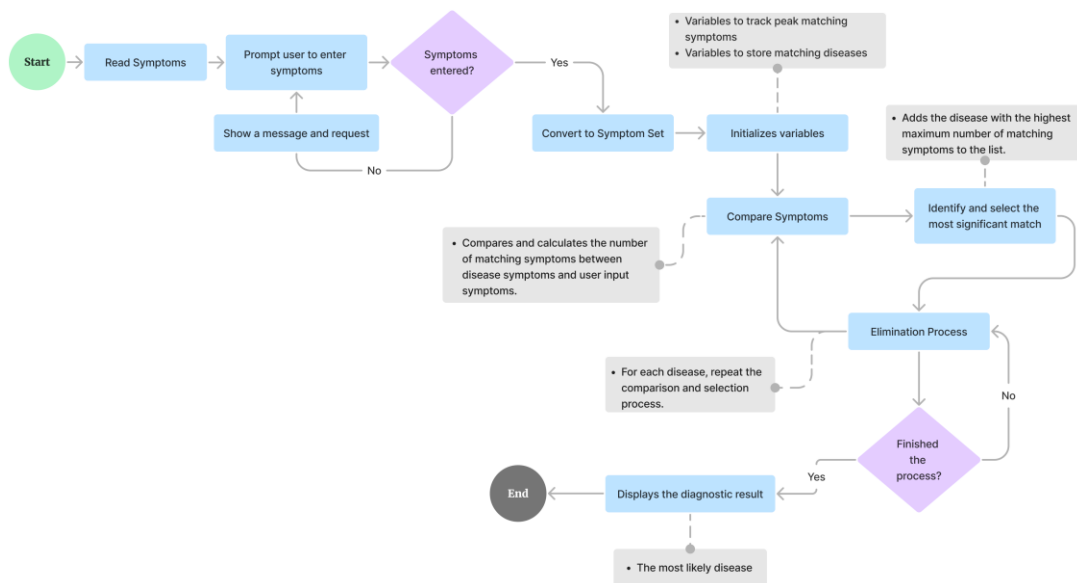


Figure 6. Unique symptom identification algorithm process

4.3.1. Reading and transformation into a set of symptoms

The system acquires the symptoms provided by the user, which are then transformed into a set of symptoms to facilitate their analysis in subsequent phases. This collection and transformation process lays the foundation for more detailed evaluation, allowing the system to accurately address the information provided and improve diagnostic effectiveness.

4.3.2. Initialization of variables

The system begins with the initial configuration of two essential variables. The first variable stores the number of symptoms that match the diagnostic guidelines, while the second variable is used to track the maximum number of matches between symptoms. This configuration of variables provides a structured basis for efficient evaluation and tracking of data during the diagnostic process, contributing to greater accuracy in identifying potential medical conditions.

4.3.3. Comparative analysis of the symptoms

Decisions are made based on the specific characteristics of each disease. The diagnostic logic focuses on the identification of common symptoms between the disease stored in the database and the symptoms provided by the user. At this critical point, a meticulous calculation is performed to determine the exact number of symptoms that match the disease under consideration and those provided by the user, thus consolidating the evaluation and allowing a more rigorous interpretation of the medical situation.

4.3.4. Identification and election of the most significant coincidence

This stage focuses on the precise identification of the disease that presents a greater coincidence of symptoms for those indicated by the user. To achieve this, the system searches for the disease that has the highest number of matching symptoms, exceeding all previously found matches. As a result, the list of diseases is refined to include only those with the highest number of matches.

4.3.5. Symptom-based elimination process

The final step involves a progressive cleansing process. Through a series of comparisons, diseases that do not match the symptoms provided by the user are eliminated. This phase involves an exhaustive analysis of all existing diseases to determine which ones match the maximum number of coinciding symptoms. Those that do not meet the predefined criteria are excluded from the final set of matching diseases. At the end of this process, the system displays the disease that has the closest symptom match based on the symptoms provided by the user.

