GAIA: AN AI DIAGNOSTIC DECISION-SUPPORT SYSTEM FOR GYNECOLOGY BASED ON ARGUMENTATION

Ioannis Yiangou

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

at the

University of Cyprus

Recommended for Acceptance

by the Department of Computer Science

June, 2021

ABSTRACT

In this thesis we present "GynAIcologicA" (GAIA), an implementation of a Diagnostic Decision-Support System [1] for Gynecology, which will provide support to the experts of the field during clinical diagnosis. The assistance in the decision-making process, the alert about possible differential diagnoses [2] and emergency conditions [3], the avoidance of overdiagnosing common diseases, as well as the avoidance of the unwilling neglect of rarer diseases are just some forms of support that the system provides. This tool is based on Argumentation [4],[5],[6] , which offers an ideal platform for developing Human-Centric [7],[8], Explainable [9],[10],[11],[12] Artificial Intelligence Systems.

Following an extended bibliographic review and collection and study of the required medical expertise, we concluded to the basic principles and requirements that the system has to abide by, as well as the goals it has to achieve. The most important of them, regarding medicine, is the enhancement of the accuracy and speed in the diagnostic process, the encouragement of broader diagnosis to avoid overdiagnosing common diseases and the complete coverage of the medical expertise in Gynecology. When it comes to the system as a tool, the most important aspects are its easy access, use and future expansion, something that is succeeded via the implementation of the system in the form of a website, its ability of naturally interacting with the user and its easily extendable modular architecture that it is based on.

Finally, a three-phased testing and evaluation program is being followed in order to identify any weaknesses and missing concepts from the system, to be implemented, with the ultimate goal to make this system an easy-to-use and useful tool that would find application in real-life cases.

ΠΕΡΙΛΗΨΗ

Σε αυτή την διατριβή παρουσιάζουμε το “GynAIcologicA” (GAIA), μία υλοποίηση ενός Συστήματος Υποβοήθησης Διάγνωσης [1] για τη Γυναικολογία, που θα προσφέρει υποστήριξη στους ειδικούς του τομέα κατά την κλινική διάγνωση. Η βοήθεια στη διαδικασία λήψης αποφάσεων, η ενημέρωση σχετικά με πιθανές διαφορικές διαγνώσεις [2] και περιπτώσεις έκτακτης ανάγκης [3], η αποφυγή υπερδιάγνωσης συχνών ασθενειών καθώς και η αποφυγή άθελης παράβλεψης σπανιότερων ασθενειών αποτελούν μορφές υποστήριξης που προσφέρει το σύστημα. Το εργαλείο είναι βασισμένο στην Επιχειρηματολογία [4],[5],[6], η οποία προσφέρει μία ιδανική πλατφόρμα για ανάπτυξη Ανθρωποκεντρικών [7],[8], Εξηγήσιμων [9],[10],[11],[12] Συστημάτων Τεχνητής Νοημοσύνης.

Μετά από εκτενή βιβλιογραφική ανασκόπηση και συλλογή και μελέτη της απαραίτητης ιατρικής εμπειρογνωμοσύνης, καταλήξαμε στις βασικές αρχές και απαιτήσεις που πρέπει να τηρεί το σύστημα μας, καθώς και στους στόχους που πρέπει να επιτυγχάνει. Οι σημαντικότερες εξ’ αυτών, όσον αφορά την ιατρική, είναι η ενίσχυση της ακρίβειας και της ταχύτητας της διαγνωστικής διαδικασίας, η ενθάρρυνση ευρείας διάγνωσης για αποφυγή υπερδιάγνωσης συχνών ασθενειών και η πλήρη κάλυψη της ιατρικής εμπειρογνωμοσύνης στη Γυναικολογία. Όσον αφορά το σύστημα ως εργαλείο, οι σημαντικότερες πτυχές είναι η εύκολη πρόσβαση, χρήση και μελλοντική επέκταση του, κάτι που επιτυγχάνεται μέσω της υλοποίησής του σε μορφή ιστοσελίδας, της δυνατότητας φυσικής αλληλεπίδρασης με το χρήστη και της εύκολα επεκτάσιμης αρθρωτής αρχιτεκτονικής στην οποία βασίζεται.

Τέλος, ένα πρόγραμμα δοκιμής και αξιολόγησης τριών φάσεων ακολουθείται, προκειμένου να εντοπιστούν τυχόν αδυναμίες και ελλείψεις του συστήματος, ώστε να υλοποιηθούν, με απώτερο σκοπό να καταστεί ένα εύχρηστο και χρήσιμο εργαλείο που θα βρίσκει εφαρμογή σε πραγματικές περιπτώσεις.

Approval Page

Master of Science Thesis

GAIA: AN AI DIAGNOSTIC DECISION-SUPPORT SYSTEM FOR GYNECOLOGY BASED ON ARGUMENTATION

Presented by

Ioannis Yiangou

Research Consultant	Dr. Antonis C. Kakas
Committee Member	Dr. Vasilios Tanos
Committee Member	Dr. Constantinos Pattichis

University of Cyprus

June, 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis advisor Dr. Antonis Kakas who was always willing to offer his guidance and advice whenever I needed it. By trusting me with this wonderful opportunity to work on this project, he helped me, not only to expand my horizons with valuable skills and knowledge, but also to become a better version of myself, by cultivating qualities such as patience and perseverance that will accompany me forever.

I would like to express my sincere gratitude to the medical expert that co-supervised this thesis, Dr. Vasilios Tanos, for his patience, the precious time he dedicated, his will to always help me and his enthusiastic involvement throughout this project, as well as for the unique experience he offered me through this close contact with the interesting field of Gynecology.

I also thank Panayiotis Tanos for contributing to this project with his research and medical knowledge. By sharing and supporting his modern ideas on combining computer technologies with medical expertise in the quest of improving healthcare services, he deeply motivated me to believe in the potential, to focus on completing this thesis and to trust the process.

Lastly, I would like to thank my family and friends for their selfless support, their encouragement and deep understanding throughout the whole process of working on my Master’s thesis. This accomplishment would have not been possible without them. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

In the recent years, there has been an increased enthusiasm in the use of Artificial Intelligence (AI), because as a tool it has the potential to dramatically increase the quality of performance while offering convenient solutions to a plethora of different problems, ranging from complex scientific problems to the most high-risk commercial decision problems, and even finding application in everyday tasks such as the recommendation of a movie or the arrangement of an appointment. This promising technology does not only serve human users, it also has the ability to think, decide and even act like a human being. This fascinating fact is used to promote the development of “Human-Centric Artificial Intelligence”, i.e. systems that can serve humanity the same way an expert in a specific domain would have done it, by simulating his/her way of thinking and decision-making in order to solve real-life problems from his/her relevant domain. One of those fields that are gradually adopting AI systems to assist them in their various tasks is the field of Medicine. Through the use of Clinical Decision-Support systems, doctors can have access to high-quality consultancy on diagnoses, drugs and medications, patient treatment and care and test result interpretations. It is indeed an interesting case of human-centric computer systems helping humans help humans, and a step towards the right direction of making AI serve humanity for the common good.

1.1 Motivation

By seeing the ever-increasing need of adopting Artificial Intelligence systems in virtually any scientific and commercial domain that could have been benefited at any degree, and by realizing the potential of contributing to this cause with our own respectable effort, we worked on developing and introducing an Artificial Intelligence system, more specifically an AI Diagnostic Decision-Support system based on Argumentation, for application in the medical field of Gynecology [13]. This particular field is relatively new to the adoption of such technologies (at least ones that are specifically related to it), yet studies reveal that AI will bring many advantages to the clinical practice of Gynecologists [1], [14], [15].

Our approach aims at covering the complete representation of the medical knowledge in Gynecology, as this would enable our GAIA system to become a real-life tool that will practically contribute to the enhancement of the quality of healthcare services.

1.2 Related Work

Artificial Intelligence offers a wide variety of modelling methodologies that are able to emulate parts of the human thinking and decision-making processes. The most prominent categorization of methodologies is based on the presence (or absence) of a Knowledge base i.e. a pool from which expert knowledge is consulted, dividing them into “Knowledge-based” and “Non-Knowledge-based” [1]. The first category includes the conventional, simple Logic Programming using inference rules with deduction [35], [36] and Argumentation technologies [4],[5],[16],[17],[18],[43],[44],[45], while the latter includes Machine Learning techniques [89],[90],[91] and Statistical Pattern Recognition [92],[93]. It is not uncommon to find Decision-Support Systems that are based on any of the above methods. Each with its advantages, disadvantages, unique traits and different goals, the above methodologies are compared below, based on features that we considered essential for the development of our own Diagnostic-Decision Support system, GAIA:

Methodology	**Features**				
	Flexibility	*Modularity*	*Explainability*	*Completeness*	*Potential for Historical Data*

Machine Learning [89],[90],[91]		✔			✔
Statistical Pattern Recognition [92], [93]		✔			✔
Logic Programming with Inference Rules & Deduction [35], [36]			✔		✔
Argumentation Technology [4],[5],[16], [17],[18], [43],[44],[45]	✔	✔	✔	✔	✔

Table 1. Comparison Table of Decision-Support System methodologies

<u>*Flexibility*</u>*:* A Diagnostic Decision-Support system has to be flexible in order to adapt to changes in the context of the current problem [5], [6]. Machine Learning and Statistical Pattern Recognition solutions require a large number of data to extract patterns for modelling the behaviour of the system [1], [89], [92], [93]. This makes it difficult for the system to adapt to new changes, since the change of the behaviour of the system and the modelling of various preferences require a number of additional data, large enough to influence the pattern extracted from the initial training set. This also poses the danger of not extracting a required, important relation between inputs and outputs, something that knowledge-based solutions are

not susceptible to, because of the explicit nature of declaring those relations using rules. Simple Logic Programming with inference rules fails to adapt to the new context, not because it does not identify new information, but because it is designed to resort to the recommendation of all possible, successful inferences without specific preferences [5], [6], [35], [36]. This means that simple scenarios, along with any of their refined versions are possible under a refined context, despite that the refined versions have more evidence supporting them (in the form of more conditions being met). Preferences are very difficult to model in simple Logic Programming, especially in large-scale systems such as Decision-Support Systems, since they require the explicit declaration of negating conditions (e.g. not(fever)) in all rules, in order to exclude the simultaneous inference of other rules and recommend refined scenarios over simpler ones. Argumentation technologies succeed in adapting to new information and handling contradicting information by implementing "defeasibility" and "user preferences" [16], [17].

Modularity: It is crucial that parts of a Decision-Support Systems can be extended without affecting other parts, thus minimizing the risks of messing-up the code of other modules or damaging the connections between them. Machine Learning, Statistical Pattern Recognition and Argumentation methodologies are modular, since new knowledge can be implemented in the form of appending more data on the training dataset (for Machine Learning and Statistical Pattern Recognition) or appending new rules and preferences among them in the knowledge base (for Argumentation). An issue with extending Machine Learning solutions is the susceptibility to adversarial attacks [88], either intended or unintended, which are able to distort extracted patterns and conclusions, through the consideration of nonsensical or invalid data that will get mixed with the valid data during the training phase. This is a possible way for error predictions, whereas in Argumentation technologies, the expert knowledge is directly declared in the knowledge base, therefore nonsensical or invalid data will not affect the quality of other conclusions as much, since this would require explicit declaration of the preference of the error piece of knowledge over the valid ones. Finally, as discussed earlier, the need for explicitly preventing the inference of some rules in order to emulate "preferences" in simple Logic Programming makes this methodology our last option when it comes to Modularity. This is because the addition of new rules in the knowledge base will not immediately capture the new knowledge, unless by first changing all other existing rules to include "safeguard" conditions, to account for "scenario preferences" [17], i.e. to explicitly append negations of conditions to the body of each rule in order to guarantee that the system concludes to the most refined scenario each time.

<u>Explainability</u>: A Decision-Support system should be able to give explanations to justify its conclusions [23], something that a "blackbox" algorithm [9] (e.g. Machine Learning, Statistical Pattern Recognition) fails to do. Simple Logic Programming can be used for simpler explanations via presentation of the conditions that led to the final conclusion however Argumentation technologies do a better job at that. Argumentation frameworks are able to, not only justify their conclusion, but also to compare it to other possible conclusions and explain why it is preferred over them [37]. It can also present alternative solutions, as well as hypothetical conclusions that could have been extracted based on specific assumptions [83].

<u>Completeness</u>: Argumentation technologies provide complete disease coverage, since we can model the whole expertise knowledge to conclude to the complete scope of diseases. The conclusion is the whole radar of diseases and for each disease the system presents if it is possible, under given inputs, or not. This is not the case for the other three approaches which only exhibit specific completeness. Machine Learning needs multiple modules to capture the complete expert knowledge [89], since a single module predicts whether or not a target disease is possible, and not which of all the disease are possible or not [47]. Statistical Pattern Recognition extracts patterns based on statistics for each target disease separately [92], [93], therefore, as in the case of Machine Learning we need statistics for all possible diseases in order to extract accurate patterns. Finally, simple Logic Programming requires explicit

declaration of all the possible diseases, and the conditions under which each one is possible or not, in the form of inference rules. With the absence of scenario-preferences, simple Logic Programming misses a very important aspect of the medical knowledge, which is to exclude a scenario in favour of another, thus only implementing partial completeness.

Potential for Historical Data: All of the above methodologies offer ways to express a patient's historical data, either by declaring facts [35] in the knowledge base (for Argumentation and Logic Programming) or by extending the fields of training data files to account for more related attributes for a diagnostic case (for Machine Learning and Statistical Pattern Recognition).

Decision-Support Systems based on Argumentation [16] provide high-quality cognitive assistance in the decision-making process, as they benefit from the ability of Argumentation to handle contradictory information well and offer reliable and essential explanations about the process and the results [17]. The implementation of concepts such as "user preferences" [17] and "defeasibility" [16] separates Argumentation from other methods, since the first one enables systems to account for a user's personal preferences and unique traits in the decision-making process, while the latter integrates a very important cognitive ability of the human mind to challenge a previous conclusion and revise prior decisions, in favour of conclusions that make more sense under the new context [18], [19].

An interesting application of Decision-Support Systems is in the field of Medicine [1], [20], [21], where they are used to provide assistance in processes that require fast, accurate and reliable decisions to be made in order to ensure the quality of the healthcare services. One of them is the clinical diagnosis, where such systems offer possible diagnoses based on given symptomatology and patient information [1], while a specific work [43] refers to the potential of an Argumentative Decision-Support system to diagnose and explain cases of breast cancer. Decision-Support Systems that are based on Argumentation, not only are able to provide such services, but they also do it by providing explanations [17], that are nowadays required [22], [23], [24] in order to persuade patients to take the necessary actions for their health and safety.

Regarding the field of Gynecology, although some initial steps are being taken to address the potential of AI systems in this domain [14], according to the medical expert that co-supervised this thesis, there is a lack of Diagnostic Decision-Support Systems, especially argumentative ones, which are focused specifically on the field of Gynecology. This could make our GAIA system one of the most complete Argumentative Diagnostic-Decision Support systems in Gynecology, being an attempt to host the complete gynecological knowledge and offer proper explanations, with the intention to make it applicable in real-life clinical trial.

1.3 Thesis Organization

Chapter 2 of this thesis presents the background from both fields of interest: Medicine/ Gynecology and Computer Science. Regarding the medical/ gynecological background, the chapter refers to the field of Gynecology in general, while regarding the background of Computer Science, the chapter refers to general concepts that the system is based on, such as Artificial Intelligence (AI), Explainability, Logic Programming and Argumentation. The two fields find a common ground in the background information that concerns the various Decision-Support Systems that find application in the field of medicine.

Chapter 3 presents the theoretical model that is used as basis for the implementation of the system. This includes methods of collecting and managing information in a structured and comprehensible manner for this large-scaled problem, the modeling of "diagnosis" as an argumentation problem and overall the identification and description of various high-level components of the system (explanations, prompting, link between components etc.).

Chapter 4 discusses all the aspects related to the development of the system and all the methods that are used to transform the theoretical model described in Chapter 3, into an actual, working computer system.

Chapter 5 gives details on the testing and evaluation process of the system.

Finally, in Chapter 6 we provide our conclusions and a list of future work recommendations that are essential for enabling GAIA to find application in real-life scenarios and become a successful commercial tool.

Chapter 2

Background

2.1 Artificial Intelligence

“Artificial Intelligence” (AI) describes the ability of computer systems to emulate human-like intelligence, cognition and behavior, with the purpose of solving real-life problems, in a similar way the human mind does [25]. The immediate response to newly-received stimuli, the recognition, understanding and adaptation to a specific context, as well as the utilization of past experiences, available examples, logical reasoning and common sense knowledge to influence the decision-making process are just some of those human mind traits implemented by such systems, which are often called “Intelligent Systems” [26].

The importance of this concept lies in, not just the sufficient, but the excellent exhibition of the aforementioned characteristics by today’s technologies, which managed to surpass human entities at equivalent tasks, in terms of accuracy, speed, efficiency and work volume [25].

The first-ever reference of such machinery, that was capable of showcasing human-like behavior and executing human tasks, can be traced back to Ancient Greek mythology, where the fire and smithing god, Hephaestus, had mechanical assistants and was also responsible for the creation of the gigantic, bronze robot “Talos” [27]. Alan Turing, the well-known English mathematician and computer scientist, made an initial reference to Artificial Intelligence with his paper “Computing Machinery and Intelligence”, back in 1950, on which he proposes his famous “Turing Test” as a mean of determining whether a computer can “think” like a human being, or not [25].

Today, “Intelligent Systems”, and generally “Artificial Intelligence”, are found in a wide range of fields. Both the scientific, as well as the industrial and commercial worlds benefit from the aid of such technologies in the processes of solving complex problems and making important, many times risky, decisions. We also come in contact with systems that use Artificial Intelligence techniques, such as image, sound, text and speech recognition, in our daily life. This may occur, either directly, e.g. personal assistants like Siri [28], [29] and Cortana [30] which help individual users with their personal daily tasks, or indirectly e.g. recommendation algorithms like the ones used by YouTube and Netflix, which suggest content that may interest their audiences [31].

2.2 Human-Centric, Ethical and Law-Abiding Artificial Intelligence

We already discussed the potential of “Intelligent Systems” to exhibit human-like intelligence; however we have to keep in mind that the human behavior is not limited to its problem-solving capabilities. Human beings follow rules and guidelines, live by specific ethical codes and abide by constitutional laws. All of the above, essentially, are the constraints that prevent communities from falling apart and diving into chaos, therefore, we cannot talk about “human behavior” without considering the acceptance, respect and compliance to such constraints. For a system to be labeled as “Intelligent”, or even “Human-Like” or “Human-Centric”, the aforementioned criteria must be met, and only then a system will, not just exhibit “humanly-possible” acts, but also “human-like” thinking and conscience [7].

However, one cannot simply navigate through life without facing at least one ethical dilemma that calls for taking an action that would either be “legal but unethical” or “ethical but illegal”.

For further understanding the concept, philosophers proposed the “Trolley Problem”, where a moving runaway trolley would run over several people standing on the tracks and you have the chance to change the trolley's direction to a sidetrack, however at the expense of a single person's life, which is standing on the side tracks [8]. The dilemma goes like this: "Do you change the direction or you let it as is?" The earlier causes the death of one person to spare the lives of a group of people but it somehow makes us feel quite more uncomfortable and guilty than the latter, by being the result of our own intervention, instead of our inaction [8].

The "Trolley Problem" raised many questions about the ethical behavior of Artificial Intelligence systems, in times where not acting at all or acting completely legally will lead to more harm than good. One of the best examples is the case of the "Autonomous Car", an emerging technological feature also known as "self-driving car", and its choice whether or not to surpass the speed limit in urban areas (something illegal) for emergency reasons e.g. saving a relative in danger, which is something ethical.

Developers must find a middle ground where both laws and morals are implemented and respected, as well as ensure that any compromise between the two will bring more benefits than disasters [8]. This will cover an integral part of the human behavior, which is essential for a system to be called "Human-like" or "Human-centric".

2.3 Explainable Artificial Intelligence

Imagine that a computer system made a decision about a very crucial issue, that we had to adopt, but in the absence of a reasonable explanation on "why", "how" and "what" influenced the machine to make that decision. We are not even referring to cases of "not suggesting alternative solutions" or "the solution not being accurate enough", but to the complete lack of essential qualities such as "security", "traceability" and "accountability" [7], something often encountered in "blackboxes" i.e. programs that solely rely on what their inputs and outputs are, thus leaving no traces of "why" and "how" they do what they do, or what techniques are involved in the medium processing steps [9]. Pair the above reasons with the "too-good-to-be-true" accuracy scores from the extreme focus in "doing things right" (usually with disregard to "doing the right thing") and we end up with systems that still have a long way from becoming "human-centric".

One step closer to reaching the goal of creating systems capable of both acting and thinking like a human being is the implementation of "Explainable Artificial Intelligence" ("Explainable AI" or "XAI"). This extended version of Artificial Intelligence offers an insight on "what" AI systems do, "how" they do it and "why" they do it, by offering solid evidence for those actions, in forms that are comprehensible and interpretable by any human user and expert, such as explanations in natural language or visualizations [10]. This lets AI stakeholders know which factors influenced the outcome of the system and for which reason, as well as what is happening before, during and after processing, something which renders the system easily traceable and transparent, thus solving any opacity issues related to "black box" algorithms [9], [11].

Major benefits of "Explainable Artificial Intelligence" include the explanations themselves, which guarantee that the system does its job as expected (no more, no less) [11]. They also provide the audience with the necessary evidences supporting any decision made, as well as promote building trust to the system, even in the case of the most skeptical individuals, something which is required for the system to be accepted and eventually be established [10], [12].

The importance of Explainability in Artificial Intelligence is reflected in the initiative of large organizations to enact laws that either discourage, or completely forbid the development of "intelligent" systems that lack the ability to offer sufficient explanations for their actions and results. One of the most prominent initiatives belongs to the European Union with its "General Data Protection Regulation" (GDPR) which generally aims to secure fundamental rights of European citizens, such as protection and privacy, pertaining to the collection and processing of their personal data [22]. GDPR includes a variety of articles which concern automated processes, such as artificially-intelligent systems, stating that European citizens have the right to object to any possession of their personal data without first being offered an explanation as to “why” this is happening, and only under extremely special circumstances, such as ones that require law enforcement, should this right be overwritten (Article 21) [23]. Article 22 refers to the right of European citizens to refuse to be subjected to any decision that was made by an automated system, as well as their right to contest any such decision and request that a human expert will be available to explain the behavior of the system [24].

2.4 Importance of context in Explainable Artificial Intelligence

Natural Language is the most common mean of delivering explanations, and for our sayings to make sense we have to ensure that they are constructed with a specific context in mind; otherwise they will just resemble a bunch of random words put together. Considering that our environment is both complex and dynamic in nature, it is difficult to construct satisfying explanations, which would still make sense or be of any use, solely by describing the cause of a result. However, there are countless explanatory theories which can be used conjunctively, in order to maximize the efficacy of explanations generated by “Explainable AI” products, even under those conditions, and with respect to the current context and some of them are presented below [32].

Explanations have to be logical. This means that explanations that lack logic, despite being grammatically and syntactically correct, are not considered adequate. Moreover, they have to be constructed in a consistent manner, much like a narration, to make them easier to understand the situation they are describing, more interesting to read and more helpful in general [32].

The wide spectrum of users which may use an AI product will yield different contexts depending on their characteristics. An example is the existence of different levels of comprehension, jargon and technical terminologies between experts of different fields (doctors vs. engineers) or even between groups of different experiences among the same field (junior doctor vs. senior doctor). Therefore, an explanation has to be adaptable to the current context, depending on both the surrounding environment and the current user, in order to appeal to its audiences [32].

Avoiding any component that would interfere with the quality and pragmatism of an explanation is of utmost importance. Constructing explanations which include personal contexts will inevitably attribute some of the following to its meaning: personal biases, outdated personal experiences and distorted memories of our personal environment that remind nothing of its current state. All those will influence the result in an undesirable way [32].

An interesting suggestion [32] is the involvement of experts from different, relevant fields in the construction of AI explanations. Social scientists, artists and philosophers would give more variety and expertise from different perspectives, thus enhancing the development of solid, admissible explanations.

2.5 Explanation Types

There are many ways to present a description of a situation, while respecting the explanatory theories mentioned earlier and their use is dependent on what is the goal we are trying to achieve with this explanation.

Below is a list of the most popular explanation types collected and given by experts (K. Cyras et al. [33], S. Chari, D. M. Gruen, O. Seneviratne, and D. L. McGuinness [34]), each with a brief description on what purpose they serve:

Actionable: offers guidelines and necessary actions to take, in order to reach to, or apply, the outcome [33].

Attributive: describes which inputs were used to infer other information (using reasoning) and the result itself, and how important were those attributes, either initial or derived, to conclude to the current result [33].

Case-based: the system remembers previous cases on which the specific explanation has also been given and presents them in order to support its decisions [34].

Contextual: use of explanation components that describe the current case, beyond the given inputs and returned outputs of the system, such as information related to the user's situation and the surrounding environment, i.e. the context of application [34].

Contrastive: the explanation describes the reasons why the resulted decision is preferred over another, expected decision, in order to support it against other options [33], [34]

Counterfactual: offers an insight into alternative scenarios which would have been occurred, had the set of inputs been different [34].

Everyday: explanations are more comprehensible and appealing to non-expert users [34].

Scientific: explanations contain scientific methods such as measurements, observations and experiment results, in order to support the decision [34].

Statistical: utilizes numerical evidence, probabilities and percentages to support decisions [34].

Simulation-based: provides a glance on what will happen if the resulted decision is applied [34].

Trace-based: offers details of processes used by the system to derive the decision, step-by-step [34].

Before we develop any explanation modules for an AI system we first need to consider our goals and the purpose of providing the explanations to the users. “Is it to present the positive effects of the decision and encourage users to apply them?” “Is it to present the negative impact of other decisions and prevent users from considering them?” “Is it targeted towards experts of a specific group or the ordinary person?”, “Will an insight into the internal process help programmers discover bugs or logical errors?” Whatever is our goal, there is an explanation type to satisfy it and if this goal is too complicated or broad, a combination of those explanation types can be considered.

2.6 Logic Programming

“Logic Programming”, as the name states, is “programming in formal logic”. It refers to the use of logic sentences to express facts and rules [35]. It is a form of “Declarative Programming”, that means the code describes the specifications of a problem we want to solve and the system itself attempts to derive a solution, through logic [36]. The sentences in a logic program do not have to adhere to a strict structure of flow, as the order of declaration does not dictate the order of execution [35].

Rules in Logic Programming are declared in the following manner:

$$P(Z) \leftarrow A_1(Y_1), A_2(Y_2), A_3(Y_3)\ldots, A_n(Y_n)$$

Interpreted as:

“If ALL $A_1(Y_1), A_2(Y_2), A_3(Y_3)\ldots, A_n(Y_n)$ are true, then $P(Z)$ is also true”

Where [35]:

1. $P(Z)$:
    - is the *head* of the rule
    - P is a predicate, parameterized by variable Z
        - P may not be parameterized, in this case: $P() \rightarrow P$
2. $A_1(Y_1), A_2(Y_2), A_3(Y_3)\ldots, A_n(Y_n)$:
    - is the *body* of the rule

- for each $A_x(Y_x)$: A_x is a predicate, parameterized by variable Y_x
  - each A_x predicate may not be parameterized, in this case: $A_x() \rightarrow A_x$
- commas denote conjunctions between the predicates that consist the body
  - *e.g. "$A_1(Y_1)$, $A_2(Y_2)$, $A_3(Y_3)$…" means "$A_1(Y_1)$ **AND** $A_2(Y_2)$ **AND** $A_3(Y_3)$"*
- each of the predicates $A_x(Y_x)$ can be:
  - a statement
    - *e.g. "The weather is rainy": weather(rainy).*
  - a negation of a statement
    - *e.g "The weather is NOT rainy": not(weather(rainy)).*

Rules that are declared without a body (i.e. only have right part) are called "Facts" because they are accepted without the need of meeting any conditions to be true [35].

Logic Programming will be mentioned later, in conjunction with other techniques, which were used to implement the idea proposed in this thesis.

2.7 Argumentation

A form of programming which is built as an extension of Logic Programming is "Argumentation". It refers to the programming of common sense, where conclusions are derived using logical reasoning through the use of arguments i.e. rules which represent claims that support (or oppose) a decision. Therefore, we could also define Argumentation as "the framework on which arguments act". It is based on the human ability to construct arguments to support opinions and positions, express thoughts and interact with other members of the society in a civilized way [4].

Argumentation possesses many benefits compared to both simple Logic Programming and any form of strict programming (where lines of code have a strict flow structure) in problems where we ask queries and expect answers based on knowledge bases and common sense. To understand this, consider the following example:

R1: cook_food ← current_day(monday).

R2: order_food ← current_day(monday), celebration.

R3: cook_food ← current_day(monday), celebration, low_budget.

The above rules are interpreted as follows:

R1: "If today is Monday, then you should cook food"

R2: "If today is Monday and you also have a celebration, then you should order food"

R3: "If today is Monday, you have a celebration, but you also have a low budget, then you should cook food"

In the case of pure Logic Programming, if today is Monday and you have a celebration, but you also have a low budget, then all three rules will be true. However, if we use Argumentation, only rule R3 will be true because any time a new piece of information becomes available, our decisions are revised and change, in favour of conclusions that are supported by more evidence [5], [6]. Thus we prefer the most refined version of a scenario possible.

Now consider that we want to re-write the above rules in a strict programming language. We would need to define IF-statements to express all the possible resulting cases and also place them in the correct order to implement the concept we described above (to prefer more refined scenarios). This will become increasingly difficult as the problem becomes more complex by including more rules and multi-parameter conditions, thus the Argumentation framework is preferred here.

2.8 Argumentation Frameworks: Graph Representation

Imagine the Argumentation framework as a graph, G<A, R>, with the following properties [37]**:**

1) G<A, R> is a directed graph where:

- *A: set of nodes depicting arguments*
- *R: set of arcs depicting "attack relations" between arguments*

2) Each argument is a claim that supports a position

- *Some arguments may support the same position*
- *Some arguments may support different or contradictory positions*

3) The direction of an arc models the following relation between two involved arguments:

- *"The destination argument (the one that is pointed-by) supports a position that opposes the position supported by the source argument (the one that is pointing-to), therefore the latter attacks the first one"*

An example of such graphic representation is the following:

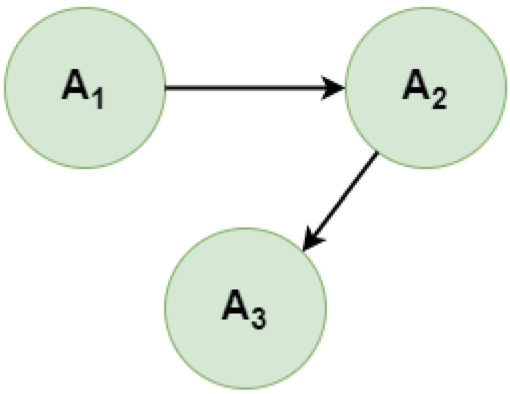


Figure 1. Simple Abstract Argumentation Graph

Interpretation: *"A1 attacks A2" & "A2 attacks A3"*

Now consider the example presented earlier, where we have the following rules:

- R_1: *cook_food* 🡐 *current_day(monday).*
- R_2: *order_food* 🡐 *current_day(monday), celebration.*
- R_3: *cook_food* 🡐 *current_day(monday), celebration, low_budget.*

This program can be graphically depicted as an Argumentation graph, as follows:

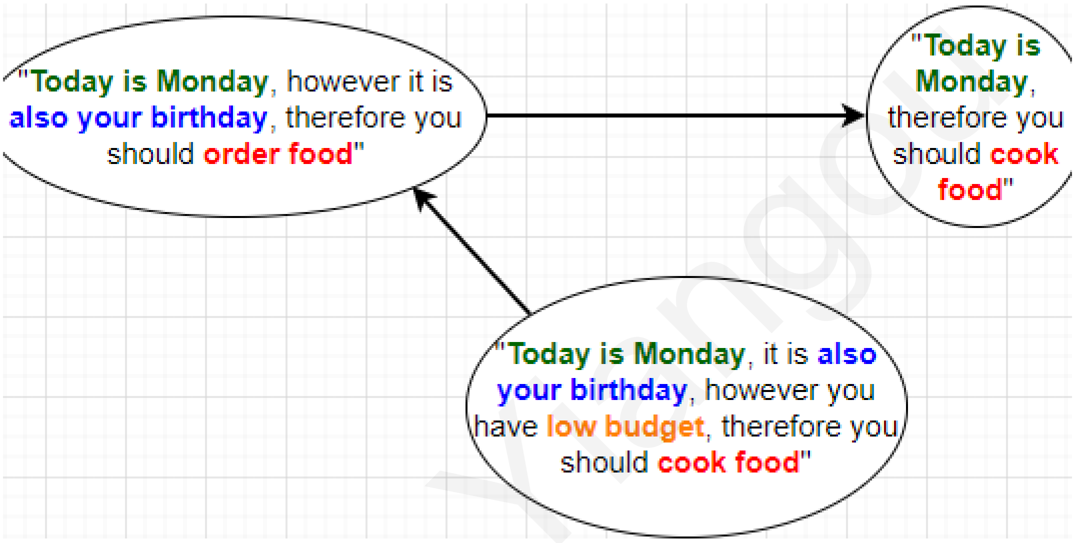


Figure 2. Argumentation Graph Example

The concept of "attack relation" can be lifted a level higher: *arguments can form support groups*. This means that arguments which support the same position, and possess no attack relation between them, can be grouped together to strengthen it and defend it against other, opposing positions [5].

Consider the following [5], [39]:

1) A subset of arguments from set A form a group, G_1 which supports decision D_1
2) Another, different subset of arguments from set A form another group, G_2 which supports decision D_2
3) Members of G_1 do not exist in G_2, and vice-versa (applies to all formed groups)
4) D_1 is contradictory to D_2

5) If at least one argument from G_1 attacks at least one argument from G_2, then [39]:

- "G_1 attacks G_2"

6) If at least one argument from G_2 attacks at least one argument from G_1, then [39]**:**

- "G_2 attacks G_1"

7) Many more groups and such relations may be formed, depending on the problem

2.9 Argumentation Frameworks: Admissibility & Explainability

The reliability of Argumentation frameworks lies in their ability to find admissible sets of arguments which strongly support the final conclusions [5]**,** without much room for doubts, coupled with the potential to offer the appropriate explanations for this decision [37]**.**

An "admissible set" or a "set of admissible arguments" is a subset A_{adm} of arguments from an Argumentation framework, which meets the following criteria [5]**,** [38]**,** [39]**:**

1) Members of A_{adm} do not possess any "attack relations" with other members of the subset → The subset is "conflict-free" [39]

2) All possible argument subsets that attack A_{adm} (i.e. counter-arguments to A_{adm}) also receive an attack from any number of arguments from A_{adm}

- Its members defend it against attacks from opposing subsets

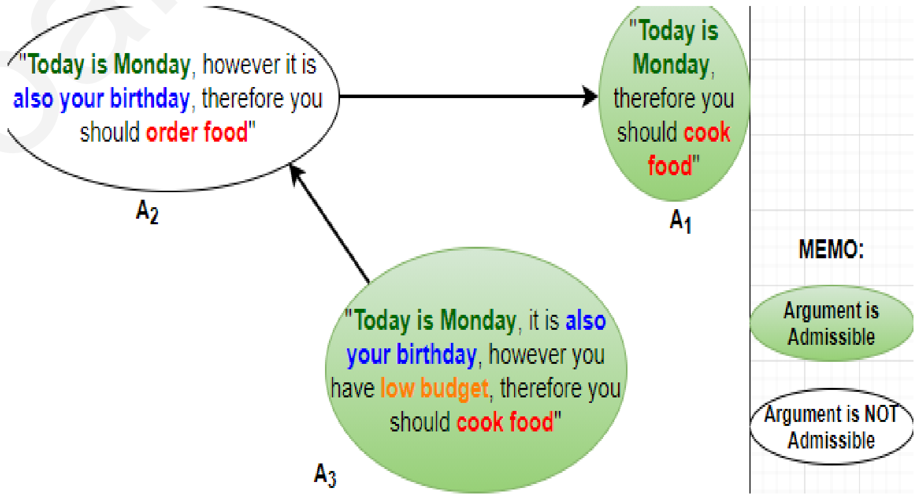


Figure 3. Argumentation Graph – Admissible Arguments

Observe Figure 3 (Argumentation Graph – Admissible Arguments). We can construct a set of admissible arguments, by collecting non-conflicting arguments that defend each other from the attacks of non-selected arguments. In this case, argument A_1 can be paired with argument A_3 to counter the attack coming from A_2 to A_1, thus A_3 protects A_1. Therefore, the group formed is admissible, since it is conflict-free **[39]** and counters all attacks against it, making the position they support ("cook food") an acceptable conclusion.

$$\textit{Set of Admissible Arguments} = \{A_1, A_3\}$$

"Admissible sets" guarantee the integrity of the resulting decision and enhance its trustworthiness, since the position they support, not only possesses no contradictions among its supporting arguments, it was also impossible to find proof (i.e. another subset of arguments) that could counter it, in favour of another decision [5], [38], [39].

Another very important property of Argumentation frameworks is the ability to generate explanations. Since any argument that belongs to an admissible set does have an explanation [37], we can construct explanations by "explaining" those arguments and presenting the output of this process in a structured manner (e.g. sentence in Natural language that makes sense and adheres to logic). This can be done by iterating through the admissible sets to retrieve the explanations of each argument and use string-manipulation techniques to tailor them together. The advantages are the enhancement of justification and transparency regarding the results and their preceding procedures [37], thus increasing the reliability of the system.

2.10 Argumentation Frameworks: Applications

The ease at which we can describe complex problems from the real world using human thinking [4], the ability to revise our decisions in the arrival of new information [5], [6], and the benefit of justifying and explaining any conclusion using reliable evidences [5], [37], [38],

[39], render Argumentation an excellent option for developing real-life applications that act in accordance with the human behaviour.

Some great examples of systems that utilize Argumentation frameworks are the Cognitive Personal Assistants, a type of “intelligent” systems that aid their human users in taking various decisions throughout their daily lives [5], such as:

- Recommending various activities such as games, sports and movies [40]
- Making reservations in dining areas, bars, cafeterias and entertainment places [5]
- Scheduling visits, tasks and appointments [5]

Those decisions are always influenced by the circumstances governing the surrounding environment and the user’s personal information and preferences [5], hence the term “Personal Assistant”.

Many fields utilize Argumentation frameworks to study and capture the way humans solve complex real-life, commercial or scientific problems from their domains, and to further apply this collected knowledge to solve problems of the same nature, in the future. We can find such systems in fields like Engineering [41], Law [42] and Medicine [43]. Because this approach aims at developing AI systems that simulate (at least partially) the complex thinking and decision-making processes that take place inside a human expert’s mind, such systems are called “Expert Systems” [40]. By using Argumentation, this is feasible by keeping knowledge bases which store the expert’s valuable, acquired experiences and knowledge over the years and declaring rules (arguments) to simulate the decision-making process *(i.e. support or opposition to possible conclusions, under a given context, and considering both gathered and stored information* [44]. The result is a model which would demonstrate a behavior, similar to its influence (*i.e. the field human expert*), when asked to solve a problem of similar nature.

2.11 Gorgias

Gorgias is an argumentation framework that successfully combines the principles of preference reasoning and abduction, while maintaining the benefits of both. It offers the ability to make reasoning, using adaptable preference policies, even in the absence of complete information and despite the dynamic nature of environments. Gorgias can model "higher order principles", by considering a user's personal preferences and settings in the decision making process [45].

e.g. a Cognitive "Call Assistant" is instructed by its human user to, generally, allow phone calls, but when at work, the assistant should deny the call [46]. *This is a policy set by the human user, and the system is adjusted to his/her personal preferences to satisfy his/her personal requirements.*

Gorgias is used as the basic argumentation framework for the development of the solution proposed in this thesis.

2.12 Clinical Decision Support Systems: What are they?

Artificial Intelligence managed to establish itself in numerous fields, both in the commercial and the scientific world, yielding benefits that otherwise would have not been possible solely by human abilities, considering the tremendous amounts of available data and the ever-increasing demand to deliver value-based services [20]. Medicine is no exception, as healthcare today is making the transition towards innovative technological solutions that will assist providers in their difficult and ever-demanding task of offering medical services. One of the greatest products of Artificial Intelligence, and an important step towards improving healthcare delivery, is the Computerized Clinical Decision Support System (CDSS), which utilizes clinical knowledge and patient-specific information to reinforce a medical expert's decision-making process [1].

A CDSS is mainly available at the point-of-care and used by experts themselves, to assess its results, combine them with their own experience and knowledge and make consolidated decisions [1] that are tailored to an individual patient's conditions [20].

Decision Support Systems (DSS) that are built upon solid, high quality verified knowledge, formed by combining available peer-reviewed studies, statistical analyses, experimental results and practical experience in a field, are called "Evidence-Based Decision Support Systems" [16]. The use of such technologies in medicine ensures that clinical decisions made by medical experts are backed by science and are acceptable in the field's community.

2.13 Clinical Decision Support Systems: Purpose and Services

Clinical Decision Support Systems are able to support and offer a wide variety of services and functions, including [1], [50]:

- medical diagnosis [1]
- medication prescriptions and drug control [1]
- disease management and treatment [1]
- risk assessment and minimization of medical errors [1]
- alarm systems and reminders for purposes such as emergency cases, availability of new possibly-useful information, prompting and guiding through the user interface, patient safety etc. [1], [50]
- filling of electronic forms, records, calculators, dialogs etc. with patient data [50]

2.14 Clinical Decision Support Systems: Design Principles

In order to design and develop CDSS that correspond to the properties mentioned above, adherence to a specific set of principles is recommended.

Firstly, since a clinical decision is influenced by a variety of parameters, and not just by the clinical knowledge of the health professional, such systems must be able to combine several sources of data [48] (e.g. laboratory/ imaging/ physical examination results, patient history, symptoms, underlying risk factors, expert’s knowledge) before concluding.

Another principle is the modelling of human cognitive abilities. The human mind has a unique way of relating information, as well as conceiving concepts, such as the possibility of emergency conditions or differential diagnoses, and the revision of a decision, as soon as a new piece of information becomes available [48]. Therefore, a CDSS is expected to exhibit the cognitive behaviour of the human expert that it is modelled after.

The third principle is related to trends and time events. The repetition of clinical procedures (blood tests, temperature measurements etc.) and the worsening or improvement of symptoms after a specific period, are just some examples of data that keep changing as a function of time. Temporal data could enhance CDSS with further abilities, such as monitoring how diseases progress over time and adjusting medication dosages according to the trends of a patient’s condition [48].

Finally, health data should be shareable [48]. Nowadays, nations and organizations worldwide show an increased interest in achieving “Interoperability” in many fields, including healthcare. As defined by the European Union, “Interoperability” is a collaborative framework which enables multiple Electronic Health (eHealth) applications to exchange and process health information, despite linguistic and cultural barriers, and even among different nations and health system jurisdictions [56]. This means that a CDSS should be able to produce shareable data (e.g. by means of using conventional file formats and terms from a universal vocabulary) and being able to share them with other professionals and systems [48].

2.15 Clinical Decision Support Systems: Methodology

Generally, systems in medicine utilize Artificial Intelligence algorithms such as Machine Learning, Natural Language Processing, Speech Processing, Robotics, Computer Vision [47] and Argumentation frameworks [16], [17]. CDSS specifically utilizes, either Machine Learning or Argumentation (the latter will be used for the solution proposed in this thesis).

Depending on the algorithms that are used to develop the system, we have two types of CDSS:

- *Knowledge-Based:* utilizes a knowledge base (also known as "theory") which holds evidence-based rules (i.e. "IF-THEN" statements) that are used to prove whether a conclusion (e.g. disease, medication, treatment) is possible or not, given a set of input data [1]. This approach uses Argumentation Frameworks, where rules act as claims that support a specific decision (arguments) and argue for its dominance over others. The final result is a set of admissible arguments which support the final decision and explain its cause.
    - The solution proposed in this thesis is knowledge-based.
- *Non Knowledge-Based:* utilizes Machine Learning ("blackbox") algorithms to extract patterns and behaviors from sets of available data, in order to recommend a decision. It is not as transparent as the Knowledge-Based approach and also demands large amounts of data from real cases, in order to perform accurately [1].

CDSS are also usually connected to an Electronic Health Record (EHR) system, which the latter acts as a pool for retrieving and utilizing patient's data that already exist [20].

2.16 Clinical Decision Support Systems: Architecture

CDSS are usually designed with an architecture that is based on the following, main components:

1) *Base* [1]: All the available data, plus the following (depending on the methodology):
   - the rules residing in the Knowledge Base (*in Knowledge-Based CDSS*), or
   - the suite of Machine Learning algorithms used (*in Non Knowledge-Based CDSS*)
2) *Inference Engine* [1]: runs the algorithms from the base, using the patient's data, and outputs the results
3) *Communication Mechanism* [1]: the user interface where inputs are given and outputs are presented

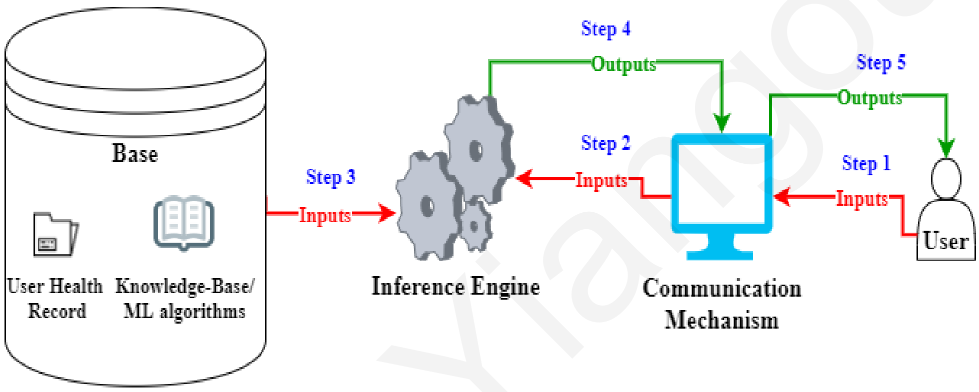


Figure 4. Sample Architecture of a CDSS

The communication mechanism could be any device such as desktop or laptop computer, or any smart handheld or wearable gadget such as a mobile phone, tablet or watch [20]. The system's interface may be available as a web, desktop or mobile application [1].

2.17 Clinical Decision Support Systems: Argumentation

We mentioned earlier the possibility of developing CDSS using Argumentation frameworks and that an example is reflected in the solution proposed in this thesis. Such systems work by declaring arguments corresponding to claims that a disease, treatment or any other decision is possibly the case, given a set of inputs. Then, they are used in a reasoning process, where strong arguments are collected as evidence that support the most relevant of the decisions [16].

Argumentation systems are able to handle contradictory information well and give reliable, quality explanations about the process, about how alternative decisions are compared to the final decision and what admissible arguments support the results [17]. They also offer the capability of declaring conditional preference rules that are tailored to a user's personal preferences, something which will influence the decision making process towards being more respectful to a user's personality and uniqueness [17]. Finally, they support the concept of "defeasibility", i.e. the possibility of a previous conclusion being challenged and even becoming invalid, when new additional information opposes it or supports other claims [16].

2.18 Clinical Decision Support Systems: Examples of Diagnostic Tools

Some examples of CDSS, which do not require visiting the point-of-care, and which lean towards the diagnostic part of healthcare (DDSS), are the following:

1) *Symptoma (https://www.symptoma.com)* [51]: a digital health assistant which accepts a number of symptoms, asks further questions, in a dialog manner, to collect more information, and returns a number of possible causes.

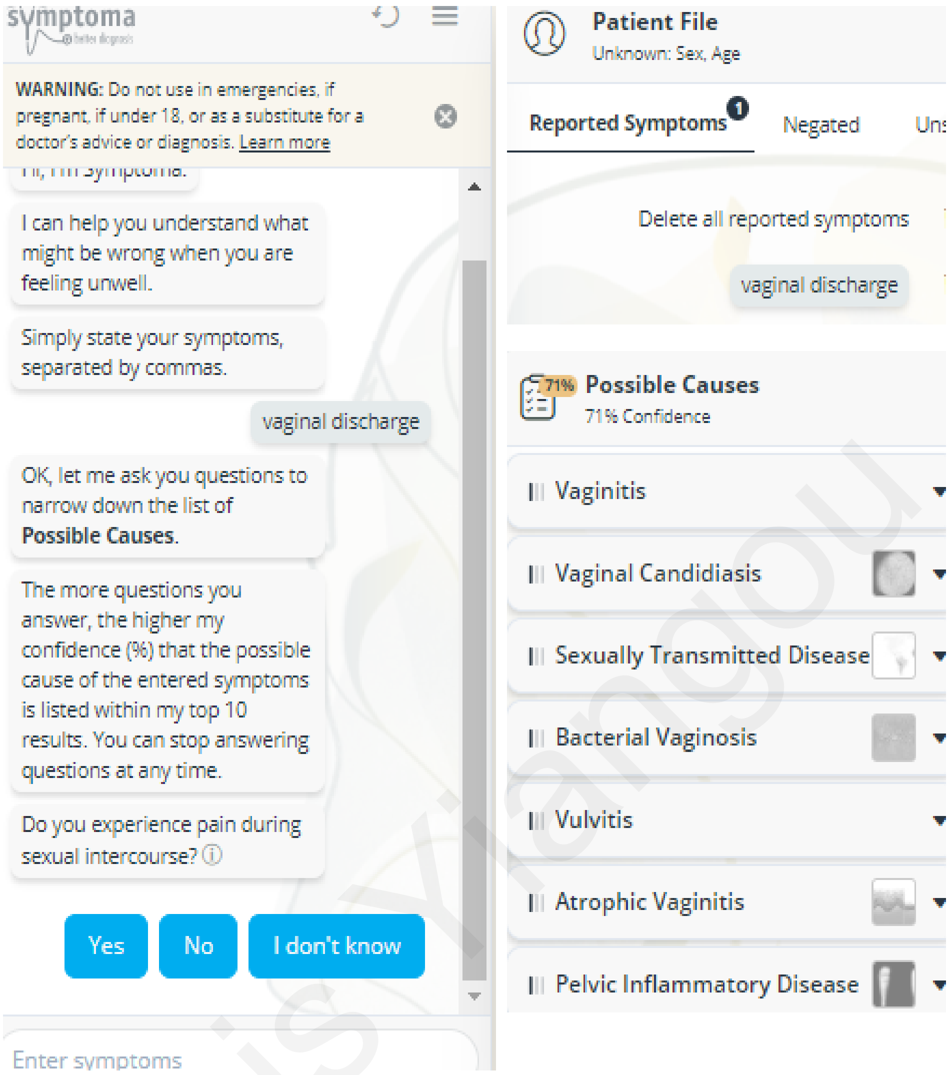


Figure 5. Symptoma – interface [59]

2) *Kunhimangalam et al.* [53]: created a DDSS for the diagnosis of peripheral neuropathies. With an increased percentage of accuracy (93%) compared to experts of the field, the system can identify the presence and type of a neuropathy

3) *DXplain* [54], [57]: acts as a diagnostic tool, as well as a medical textbook

- *As diagnostic tool:* Accepts a set of clinical information and returns a ranked list of diagnoses with explanations and suggestions for information to collect subsequently
- *As a medical textbook:* offers information for diseases such as symptoms, pathology and references to other sources

4) *Babylon AI* [52]**:** Does not provide medical diagnosis but information about possible common medical conditions, based on input risk factors and statistics.

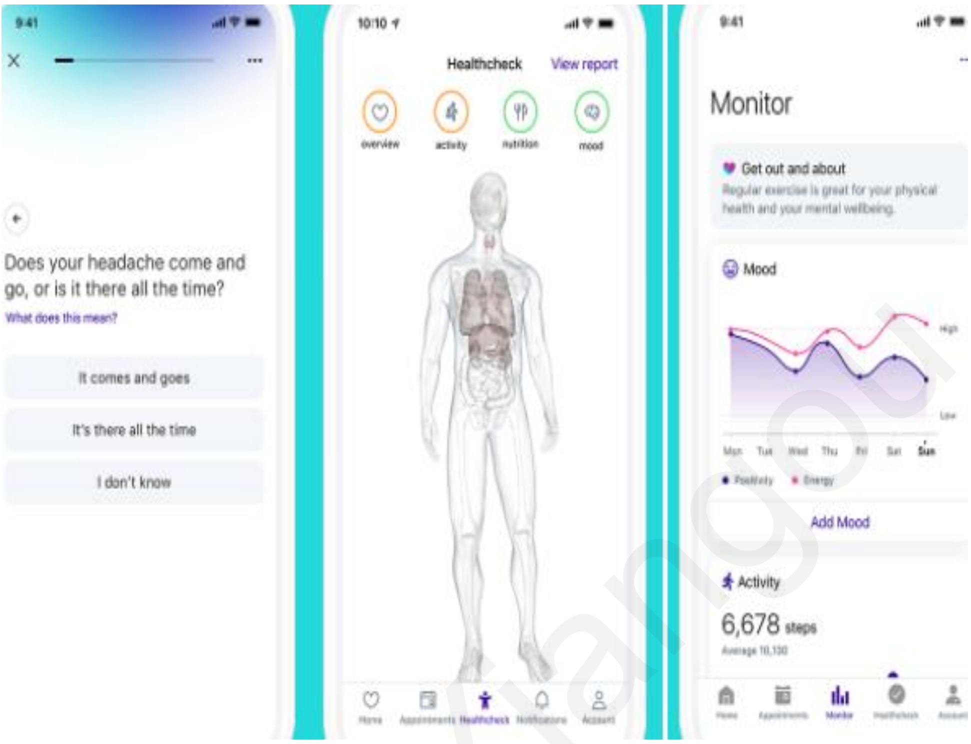


Figure 6. Babylon AI – interface [58]

2.19 Clinical Decision Support Systems: Benefits and Impact

A study conducted [21] (T. Porat, B. Delaney, and O. Kostopoulou, 2017) to investigate the impact that diagnostic decision support systems have on clinical consultation, from both the general practitioners', as well as the patients' perspective. Experts compared the consultation practice using only Electronic Health Record (EHR) systems, versus using an integrated DSS prototype version. While patients found both methods equally satisfying, the majority of the doctors admitted that more diagnoses were considered, as well as more targeted questions were asked, when using the DSS [21], since such systems remind doctors about important aspects of decision making, even when investigating stressful, complex or tiring cases that make them forget something.

2.20 Clinical Decision Support Systems: Challenges and Drawbacks

It goes without saying that CDSS also have some drawbacks and their implementation, evaluation and wide acceptance are filled with challenges that we need to overcome. From the doctor's perspective (who presumably is also the user) some of the most prominent drawbacks are the following:

1) *"alert fatigue"*: several and frequent alarms are becoming tiring and irritating [20]
2) requiring big efforts to use, especially when lacking a user-friendly interface, to interact with [20]
3) need to adapt their consultation styles to the new conditions and accept changes [21]
4) *Patient-centered consultation*: focus on using the system during the whole process rather than at the end, all while trying to maintain the integrity of the dialog with the patient (not ignoring them, not forgetting important verbal information etc.) [21].

From a patient's standpoint, drawbacks include the need to arrange an appointment with a health professional and visit the point-of-care, in order to consult a professional, more accurate CDSS [20]. Also, patients have to show trust to the system's results, and only with the system's improved diagnostic accuracy and convincing explanations, the patient's acceptance will be granted [21].

Finally, from the developers' view, it seems like many challenges and obstacles are associated with the evaluation part of development [49]. Generally AI systems in healthcare manage enormous amounts of data, that are dynamic in nature, and therefore, while crucial, it is difficult to evaluate their safety and effectiveness [49].

2.21 Clinical Decision Support Systems: Conclusion and Motivation

Considering the numerous benefits that Clinical Decision Support Systems offered to the field of medicine, there is an increased hope that IT and AI solutions will further enhance the efficacy of clinical processes [55]. There is also an undeniable need to overcome the drawbacks and challenges mentioned earlier, and this only serves as motivation, in order to start working towards the fulfilment of this goal.