5. RESULTS

5.1. User interaction with the developed system

The visual representation provided in Figure 7 graphically and sequentially details the various phases inherent in the disease diagnosis process, providing a more complete and detailed view of the stages involved in this diagnostic process. Figure 7(a) shows the user interaction with the system, emphasizing its diagnostic capacity. The user can enter the patient's symptoms into the system, which then performs a comprehensive analysis to identify possible diseases associated with the symptoms. On the other hand, Figure 7(b) shows the results of the diagnostic process, presenting potential diseases based on the provided symptoms. Additionally, the system allows users to save these results in PDF format, preserving critical information such as the disease identified and the symptoms provided. It's important to emphasize that this diagnosis is preliminary and should not replace a healthcare professional's evaluation.

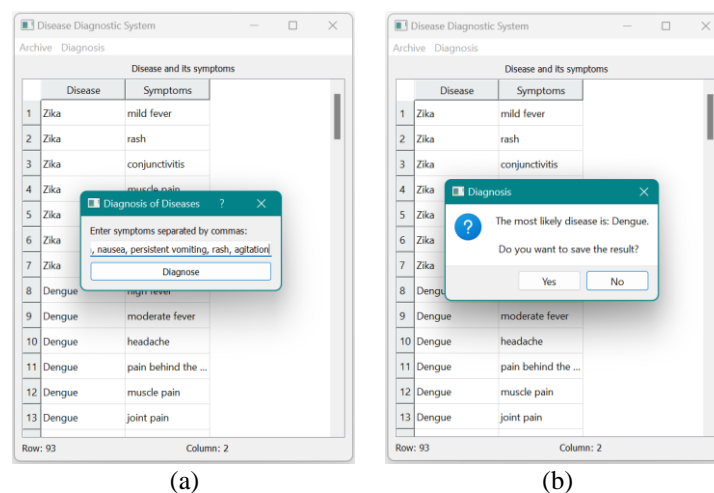


Figure 7. User interface (a) input of symptoms for diagnosis and (b) diagnostic result

5.2. Database and selection of evaluation diseases

5.2.1. Evaluation of disease size

The evaluation sample consists of 35 cases, divided into 21 positive cases and 14 negative cases, covering seven different diseases: dengue, zika, chikungunya, COVID-19, influenza, monkeypox, and the common cold. Each disease is represented by three positive cases, allowing for an equitable analysis of diseases with similar symptoms. In addition, two negative cases are included for each of the seven diseases. This sample size design has been carefully designed to provide a broad and diverse representation of the diseases under study, allowing for a robust and informed evaluation of the performance of the diagnostic system in contexts with similar symptoms.

5.2.2. Test database

The database has been built through the collection of symptoms associated with the diseases considered in the evaluation. These symptoms have been obtained from highly reliable sources, specifically the official websites of two widely recognized health organizations: the WHO and the pan american health organisation (PAHO). The database includes a total of 93 different symptoms, with a varying number of symptoms associated with each disease. In addition, the database is an essential resource for the analysis and evaluation of the diagnostic system, allowing rigorous validation of the system's performance based on its ability to identify and associate symptoms with specific diseases.

5.3. Performance evaluation in system diagnostics

In this process, the 35 selected cases covering the 7 diseases under consideration were provided to the system, to have the system identify the corresponding disease in each of these cases. The results obtained were subjected to an exhaustive analysis through the confusion matrix and the evaluation of various performance metrics. We will carry out an analysis of them.

5.3.1. Confusion matrix analysis

Figure 8 shows the analysis of the evaluation using the confusion matrix. As can be seen, of the 21 positive cases, the system correctly identified 19 as true positives but identified two cases as false negatives. Similarly, of the 14 negative cases, the system identified 13 as true negatives but identified one as a false positive. These results indicate that the system has good performance in correctly identifying diseases with similar initial symptoms.