2.22 Gynecology

Gynecology is a field of medicine which is responsible for the health of the female reproductive system, including the breasts, the fallopian tubes, the ovaries and the uterus. A doctor that specializes in this domain is called "Gynecologist" [13] and offers services that aim towards prevention, diagnosis and treatment of both reproductive and sexual health issues. Such services include:

1) Examinations (Pelvic, Breast, Vulvar etc.)

2) Laboratory tests (PAP smear, Vaginal Culture [60] etc.)

3) Imaging tests (Cancer screening, Ultra-Sound Scanning, Mammography etc.)

4) Diagnosis, prevention and treatment for reproductive tract disorders (e.g. infertility), diseases (e.g. endometriosis), cancers (e.g. breast, ovarian, endometrial), abnormal discharge and others

5) Care during pregnancy and childbirth

- done by experts in both gynecology and obstetrics (called OB-GYNs, or obstetrician-gynecologists)

6) Consultancy on topics such as: birth control, pregnancy, sexual health, contraception, menstrual changes, menopause

7) Referrals to other professionals for problems that concern other specialties of medicine

2.23 Artificial Intelligence and Challenges in Gynecology

Over the course of the time, more and more fields of medicine implement Artificial Intelligence technologies to their practices. In the case of Gynecology (and Obstetrics) there exist a number of challenges, which are related to the nature of the field, and which render the adoption of Artificial Intelligence technologies even more crucial.

Researchers [14] reveal that both Gynecology and Obstetrics are highly prone to legal and ethical issues. Any medical error could mean complications such as severe discomfort (at best), a permanent damage, or even death, but in Gynecology and Obstetrics, not only the patient is exposed to those dangers, but also their unborn child or fetus, in case of a pregnancy. Another serious issue is the failure to diagnose malignant cases early on [14], thus the treatment process does not start on time, resulting in fatal consequences. All those errors also contribute to the build-up of socioeconomic consequences [14] that stigmatize the gynecologist or obstetrician.

Adding to the above, there are also several limitations when it comes to providing individualized treatment [14]. Also, keeping the practice up-to-date is hindered by many factors, as evidence-based methodologies require frequent clinical trials and reviews that are expensive, monetarily, as well as temporally, in order to keep up with the new trends and practices in the field [14].

Finally, assisted reproduction raises ethical and practical issues that touch topics such as Embryo selection and In Vitro Fertilization success rates [14].

Artificial Intelligence could be a promising tool for both gynecology and obstetrics, since it can be used in ways that will lead to the overcoming of the above obstacles. Detecting malignancies on an early stage [15], providing more accurate diagnoses [1], reducing medical errors [1] and utilizing explanations generated by argumentation reasoning [37] in order to justify decisions and ensure compliance to ethical and legal codes [22], [23], [24], are all solutions that are viable under the scope of Artificial Intelligence. As long as it is used in the correct way, and that is to extend the knowledge of the medical expert and assist him in the decision-making process [14], Artificial Intelligence is an ally towards a better healthcare.

Chapter 3

Theoretical Model

3.1 Communication & Information Collection

It is not sufficient to only have the necessary computer tools and knowledge to design a solution; we also need to have the knowledge associated with the context of the problem that we are trying to solve. As discussed earlier in this work, the system that is proposed is a Cognitive Gynecological Assistant, aimed to provide diagnostic decision-making support to gynecologists. Therefore it is required for its development that we, at least, collect the appropriate knowledge from the field of Gynecology, with the ideal goal being to comprehend it and become familiarized with it too. This enables us to formulate valid and reasonable questions that will guide the discussion with the expert towards a level of dialogue that we can communicate, while also helping us collect pieces of essential information, that the doctor may have ignored, forgotten about, or not considered them to be important.

The knowledge that is associated with the field of Gynecology has been provided by an expert obstetrician-gynecologist (who is also the co-supervisor of this thesis), with long years of experience in his specialization, both as a clinician and as an academic. Therefore, every piece of information that is collected has been evaluated by the doctor himself to ensure that it follows well established and internationally accepted guidelines and that is supported by the latest medical bibliography, as well as by his empirical observations and findings. We had a series of systematic meetings and regular discussions, in the form of interview, with the doctor in order to understand the nature of the problem and collect the associated information.

For any further questions, problems, small changes and misunderstandings that occurred, the doctor was always willing to help, either via electronic mail or by phone. This close contact with him enabled us to become familiar with the aspects of Gynecology that are essential for the solution and for having a clear vision on how to design and implement it.

In addition to the verbal collection of information, we also constructed a number of comprehensible data structures for collecting information in a standard manner, which will be discussed in this chapter.

3.2 Medical Concepts in Gynecology

The first important step that we needed to take in order to understand the context of the problem was to define the medical concepts that govern the field of Gynecology. As any other specialization in medicine, Gynecology is associated with a set of diseases, symptoms, findings and other parameters that are encountered during clinical practice. One of the challenges that we had to overcome in this thesis was to list them down and make sure that they are complete, accurate and reflect the whole range of possibilities that a gynecologist might have to deal with.

Based on our discussion with the expert, the most prominent medical concepts that we needed to know about, before beginning the system development cycle, are the following:

Diseases: Gynecology deals with a number of diseases that affect the female reproductive system [13].

Groups of Diseases: Diseases are divided into groups based on traits such as:

- their nature (e.g. cancers, transmitted through sexual contact)
- their relation to specific indicative symptoms (e.g. bleeding diseases)

- the part or the process of the reproductive system that they affect (e.g. vulva pathology, infertility, endocrinology)

After the discussion with the doctor, we ended-up with the following groups of diseases:

- *Group #1: Sexually Transmitted Infections*
  - Diseases and conditions that share the attribute of being transmitted through sexual contact
- *Group #2: Bleeding / Amenorrhea / Coagulation / Vascular Problems*
  - Diseases and conditions related to blood (coagulation, abnormal bleeding, no menstruation etc.) and vascular (thromboembolism, varicose veins) problems
- *Group #3: Pelvic Pain*
  - Diseases and conditions related to pain in the pelvic girdle area.
- *Group #4: Urogynaecology and Prolapses*
  - Diseases and conditions that affect parts of the reproductive system that are also parts of the urinary tract, and various prolapses of gynecological organs
- *Group #5: Infertility*
  - Reasons that a woman cannot conceive a child
- *Group #6: Gynecological Cancer*
  - Cancerous conditions & tumors in the female reproductive system
- *Group #7: Endocrinology*
  - Diseases and conditions that affect a patient’s hormone production
- *Group #8: Vulva Pathologies*
  - Diseases and conditions that affect the outer part of the female reproductive system (vulva), and are mostly skin conditions

The existence of groups of diseases helps the doctor focus on a specific, limited number of diseases that are the most relevant to a patient’s condition.

Symptoms: There exist a number of symptoms that a patient may present when suffering from a condition. Some of them are more indicative of a gynecological condition (e.g. symptoms related to menstruation or reproduction) than others (e.g. nausea and fever). They are usually collected by verbal diagnosis and are either observable (can be seen) or noticeable (can be felt) by the patient, or the doctor may discover their presence by asking questions during the diagnosis, to help the patient express details of their condition that they may ignored or forgot about. Symptoms are physical in nature and in many cases they are the sole piece of information that is needed to diagnose some diseases *e.g. "non-offensive, curdy vaginal discharge" is enough to diagnose Vulva Candidiasis.* Finally, some symptoms can be further described by the patient, so that the doctor can understand the cause, as well as the progression of the condition, more accurately:

e.g. "Pelvic Pain" can be reported, along with its severity, duration and localization:

"Severe Acute Pelvic Pain on right side" or "Mild Chronic Pelvic Pain at the center"

Worries and Concerns: Patients might visit a doctor out of pure concern about something that may, or may not have physical manifestation

E.g. had unprotected sexual intercourse and now worries about catching a Sexually Transmitted Disease (STD)

The patient usually seeks to relieve those worries and take further actions if needed (consultancy, treatment etc.)

Risk Factors: A patient's trait that makes them more vulnerable to specific diseases [61]. This trait might be anything such as:

- a bad habit (*e.g. increased weight, alcohol / tobacco / drug abuse, sedentary lifestyle*)
- a previous infection (*e.g. some diseases are more likely to re-emerge in previously infected patients*)
- hereditary condition (*e.g. genes related to the development of cancer*)

- previous surgeries, medications or other treatments (*e.g. bleeding from a wound caused by a past surgery*)

Risk factors alone are not adequate to accurately diagnose a disease, but they can be used in conjunction with more concrete information to further strengthen a diagnosis:

e.g. Being a "Heavy Smoker" does not necessarily imply that the patient has Cancer, however, with all other things being equal, it is more likely to develop Cancer, compared to another, non-smoker patient [62]

Findings: Results of physical, clinical, laboratory and imaging examinations are also very important for the diagnostic process and may be used, either as input for the process, or may be a result of it, depending on the occasion

Presenting Complaints: These are any symptoms and/or concerns that lead a patient to arrange an appointment with a gynecologist, in order to resolve it

Referrals: A possible reason that a patient might visit their gynecologist is the referral from another doctor. If another doctor encounters abnormal findings or markers in an examination, or even a doctor of another specialization has reasonable suspicions that the patient has a gynecological problem, they refer the patient to a gynecologist.

Diagnosis: A doctor utilizes a plethora of sources of information, such as symptoms, risk factors, patient history and examination results, in order to determine which diseases or conditions correspond to these information, as accurately as possible.

Patient Record: a record containing all the information about the patient and their medical history [63]. More specifically, it contains important information such as:

- demographics (name, age, gender, ethnicity etc.)
- history and past diagnoses

- family history and hereditary conditions
- laboratory / imaging / clinical & physical examination results
- recent or past medications / operations / treatments
- menstruation details
- sexual health & use of contraception
- habits: nutrition, activity, smoking, drug-use, alcohol-use

Information from a patient's record is of utmost importance during the diagnostic process, as they provide an insight to a patient's underlying conditions, which the doctor would otherwise not be able to know, had it not been all collected in a place:

E.g. imagine if the doctor had to ask a large number of questions, to end up collecting an extremely specific piece of information such as "Low Blood Cortisol". Considering all the possible parameters, this would yield a large number of combinations of questions to ask if done verbally and not collected by a structured document like the patient record.

Differential Diagnosis: a set of inputs (symptoms, patient information etc.) might lead to the suspicion of more than one possible diseases and conditions at a time, thus requiring more accurate information (e.g. more specific symptoms or test results) to exclude some options [2]

Emergency or Life-Threatening Conditions: some diseases are considered emergency conditions, in the sense that they pose an immediate threat to the patient's long-term health, or even their life (life-threatening), and in the case of a pregnant women, they can even harm the unborn child. Doctors give an increased priority to these cases, compared to regular and less harmful ones, in order to avoid the worst case scenarios [3]. *An example is a Pelvic Pain that is described as "Sudden, Acute and Severe", indicating an urgent condition that requires immediate action.*

Pathognomonics: some symptoms are very characteristic for some diseases and their presence greatly increase their possibility of being diagnosed [64]

By listing the above, high-level concepts of the medical practice first, we ensure that we have a brief understanding of the field of Gynecology and we are then able to proceed towards collecting more specific knowledge for the development of the system, i.e. the different instances that correspond to the above concepts, such as specific diseases, specific symptoms etc. After this is done, the next step is to organize the information in a way that is useful for the modeling of the diagnostic decision problem we are called to solve.

3.3 Diagnosis as an Argumentation problem

Since the goal of this thesis is to propose a diagnostic decision-making tool that assists gynecologists of various levels of experience, and given all the necessary information related to the context of the problem, the next step is to model the “diagnosis”.

Abstractly, a diagnosis can be described as a process of mapping a set of inputs, to a set of outputs that correspond to those inputs. Under the context of medicine that we discussed above, more specifically, in a diagnosis we map a set of patient information and symptoms, to a set of diseases that are possible under them [65]. To model the diagnostic process, we use an Argumentation framework, as described in a previous chapter.

The inputs may come from various sources such as symptoms, patient record information, risk factors and examination results. They will be used by an Argumentation framework to construct the arguments that participate in the decision-making process. Therefore, the inputs can be used to create claims which support the existence of a specific disease or condition.

The outputs are a number of diseases, findings or conditions that are relevant to the inputs. They will be used by an Argumentation framework to construct the decision options (i.e. the possible conclusions) of the decision-making process. Therefore, the outputs can be used to create the possible conclusions, which their existence is supported by a set of input symptoms and other patient information.

Once the various symptoms and patient information are related to the possible diseases and conditions by forming arguments, the next step is to reason using Argumentation. In this process, the admissible arguments are collected in an "admissible set", and any claim that is supported by its arguments is considered a possible disease or condition that corresponds to the inputs.

The "admissible set" has two main properties. The first one is that its arguments do not counter each other. This means that the claims (i.e. diseases and conditions) they support are not contradictory, under the given inputs, and therefore can be considered equally possible (the set is conflict-free) [39]. The second property is that this set is either not attacked by any arguments that are not members of it, or it does get attacked but also attacks back all arguments that attack it [5], [38]. The importance of this step is that it ensures that, given a specific set of inputs, any decision option we accept as conclusion is strongly supported by solid arguments. Also, it does not contradict any other acceptable decision option and there is no possible way of countering it with other arguments and render it invalid. This leaves no doubts about the admissibility of the results.

Depending on the case, many possible diagnoses could be relevant, therefore, more than one admissible sets may exist, which are not contradictive to each other and the claims they support are equally possible. This way, we implement the concept of "Differential Diagnosis", where, not just one, but many diseases or conditions could be possible [2], under given patient information. Also, because all resulting conclusions are equally possible, we avoid "overdiagnosing" common diseases, where a medical expert is used to encountering a number of diseases that occur much more often than others, and unconsciously falls into the trap of always resorting in one of them when diagnosing a patient. This defies the existence of other, less common, but still possible diseases, a problem that our implementation is able to sort out.

Another concept of Cognitive AI that is implemented is the revision of a decision. As the dialog between the doctor and the patient progresses, more additional information becomes available and this will lead to the consideration of new diseases/conditions or the ruling-out of some previously relevant disease/condition or even both. This is a cognitive ability of the human mind, as any time a new piece of information becomes available, our decisions are revised and change, in favour of conclusions that are supported by more evidence [5], [6]. This is captured by our implementation, by executing the argumentation reasoning step any time a new input is given, thus giving the chance to the system to re-consider its final conclusions, under the new context.

A simple graphical example of an argumentation framework that describes a diagnostic process is the following:

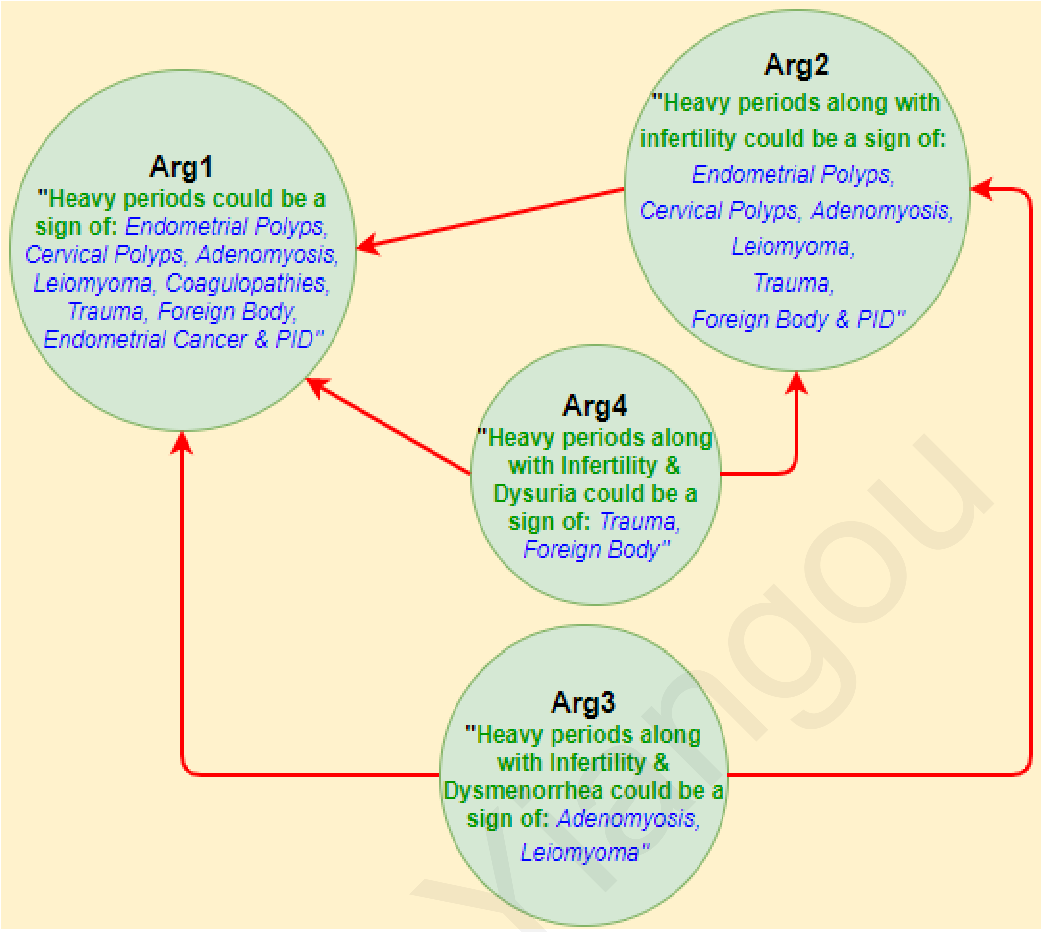


Figure 7. Diagnosis as an Argumentation Problem

The figure above (Figure 7.) describes an argumentation framework that represents a part of the knowledge that is used in the diagnostic process. More specifically, it shows the argumentation reasoning between four different arguments (Arg1, Arg2, Arg3, and Arg4), where the following applies:

- For Arg1:
  - Considered when the set of inputs contains: {*Heavy Periods*}
  - Conclusions: {*Endometrial Polyps, Cervical Polyps, Adenomyosis, Leiomyoma, Coagulopathies, Trauma, Foreign Body, Endometrial Cancer, PID*}
  - Attacks (and is preferred over) arguments: *None*
- For Arg2:

   - Considered when the set of inputs contains: {*Heavy Periods, Infertility*}
   - Conclusions: {*Endometrial Polyps, Cervical Polyps, Adenomyosis, Leiomyoma, , Trauma, Foreign Body, PID*}
   - Attacks (and is preferred over) arguments: *Arg1*
- For Arg3:
   - Considered when the set of inputs contains: {*Heavy Periods, Infertility, Dysmenorrhea*}
   - Conclusions: {*Adenomyosis, Leiomyoma*}
   - Attacks (and is preferred over) arguments: *Arg1, Arg2*
- For Arg4:
   - Considered when the set of inputs contains: {*Heavy Periods, Infertility, Dysuria*}
   - Conclusions: {*Trauma, Foreign Body*}
   - Attacks (and is preferred over) arguments: *Arg1, Arg2*

To understand the concept of adapting to new information, consider the following sets of inputs:

$$I_1 = \{\textit{Heavy Periods}\},$$

$$I_2 = \{\textit{Heavy Periods, Infertility}\}$$

If I_1 is given as input for the above Argumentation Framework, the set of admissible arguments will only contain *Arg1*, since, under this context, no arguments that attack *Arg1* are considered, and therefore it was not possible to counter the decision option that *Arg1* supports. At the end of this process, we end up with the following results:

Admissible Set = {Arg1},

Conclusions = {Endometrial Polyps, Cervical Polyps, Adenomyosis, Leiomyoma, Coagulopathies, Trauma, Foreign Body, Endometrial Cancer, PID}

Now imagine that during the diagnostic process, the doctor learns from the patient that she tries to get pregnant but cannot i.e. infertility. The new set of inputs is I_2, and since we have a new piece of information available, the argumentation reasoning process is executed once more, to revise the decisions that were made before. Under the new context, we also consider Arg2 in the decision-making process, which successfully attacks the position supported by Arg1, without opposition from other arguments. At the end of this process, we end up with the following results:

Admissible Set = {Arg2},

Conclusions = {Endometrial Polyps, Cervical Polyps, Adenomyosis, Leiomyoma, Trauma, Foreign Body, PID}

In this example, the doctor ruled-out Coagulopathies and Endometrial Cancer, by making a new decision when the information "Infertility" became available, on top of the already known "Heavy Periods". Further refinements can be made, by learning more information that concerns the patient's condition.

To understand the concept of differential diagnosis, consider the following set of inputs:

I_3*= {Heavy Periods, Infertility, Dysmenorrhea, Dysuria}*

This set enables all arguments (Arg1, Arg2, Arg3 and Arg4) to participate in the reasoning process, since it contains all the pieces of information that each one requires. By giving this input to the above Argumentation framework, we end up with the following results:

Admissible Set = {Arg3, Arg4},

Conclusions = {Adenomyosis, Leiomyoma, Trauma, Foreign Body}

Arg1 and Arg2 are not admissible because they are both attacked by other arguments without countering the attacks. Arg3 and Arg4 attack all other arguments, while receiving no attacks. They also do not attack each other, therefore they are not contradictory. This means that we can group them together in an admissible set that supports the union of their conclusions.

In this example, given the information “Heavy Periods” and “Infertility”, the doctor considers Adenomyosis and Leiomyoma to be possible diagnoses because of the additional information “Dysmenorrhea” (Arg3), and also considers Trauma and Foreign Body to be possible diagnoses, because of the additional information “Dysuria” (Arg4).

3.4 Scenario-Based Preferences

Once it is determined how to model the diagnostic process in an Argumentation framework, the next step is to collect the inputs and outputs in a way that we have a clear understanding of the relation between them, in what we call “Scenario-Based Preferences”. This relation between inputs and outputs describes scenarios from the real world, where a doctor comes across a set of patient information and symptoms (inputs) and makes a diagnostic decision (outputs) based on them.

While a computer scientist might be more comfortable with a structured, standardized way of representing data (such as code or diagrams), scientists of other fields might be more inclined towards describing a case in a more “natural way” (free-text or semi-free text or speech). In our case, Gynecology, although following some standards has more freedom of expression than computer science therefore we needed to find a way to represent our data (symptoms, conditions and diagnosis) that would enhance the communication between the two fields. Therefore, in order to collect the various Scenario-Based Preferences, it was very important that a structured manner would be used, which would make it clear how the inputs are related to various possible outputs of the system and would allow the use of natural language in the description, while retaining a proper structure. This would help both the doctor to fill them correctly without confusion and the developer of the system to understand how to program the scenarios even with minimal knowledge of gynecology.

The preferred approach was the “Scenario Tables”, two dimensional arrays which showcase the diagnostic process hierarchically.

The structure of a Scenario Table looks like this:

Scenario	**DC 1**	**DC 2**	**…**	**DC n**
Scenario 1			…	
…	…	…	…	…
Scenario k			…	

Table 2. Scenario Table - Structure

The columns represent the necessary fields to describe a diagnosis:

Scenario	**DC 1**	**DC 2**	**…**	**DC n**

Table 3. Scenario Table - Columns

The first column (Scenarios) shows the symptoms and other patient information that describe a scenario, expressed in controlled, semi-standard Natural Language. All other columns show the diseases or conditions that the inputs in the first column could lead to, expressed with a “tick” that indicates the possibility of the respective disease or condition to be diagnosed, under given inputs.

Each row is a record of the table which describes a diagnosis

Scenarios	**DC1**	**DC2**	**DC3**
PatientInfo1	✔		✔

Table 4. Scenario Table - Rows

Table 4. describes the following:

“DC1 and DC3 are both possible, under PatientInfo1”

The scenario tables follow a hierarchical structure. This means that each subsequent row is a refined version of the scenarios presented on each preceding row of its table. This means that a row describes a case where the conditions of each previous row on its table are still present, but in conjunction with other extra information.

E.g.:

Scenarios	**DC1**	**DC2**	**DC3**
PatientInfo1	✔		✔
PatientInfo1 + *PatientInfo2*	✔		

Table 5. Scenario Table - Hierarchy

The 1st row denotes that, under PatientInfo1, both DC1 and DC3 are possible. When having PatientInfo1 along with an extra piece of information PatientInfo2 (2nd row), we can safely exclude DC3, making a more accurate diagnosis which leads to DC1 only.

Usually, the higher we are on the table (upper rows), the more diseases are possible to be the case, since not many information is collected yet. The lower we visit a table (lower rows), the more the available knowledge, therefore the more diseases we can exclude, ending up with the few (usually one or two), more accurate diseases that could be the case, by the end of the table.

The hierarchies in the scenario table model the knowledge adaptation and decision revision discussed in an earlier chapter, where the introduction of additional information cause us, humans, to re-consider our initial decisions, by making new, more relevant ones based on the new context [5], [6].

In addition to all the above, a specific notation using mathematical and logical symbols is used to enhance the expressiveness of the scenario tables, consisting of the following:

Conditional OR (/):

When either *I1* OR *I2* (OR *both*) exist, DC1 and DC3 are possible

Scenarios	DC1	DC2	DC3
I1 / I2	✔		✔

Table 6. Scenario Table – Conditional OR

Conditional strengthening OR (+/):

When either *I1* OR *I2* (OR *both*) exist, DC1 and DC3 are possible, and if both do exist, they further strengthen the decision, making us more confident that the diagnosis is correct.

Scenarios	DC1	DC2	DC3
I1 +/ I2	✔		✔

Table 7. Scenario Table – Conditional Strengthening OR

Depending on its use and position in the hierarchy, it might not greatly differ from the simple Conditional OR (/).

Conditional AND (+): When both *I1* and *I2* exist, DC1 and DC3 are possible

Scenarios	DC1	DC2	DC3
I1 + I2	✔		✔

Table 8. Scenario Table – Conditional AND

Conditional strengthening AND (++):

When both *I1* and *I2* exist, DC1 and DC3 are possible and this decision is further strengthened, making us more confident that the diagnosis is correct.

Scenarios	DC1	DC2	DC3
I1 ++ I2	✔		✔

Table 9. Scenario Table – Conditional Strengthening AND

Depending on its use and position in the hierarchy, it might not greatly differ from the simple Conditional AND (+).

Below is a small sample from the collected tables for the Sexually Transmitted Infections:

Scenario	Bacterial Vaginosis	Tricho moniasis	Vulva Candi diasis	Neisseria Gonor rhoea	Chla mydia	HSV II	Ano genital Warts
Burning / Itching	✔	✔	✔	✔	✔	✔	✔
(Burning / Itching) ++ (Inter-menstrual bleeding / Post-coital bleeding)			✔	✔	✔		✔
(Burning / Itching) ++			✔	✔	✔		

(Inter-menstrual bleeding / Post-coital bleeding) ++ Vaginal Discharge							

Table 10. Scenario Table – STIs Example #1

Table 10 describes the following:

- If patient feels a burning or itching sensation in their genitals, the following diseases and conditions are possible: *Bacterial Vaginosis, Trichomoniasis, Vulva Candidiasis, Neisseria Gonorrhoea, Chlamydia, HSV II, Anogenital Warts*.
- However, if the patient also experiences inter-menstrual or post-coital bleeding (along with the above symptoms), the following diseases and conditions are more possible: *Vulva Candidiasis, Neisseria Gonorrhoea, Chlamydia, Anogenital Warts*.
- Finally, if the patient experiences burning or itching sensation, along with inter-menstrual or post-coital bleeding, and also experiences vaginal discharge, the following diseases and conditions are more possible: *Vulva Candidiasis, Neisseria Gonorrhoea, Chlamydia*.

Notice that, the lower we visit the Scenario Table, more knowledge and at the same time fewer diseases become available.

Also, notice in Table 10 how the existence of "Vaginal Discharge" in the 3rd row rules-out *Anogenital Warts*. In another, similar table (Table 11), instead of "Vaginal Discharge" there is another piece of information in the very same place in the 3rd row (specifically "Small Cauliflower-Shaped Lumps"), supporting the possibility of Anogenital Warts over any other disease and condition. Under a common set of information (here it is "burning" or "itching", along with "inter-menstrual bleeding" or "post-coital bleeding"), different next pieces of information render different diseases possible.