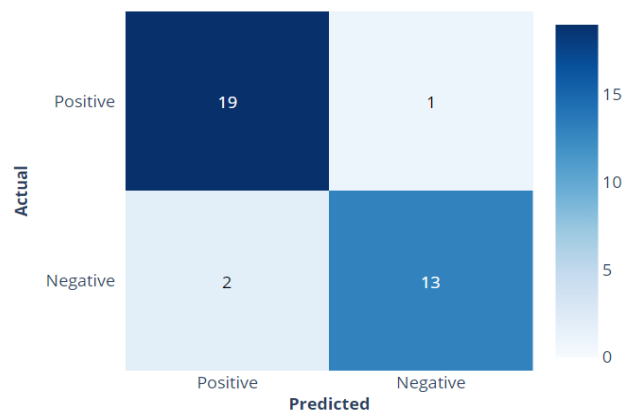


Figure 8. Confusion matrix analysis

5.3.2. Performance metrics evaluation

The results of the system evaluation show very promising performance in terms of its ability to identify and categorize positive and negative cases, as shown in Figure 9. With an outstanding sensitivity of 90%, the system demonstrates a significant ability to accurately detect the vast majority of positive disease cases evaluated in the test. This ability is critical to ensuring early and effective diagnosis, which can have a profoundly positive impact on patient health. Similarly, the 93% specificity indicates that the system can correctly distinguish the majority of negative cases, minimizing the likelihood of false positives. In addition, the robust precision of 95% emphasizes that when the system makes a positive diagnosis, there is a high probability that the diagnosis is correct.

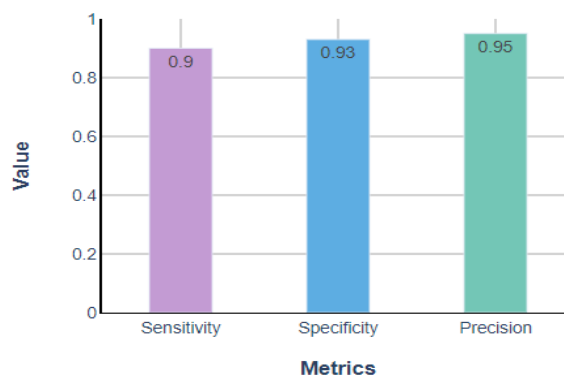


Figure 9. Performance metrics

5.4. System technical performance evaluation

The evaluation of the technical performance of the system was carried out by a group of five experts, who gave ratings on a scale ranging from 1 to 5, where 1 indicates very bad and 5 indicates very good. This evaluation process was carried out through an online questionnaire that consisted of 10 predefined technical aspects, as detailed in Table 1. Subsequently, the results were analyzed, calculating both the average score and the standard deviation of the ratings.

5.4.1. Analysis of mean and standard deviation

Table 2 shows the average of the scores assigned to each of the technical aspects evaluated. It can be seen that the average of the ratings for each technical aspect exceeds the value of 4 on a scale of 1 to 5. Similarly, the low variability of the ratings reflects the low standard deviation of the responses provided by the experts for each of the technical aspects evaluated. In particular, the overall average rating for the chatbot is 4.78, indicating a positive and widespread perception among experts. This score suggests that the chatbot has proven to be effective and satisfactory in terms of interaction, functionality, comprehension, and delivery of relevant answers, among others. However, the standard deviation of 0.42 shows moderate variability in the experts' responses, suggesting that although most evaluations are positive, there are some differences of opinion among evaluators on certain aspects. These results support the effectiveness of the chatbot as a valuable tool for supporting mental well-being, with room for further refinement and improvement in specific areas identified by the experts.

Table 2. Evaluation result of technical aspects of the system

Id	Technical aspect	Average	SD
1	Data import capability	5.00	0.00
2	System performance	5.00	0.00
3	User experience optimisation	5.00	0.00
4	Diagnosis accuracy and clarity	4.44	0.49
5	Symptom diagnosis algorithm	4.44	0.49
6	Error handling	4.80	0.45
7	Data loading validation	5.00	0.00
8	Result export functionality	5.00	0.00
9	Symptom input validation	5.00	0.00
10	System scalability	4.60	0.55

6. DISCUSSION

6.1. Main findings

This study evaluated a preliminary diagnosis system for various diseases based on a unique symptom identification algorithm. Although previous research has evaluated disease diagnosis systems based on rules, and fuzzy logic, none has specifically addressed a system capable of diagnosing multiple diseases. The study's results show that the system, created with Python and PyQt5, achieved a precision of 95%, a sensitivity of 90%, and a specificity of 93% when assessing 35 clinical cases involving seven different conditions with similar initial symptoms. Comparing these findings with past research, it is clear that the proposed system excels in terms of accuracy when compared to similar tools like the Malaria Screener [11], which measured 74.1% accuracy in detecting *Plasmodium falciparum*. Additionally, the proposed system

surpasses other tools that have utilized innovative methods, such as logical criteria fusion and variable evocation strength [12], which averaged 75% precision for initial diagnosis. Likewise, a fuzzy system implemented in MATLAB for analyzing the infection level in viral diseases achieved a precision rate of 86.6 [14], indicating that the system developed in this study performs comparably in the early detection of viral diseases. However, the rule-based expert system for arbovirus diagnosis exhibited a 96.88% precision rate [16], surpassing the success rate of medical professionals by a significant margin. In this context, the accuracy of the system in this study is similar, reaching a level of 95% precision. However, the accuracy, sensitivity, and specificity results of the system in this study are based on clinical cases, highlighting the importance of conducting evaluations in real-world settings with patients to validate the effectiveness and reliability of the system in clinical practice. Future studies can clinically validate the system in real-world settings, compare it with other diagnostic methods, evaluate its integration into clinical practice, and study the clinical and economic impact of the system. These actions will help to strengthen the validity and applicability of the preliminary diagnostic system, thus contributing to its effective implementation in the medical field.

Finally, the findings suggest that the approach used in the system developed in this study, based on unique diagnostic algorithms for symptom identification, is promising in terms of diagnostic accuracy. In addition, the expert evaluation of the technical aspects of the system also supports its effectiveness, with an overall mean score of 4.78 and an overall standard deviation of 0.42. Furthermore, these results indicate that the developed system is an effective tool for the early detection of diseases with similar symptoms. Comparative results with previous research highlight its superior performance in terms of diagnostic precision, suggesting its usefulness in clinical settings and its potential to guide future research in this area.

6.2. The novelty of the study compared to previous research

The proposed diagnostic system can diagnose any type of disease based on its symptoms. This greatly expands the range of applicability, as it is versatile and adaptable to a wide variety of medical conditions. This differs from other studies that focus on specific diseases such as malaria [11], COVID-19 [13], influenza [17], or chikungunya [15]. The developed system is notable for being lightweight and compatible with the Windows operating system. Its implementation uses the PyQt5 library and a unique symptom identification algorithm in Python. This feature makes it accessible to a large number of users. In contrast, some studies explore different platforms, including the development of specific applications for smartphones [16], [17]. This variation in technological choices highlights the diversity of strategies adopted in research, where the mobility and accessibility of applications for smartphones and computers play a relevant role in the implementation of innovative solutions in the field of health and technology. The developed system takes advantage of the ability of health professionals and experts to organize disease and symptom data in CSV or Excel files, which facilitates the diagnosis of a variety of diseases. In contrast, none of the other studies mention the use of CSV or Excel files to store disease and symptom data in the diagnosis process. This feature potentially makes the developed system more accessible and friendly to healthcare professionals who are already familiar with these tools, especially in resource-limited settings. Also, the current study introduces a novelty by employing a unique symptom identification algorithm developed in Python. This approach improves the system's ability to distinguish between different diseases based on unique combinations of symptoms. This specific technique, which focuses on symptom-specific logic and data processing, can contribute to diagnostic accuracy. Meanwhile, some studies opted for the implementation of rule-based expert systems [16] and the application of machine learning models, such as the SVM [18].