Scenario	**Bact erial Vaginosis**	**Tricho moniasis**	**Vulva Candi diasis**	**Neisseria Gono rrhoea**	**Chla mydia**	**HSV II**	**Ano genital Warts**
Burning / Itching	✓	✓	✓	✓	✓	✓	✓
(Burning / Itching) ++ (Inter-menstrual bleeding / Post-coital bleeding)			✓	✓	✓		✓
(Burning / Itching) ++ (Inter-menstrual bleeding / Post-coital bleeding)							✓

++ Small Cauliflower-Shaped Lumps							

Table 11. Scenario Table – STIs Example #2

For each diagnosis that builds on a specific initial symptom or patient information we create a new scenario table, e.g. a table exists which describes scenarios that begin with the patient reporting “Heavy Periods”, while another table exists which describes scenarios that begin with the patient reporting “Pelvic Pain”. Every subsequent row builds on that initial information, describing different scenarios, hence the need for new tables for each different beginning of a diagnosis.

3.5 Pelvic Pain Group of Diseases

A unique case that was challenging to handle was that of the “Pelvic Pain” group. Creating simple scenario tables, like the ones described earlier, is difficult and this is mainly attributed to some characteristics not found in many of the other groups.

Firstly, Pelvic Pain itself has various parameters to describe it such as intensity, duration, localization, side, onset and character [66]. All those needed to be taken into account when creating scenario tables because of their importance. A pain description enables us to uncover possible emergency or even life-threatening conditions (i.e. cases where pain is “Severe, Acute and Sudden”) and to determine the exact profile of the pain in order for the doctor to proceed to more accurate targeted questions. However all the potential combinations of those parameters would yield a large number of valid pain descriptions and therefore an increased number of scenario tables would be required to express them, before even proceeding to the addition of other symptoms.

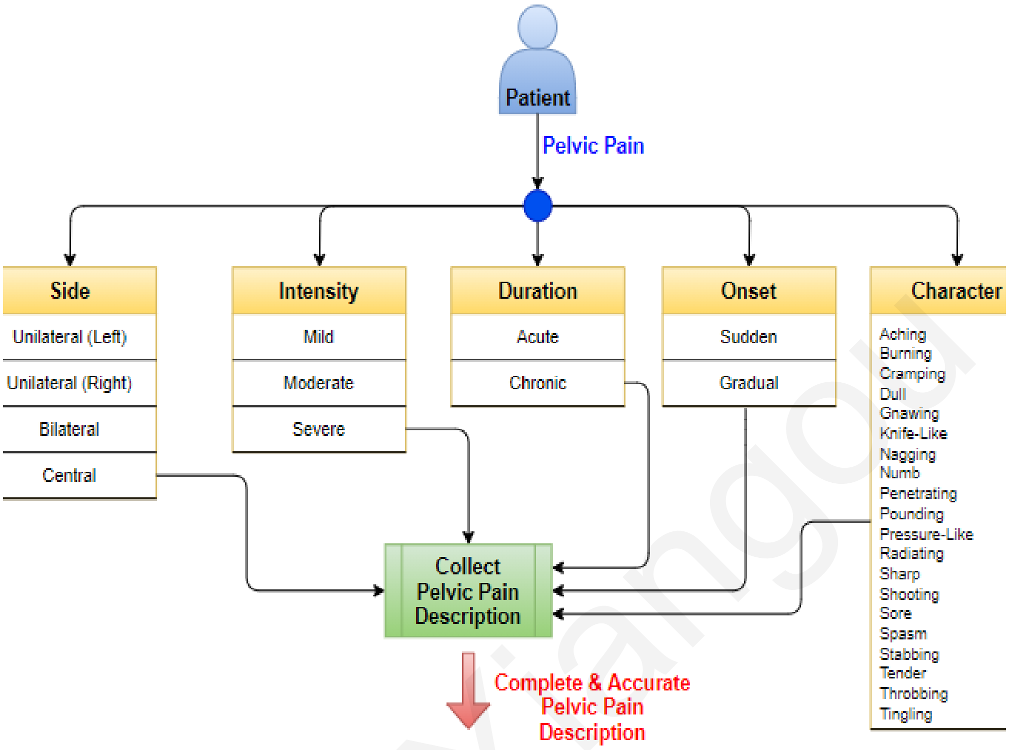


Figure 8. Pelvic Pain Description Parameters

In the second place, the number of diseases in this group is too big and broad compared to other groups. There are many possible differential diagnoses that might be the cause of pain in the pelvic girdle, and that may not even be of gynecological nature. Urological / urinary, gastrointestinal, neurological, psychological and musculoskeletal conditions, as well as pregnancy complications are possible causes of pelvic pain and all of them have to be taken into account [67]. The reason is that a gynecologist has to investigate a case of “Pelvic Pain”, even if it is not caused by a gynecological issue, in order to refer the patient to the appropriate doctor (*e.g. if “Pelvic Pain” is related to Inflammatory Bowel Disease (IBD) the doctor still needs to investigate, in order to be able to justify his/her suspicions and refer the patient to a gastroenterologist).* The problem here is that given the presenting complaint “Pelvic Pain”, a large number of diseases that are mostly unrelated would be admissible decisions in the first rows of the table. Then, in subsequent rows we would need to consider symptoms that would differentiate between them from a large number of symptoms that originated from all the fields described above and this problem is accumulated once we proceed onwards.

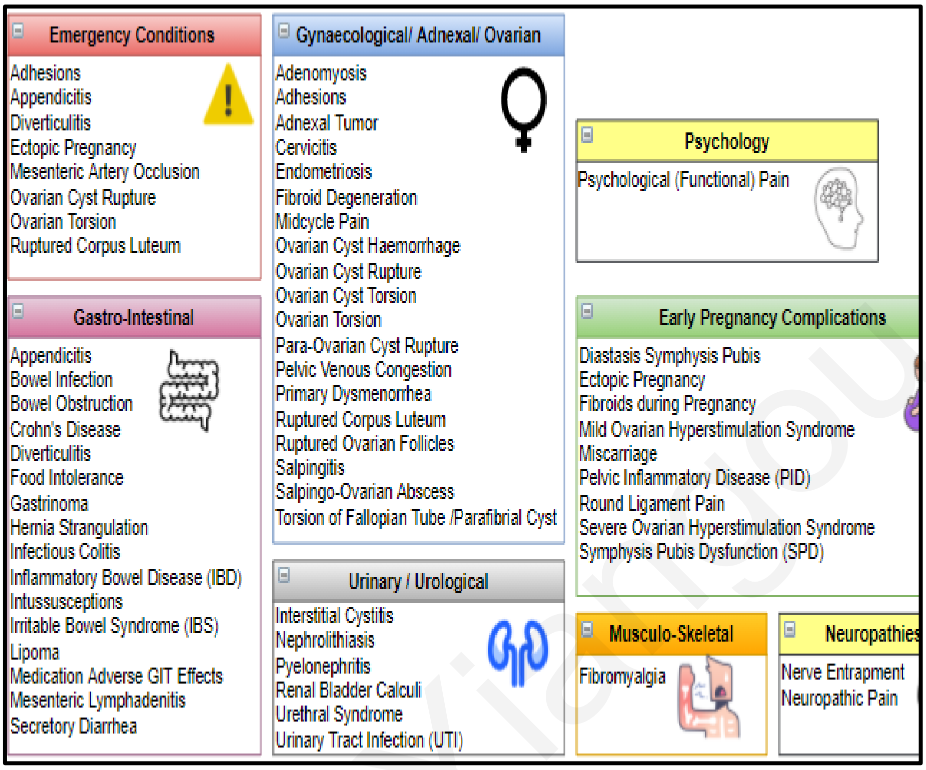


Figure 9. “Pelvic Pain” diseases grouped by their nature

All of the above contribute to the increased size of the navigation space of the group and that requires a large number of scenario tables to express it and to consider all the possible combinations and refinements of scenarios.

To address this challenge a high-level parameterization is introduced that is based on the concept of “Pathognomonics”. We know that some symptoms and patient information are very indicative of some diseases beyond any doubt [64]. The proposed approach utilizes the Pain descriptions along with symptoms that are more frequent in a disease and other strong information (e.g. test results) to construct scenario tables. This solution is not ideal, but it is an acceptable compromise to handle the enormous investigation space, while still retaining the philosophy of the scenario tables.

Firstly, this parameterization into high-level groups of information separates the pelvic pain description. There is one or more pelvic pain descriptions that are associated with each disease, depending on its progression, its location etc. These differ for each disease:

e.g. Ovarian Torsion could cause Pelvic pain described as:

- *Right-Sided or Left-Sided (unilateral)*
- *Acute*
- *Sudden*
- *Severe*
- *Any combination of the above (sudden & acute, severe & sudden…)*

The parameterization then considers all the additional symptoms (other than pelvic pain) and patient information that are more relevant and frequent to each disease, let us call this set *L2 for each disease*.

e.g. *Ovarian Torsion is associated with the following information, based on their relevance and/or frequency of appearance:*

- *Finding: Adnexal Mass/ Tumor*
- *Symptom: Fever / Shivers*
- *Symptom: Nausea / Vomiting*
- *Risk Factor: Pregnancy*
- *Risk Factor: History of Pregnancy*
- *Risk Factor: History of Ovarian Cyst*
- *Symptom: Vaginal Bleeding*
- *Risk Factor: Age between 20 and 40*

Finally the parameterization considers all pieces of information that indicate the existence of each disease beyond any doubt; let us call this set *L3 for each disease*. For this application we consider them to be test results from various imaging / laboratory / clinical / physical examinations or other strong patient information

e.g. Ovarian Torsion is associated with the following information in this level:

- *Transvaginal Ultra-Sound Scan showing Ovarian Swelling*
- *Color Doppler Ultra-Sound Scan shows Ovarian Artery Occlusion*
- *Ultra-Sound Scan shows Ovarian Torsion*
- *Laparoscopy shows Ovarian Torsion*

Given the above parameterization, in order to narrow down the number of possible diseases, we apply the following policy:

1) If Pelvic Pain is given, consider all diseases and conditions
2) Then, when a specific description of Pelvic Pain is given, prefer only diseases and conditions that could correspond to it
3) On top of the pelvic pain description, whenever a piece of information other than pelvic pain is collected then:
    a. Prefer all diseases that list it in their L2 set and consider them “activated by L2”

       i.e, only consider diseases in the diagnostic process, that they have been activated by L2. This narrows down the diseases to the ones that managed to be supported by more information than just the pain description.
4) On top of the pelvic pain description, whenever a piece of information that indicates a specific disease(s) becomes available, then:
    a. Prefer the few diseases that list it in their L3 set and consider them “activated by L3”

       i.e. only consider diseases and conditions in the diagnostic process, that they have been activated both by L2 and by L3. This narrows down the diseases to the ones that managed to be supported by more information than just the pain description, along with information that is considered strong evidence.

Using the above parameterization and policy formation, we can create scenario tables in the manner displayed in the following examples. The tables below show, in a simplified way, how we can depict the information:

Scenario	DC1	DC2	DC3	DC4	DC5
Pelvic Pain	✔	✔	✔	✔	✔
** **Pelvic Pain Description** A**	✔		✔	✔	
** **Pelvic Pain Description** A** ++ ** combination of **info from list L2 of DC1****	✔				
** **Pelvic Pain Description** A** ++ ** combination of **info from list L2 of DC1**** ++ ** combination of **info from list L3 of DC1****	✔				

Table 12. Pelvic Pain Example DC1

Scenarios	DC1	DC2	DC3	DC4	DC5
Pelvic Pain	✔	✔	✔	✔	✔
** Pelvic Pain Description A**	✔		✔	✔	
** Pelvic Pain Description A** ++ ** combination of info from list L2 of DC3**			✔		
** Pelvic Pain Description A** ++ ** combination of info from list L2 of DC3**			✔		

++ *combination of info from list L3 of DC3*					

Table 13. Pelvic Pain Example DC3

Scenarios	DC1	DC2	DC3	DC4	DC5
Pelvic Pain	✓	✓	✓	✓	✓
** Pelvic Pain Description A**	✓		✓	✓	
** Pelvic Pain Description A** ++ ** combination of info from list L2 of DC4**				✓	
** Pelvic Pain Description A** ++ ** combination of info from list L2 of DC4** ++ ** combination of info from list L3 of DC4**				✓	

Table 14. Pelvic Pain Example DC4

It is also acceptable to prefer diseases, which have been activated by L3 but not by L2, since information from L3 is indicative of the disease *e.g. the description of the "Pelvic Pain" along with a laboratory test that indicates a disease, are adequate information to support the existence of this particular disease, even in the absence of further symptoms.* It is also possible that, information from L3 is collected first (therefore activated by L3 first) and later information from L2 arrives (therefore activated by L2 after being activated by L3). These cases are displayed in the following, simplified example table:

Scenarios	DC1	DC2	DC3	DC4	DC5
Pelvic Pain	✔	✔	✔	✔	✔
** Pelvic Pain Description A**	✔		✔	✔	
** Pelvic Pain Description A** ++ ** combination of **info from list L3 of DC4****				✔	
** Pelvic Pain Description A** ++ ** combination of **info from list L3 of DC4**** ++ ** combination of **info from list L2 of DC4****				✔	

Table 15. Pelvic Pain Example DC4

To further understand how this method is related to the Scenario Tables above, we can interpret it like this: "from all the Scenario Tables of diseases that contain a specific Pelvic Pain description, find the one that is currently in the lowest hierarchy possible under given inputs. Then, only consider diseases which their table is currently in the same hierarchy as the table in the lowest possible hierarchy, under given inputs". This ensures that if more than one disease is a possible diagnosis, they are equally supported by evidence of the same level.

Finally, observe the following table:

Scenarios	DC1	DC2	DC3	DC4	DC5

Pelvic Pain	✔	✔	✔	✔	✔
** Pelvic Pain Description A**	✔		✔	✔	
** Pelvic Pain Description A** ++ ***specific info from list L2 of DC4***			✔	✔	
** Pelvic Pain Description A** ++ **specific info from list L2 of DC4** ++ ** combination of info from list L3 of DC4**				✔	

Table 16. Pelvic Pain Group: More than 1 disease possible

In Table 16 we have a scenario table where the hierarchies describe scenarios of gradual collection of information related to disease DC4. The 3rd row shows that the collected information supports two possible diseases, DC3 and DC4. Among the information there is a specific piece from the set L2 of disease DC4. This fact denotes that this particular symptom is also common in DC3 (since it is also supported) therefore it is also in the set L2 of disease DC3. In the last row of the table, only DC4 remains as a possible decision option, since the new information from the set L3 of DC4 rules-out DC3.

A real example of the above is the following:

Scenarios	**Adhesions**	**Appendicitis**	**…**
Pelvic Pain	✔	✔	✔
Sudden, Severe Acute Pelvic Pain *on the Right-side*	✔	✔	

Sudden, Severe Acute Pelvic Pain on the Right-side ++ ***Constipation***	✓	✓	
Sudden, Severe Acute Pelvic Pain on the Right-side ++ ***Constipation*** ++ ***Laparoscopy shows Appendicitis***		✓	

Table 17. Pelvic Pain Group: More than 1 disease possible

Observe the 2nd row of Table 17**.** The Pelvic Pain description matches both Adhesions and Appendicitis, therefore those two decision options are supported, along with any other that may correspond to this description.

Observe the 3rd row of Table 17. Along with the Pelvic Pain description, the patient reported "Constipation". Since both Adhesions and Appendicitis are still possible options, we realize that "Constipation" is a symptom that is common to both conditions (exists in set L2 of both of them).

Finally, observe the 4th row of Table 17. Along with the Pelvic Pain description and Constipation, an examination result becomes available ("Laparoscopy shows Appendicitis"). This information strongly indicates that the most possible condition is "Appendicitis" therefore "Adhesions" is ruled-out. This means that this specific piece of information exists only in set L3 of "Appendicitis".

3.6 Presenting Complaints, Referrals & Initial Investigation

Normally, a patient has a reason that leaded her to arrange an appointment with a gynecologist, either because she is concerned about a symptom or condition or because is referred by another doctor to a gynecologist. If a patient's reason for visiting is a concern about a symptom or condition, then we refer to it as "Presenting Complaint" [68]. In the case another doctor found something abnormal during an examination (imaging / laboratory / physical) or decided that the problem is related to the patient's reproductive system and referred her to the gynecologist, we simply refer to this information as "Referral" [69].

The reason a patient visits the gynecologist is the first piece of information that helps the doctor understand how to approach the diagnostic process, since it lets him have a first impression on what the problem is. Once the presenting complaint(s) or referrals(s) is/are reported, the doctor will choose a number of *Groups of Diseases*, among the ones mentioned in the *"Medical Concepts in Gynecology"* section, that are the most relevant to this information and focus on investigating them separately. The result of this step is the elimination of many diseases from the decision-making process that are not yet supported by any evidence at this level (therefore not relevant), and the division of the problem to smaller ones by encouraging the investigation to be continued inside the groups separately which helps the medical expert focus at a smaller investigation space each time.

After multiple interviews with the gynecologist we came up with a solution to approach this diagnostic step. Firstly, the doctor helped us identify and list-down all the possible presenting complaints and referrals that a patient could report if she would visit a gynecologist. Considering the 8 Groups of Diseases that are mentioned in the *"Medical Concepts in Gynecology"* section, next we constructed a 2-dimensional table structure that maps each listed possible presenting complaint or referral to its more relevant Groups of Diseases. Then, each time a patient case is being investigated its Presenting Complaints and Referrals are related to the appropriate entries of this table and the corresponding groups of diseases are retrieved and suggested for investigation. *To observe this table, please refer to APPENDIX A.*

This approach is used under some assumptions and considerations:

- We assume that the patient can visit the doctor with any number of presenting complaints and referrals symptoms to report
- It is possible that more than one groups of diseases may be relevant to a single presenting complaint or referral, because there may be various suspicions about the condition that causes it, even of different nature of one from another
    - *e.g. Heavy Bleeding could be a sign of both an Abnormal Bleeding condition (1st suspicious group for investigation) and a Gynecological Cancer (2nd suspicious group for investigation)* [70], [71]
    - We define this as Cross-Group overlap
    - It is up to the doctor to choose whether to accept or ignore the suggestion of those relevant groups.
    - It is also up to the doctor to choose in which order he/she will investigate the groups
- The results of this approach are accumulated, in the sense that, if a patient reports two or more different presenting complaints or referrals the final result will be the suggestion of all the groups that are relevant without exclusions, *e.g.*
    - *Presenting Complain PC1 is reported along with referral R1*
    - *PC1 is related to 2 groups of diseases*
    - *R1 is related to 1 other group of diseases*
    - *The result is the suggestion of all 3 groups of diseases for investigation*
- Considering the large number of dangers lurking in the field of medicine, including the failure to identify differential diagnoses and all possible causes [14], we prefer to be as inclusive as we can, when deciding groups of diseases to work with. More specifically, presenting complaints and referrals are collected, not with the purpose of

permanently excluding a specific group of diseases, but with the purpose of including as many relevant groups of diseases as possible to the investigation, in order to not miss any possible alternative scenarios. In fact, groups could be enabled for investigation even after this step, since they might be proved relevant at any time, when more information becomes available later on.

The presenting complaints and referrals are used to start new hierarchies in the Scenario Tables discussed earlier, as they form the foundation on which a possible diagnosis is built on. By knowing the reason behind a patient's visit, the doctor can ask further questions based on them. Using this logic, we create Scenario Tables, with their very first row describing the case of only having the reason of visiting, without any additional information yet. Each reason of visiting requires a new Scenario Table to distinguish them. With every subsequent row that builds on its previous one, we are recursively building on the presenting complaint or referral that lies in the very first row of the Scenario Table.

Scenarios	**Polycystic Ovarian Syndrome**	**Hypo Thyro idism**	**Hyper Thyro idism**	**Cush ing's**	**Hyper Prolacti nemia**	**Corpus Luteum Insuffi ciency**
Amenorrhea	✔	✔	✔	✔	✔	✔
Amenorrhea ++ *Weight Gain*	✔	✔		✔		✔

Table 18. Endocrinology Scenario Table

In Table 18, we observe a Scenario Table from the group of Endocrinology. The patient reports that she missed a menstrual period (*Amenorrhea*) and this is the presenting complaint which starts a new Scenario Table, and therefore a new hierarchy (*notice that it occupies the very first cell of the table*). This specific presenting complaint leads the doctor to the group of Endocrinology, where inside this group he continues his/her investigation by further building upon this presenting complaint (*notice the 2nd row of the table*), until a diagnosis is reached.

3.7 Cross-Group Overlap

Another concept that is implemented is the “Cross-Group Overlap”. It refers to the potential of more than one group to be recommended for investigation. This happens for various reasons such as:

- A patient initially reports a number of presenting complaints and/or referrals that are relevant to many groups of diseases
    - *e.g. patient complains about “Itchiness” in her vulvar area, a symptom which is related to both groups of “Sexually Transmitted Infections” and “Vulva Pathologies” , thus both are recommended for investigation.*
- During the investigation of a specific group, a piece of information has been extracted from the dialog with the patient, that is also related to other groups of diseases, therefore those groups are also suspicious for investigation
    - *e.g. patient complains about “Vaginal Discharge”, a symptom which is related to the group of “Sexually Transmitted Diseases”* [72], *which is recommended for investigation. The doctor visits this group and later, during the investigation, he/she learns from the patient that she also experiences bleeding after having sexual intercourse (“Post-Coital Bleeding* [73]*”). This newly extracted symptom is related to another group of diseases: “Bleeding / Amenorrhea / Coagulation & Vascular Problems”, which is immediately recommended for investigation.*

When two or more groups are suspicious for investigation, the doctor investigates one of them and once he/she reaches to a point where he believes the diagnosis is not sufficient enough or that the current group has nothing else to offer to the process, or desires to check or eliminate alternative cases, he/she can visit another group. Any information that was collected initially (i.e. presenting complaints and referrals) or from investigating previous groups are still at effect.

The same table structure mentioned earlier, which maps each listed possible presenting complaint or referral to its more relevant Groups of Diseases is used for this *cause (To observe this table, please refer to APPENDIX A)*. Every time a new piece of information becomes available, either in the initial level (presenting complaints & referrals) or inside the groups (symptoms and patient information), its corresponding groups of diseases are found from this table and are recommended for investigation.

3.8 Decision-Making Theoretical Model

The system follows the natural way a diagnostic process progresses in the real world, as explained by the doctor.

The very first step is to collect the patient's information for his/her health record, which includes:

- *Demographics*
- *Personal & Familial History*
- *Laboratory & Imaging Tests*
- *Clinical & Physical Examinations*
- *Past Operations*
- *Medications & Treatment*
- *Menstruation (last menstrual period, days of a normal menstrual cycle etc.)*

- *Sexual Health, History and Contraception*
- *Nutritional, Activity and Exercise habits*
- *Smoking / Alcohol / Drugs / Needle use*

Any new addition to the patient record, along with older entries will be utilized in the diagnostic decision-making process, in conjunction with the reported symptoms and complaints.

Normally a patient has a reason that desired to arrange an appointment with a gynecologist, either because is concerned about a symptom or condition or because is referred by another doctor. Therefore, the second step is to collect the Presenting Complaints and Referrals, as described in the "*Presenting Complaints, Referrals & Initial Investigation*" section. The result of this step is the recommendation of all the relevant Groups of Diseases to the doctor, in order to choose from where to begin his/her investigation. Those Groups of Diseases occurred from the methodology described in the "*Presenting Complaints, Referrals & Initial Investigation*" section, using a table that maps presenting complaints and referrals to their relevant groups of diseases.

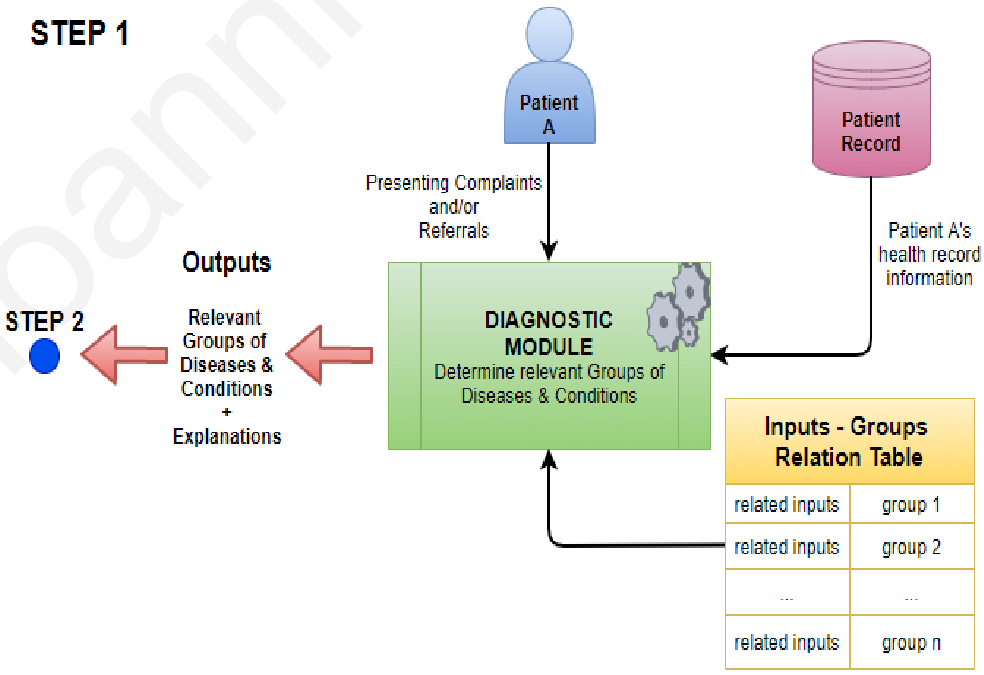


Figure 10. Decision-Making Theoretical Model – Step 1

Once the doctor enters a Group of Diseases for investigation, the next step is a more focused diagnostic process, where more additional symptoms and patient information are collected, this time to narrow-down the number of possible diseases inside a group, to the few, most relevant ones. We consider the entire set of inputs collected up to this point for this decision (from the patient record data and the presenting complaints / referrals, to the latest reported symptom). As the dialog between the doctor and patient progresses, more and more information becomes available and each time, the decision-making process rules-out even more diseases. This would help the doctor feel more confident to ask more targeted questions that lead towards the ruling-out or the diagnosis of a specific, targeted disease or condition. This step is repeated until the doctor is satisfied with the results.

- This last step does not aim at returning a perfectly accurate diagnosis of a condition or disease, but rather at recommending a number of possible diseases or conditions that are considered to be more relevant to the case being described
- Diseases from other groups may also become relevant for investigation along the way

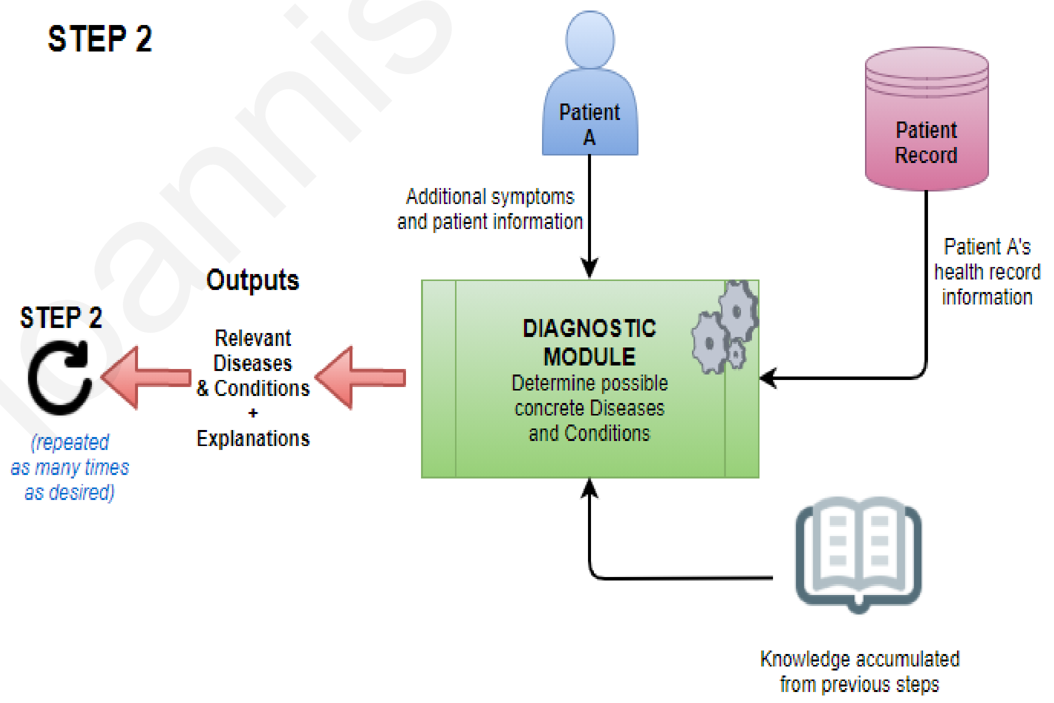


Figure 11. Decision-Making Theoretical Model – Step 2

3.9 Common Sense Knowledge

Some other concepts that the system approaches do not necessarily originate from, or are related directly to the field of medicine or gynecology per se. Those concepts are related to the human common sense knowledge, dealing with issues such as how human beings interpret information and how low-level (e.g. numeric) data are perceived and processed by our mind in order to extract comprehensible high-level concepts (e.g. terms).