7. CONCLUSION

In conclusion, this study successfully achieved its main objective of developing and evaluating the effectiveness and accuracy of the preliminary diagnostic system based on CSV and Excel files. Through the implementation of a diagnostic algorithm focused on the unique identification of symptoms and the creation of a graphical interface with PyQt5, the system was evaluated using a set of 35 cases, including seven diseases with similar symptoms. The results show that the system achieved a precision of 95%, a sensitivity of 90%, and a specificity of 93%. These results underline the strong performance of the system in the early detection of diseases with similar symptoms, making it a promising tool to support the medical diagnosis process. Furthermore, the relevance of these results lies in the improvement of diagnostic accuracy, which can lead to more efficient medical care and the early detection of diseases, especially in settings where medical resources are limited. In addition, the positive evaluation by experts on technical aspects supports the feasibility and usefulness of the system, paving the way for its possible implementation in conventional medical practice. It's important to emphasize that the system plays a complementary role in the preliminary diagnosis and does not replace the evaluation of the healthcare professional.

However, one of the fundamental limitations of this disease diagnosis system is that it focuses exclusively on the presentation of the diagnosed disease without providing recommendations on how to treat the disease or any other relevant information. Furthermore, its accuracy is inextricably linked to the organizational quality of the disease data and the symptoms stored in a CSV or Excel file. If this data is poorly organized or has structural flaws, the system tends to have significantly reduced accuracy in its diagnoses. For future work, it is recommended to incorporate treatment recommendations into the diagnostic system and to improve accuracy by applying machine learning techniques. It is also proposed to integrate real-time data into these systems, which could streamline decision-making based on up-to-date information, especially in rapidly evolving clinical situations or medical emergencies. These areas of future research have the potential to significantly improve the utility and applicability of diagnostic systems in clinical practice and disease management, thus overcoming the current limitations.

ACKNOWLEDGEMENTS

Thank all the people who made the project possible and in this way be able to contribute through the research.