More specifically, the system is implemented in a way that it considers the following issues when executing a diagnostic process:

- Complements (blocking information/arguments): Some pieces of information cannot co-exist because in the real world they contradict each other
    - Explicit example: a patient cannot be at a state of both being pregnant and non-pregnant at the same time
    - Implicit example: patient cannot have history of nulliparity (i.e. condition on which a woman never carried a pregnancy) if she has already given birth before
- Age Groups:
    - The number of years determines the age group that a patient belongs to
        - e.g. 13 <= Age (in years) <= 19 → patient is "teenager"
        - e.g. 53< Age (in years) → patient is "at postmenopausal age"
    - A patient may belong to more than one age groups, depending on her/his age
        - e.g. If a patient is 17 years old, she is both a "teenager" and "at a reproductive age" [74]
    - Age groups are crucial for addressing possible risk factors and strong differentiations between scenarios, that are associated with age [75]
        - e.g. for a 60-years old woman that does not have menstruation, her condition is most probably related to menopause and not to other diseases

or conditions, whilst this would be a completely different case if the patient was at her teen ages.

- Using the Body Mass Index (BMI) to be scientifically accurate [76], we translate the height and weight measurements to high-level concepts to describe the relation between a patient's weight and height, in a way that a human could understand easier (e.g. obese, overweight). This is important for determining risk factors that have to do with a patient's weight, with respect to his/her height
    - e.g. Obesity is a risk factor for Ovarian Cancer [77]
- A woman that reports zero number of children to her doctor, is also considered to never having given birth or carried a pregnancy before, therefore she has history of "nulliparity".

3.10 Explanations, Prompting & User-System Dialogs

At the end of each diagnostic step, along with the results, the system offers explanations for each one of the resulted final decisions. Each explanation describes the reason we concluded to a specific group or disease. This is achieved by collecting all the admissible arguments, and then by extracting explanations for each one of them we tailor those pieces to form comprehensible, reasonable and convincing explanations in Natural Language [37].

In order to make the expressions more detailed and at the same time more colorful, a methodology is used for the explanations of diseases, which distinguishes between the various admissible arguments that consist them. More specifically, the explanation is consisted of smaller explanations, much like paragraphs, in the following manner:

1) Paragraph #1: Explains which were the very first pieces of information that made the disease, even slightly, possible
    - usually, it is a presenting complaint or referral that is listed in here

- a simple statement that the corresponding disease is a possible decision option (i.e. possible cause) is also added here

2) Paragraph #2: Explains which other pieces of information support this decision
3) Paragraph #3: Explains which pieces of information leave no doubt that this decision is possible, when they are present with the above information.

e.g.

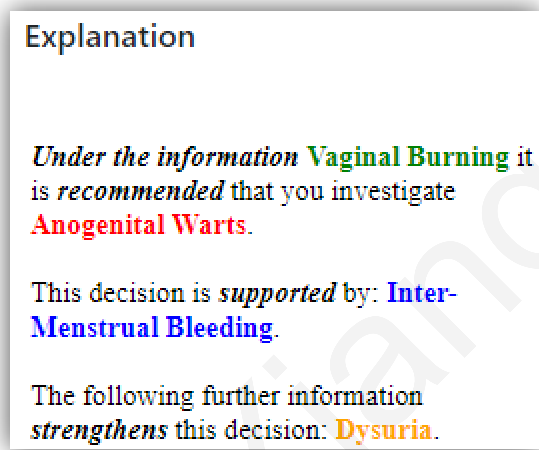


Figure 12. Explanation from the system

In Figure 12 the disease of interest is “Anogenital Warts”. The presenting complaint is “Vaginal Burning”. The symptom “Inter-Menstrual Bleeding” supports the belief that this specific disease is possible when “Vaginal Burning” is present. The additional symptom “Dysuria” leaves no doubt that, when “Vaginal Burning” and “Inter-Menstrual Bleeding” also exist, this specific disease is possible.

3.11 Scale of the Problem

From theory to development, and from medicine to computer science, one thing remains certain: the scale of the problem we are tasked to analyze is enormous. In order to comprehend this fact, a collection of numerical values has been created, representing the sizes of different information structures and coding documents that were necessary for the complete coverage of the problem. Such metrics include:

<u>Totals:</u>

1) Total input information of Patient Record (i.e. the complete vocabulary of historical information), per section (e.g. Demographics, Laboratory Tests etc.)

2) Total number of symptoms, complaints and additional information (i.e. complete vocabulary of user inputs, excluding the patient record information)

- *This includes the different descriptions of symptoms too, e.g. "Pelvic Pain" is considered different than "Right-Side Pelvic Pain": though the first is a generalization of the latter in medical terms, a computer system understands those terms as two different vocabulary entries*

3) Total number of diseases/conditions and groups of diseases (i.e. decision options)

<u>Per Group of Diseases:</u>

1) Number of total rows of Scenario Tables

2) Number of lines of code in Gorgias

3) Number of diseases

4) Number of related symptoms and background information

Total input information of Patient Record (per section)		
Section	**Value**	**Material**
Demographics	51	*Free-text, Numerical & Date fields + Drop-Down*
History (Past Diagnosis & Existing Conditions)	98	*Drop-Down List Options (per list)*
Family History (Past Diagnosis & Hereditary Conditions)	30	*Drop-Down List Options (per list)*
Laboratory Tests	65	*Drop-Down List Options (per list)*
Imaging Tests	80	*Drop-Down List Options (per list)*
Clinical & Physical Examinations	32	*Drop-Down List Options (per list)*
Past Operations	19	*Drop-Down List Options (per list)*
Medications & Past Treatments	14	*Drop-Down List Options (per list)*
Menstruation	16	*Free-text, Numerical & Date fields + Checkboxes*
Sexual Health, History & Contraception	21	*Drop-Down List Options (per list)*
Nutritional Habits & Eating Disorders	11	*Drop-Down List Options (per list)*
Exercising & Physical Activity	6	*Drop-Down List Options (per list)*
Smoking	2	*Drop-Down List Options (per list)*
Drugs & Needles use	3	*Drop-Down List Options (per list)*
Alcohol	2	*Drop-Down List Options (per list)*
TOTAL VOCABULARY:	**450**	**terms & concepts**

Table 19. Patient Record Vocabulary size

Total Number of Information (excluding Patient Record)		
Section	**Inputs**	**Decision Options**
GROUP #1: Sexually Transmitted Infections (STIs)	49	10 diseases
GROUP #2: Bleeding / Amenorrhea / Coagulation / Vascular Problems	152	59 diseases
GROUP #3: Pelvic Pain	151	55 diseases
GROUP #4: Urogynecology & Prolapses	94	9 diseases
GROUP #5: Infertility	76	4 diseases
GROUP #6: Gynaecological Cancer	118	9 diseases
GROUP #7: Endocrinology	122	28 diseases
GROUP #8: Vulva Pathologies	18	3 diseases
TOTAL (including overlap)	**780**	**177**
TOTAL (excluding overlap)	**507**	**139**
INITIAL DIAGNOSTIC STAGE (Group selection)	**84**	**8 groups of diseases**

Table 20. Total Number of Input and Output Information per section

Rows of Scenario Tables (for Groups of Diseases)	
Section	**Rows**
GROUP #1: Sexually Transmitted Infections (STIs)	81
GROUP #2: Bleeding / Amenorrhea / Coagulation / Vascular Problems	127
GROUP #3: Pelvic Pain	*275
GROUP #4: Urogynecology & Prolapses	66
GROUP #5: Infertility	12
GROUP #6: Gynaecological Cancer	22
GROUP #7: Endocrinology	23
GROUP #8: Vulva Pathologies	6
TOTAL	**612**
**modeled using the high-level parameterization described in Chapter 2, the estimated number of rows without it would have been much larger*	

Table 21. Total Number of rows in Scenario Tables

Lines of *Code in Gorgias (for Groups of Diseases)	
Section	**Lines**
GROUP #1: Sexually Transmitted Infections (STIs)	864
GROUP #2: Bleeding / Amenorrhea / Coagulation / Vascular Problems	2040
GROUP #3: Pelvic Pain	2435
GROUP #4: Urogynecology & Prolapses	400
GROUP #5: Infertility	214
GROUP #6: Gynaecological Cancer	549
GROUP #7: Endocrinology	795
GROUP #8: Vulva Pathologies	119
TOTAL	**7416**
**only considering non-empty lines of code*	

Table 22. Total Number of lines of code in Gorgias

The above tables show an estimate of the magnitude of the problem we are analyzing. The lines of code in Gorgias represent the size of the representation of the theory in argumentation. It would have not been possible to code all this knowledge, rule-by-rule, without first bringing the problem to a manageable condition that we could control, else this would require a bruteforce approach of adding knowledge. Using simple and comprehensible, yet expressive, data structures such as the Scenario Tables, we were able, not only to more easily control, store and process the large amounts of information we have been given, but also to promote modularity in the development of the project. The latter is of utmost importance for the achievement of our goal of collecting the complete expert knowledge, as it would allow future expansion work to be implemented easier, just with the addition of new rows on scenario tables.

Chapter 4

Requirements and Implementation

4.1 Requirements and Specifications

The overall aim of the project is to provide support to both senior and junior doctors, as well as other types of users, by assisting them in different ways depending on their level of practical experience in the field and their level of medical knowledge.

Although consultant gynecology experts and senior gynecologists are expected to always perform optimally, given their long years of experience in their field (both in theory and in practice), there is still room for additional support to guarantee the quality of their service. The system aims at assisting this focus group on refreshing their memory, honing their skills and reminding them of cases they might forgot to examine, given the extreme pressure and stressful environment of their job. Also, if the system gets constantly updated to host the newest trends and practices in the medical field, it will help the doctors of this group to stay updated in a process of life-long learning, while also encouraging them to incorporate computer/AI assistance to their practices, thus keeping up with the technological advancement of our era.

In the case of junior and/or trainee doctors, the system will be able to enhance their training and learning process in gynecology. This is done by helping them understand how different parameters influence the decision making process, what piece of information could lead to the need of investigating a new group and how the dialog with the patient guides the diagnostic process (e.g. what to ask/exclude next, how many clues are "sufficient" for a final diagnosis).

Finally, various other interest groups such as nurses/paramedics and simple users with no medical expertise could be benefitted from the system, mainly by having a first glance at the conditions that various symptoms or other patient information lead to. This would help nursing staff identify the urgency of the case of a patient in order to act accordingly (arrange an appointment immediately, or for a later date etc.), as well as simple users to form an initial opinion on whether their condition really calls for visiting a doctor.

Considering the need to make the system respond to the above overall aim, a list of expected requirements and specifications has been extracted where we describe more specifically what is needed to be done, how to do it and what are the means to achieve it. Making the system respond to these requirements and specifications guarantee that it works as it is expected (offer support to doctor users), while providing users with a list of the implemented features and the way they enhance their experience with the system.

Requirements and Specifications:

1) *Host a complete knowledge of medical expertise, as practiced by gynecologists today, and keep it updated over the course of the time*

This is succeeded by interviewing the doctor and collecting all the medical background information that the system needs to act as expected. This includes the importance of the patient record, the symptomatology, the various types and causes of diseases and conditions, the examinations related to each disease, the possible risk factors and other necessary medical/gynecological information.

Regarding the aim to keep this knowledge updated as it evolves over time, it can be achieved via reviewing it systematically with the help of many experts on the field in set intervals, e.g.

every two or three years. This will provide a sufficiently frequent information update and can be done by gather and documenting the inputs and expected outputs given by experts, in order to identify the change in the expertise and implement it to the actual system, in a modular manner.

2) *Formalize Medical expertise, within Argumentation logic, in order to become material ready for coding scenario preferences*

This is succeeded by constructing a formal vocabulary of predicates that would represent each piece of knowledge (inputs: symptoms and patient information, decision options: diseases or whole groups of diseases), while covering important attributes of information such as specific descriptions (e.g. "Severe & Acute Pelvic Pain") or origin of information (e.g. "High Blood Pressure" comes from the patient record, specifically from Laboratory tests).

3) *Enhance the accuracy and speed of a doctor's diagnostic process*

Enhancement of diagnostic accuracy is succeeded through alerting the user to the broad range of possible diagnoses, by providing him/her with full radar of diseases at all times, which dynamically changes as more information becomes available. This is achieved by:

1. Alerting users if new groups of diseases are available, given a new piece of information
2. Showing all diseases that are possible under a given set of inputs in order to ensure that the user is well informed about possible Differential Diagnoses
3. Automating the diagnostic process by using an Argumentation framework which has the ability of revising previous decisions when presented with new information

Enhancement of diagnostic speed is succeeded through guiding the user into collecting further relevant information

This is achieved by providing targeted prompts or questions to new relevant and differentiating/confirming information, in the form of interactive dialogs with the interface

4) *Encourage broader diagnoses to reduce the "over-diagnosis" of common diseases out of convenience*

The coverage of all gynecological conditions, the flagging of all possible diseases with shared symptoms and the ability to inform the user when a group of investigation becomes relevant under a new, specific information (cross-group overlap) are the means to combat the phenomenon of "over-diagnosing" common diseases, where a doctor, usually for the ease of convenience, tends to diagnose cases that occur quite often, while ignoring rarer cases that could still be possible.

Another mean of achieving this goal is the stepwise guidance to further relevant information via prompting, where highlighting the next important information to gather, or explicitly asking the user for it, ensures that the diagnosis is taking a path towards uncovering all the relevant diseases, before concluding to one.

5) *Alert when a life-threatening situation is possible*

By separating out diseases that could be life-threatening and giving them priority in investigation, we ensure that this requirement is covered

This is achieved by bringing emergency diseases to the foreground attention of the doctor, through the use of prompting dialogs, which ask the doctor relevant questions in order to rule-out or consider life-threatening conditions first, before proceeding

6) *Reduce the over-use of resources*

This requirement is covered by satisfying the above requirements, mainly because they implicitly promote the correct use of resources, whether this means reducing the number of tests and visits to the doctor, or sparing the doctor's precious time and reducing his/her cognitive load

The system promotes a comprehensive and structured diagnostic analysis that helps in organizing the thoughts of the doctors, allowing them to exhibit greater concentration on their work, thus implicitly improving their performance

7) *Follow a Natural Interaction Process with users, that simulates a diagnostic process in real-life*

The interaction with the users should follow an iterative process through cycles of information gathering, following this sequence:

1. Collect information through dialog
2. Focus on the possible diagnoses
3. Provide explanations that justify the possibilities
4. Repeat until satisfied with the diagnostic results

The interaction process should represent a real-life diagnostic process, where:

1. Firstly, the doctor collects historical information about the patient
2. Secondly, the patient reports his/her initial complaint or presents another doctor's referral to the gynecologist. This step leads the initial focus on possible groups of diseases.
3. Thirdly, the doctor collects strong differential symptoms and/or information within each possible group. This step leads the focus on possible diseases within suspicious groups.
4. Fourthly, the doctor continues by collecting further differential information within each possible group, in order to narrow-down the number of possible diseases to the most suspicious ones, given the new information.
5. Finally, the doctor can collect supportive symptoms for each possible disease, in order to ensure/reinforce his/her final decisions

- In each of the above steps, the possible diseases within the current group of investigation are highlighted, while informative explanations for each possible disease are available.

- The general interaction and interface of the system with the user follows the natural dialog with a doctor and offers highlighting, pop-up windows and other methods of prompting to guide this process

8) *Design of User-Friendly Interfaces*

The interfaces should be designed by following a comprehensive structure. This is achieved by separating the screen to two areas:

- Left-side of the screen: information to collect such as symptoms, history information, risk factors (i.e. user inputs)
- Right-side of the screen: decision options and results such as groups of diseases or diseases, conditions and other findings (i.e. system outputs)

This method is quite appealing to younger doctors since it follows the way many applications are designed today, with distinct separation inputs and outputs. This structure is also easily understandable by senior doctors, thus promoting the incorporation of computer assistance to their practices

9) *Incorporate methods of improving system performance*

By using Query Optimization techniques we can make the automated diagnostic process more effective, by selectively consulting only a piece of the knowledge that is related to the current known information, thus minimizing execution time

10) *Implement the concept of "Explainable Artificial Intelligence" (XAI)*

This is achieved by providing explanations in a comprehensive manner, at each step of the diagnostic process, for possible groups of diseases / possible diseases (depending on the current step). The structure of those explanations will follow the principles of natural language, where multiple pieces of information are tailored together using string-manipulation techniques and the appropriate prepositions and linking words, in order to form structurally and logically valid sentences in natural language. The explanations will follow the form of "justifications", meaning that they will be constructed in a way that a set of input information (i.e. symptoms and other patient information) justify the final decision returned by the system (i.e. group of diseases or diseases and other conditions)

***11)** Follow a Modular Architecture*

It is important that the system is implemented in a way that is easy to make changes, to replicate parts of it or to extend it with additional functionality or knowledge, without affecting the operation of the rest of the system. This is done by separating the processes in front-end (user interface and user-system interactions), middleware (running processes, handling requests and linking the various pieces together) and back-end (Argumentation reasoning, running Prolog and Gorgias files)

***12)** Preserve user's freedom by "recommending" options and not "forcing" them*

It is important that the system preserves the user's right to not follow the recommended solutions or prompts. This is especially important since doctors may have their own, unique approach on diagnosis and molding their style of practice into the one promoted by the system should not be mandatory. This is achieved by the following means:

1. The system gives the opportunity to the user to close the dialogs, or continue the diagnosis without answering any questions
2. The system lets the users go back to previous pages and visit other groups of diseases or stages of the diagnostic process (e.g. patient record screen) even if the diagnosis was abandoned midway
3. Disease and group buttons are "clickable" at any time. It is up to the doctor to choose what group to visit or what diseases to examine. All groups and diseases are available for investigation and the system cannot force any of them, it can only recommend the most relevant ones

4.2 Technologies and Tools

Many implementation techniques have been incorporated, since the system uses various components that require different approaches to coding.

The *website* that acts as the Graphical User Interface of the system was designed and implemented by combining various popular web development tools, specifically the following:

- *HyperText Markup Language (HTML):* client-side markup language for designing the components of the pages
- *Cascading Style Sheets (CSS):* client-side styling language for styling the pages
- *Bootstrap:* an open source framework which enhances the styling abilities of CSS
- *JavaScript (JS):* client-side scripting language for writing scripts that make the page interactive, such as popping up windows, showing alerts, changing colors upon a specific action etc.
- *jQuery:* a JavaScript library which enhances the event-handling, as well as other abilities of the scripting language
- *PHP:* server-side scripting language used to program the dynamic changes in the website, the redirecting from one page to another and services for printing necessary files and calling other processes (such as JAR files or Prolog files)

Additional tools for implementing and hosting the website include the following:

1) *Apache server* for hosting the website
    - To host it locally: XAMPP was used
    - To host it online: a repository on a web server was offered by the department
2) *MySQL database* for hosting user credentials and other information
    - Local database: XAMPP offers its own MySQL database (MariaDB)
    - Online database: a database (MariaDB) was offered by the department

The *Argumentation Framework* is developed in Gorgias, which is an extension of the logic programming language Prolog that enables programmers to write rules/arguments, express scenario-based preferences and derive explainable conclusions that resulted from applying common sense and argumentative reasoning. Prolog was used for various, general purposes that required Logic Programming, such as list manipulation and fact declarations.

In order to connect the Argumentation module with the website, a middleware program was written in Java programming language, by utilizing the JPL library, which enables Java programs to run Prolog queries (therefore Gorgias queries too). The JPL library links the Argumentation module to the middleware, while a PHP service that runs extracted Java programs (JAR files) was used to link the webpage with the middleware, thus achieving an implicit connection between the webpage and the Argumentation module. The result of this connection is the ability to ask Gorgias queries and retrieve results and explanations, all from a user-friendly graphical environment, thus bypassing complex Prolog commands or the use of low-level terminal consoles.

4.3 Modeling the Scenario Tables in Gorgias

As mentioned in a previous chapter, all the necessary information related to the various diagnostic scenarios has been collected using the Scenario Tables. This structure allows us to easily translate the information acquired to Gorgias code, in order to model the various scenario preferences that participate in the Argumentative reasoning process.

In order to model a Scenario Table in an Argumentation Framework, we need to specify the following first:

Decision Options: those are the possible conclusions of an argumentation process. In our case, the decision options indicate whether or not a particular disease can be diagnosed with the given knowledge, or not, therefore we have two options:

- A disease is *possible* under a set of given patient information (*poss)*
- A disease is *not possible* under a set of given patient information (*notPoss)*

Object-Level Rules: they use the decision options mentioned above, in order to model the possibility for each disease, when given a set of symptoms and/or patient information. For each disease *D,* we create the following two Gorgias rules based on the above two decision options:

- *rule(poss(D), dis(D), []).*
  - "*Disease D is possible*"
- *rule(notPoss(D), neg(dis(D)), []).*
  - "*Disease D is not possible*"
  - *neg(P):* the negation of predicate *P*

Priority Level Rules: This type of rules enables us to program preferences for a decision ("Possible" or "Not Possible"), under certain circumstances. They are used to model the scenarios (rows/ records of the Scenario Table).

Priority Level Rule - General Notation:	**Explanation:**
rule(rName(X,E), prefer(d1(X), d2(X)), []):- C.	*"Rule rName states that, for disease X, decision d1 holds, and is preferred over decision d2, when set of conditions C is met. A list of explanations E is returned that justify this preference"*

Table 23. Priority Level Rule

When diagnosing a disease, we want to exclude any other disease that does not hold based on given context. Therefore we need two priority level rules for each row of the Scenario Table:

- 1 rule for *including* a disease to our decision, by specifically using its name
- 1 rule for *excluding* a disease from our decision, by specifically using its name
- The two rules have to have the same name, in order to overwrite one of them, when the other holds

e.g.

rule(p1(X), prefer(poss(X), notPoss(X)), []):-c1, c2, c3, (X=x1 ; X=x2).

rule(p1(X), prefer(notPoss(X), poss(X)), []):-c1, c2, c3, not(X=x1 ; X=x2).

The 1st rule states that: *"disease X is preferred to be possible (rather than not possible), if conditions c1, c2 and c3 are all met, and this disease is specifically, either x1 or x2".*

The 2nd rule states that: *"disease X is preferred to be not possible (rather than possible), if conditions c1, c2 and c3 are all met, and this disease is specifically, neither x1 nor x2".*

The above technique ensures that:

- A disease is diagnosed only with the appropriate collection of knowledge
  - For the above example: there cannot be a diagnosis with this specific conditions only (c1, c2, c3), that do not include x1 or x2 to the solution
- The name of a disease is collected and stored in a variable (X) to be used later

<u>*High Order Priority Rules:*</u> This type of rules describes which scenarios are preferred over others, in case of overlap. These rules are used to implement the Scenario Table Hierarchies, meaning that the preference of refined scenarios over their simpler versions can be implemented this way.

High Order Priority Rule General Notation:	**Explanation:**
rule(h1(X,_),prefer(r1(X,_), r2(_,_)),[]).	*"Rule h1 states that, whenever the*

	scenarios described by rules r1 and r2 overlap, prefer the decision of r1"

Table 24. High Order Priority Rule

<u>Translating Conditional OR (/):</u> there are 2 possible notations in Prolog/Gorgias to express it

Scenario	d1	d2	d3
s / e	✔		✔

Table 25. Translating Conditional OR

- Notation #1: Using Prolog's available conditional OR operator (';'), e.g.:

  <u>% rules for s and e, together (s;e)</u>

  rule(p1(X,E), prefer(poss(X), notPoss(X)), []):-(s ; e), (X=d1 ; X= d3).

  rule(p1(X,E), prefer(notPoss(X), poss(X)), []):-(s ; e), not(X=d1 ; X= d3).

- Notation #2: Using Prolog's common practice of inferring a new rule for each part of an OR operation (in our case: 2 rules for the left part of the operator, and 2 for the right part, considering the steps described in Priority-Level Rules section)

  <u>% rules for s</u>

  rule(p1(X,E), prefer(poss(X), notPoss(X)), []):-(s), (X=d1 ; X= d3).

  rule(p1(X,E), prefer(notPoss(X), poss(X)), []):-(s), not(X=d1 ; X= d3).

  <u>% rules for e</u>

  rule(p1(X,E), prefer(poss(X), notPoss(X)), []):-(e), (X=d1 ; X= d3).

  rule(p1(X,E), prefer(notPoss(X), poss(X)), []):-(e), not(X=d1 ; X= d3).

<u>Translating Conditional strengthening OR (+/):</u> Same approach as the normal conditional OR operator, with the addition of High-Order Priority Rules to define the preference for a stronger (more refined) scenario

- Since we have a strengthened scenario, this means that it is a refined version of other scenarios in its hierarchy therefore we set it as more preferred over any previous ones in the hierarchy, by using High-Order Priority Rules.

  E.g. using Prolog's OR operator (can be implement with the other option too)

Scenario	**d1**	**d2**	**d3**
s	✔		✔
s +/ e	✔		

Table 26. Translating Conditional Strengthening OR

% rules for s

rule(p1(X,E), prefer(poss(X), notPoss (X)), []):-s, (X=d1 ; X= d3).

rule(p1(X,E), prefer(notPoss(X), poss(X)), []):-s, ***not****(X= d1; X= d3).*

% rules for s +/ e

rule(p1_refined(X,E), prefer(poss(X), notPoss (X)), []):-(s **;** *e), (X= d1).*

rule(p1_refined (X,E), prefer(notPoss(X), poss(X)), []):-(s **;** *e),* ***not****(X= d1).*

% prefer refined scenario s +/ e *over simpler version* s

rule(h1(X,_),prefer(p1_refined(X,_), p1(_,_)),[]).

Translating Conditional AND (+): Using Prolog's conditional AND operator ('**,**')

Scenario	**d1**	**d2**	**d3**
s + e	✔		✔

Table 27. Translating Conditional AND

% rules for s + e

rule(p1(X,E), prefer(poss(X), notPoss(X)), []):-(s **,** *e), (X=d1 ; X= d3).*

rule(p1(X,E), prefer(notPoss(X), poss(X)), []):-(s **,** *e), not(X=d1 ; X= d3).*

Translating Conditional strengthening AND (++): Same approach as the normal conditional AND operator, with the addition of High-Order Priority Rules.

- Since we have a strengthened scenario, this means that it is a refined version of other scenarios in its hierarchy; therefore we set it as more preferred over any previous ones in the hierarchy, by using High-Order Priority Rules.

Scenario	**d1**	**d2**	**d3**
s	✓		✓
s ++ e	✓		

Table 28. Translating Conditional strengthening AND

% rules for s

rule(p1(X,E), prefer(poss(X), notPoss (X)), []):-s, (X=d1 ; X= d3).

rule(p1(X,E), prefer(notPoss(X), poss(X)), []):-s, ***not****(X= d1; X= d3).*

% rules for s ++ e

rule(p1_refined(X,E), prefer(poss(X), notPoss (X)), []):-(s **,** *e), (X= d1).*

rule(p1_refined (X,E), prefer(notPoss(X), poss(X)), []):-(s **,** *e),* ***not****(X= d1).*

% prefer refined scenario s +/ e *over simpler version* s

rule(h1(X,_),prefer(p1_refined(X,_), p1(_,_)),[]).