REFERENCES

- [1] S. Minić, D. Trpinac, I. Novaković, N. Cerovac, D. D. Vukojević, and J. Rosain, "Challenges in rare diseases diagnostics: incontinentia pigmenti with heterozygous GBA Mutation," *Diagnostics*, vol. 12, no. 7, p. 1711, Jul. 2022, doi: 10.3390/diagnostics12071711.
- [2] R. Jáuregui-Velarde, P. Molina-Velarde, C. Yactayo-Arias, and L. Andrade-Arenas, "Evaluation of a prototype mobile application based on expert system for the diagnosis of diseases transmitted by the aedes aegypti mosquito," *International journal of online and biomedical engineering*, vol. 19, no. 13, pp. 72–91, Sep. 2023, doi: 10.3991/ijoe.v19i13.41431.
- [3] M. F. Á. Mejía, P. Y. Shu, and D. Der Ji, "Accuracy of dengue, Chikungunya, and Zika diagnoses by primary healthcare physicians in Tegucigalpa, Honduras," *BMC Infectious Diseases*, vol. 23, no. 1, p. 371, Jun. 2023, doi: 10.1186/s12879-023-08346-1.
- [4] M. A. Carrillo, R. Cardenas, J. Yañez, M. Petzold, and A. Kroeger, "Risk of dengue, Zika, and chikungunya transmission in the metropolitan area of Cucuta, Colombia: cross-sectional analysis, baseline for a cluster-randomised controlled trial of a novel vector tool for water containers," *BMC Public Health*, vol. 23, no. 1, p. 1000, May 2023, doi: 10.1186/s12889-023-15893-4.
- [5] P. Somboonsak, "Development innovation to predict dengue affected area and alert people with smartphones," *International journal of online and biomedical engineering*, vol. 16, no. 2, pp. 62–79, Feb. 2020, doi: 10.3991/ijoe.v16i02.12425.
- [6] J. Xavier *et al.*, "Increased interregional virus exchange and nucleotide diversity outline the expansion of chikungunya virus in Brazil," *Nature Communications*, vol. 14, no. 1, p. 4413, Jul. 2023, doi: 10.1038/s41467-023-40099-y.
- [7] M. D. Sinaga, F. Tambunan, C. J. M. Sianturi, A. Syahputra, F. Tahel, and S. Aliyah, "An expert system for diagnosing leptospirosis disease using forward chaining and bayes theorem," in *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*, IEEE, Nov. 2019, pp. 1–4. doi: 10.1109/CITSM47753.2019.8965338.
- [8] B. Hnusuwan, S. Kajornkasirat, and S. Puttinaovarat, "Dengue risk mapping from geospatial data using GIS and data mining techniques," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 11, p. 57, Oct. 2020, doi: 10.3991/ijoe.v16i11.16455.
- [9] B. Chala, "Advances in diagnosis of schistosomiasis: focus on challenges and future approaches," *International Journal of General Medicine*, vol. 16, pp. 983–995, Mar. 2023, doi: 10.2147/IJGM.S391017.
- [10] A. Bernabé-Ortiz *et al.*, "Diagnostics and monitoring tools for noncommunicable diseases: a missing component in the global response," *Globalization and Health*, vol. 17, no. 1, p. 26, Dec. 2021, doi: 10.1186/s12992-021-00676-6.
- [11] H. Yu *et al.*, "Patient-level performance evaluation of a smartphone-based malaria diagnostic application," *Malaria Journal*, vol. 22, no. 1, p. 33, Jan. 2023, doi: 10.1186/s12936-023-04446-0.
- [12] S. Yazdani, C. Lerner, D. Kulkarni, A. Kamzan, and R. C. Henry, "A new expert system with diagnostic accuracy for pediatric upper respiratory conditions," *Healthcare Analytics*, vol. 2, p. 100042, Nov. 2022, doi: 10.1016/j.health.2022.100042.
- [13] S. E. Dansy, A. Yani, A. Manaf, A. S. Abdulbaqi, and N. Iksan, "Expert system for diagnosis coronavirus disease," *International Journal of Artificial Intelligence*, vol. 10, no. 1, pp. 39–44, Jun. 2023, doi: 10.36079/lamintang.ijai-01001.537.
- [14] N. J. Melendez-Acosta, D. M. Espinoza-Solis, and J. A. Leon-Borges, "Expert fuzzy system determining dengue, Zika, Chikungunya and yellow fever infection," *Computacion y Sistemas*, vol. 26, no. 1, pp. 461–472, Mar. 2022, doi: 10.13053/CyS-26-1-4181.
- [15] Z. Sultana, L. Nahar, N. Basnin, and M. S. Hossain, "Inference and learning methodology of belief rule based expert system to assess Chikungunya," in *Communications in Computer and Information Science*, vol. 1435, 2021, pp. 3–16. doi: 10.1007/978-3-030-82269-9_1.
- [16] A. P. R. de Araújo, M. C. M. de Araujo, T. C. Cavalcanti, C. F. de Lacerda Vidal, and M. G. N. M. da Silva, "DZC DIAG: mobile application based on expert system to aid in the diagnosis of dengue, Zika, and chikungunya," *Medical and Biological Engineering and Computing*, vol. 58, no. 11, pp. 2657–2672, Nov. 2020, doi: 10.1007/s11517-020-02233-6.
- [17] M. Kim, S. Yune, S. Chang, Y. Jung, S. Ok Sa, and H. W. Han, "The fever coach mobile app for participatory influenza surveillance in children: Usability study," *JMIR mHealth and uHealth*, vol. 7, no. 10, p. e14276, Oct. 2019, doi: 10.2196/14276.
- [18] R. Karsi, M. Zaim, and J. El Alami, "Early detection of severe flu outbreaks using contextual word embeddings," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 2, pp. 212–219, 2021, doi: 10.14569/IJACSA.2021.0120227.
- [19] R. Jauregui-Velarde, F. G. C. Arias, J. L. H. Salazar, M. Cabanillas-Carbonell, and L. Andrade-Arenas, "Mobile application design: sale of clothes through electronic commerce," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 7, pp. 801–809, 2022, doi: 10.14569/IJACSA.2022.0130793.