Collecting Explanations: Priority Level rules are extended with an additional variable E, to account for the storage of explanations in a place where we can easily extract them later on. Since the type of explanations we desire are the justifications of our decisions as to why a disease is possible, we need to concentrate all the reasons that are used to justify the decision in a list variable. Reason for a disease to be possible are symptoms and/or patient information therefore we need to collect the conditions on the right side of the rule (i.e. from the rule body) that correspond to this kind of information, and store them in variable E as a list, e.g.

*rule(p1(X,****E****), prefer(poss(X), notPoss (X)), []):-****c1, c2****, (X=d1 ; X= d3),* ***E = [c1, c2].***

*rule(p1(X,****E****), prefer(notPoss(X), poss(X)), []):-****c1, c2****,* ***not****(X= d1; X= d3),* ***E=[c1, c2].***

The above code considers that the combination of conditions c1 and c2 lead to the diagnosis of diseases d1 and d3, therefore the value of the explanation variable E is equal to a Prolog list which contains those conditions ([c1, c2]). In a later stage, those conditions will be extracted from this list, translated to Natural Language and utilized to construct explanations.

For further examples of translating a Scenario Table to Gorgias code, please refer to APPENDIX A

4.4 Graphical User Interface & Underlying Processes

The Graphical User Interface (GUI) of the system is comprised of a website that was built in order to capture all the aspects and simulate the major steps of a diagnostic process, as described in the Theoretical Model of this work (Chapter 2). In this section, we go through the GUI, while describing some implementation issues that are related to each screen presented.

The first screen of the GUI is the Log-In Screen. A registered user gives his/her credentials (username and password), which are sent to a MySQL database, which stores the credentials of all uses, in order to become validated. If are correct, authorization to proceed is granted, else access is denied.

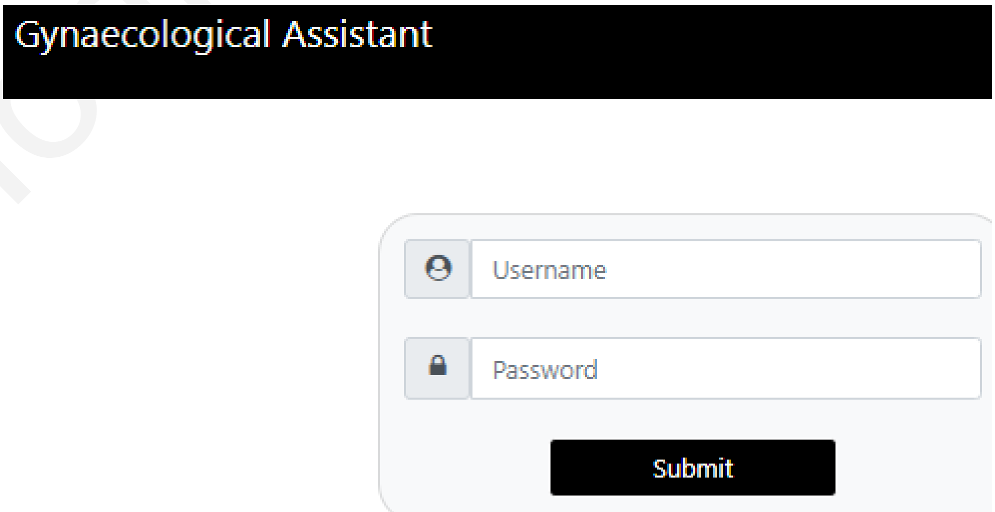


Figure 13. Login Page

The second screen presented is the “Patient Record” screen. In this page, the user can add information to the patient record of a hypothetical patient, in order to simulate information that influences the diagnostic progress in the background [63].

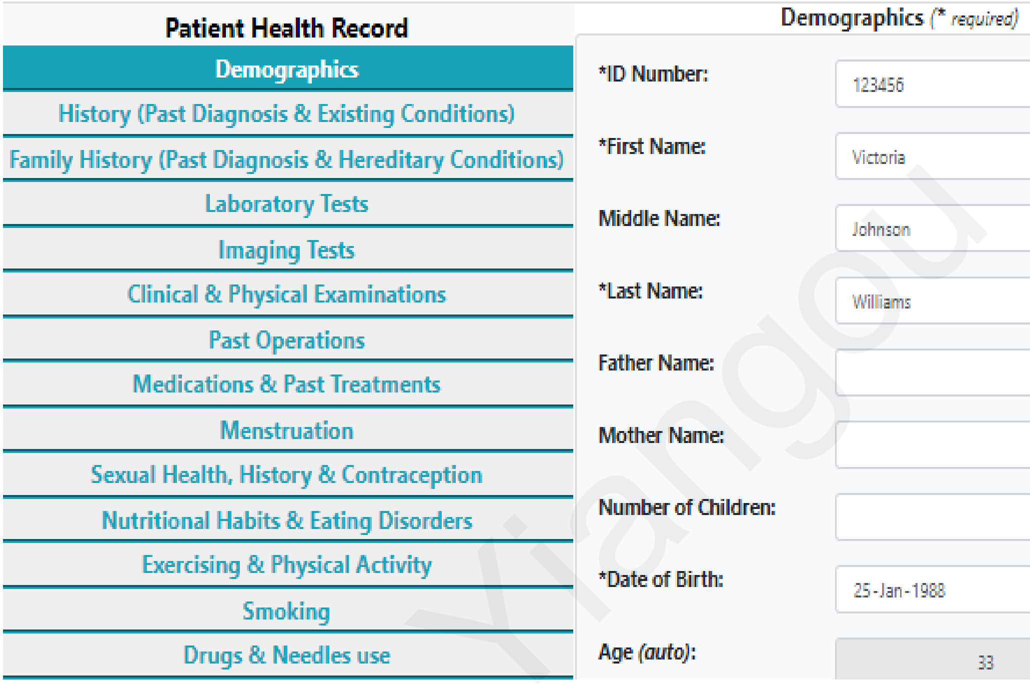


Figure 14. Patient Record Page - Demographics

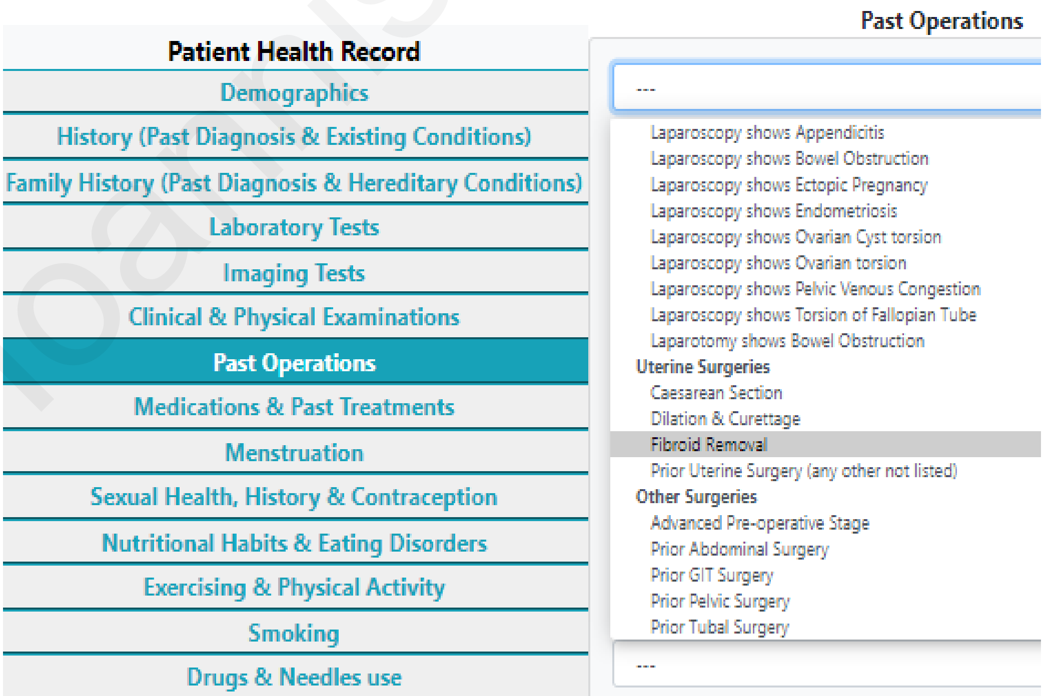


Figure 15. Patient Record Page – Past Operations

The user can fill a patient record here by checking checkboxes, selecting options from drop-down lists, filling text areas etc.

This step is important for the argumentation process, since it lets the user load patient information that may cause different outputs if considered along with any further symptoms and other information given later.

Once the user completes the required fields and any other field they desire, by clicking Submit, a Prolog file will be created, which contains the knowledge presented in the patient record, for the current user.

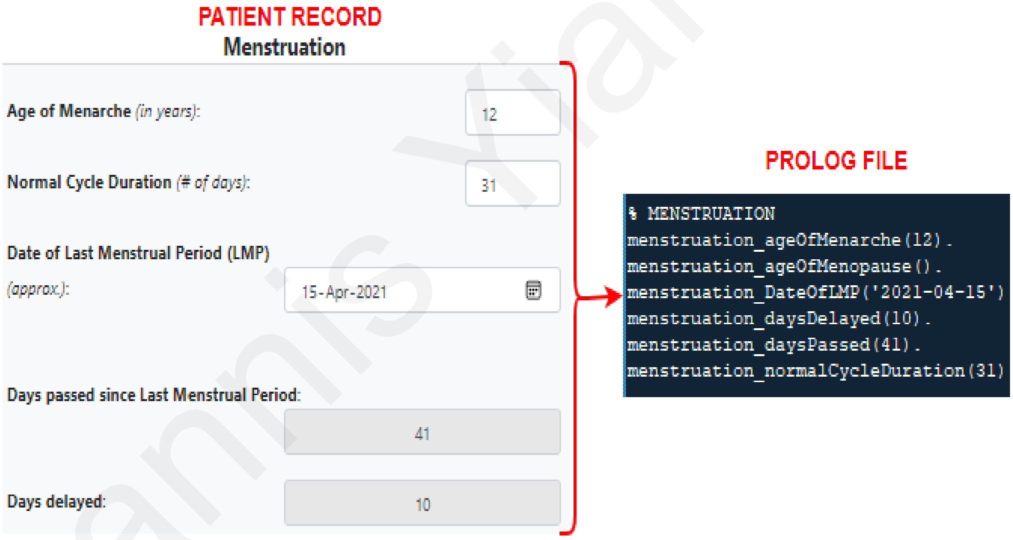


Figure 16. Patient Record to Prolog – Menstruation

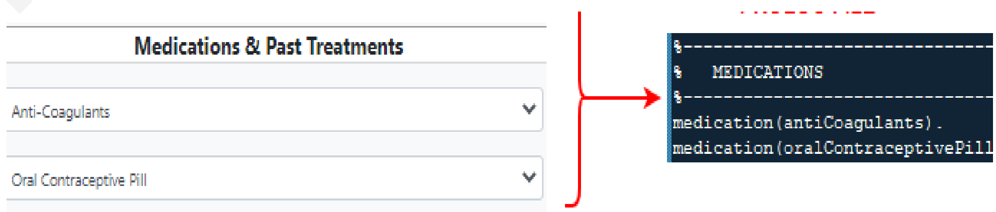


Figure 17. Patient Record to Prolog - Medications

Following the “Patient Record” screen is the “Presenting Complaints & Referrals” screen. This is where the user gives the patient’s reasons for visiting the gynecologist, in order for the system to recommend the most relevant groups of diseases to investigate [68], [69].

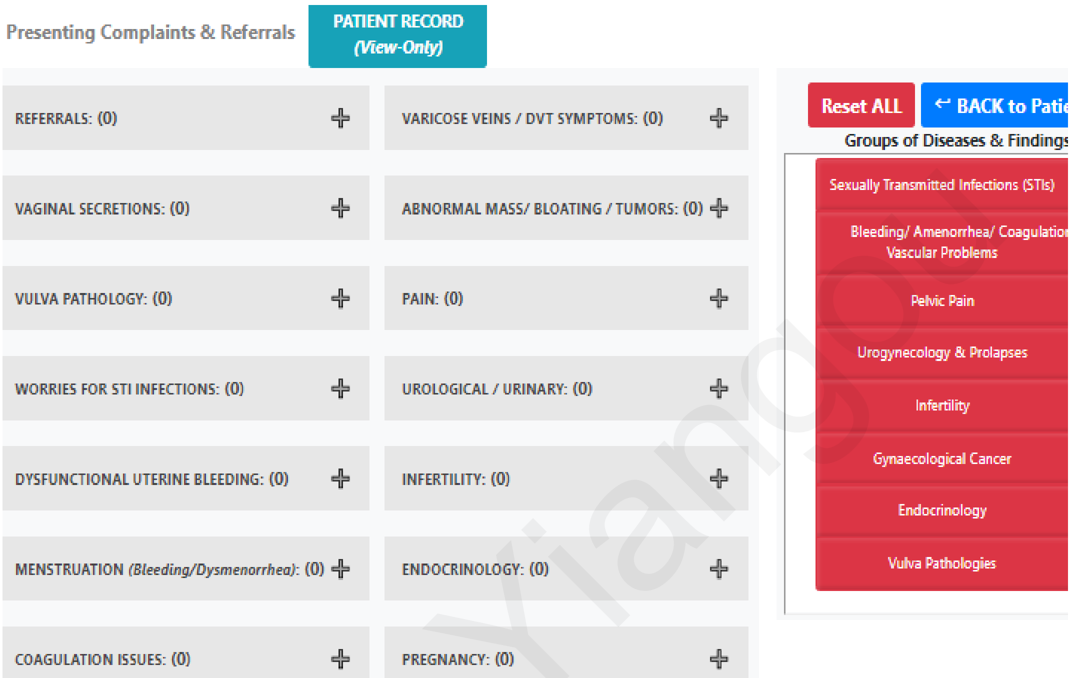


Figure 18. Presenting Complaints & Referrals screen

This screen is split in two areas:

1) the left-side is the “input panel”, where the user gives his/her inputs

- The inputs are divided into multiple categories for easier management, based on their nature e.g. one category for referrals only, urological / urinary symptoms, pain symptoms, complaints related to coagulation issues etc.

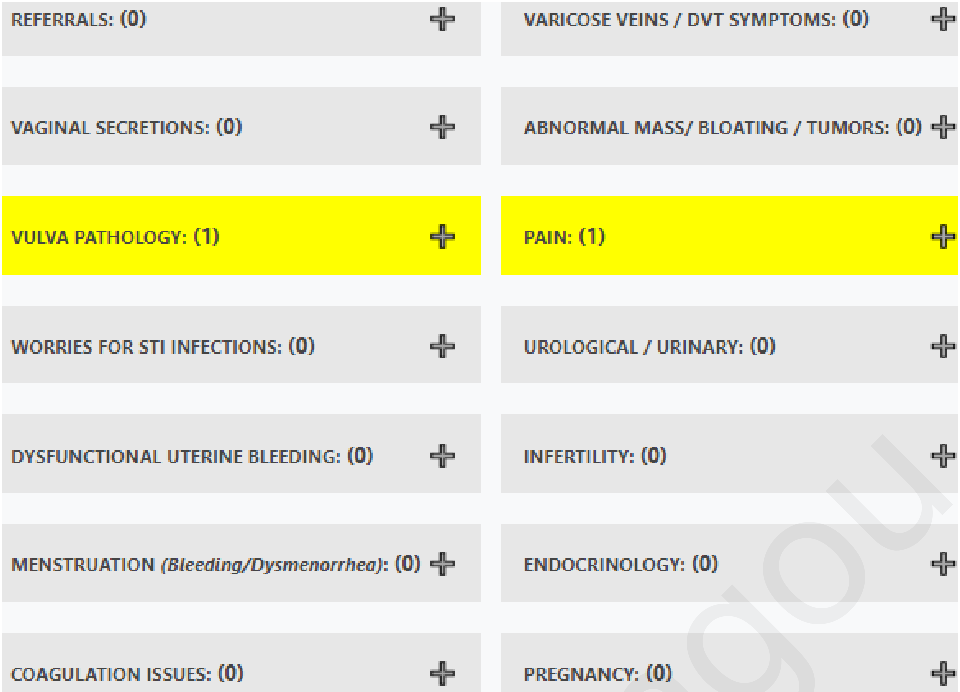


Figure 19. Presenting Complaints & Referrals – Input Categories

- Each category is inside a collapsible panel, for easier handling

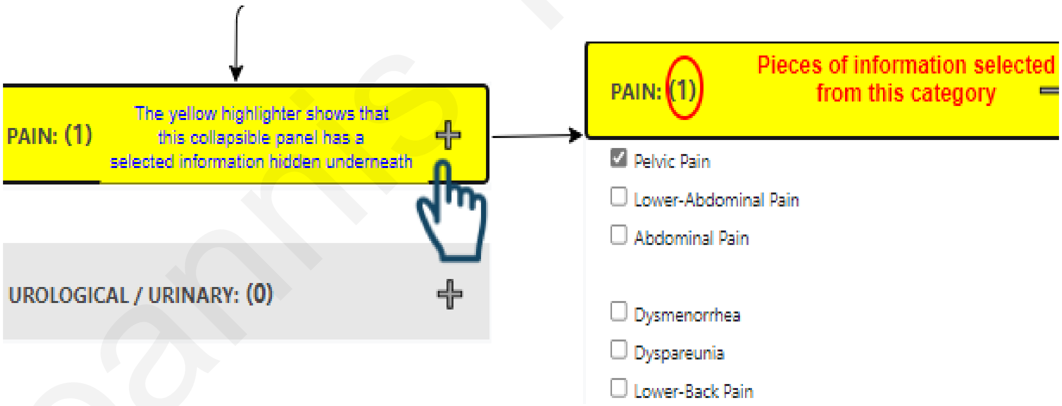


Figure 20. Presenting Complaints & Referrals - Collapsible Panels

2) the right-side is the "results panel", where the system shows the results of the initial diagnostic cycle (relevant groups, colored buttons and explanations)

a. There are 8 click-able buttons (which are initially red if no inputs have been given), which represent the 8 different groups of diseases that are available for investigation

b. Green Colored buttons denote that it is relevant group to investigate, based on given information

- The label "NEW!" denotes that it is a group that is newly-available, therefore not investigated yet

c. Red Color buttons denote that it is not a relevant group to investigate, based on given information

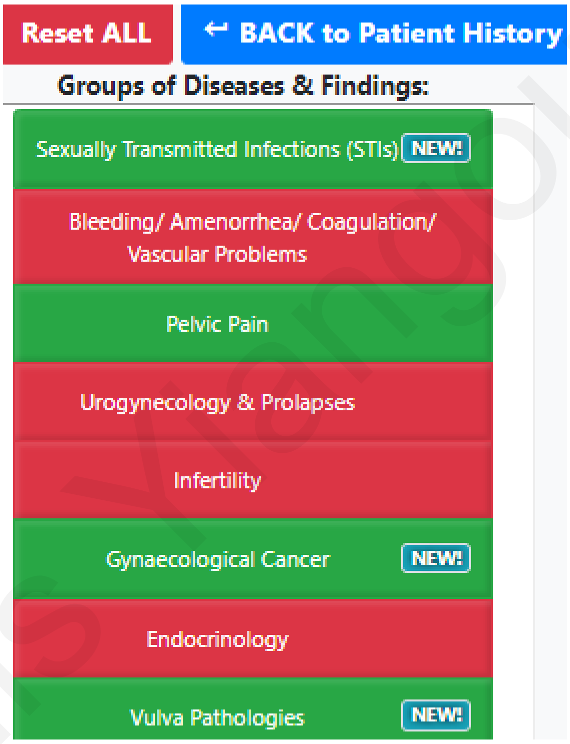


Figure 21. Presenting Complaints & Referrals – Results Panel

By clicking a green colored button, the appropriate explanation that justifies the respective decision is presented. By further clicking "Investigate", the user enters inside the specified group for investigation.

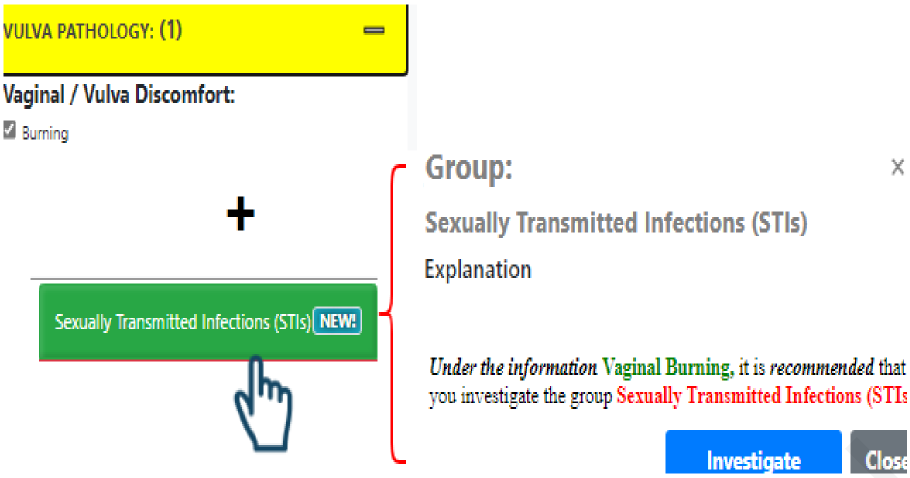


Figure 22. Presenting Complaints & Referrals – Clicking a Group Button

Once an input (presenting complaint or referral) is given, if the inputs lead to the suspicion of a specific group of diseases for investigation, the button of this group becomes green. The button remains green, as long as inputs that are related to it are still selected, otherwise it becomes red. If the group has not been investigated yet (not been clicked to enter), it has a label attached that denotes it is a NEW, uninvestigated yet, group. If visited at least once, then this label is removed. This lets the user know which groups have not yet been investigated, in order to not forget about them.

Some implementation issues that have to be mentioned in this step are the following (for the "Presenting Complaints & Referrals" screen):

Once an input is clicked, the program automatically submits the form (no need for clicking a "Submit" button), the following happens in this particular order:

1) *A Prolog file with the given inputs is created*

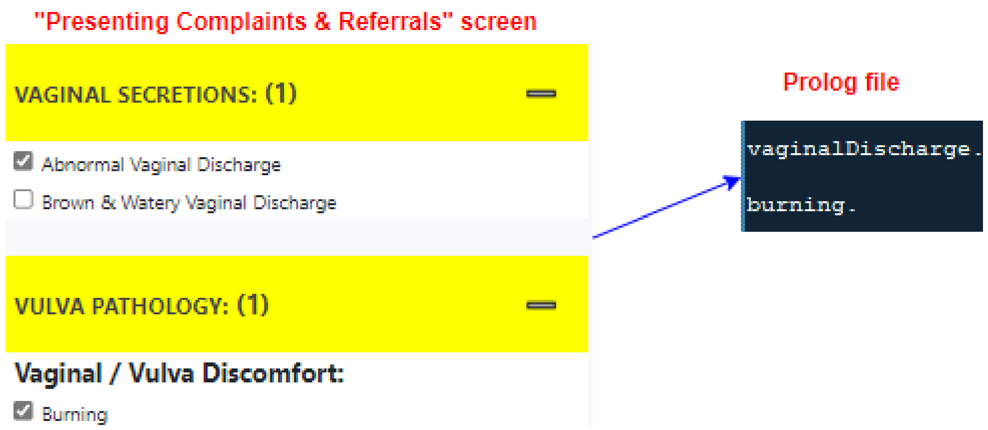


Figure 23: Presenting Complaints & Referrals – Prolog Input File

2) *A table that relates presenting complaints & referrals to their relevant groups of diseases (the table mentioned in Chapter 2) is used to find the groups that are relevant to the current inputs*

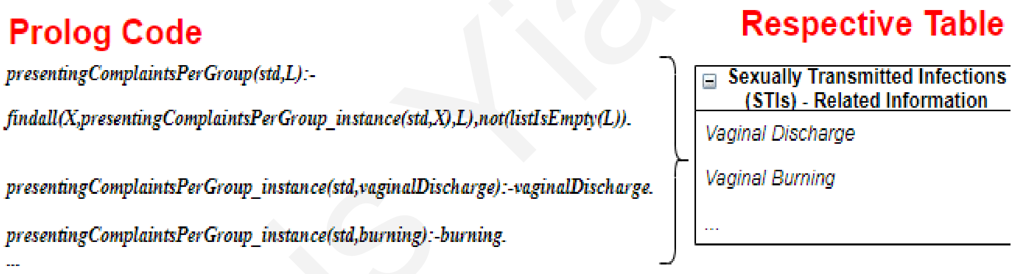


Figure 24. Presenting Complaints & Referrals – Info-Group Relation Table

- *The rest of the groups and information are also coded in the same way (using Prolog lists and inference rules) presented in Figure #screens9, in order to make such tables for all groups of diseases*
- *The above code creates a list of each group containing its relevant information that is given by the user. Any group that does not have an empty list (i.e. at least 1 piece of relevant presenting information was given), is considered relevant for investigation, else the group is not recommended. For each recommended relevant group for investigation, its button becomes*

green. If the group was not available before, the program also attaches it a "NEW" label.

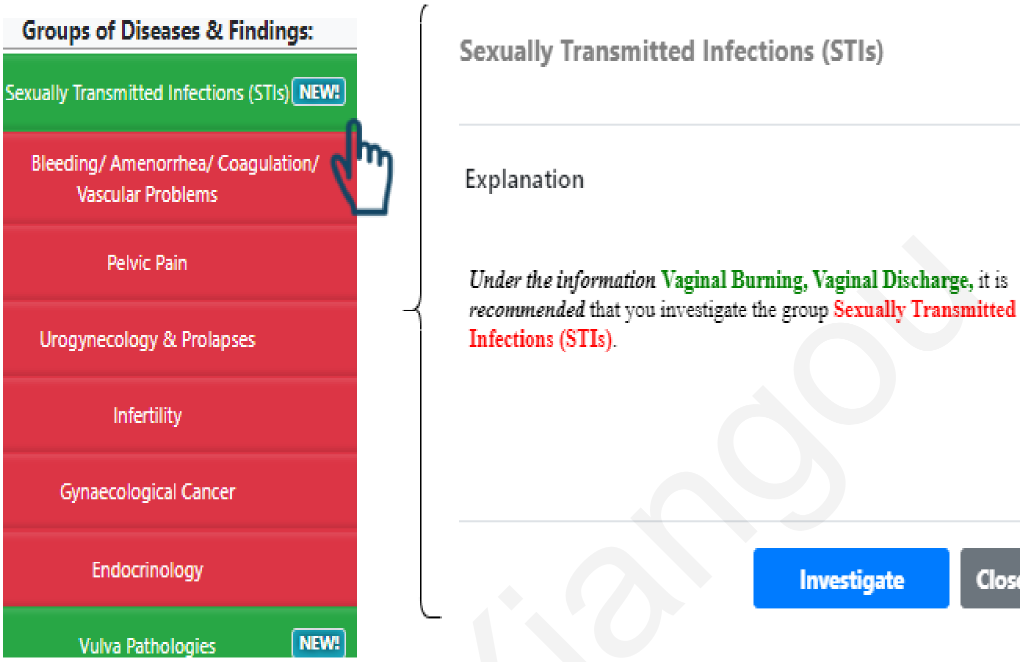


Figure 25. Presenting Complaints & Referrals – Relevant Groups

3) ***EXPLANATIONS:*** *By clicking on a group button, an explanation is offered to justify why this group has been selected (only for green buttons), by listing the related inputs that contributed to the suspicion of this group*

4) ***INVESTIGATE:*** *To enter the group for investigation, click on the button of the group and then click "Investigate".*
   - *Notice that even red buttons can be clicked for investigation, this happens because the system does not forbid the doctor from investigating any group he/she desires, the system only recommends what it is supposed to be relevant for investigation, given a set of presenting complaints and/or referrals*

Once a user selects to investigate a specific group of diseases the screen changes to that of the selected group. The structure of the page is similar to that of the "presenting complaints & referrals" screen:

Left-side of the screen: Symptoms and further patient information that do occur in diseases of the group. Most of them are checkboxes; some groups also contain drop-down lists for description of specific symptoms (e.g. "Sexually Transmitted Infections" and "Pelvic Pain")

Right-side of the screen: Diseases and Conditions (the same philosophy about green/red buttons apply as the groups of diseases in the previous, "Presenting Complaints & Referrals" screen, but this time with possible Diseases and Conditions).

- We can see that each group of diseases, while being faithful to the form structure we introduced earlier (left-side: inputs, right-side: results), the input panel and the result panel differs from group to group, since different numbers and kinds of symptoms and patient information are involved in each group, as well as different numbers, categories and kinds of diseases and conditions.
    - Below are some screenshots from some of the 8 groups. Observe the differences in appearance

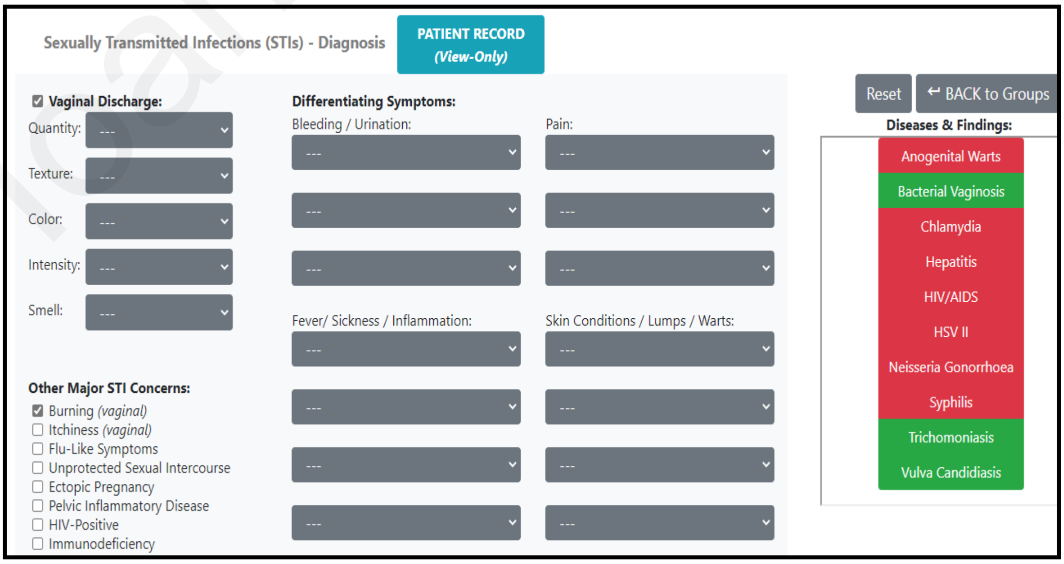


Figure 26. Investigating Groups – Sexually Transmitted Infections group

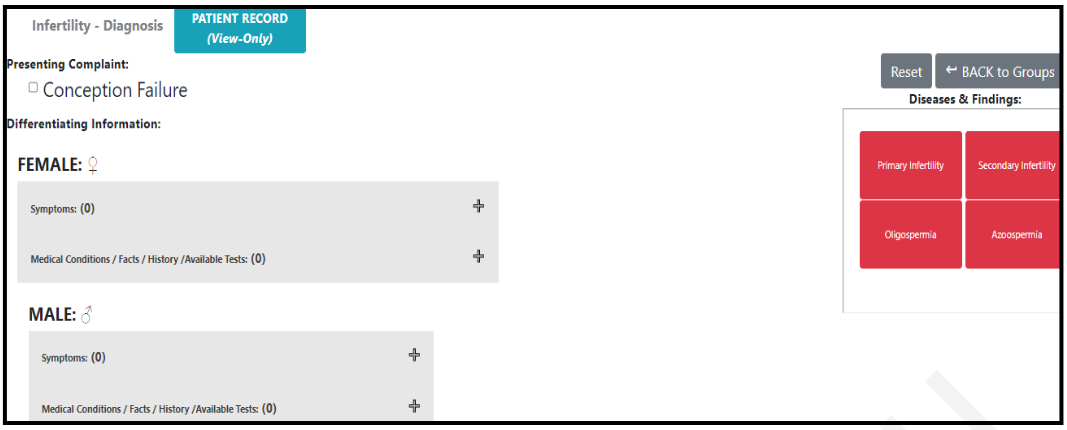


Figure 27. Investigating Groups – Infertility group

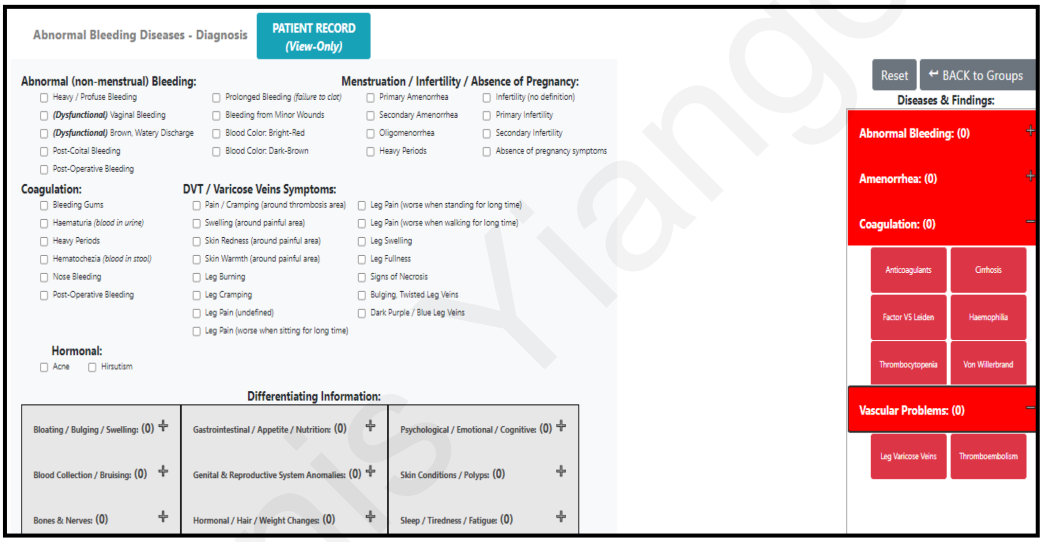


Figure 28. Investigating Groups – Bleeding/Amenorrhea /Coagulation /Vascular Problems

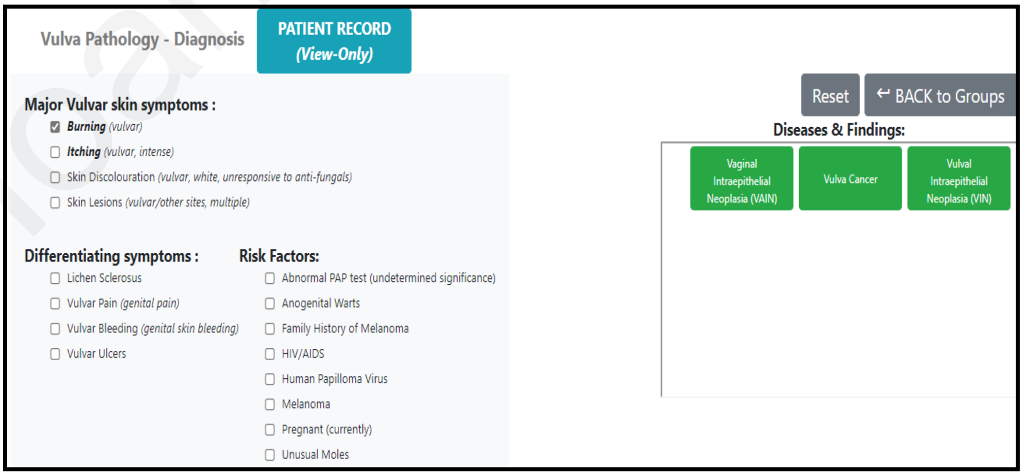


Figure 29. Investigating Groups – Vulva Pathologies group

Once an input is clicked, the program automatically submits the form (no need for clicking a "Submit" button), the following happens in this particular order:

1) *Create the knowledge in Prolog (write selected inputs to file)*
2) *Consult all relevant Gorgias theory files and Prolog input files, to consider all the information collected so far, this means all the following:*
   - *Patient Record Prolog input file*
   - *Presenting Complaints & Referrals Prolog input file*
   - *Any file that contains inputs given in other groups before this investigation (since the patient is the same and his/her known conditions so far are tested against any group and are accumulated)*
3) *Create arguments as explained in the section "Modeling the Scenario Tables in Gorgias" of this chapter*
4) *Retrieve solutions from Gorgias' Delta list of admissible arguments. Solutions are:*
   - *Decision options, i.e. the diseases that are possible*
   - *The rules that contributed to preferring a specific disease over others*
5) *Translate each rule name from the arguments collected, to natural language (further explained in section "Explanations" of this chapter) to create Explanations*
6) *Print all results to output files that will be consider in the next step in order to know which changes are made (button colors, explanations etc.)*
7) *Refresh the current page in order to present the results, such as:*
   - *Changes in button colors (green for preferred, red for not preferred diseases)*
   - *Explanations (upon clicking on a green button)*
   - *Appropriate messages (will be explained further in the section "***Prompting & Dialogs***" of this chapter*

4.5 Pelvic Pain Approach

As mentioned in Chapter 2, the group of diseases related to “Pelvic Pain” has been a great challenge, since it combines symptoms, patient information, diseases and findings from various fields of medicine, besides Gynecology, and since one of the tasks of a Gynecologist is to be able to identify such cases (even from another field) in order to offer the appropriate referrals and therapies, we had to find a way to implement them all [67].

Pain Description

First of all, one of the most important aspects of a Pelvic Pain is its description [66]. Solely based on the description of the pain, a doctor is able to differentiate between many diseases that could be possible. In order to implement the descriptions, we considered the element of “progress” of a disease, meaning that a disease may be known to be “severe” in intensity, but in primary stages of occurrence, it may not have progressed to this level yet, and the patient describes it as “mild” or “moderate”. Also, some diseases that are known to be of acute duration could also be chronic, depending on how a patient experiences them. Therefore, the descriptions may vary between diseases, and a disease may have more than one possible descriptions attributed to it.

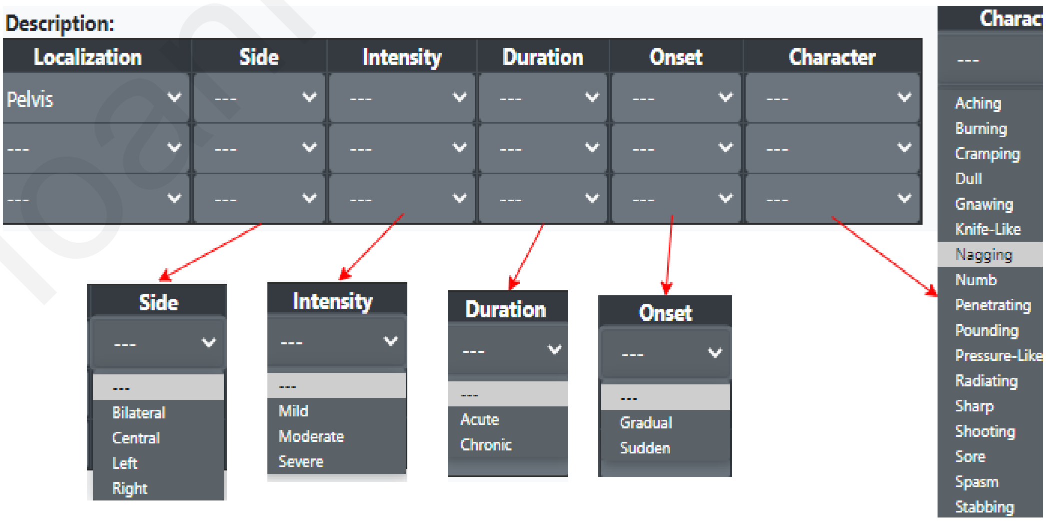


Figure 30. Pelvic Pain Description Fields

The program offers, within this group, a comprehensive way of describing a patient's Pelvic Pain, Lower-Abdominal Pain or Abdominal Pain, which are all critical for this group of diseases, with a level of expression that is acceptable by the doctors too, in the form of drop-down lists, as presented in Figure 30.

The program creates all possible pain description combinations that are found to occur in each disease, and they are used in the first level to differentiate between diseases, before moving to symptoms and information that are not related to pain. Below, a small piece of Prolog code is presented, which shows how a description is implemented by inferring logical predicates which specify the decisions that are attributed to the description:

% A "gynecological pain on the right side" is a "right-sided Lower-Abdominal Pain", and/or a "right-sided Pelvic Pain", and/or a "right-sided Abdominal Pain":

rightSideGynaePains_instance(contextARG_rightSideLowerAbdominalPain):-

rightSideLowerAbdominalPain.

rightSideGynaePains_instance(contextARG_rightSidePelvicPain):-rightSidePelvicPain.

rightSideGynaePains_instance(contextARG_rightSideAbdominalPain):-

rightSideAbdominalPain.

% Create predicate which lets us know which diseases could be "right-sided":

% rightSideGynaePains(D,L): disease D is related to right-sided pain, and patient presented the pain(s) (pelvic, lower-abdominal, and/or abdominal) in list L, which justify this

*rightSideGynaePains(**adenomyosis**,L):-*

findall(X,rightSideGynaePains_instance(X),L),not(listIsEmpty(L)).

*rightSideGynaePains(**adnexalTumor**,L):-*

findall(X,rightSideGynaePains_instance(X),L),not(listIsEmpty(L)).

*rightSideGynaePains(**appendicitis**,L):-*

findall(X,rightSideGynaePains_instance(X),L),not(listIsEmpty(L)).

...

In the above piece of code, we conclude that Adenomyosis, Adnexal Tumor, Appendicitis…, are diseases, which are related to gynecological (pelvic/lower-abdominal/abdominal) pains that occur on the right-side. The same way we implement left-sided/central/bilateral pains, severe/mild/moderate pains, acute/chronic pains etc. Their combinations are used in order to express complex descriptions of pains:

left_chronic_sudden_severe(D,L):-
leftSideGynaePains(D,L3), chronicGynaePains(D,L2), suddenGynaePains(D,L7), severeGynaePains(D,L6), L=[L3,L2,L7,L6].

The above piece of code works like that: "Disease D is related to left-sided, chronic, sudden and severe gynecological pain, if Disease D is related to left-sided pains, chronic pains, sudden pains and severe pains, separately".

Using the above coding examples, we exclude diseases which do not match the descriptions, in the first level, and also, when a more accurate description is reported, it overwrites a previous description, e.g.: *If we know that a patient has Severe and Acute Pelvic Pain, prefer diseases that present this specific pain (not "severe pelvic pain" or "acute pelvic pain" only).*

<u>High-Level Parameterization based on Pathognomonics</u>

In Chapter 2, we presented a high-level parameterization based on the concept of Pathognomonics [64] (i.e. symptoms and information that are closely related to a disease) and in this chapter, we present the way it is implemented in the program.

First, the parameterization differentiates by using the pain descriptions explained above.

Then, for each disease, we created a list (*L2*) with its related symptoms and patient information, which are not related to the pain described earlier. An example is in the following piece of Prolog code:

```
pathognomonics_instance(adenomyosis,prs2,supportingARG_clinicalExam(palpationSymmet
ricallyEnlargedTenderUterus)):-clinicalExam(palpationSymmetricallyEnlargedTenderUterus).
pathognomonics_instance(adenomyosis,prs2,supportingARG_dysmenorrhea):-dysmenorrhea.
pathognomonics_instance(adenomyosis,prs2,supportingARG_dyspareunia):-dyspareunia.
pathognomonics_instance(adenomyosis,prs2,supportingARG_heavyPeriods):-heavyPeriods.
pathognomonics_instance(adenomyosis,prs2,supportingARG_heavyProfuseBleeding):-
heavyProfuseBleeding.
pathognomonics_instance(adenomyosis,prs2,supportingARG_prolongedBleeding):-
prolongedBleeding.
```

The above code states that “Adenomyosis” is a disease that is related to the following information: *Symmetrically Enlarged Tender Uterus in examination, Dysmenorrhea, Dyspareunia, Heavy Periods, Heavy & Profuse Bleeding and Prolonged Bleeding.*

Then, for each disease, we created another list (*L3*) with its related examination results and historic information that are strong evidence that the disease is possible. An example is in the following piece of Prolog code:

```
pathognomonics_instance(adenomyosis,prs3,strengtheningARG_familyHistory(adenomyosis)
):-familyHistory(adenomyosis).
pathognomonics_instance(adenomyosis,prs3,strengtheningARG_history(highParity)):-
history(highParity).
pathognomonics_instance(adenomyosis,prs3,
strengtheningARG_imagingTest(mriAdenomyosis)):- imagingTest(mriAdenomyosis).
```

```
pathognomonics_instance(adenomyosis,prs3,strengtheningARG_imagingTest(uss4DColourD
opplerAdenomyosis)):- imagingTest(uss4DColourDopplerAdenomyosis).
pathognomonics_instance(adenomyosis,prs3,
strengtheningARG_pastOperation(caesareanSection)):-pastOperation(caesareanSection).
pathognomonics_instance(adenomyosis,prs3,
strengtheningARG_pastOperation(dilationCurettage)):-pastOperation(dilationCurettage).
pathognomonics_instance(adenomyosis,prs3,
strengtheningARG_pastOperation(fibroidRemoval)):-pastOperation(fibroidRemoval).
pathognomonics_instance(adenomyosis,prs3,
strengtheningARG_pastOperation(priorUterineSurgery)):-
pastOperation(priorUterineSurgery).
```

The above code states that “Adenomyosis” is a disease that is strongly supported by the following information to be possible: *Family History of Adenomyosis, History of High Parity, MRI or Ultra Sound Scan-4D Color Doppler showing Adenomyosis and History of Past Uterine operations (Caesarean Section, Dilation & Curettage, Fibroid Removal or any other uterine surgery).*

By utilizing the three layers mentioned above, a solution was designed to approach this special group of diseases. Although not ideal, the solution managed to sort the symptoms and patient information into different categories and utilize them in a logical sequence (e.g. first check the pain, then other symptoms, then additional information), as well as handled the large navigation space of the group.

Using preferences between Gorgias scenarios, the high-level parameterization we proposed as solution in Chapter 2 is implemented.

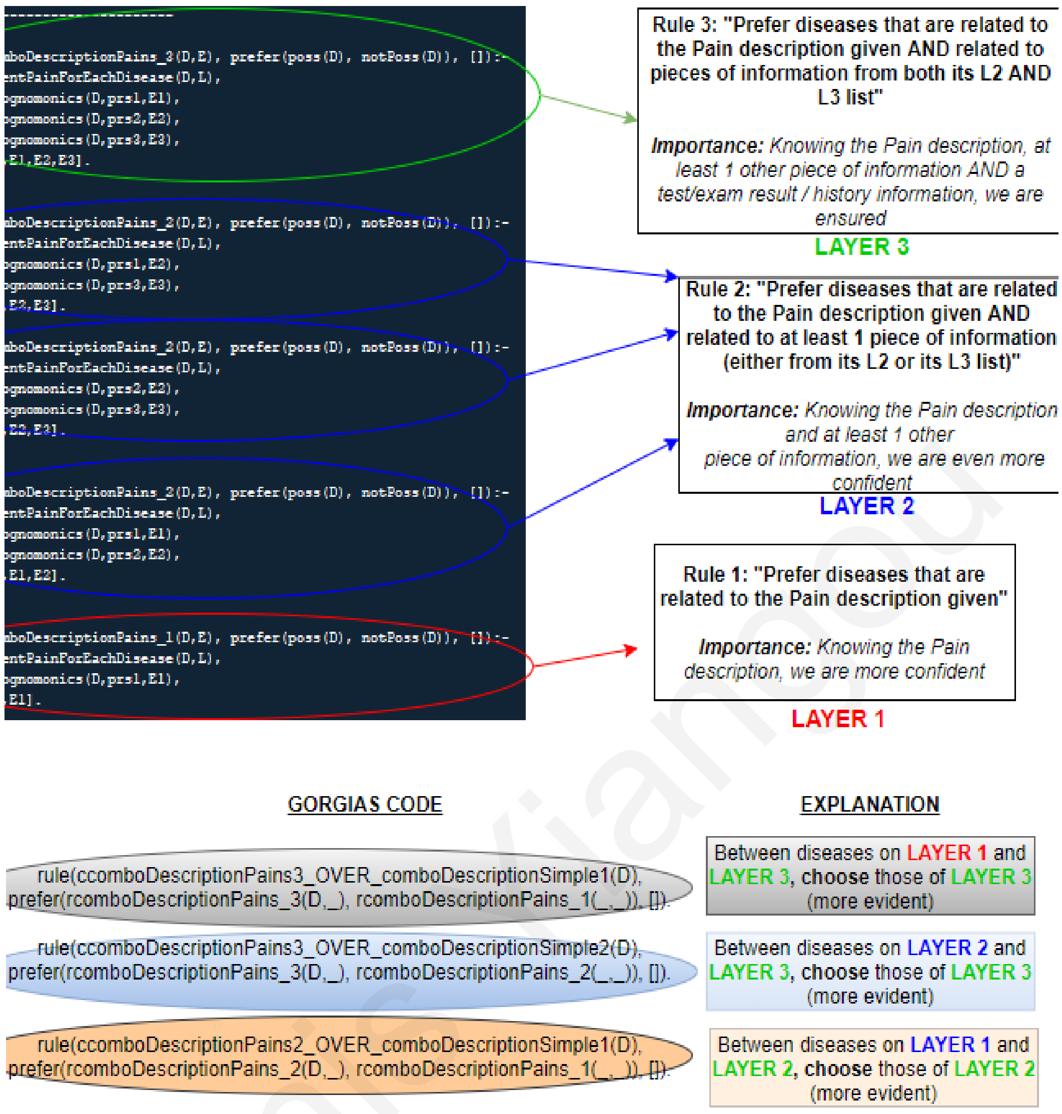


Figure 31. Pelvic Pain – Parameterization Layers

Prompted Guidance

Since we specified a standard way of collecting information in the group of diseases related to "Pelvic Pain", prompted assistance using pop-up windows and user-system dialogs was used in order to guide the user into a sequence of information seeking that would make good use of the parameterization described earlier.

The most important aspects implemented are the following:

1) ***First, we need to exclude the possibility of an Emergency / Life-Threatening condition*** [3]

- **SOLUTION:** Ask whether or not the pain is described as "Severe, Sudden and Acute". If yes, then focus only on treating the Emergency condition (***skip step 2***), otherwise learn more about the pain description and consider other diseases too
- **HOW:** If the "Pain Description" fields are empty, then the program asks the user about emergency conditions first.

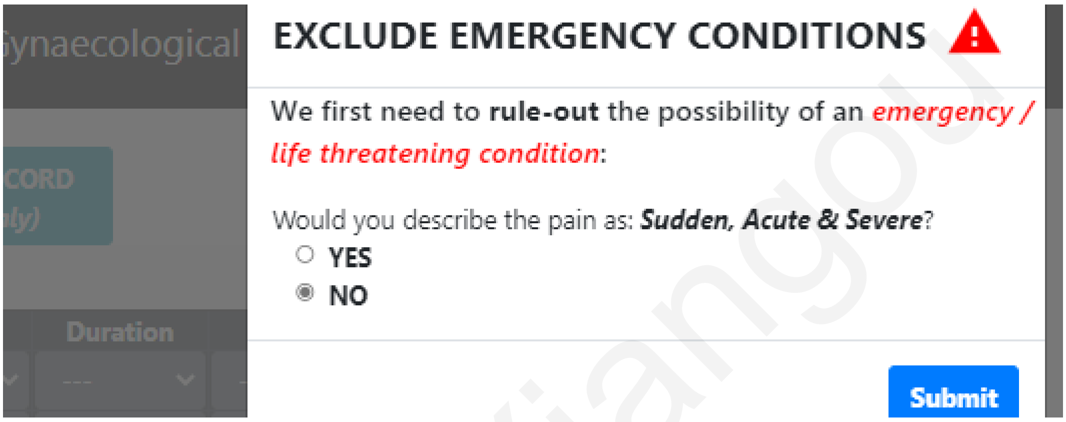


Figure 32. Prompting for Emergency Conditions 1

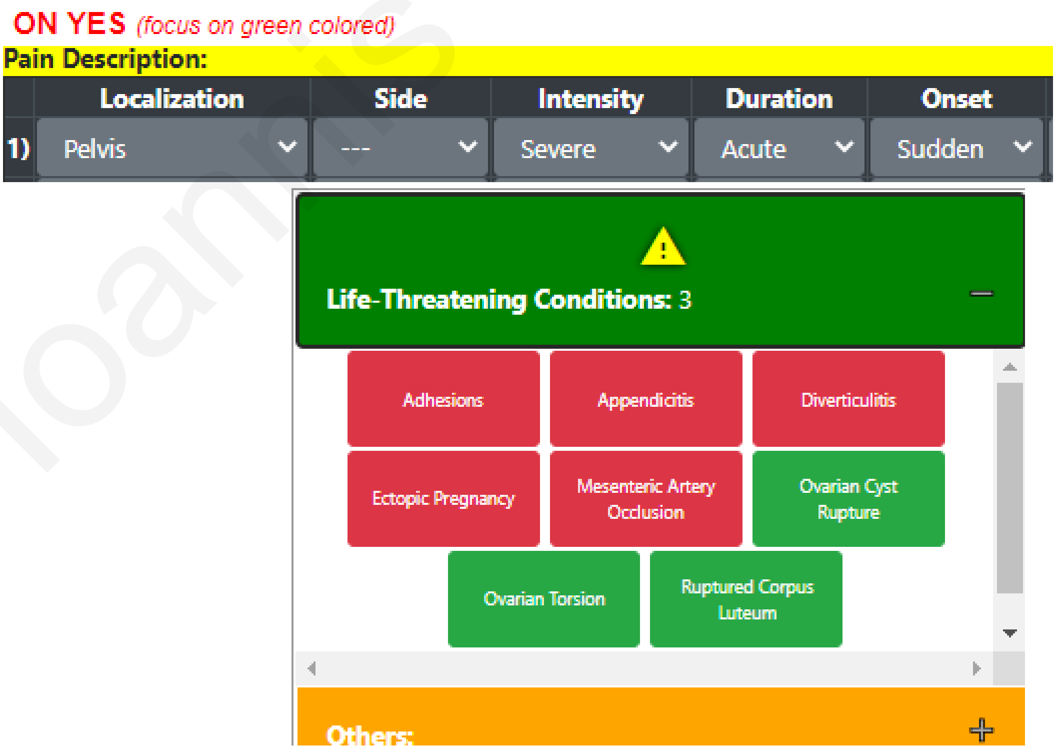


Figure 33. Prompting for Emergency Conditions 2

2) Then, if the patient does not report an Emergency condition, he/she is asked to describe his/her pain as accurately as possible

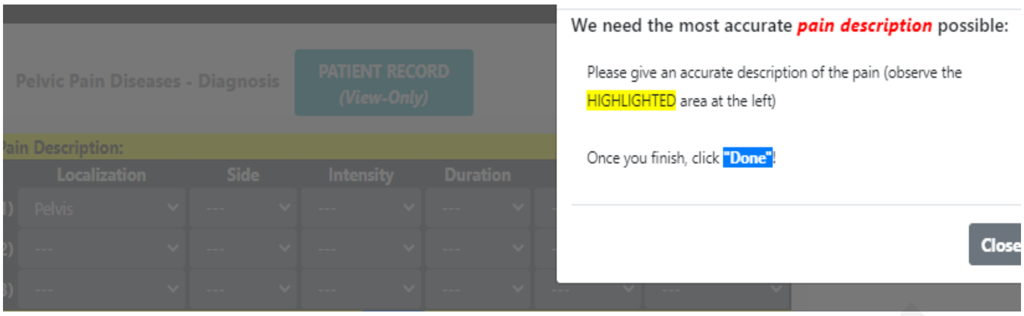


Figure 34. Prompting for Pain Description

3) Finally, when all information related to Pelvic / Abdominal / Lower-Abdominal pain is given, then the system asks for further information (symptoms, patient information etc.) that are not related to pain

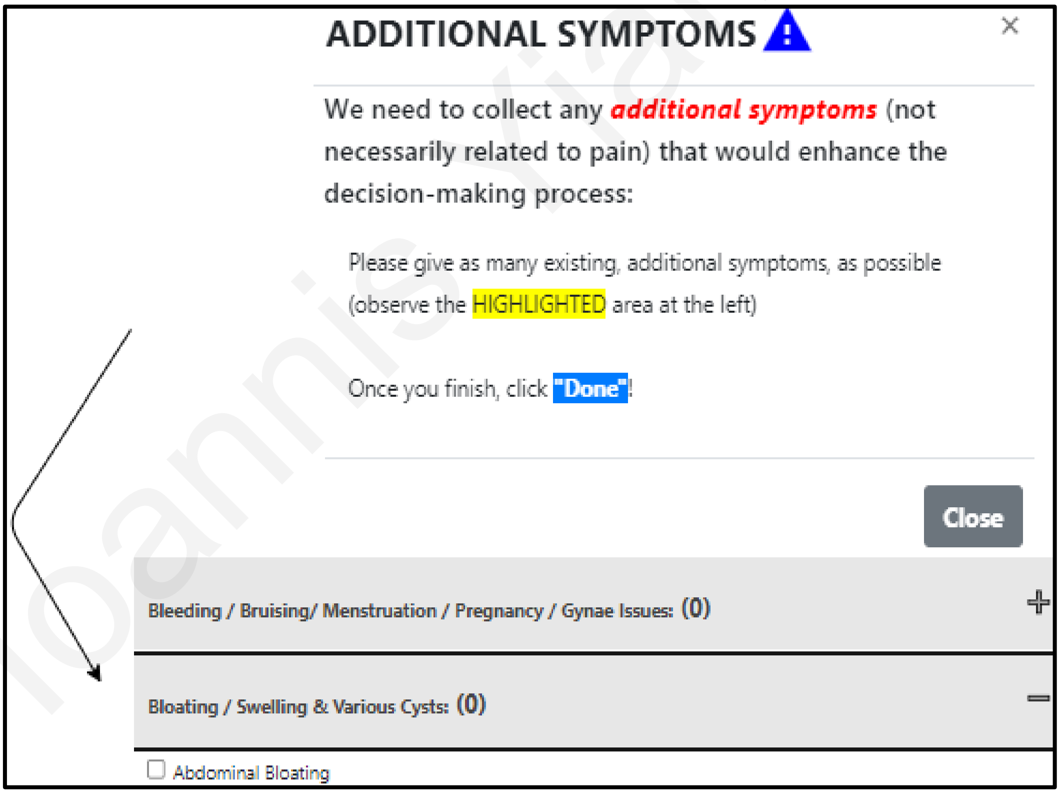


Figure 35. Prompting for Additional Information

The above prompting guidance can be ignored, or information can be given from different fields despite that the system highlights a different area on the screen. This is to ensure that the user has freedom of use and is not obliged, but is rather being assisted by the system, to follow a specific sequence of use.