- [20] M. A. Elashmawy, I. Elamvazuthi, L. I. Izhar, S. Paramasivam, and S. Su, "Detection of tuberculosis based on hybridized pre-processing deep learning method," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 69–76, 2023, doi: 10.14569/IJACSA.2023.0140808.
- [21] D. Mamadou, K. J. Ayikpa, A. B. Ballo, and B. M. Kouassi, "Cocoa pods diseases detection by MobileNet confluence and classification algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 344–352, 2023, doi: 10.14569/IJACSA.2023.0140937.
- [22] P. Utomo, K. Adi, and O. D. Nurhayati, "Success factors for citizen-based government decision making using K-means fuzzy learning vector quantization," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 1, pp. 506–516, Oct. 2023, doi: 10.11591/ijeecs.v32.i1.pp506-516.
- [23] U. B. Patayon and R. V. Crisostomo, "Peanut leaf spot disease identification using pre-trained deep convolutional neural network," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 3, pp. 3005–3012, Jun. 2022, doi: 10.11591/ijece.v12i3.pp3005-3012.
- [24] K. S. Ayomide, T. N. M. Aris, and M. Zolkepli, "Improving brain tumor segmentation in MRI images through enhanced convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, pp. 670–678, 2023, doi: 10.14569/IJACSA.2023.0140473.
- [25] K. Omari, "Comparative study of machine learning algorithms for phishing website detection," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 417–425, 2023, doi: 10.14569/IJACSA.2023.0140945.
- [26] R. Jáuregui-Velarde, D. H. Celis, P. Molina-Velarde, C. Yactayo-Arias, and L. Andrade-Arenas, "Evaluation of a chatbot powered by ChatGPT for the preliminary diagnosis of dengue," *International journal of online and biomedical engineering*, vol. 20, no. 1, pp. 58–73, Jan. 2024, doi: 10.3991/ijoe.v20i01.45029.
- [27] R. Zaitoon and H. Syed, "Enhanced brain tumor detection and classification in MRI scans using convolutional neural networks," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 9, pp. 266–275, 2023, doi: 10.14569/IJACSA.2023.0140929.
- [28] G. Prasetyadi, U. T. Hantoro, and A. B. Mutiara, "Singkat: a keyword-based URL shortener and click tracker package for Django web application," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 9, pp. 118–122, 2018, doi: 10.14569/ijacsa.2018.090916.
- [29] Y. M. Alkhanafseh and T. C. Akinci, "A python-based interface design for electric power system education," *International Journal of Smart Grid and Sustainable Energy Technologies*, vol. 4, no. 1, pp. 163–168, May 2021, doi: 10.36040/ijsgset.v4i1.3905.
- [30] D. V. Bhargav, M. M. R. Nandan, P. V. Kowshik Bharadwaj, and P. Rekha, "A simple approach for scripting in air and display using computer vision," in *2021 IEEE International Conference on Mobile Networks and Wireless Communications, ICMNWC 2021*, IEEE, Dec. 2021, pp. 1–4. doi: 10.1109/ICMNWC52512.2021.9688554.

BIOGRAPHIES OF AUTHORS



Dr. Laberiano Andrade-Arenas     is a Doctor in Systems and Computer Engineering. Master in Systems Engineering. Graduated from the Master's Degree in University Teaching. Graduated with a Master's degree in accreditation and evaluation of educational quality, research Professor with publications in SCOPUS-indexed journals. He has extensive experience as the University Chair in face-to-face and blended classes at different undergraduate and postgraduate universities in Lima. He can be contacted at email: landrade@uch.edu.pe.



Cesar Yactayo-Arias     is a professional in administration and Master's in university teaching. Extensive teaching experience in higher education. In addition, experience in educational management. Doctoral study in administration. He can be contacted at email: cyactayo@continental.edu.pe.