4.6 Explanations

One of the biggest advantages of this work, and generally of Cognitive Systems that utilize Argumentation frameworks, is the ability to offer high quality, comprehensible, reasonable and convincing explanations in Natural Language in order to justify the conclusions of the system.

As mentioned earlier, in the section "Modeling the Scenario Tables in Gorgias" of this Chapter, we extend Priority Level rules with an additional variable E, to account for the storage of explanations in a place where we can easily extract them. The justification of a decision option is the set of inputs that corresponds to it therefore we need to store those inputs inside this variable E, for each decision option X, as follows:

*rule(p1(X,**E**), prefer(poss(X), notPoss (X)), []):**-c1, c2**, (X=d1 ; X= d3),* ***E = [c1, c2].***

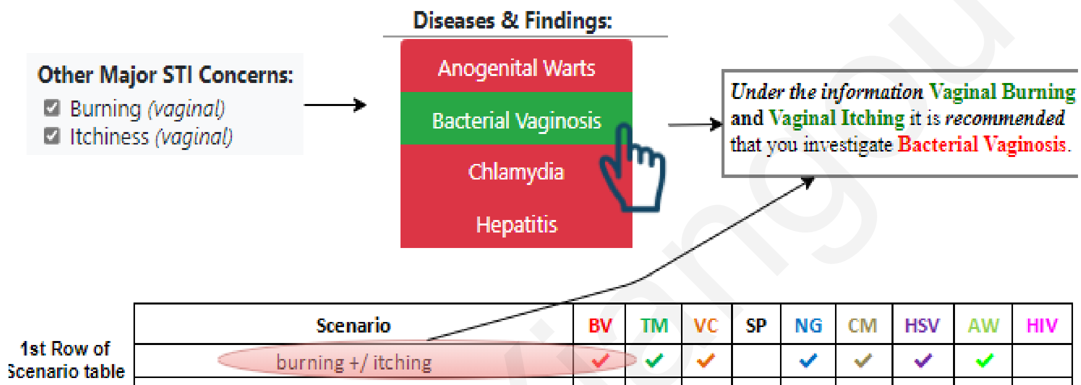
*rule(p1(X,**E**), prefer(notPoss(X), poss(X)), []):**-c1, c2**,* ***not****(X= d1; X= d3),* ***E=[c1, c2].***

Considering the above piece of code, if disease X is possible, because of rule *p1(X,E)*, then it is justified by using *c1* and *c2 (e.g. "Because c1 and c2, X is possible"),* because *c1* and *c2* are the components of the explanation list E.

To add more depth and color to the explanations, as well as to let users know what was the contribution of each information to the decision, three different categories of explanation components have been identified, which are the following:

1) ***Context information:*** components which came from a very simple version of the scenario table (from the very first rows). They are less indicative of a disease however they present the "context" we are investigating in, meaning the initial symptoms /

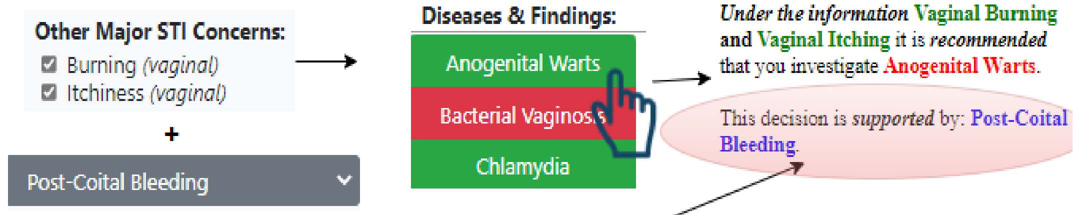
presenting complaints / referrals that first made the consideration of the disease possible, along with others. It shows the piece of information that helped us have a first focus on a subset of relevant diseases. They represent the first piece of information that is found in Scenario Tables (in the first rows) therefore we could also define the "context" as the "hierarchy" of table we are working in, and this piece of information as "hierarchy-starting", e.g.



Scenario	BV	TM	VC	SP	NG	CM	HSV	AW	HIV
burning +/ itching	✓	✓	✓		✓	✓	✓	✓	
(burning +/ itching) ++ (intermenstrual_bleeding +/ postcoital_bleeding)			✓		✓	✓		✓	
(burning +/ itching) ++ (intermenstrual_bleeding +/ postcoital_bleeding) + dysuria			✓		✓	✓		✓	

Figure 36. Relation of Scenario Tables with Explanations – Context

2) ***Supporting information:*** components which helped us exclude a good number of diseases, leaving us with a limited set of the most relevant decision options. They usually represent information that is used to make refinements of previous scenarios therefore they are usually presented in the next 1-2 rows of the Scenario Table, e.g.



Scenario	BV	TM	VC	SP	NG	CM	HSV	AW	HIV
burning +/ itching	✓	✓	✓		✓	✓	✓	✓	
(burning +/ itching) ++ (intermenstrual_bleeding +/ postcoital_bleeding)			✓		✓	✓		✓	
(burning +/ itching) ++ (intermenstrual_bleeding +/ postcoital_bleeding) + dysuria			✓		✓	✓		✓	

2nd Row of Scenario table

Figure 37. Relation of Scenario Tables with Explanations – Support

3) ***Strengthening information:*** information that are used by the end of the diagnosis to strengthen the decision that the system is about to make. Changes might not occur in the number of diseases being available, but the explanation gives this additional strengthening information to inform the user that we have no doubts about the decisions made, under a specific context. Information here might be test results, examination results, patient record information or symptoms that are very characteristic of some diseases. They are usually presented by the end of the Scenario Tables, e.g.

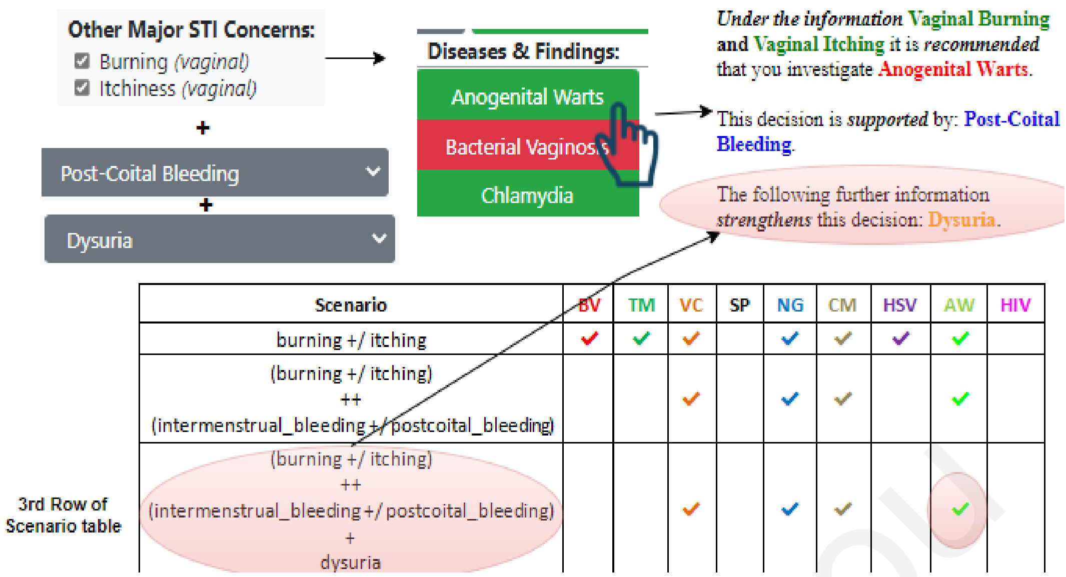


Figure 38. Relation of Scenario Tables with Explanations – Strengthening

In order for the system to understand these categories of explanation components, we append an identification string in the Gorgias code, in front of the information when added to the list of explanations E (e.g. *E=[L1| [L2,strengtheningARG_dysuria]]*), which will be processed later by a Java explanation module to extract the information and categorize it appropriately, e.g.

- contextARG_burning → *Vaginal burning is a context information*
- supportingARG_postcoitalBleeding → *Post-coital bleeding is supporting information*
- strengtheningARG_dysuria → *Dysuria is strengthening information*

Once Gorgias solves a query and returns the Delta list (i.e. the list of rules collected that contributed to the decision of an option *or* the admissible set [37]), a string-manipulation algorithm written in Java, is applied to extract and process the Explanations, as follows:

1) *Iterate Delta and extract the rule name, e.g.*
    - *Delta = [p1(X,[c1, c2, c3]), ...] → p1(X,[c1, c2, c3])*
2) *Extract the decision option X and the list of conditions and store them in variables*
    - *p1(X,[c1, c2, c3])*
        - *solvedFor = X;*

○ *solvedForExplanation = [c1, c2, c3]*

3) *Run through the extracted list of conditions:*

- *Extract the component category string and categorize in different lists the conditions, based on their category*
    - ○ *E.g. if c1 = "supportingARG_postcoitalBleeding", remove the identifying "supportingARG_" and save the string "postcoitalBleeding" to a list of supporting information*
- *Using a large vocabulary map designed specifically for this cause, translate the predicate name to natural language phrase*
    - ○ *E.g. take "postCoitalBleeding" and search the vocabulary map to translate it to the more natural-looking format "Post-Coital Bleeding"*

4) *Tailor the decision option and the translated explanations in a way that form comprehensible and reasonable sentences*

The complete process can be described using the following pipeline:

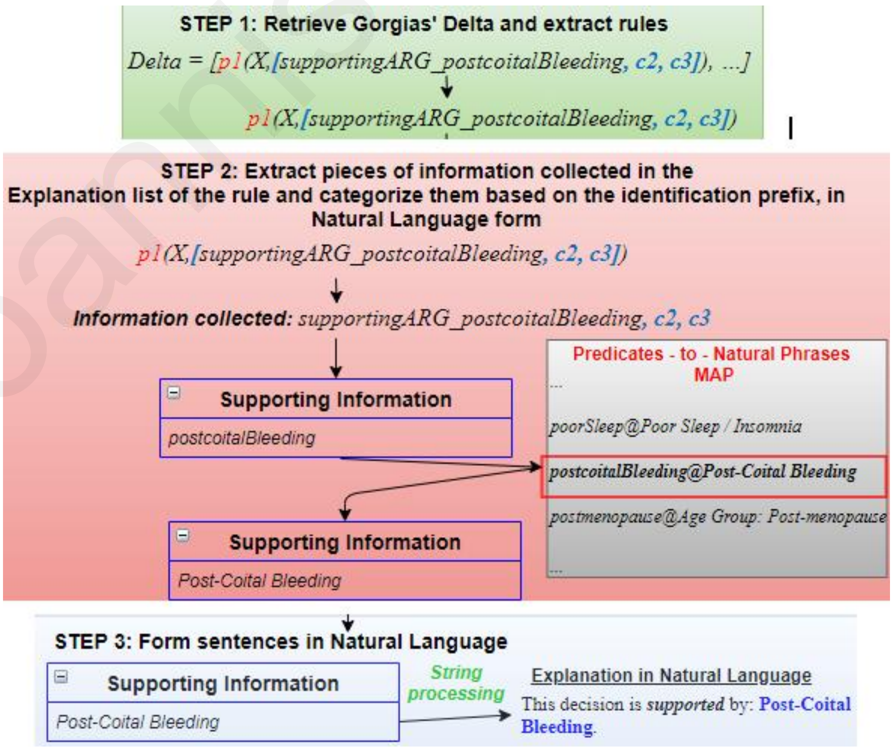


Figure 39. Explanation Construction Pipeline

4.7 Prompting & Dialogs

The system relies on human-computer interaction in order to simulate a real-life diagnostic scenario as accurately as possible. This is achieved by offering various prompting approaches, including diagnostic guidance and special event alert (either via highlighting or dialog boxes), which aim towards engaging the user into the process and encouraging him/her to learn better.

More specifically, the system offers the following prompting utilities:

Case 1: Guidance through diagnosis in a dialog manner (for the group of "Sexually Transmitted Infections" and "Pelvic Pain" diseases), where the system collects targeted information, that are adapted to correspond to the previous answers. This is done by checking the "Scenario Tables" and trying to collect information that would lead to further refinement of the current scenario, i.e. to move further lower in the rows/hierarchies of the table, as this leads as to a more accurate diagnosis, e.g.

Scenario	BV	TM	VC	SP	NG	CM	HSV	AW
burning +/ itching	✓	✓	✓		✓	✓	✓	✓
(burning +/ itching) ++ (intermenstrual_bleeding +/ postcoital_bleeding)			✓		✓	✓		✓
(burning +/ itching) ++ (intermenstrual_bleeding +/ postcoital_bleeding) + dysuria			✓		✓	✓		✓

Table 29. Scenario for prompting

If the current scenario is in the 1st row of the Scenario Table above (e.g. we have "Vaginal Burning"), then the system will first try to learn if we also have "Itching". Then, it will attempt to go lower in the hierarchy, by asking about "Inter-Menstrual Bleeding" and "Post-Coital Bleeding". If this is the case, then it will ask for "Dysuria" next.

The system asks information that are relevant at each hierarchy, from every relevant, parallel Scenario Table e.g. a table similar to the above ends with "Conjunctivitis" instead of "Dysuria" therefore at this level, the system will ask for both, since both Scenario Tables have scenarios that are relevant to the case so far, given the set of inputs.

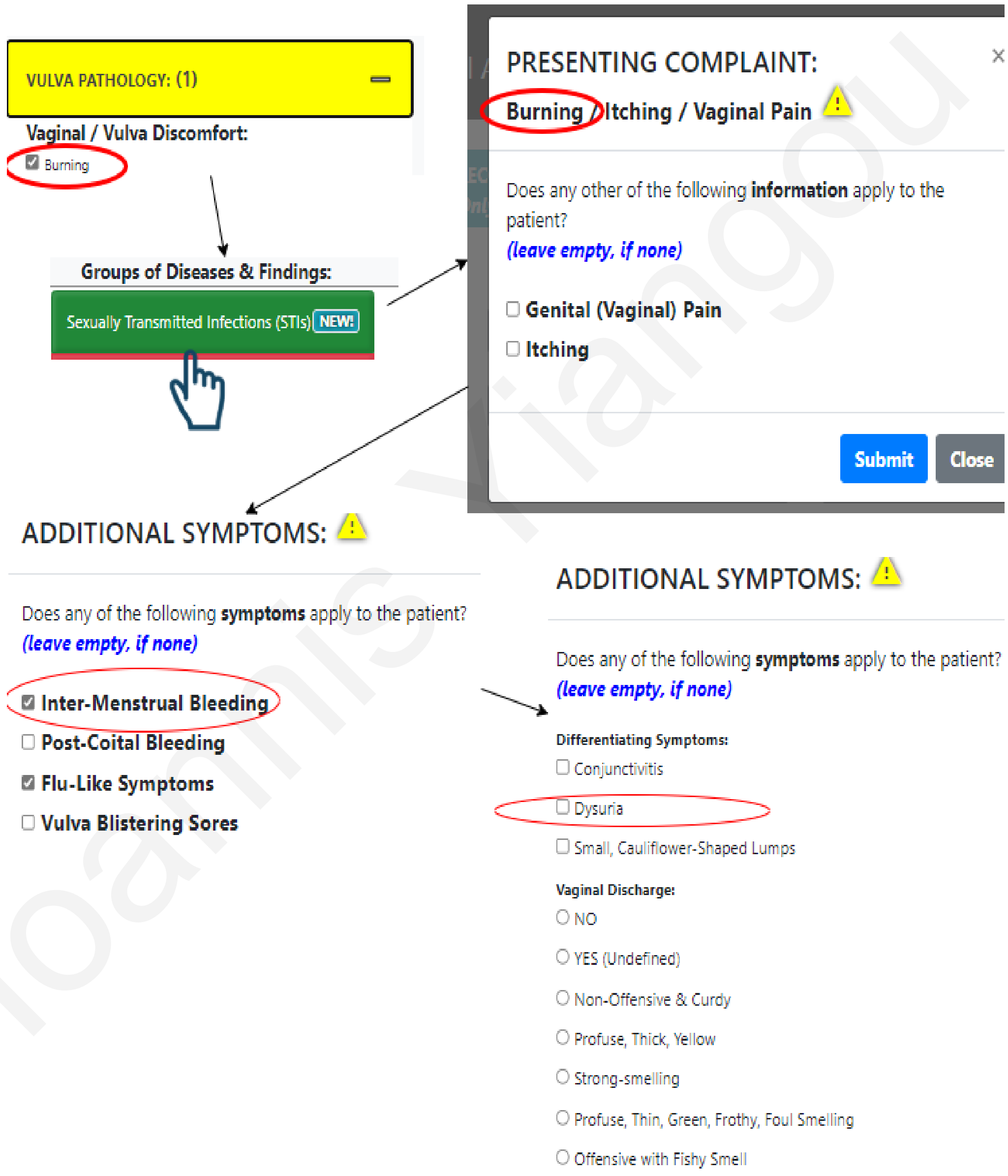


Figure 40. Dialog Prompting

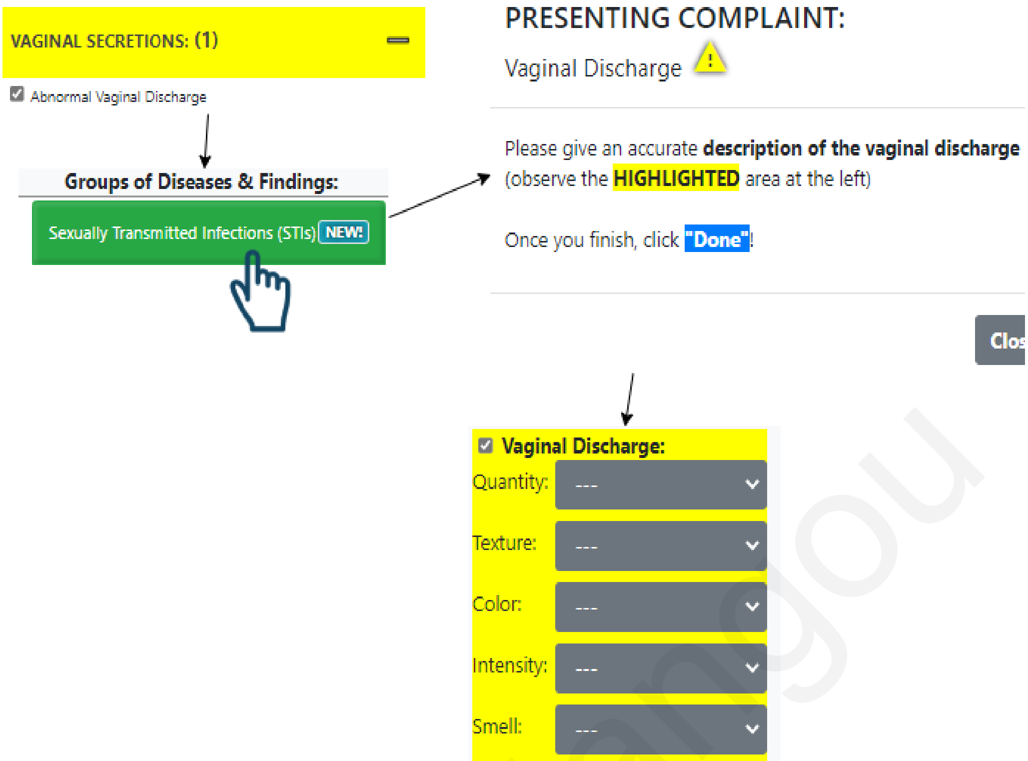


Figure 41. Highlighting Prompting

The above method simulates a (almost) "natural" human dialogue, where subsequent questions are influenced by the flow of the discussion preceded, between the two parties (the doctor and the patient).

<u>Case 2: Newly-available groups</u>, where the system checks whenever a new group of diseases becomes available, by identifying groups that were not recommended in the previous collection of information, but they are now, given a new piece of information, and alerts the user about the new investigation possibilities.

The following figure presents an example where a piece of information ("Inter-Menstrual Bleeding") became known while investigating the "Sexually Transmitted Infections" group, something that triggered the alert of newly-available groups, since this particular symptom is also related to Cancer and Bleeding diseases and conditions.

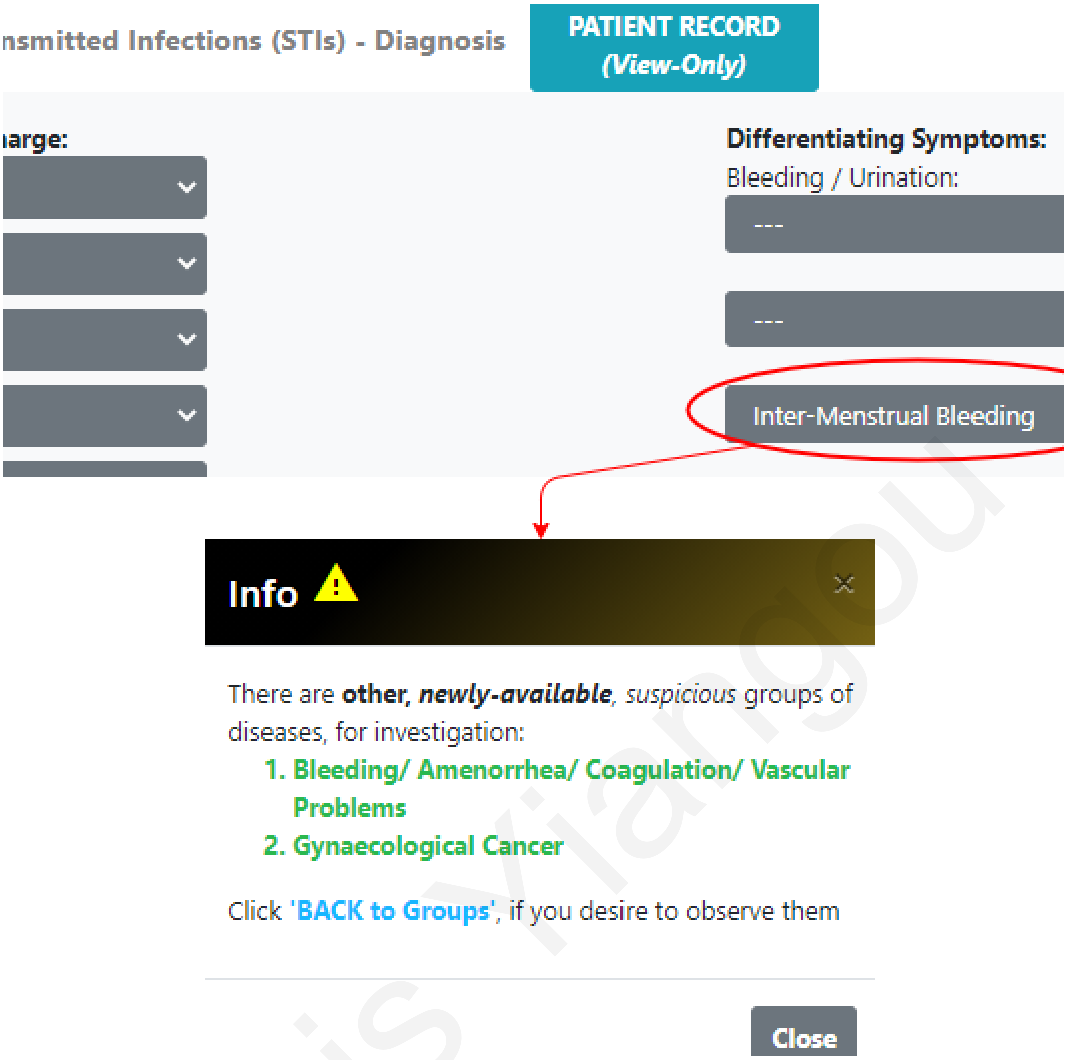


Figure 42. Alert for newly-available groups, inside another group

<u>Case 3: Emergency conditions ruling-out first</u> (for the "Pelvic Pain" group of diseases), where the system prompts the user to eliminate the possibility of an Emergency or Life-Threatening condition. In the case of "Pelvic Pain", we are referring to "Sudden, Severe and Acute Pelvic or Lower-Abdominal Pain" as "Emergency Conditions", and is the first thing that concerns the system, once the group is visited for investigation [3].

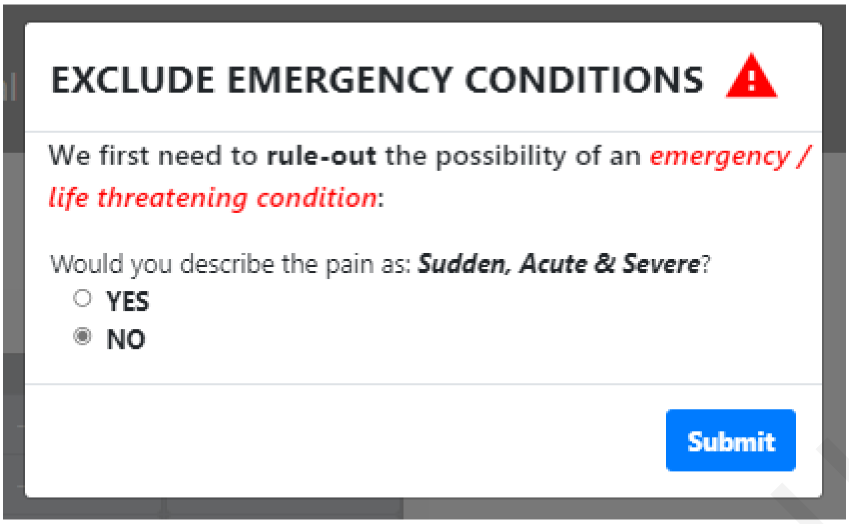


Figure 43. Trying to rule-out Emergency Pelvic Pain

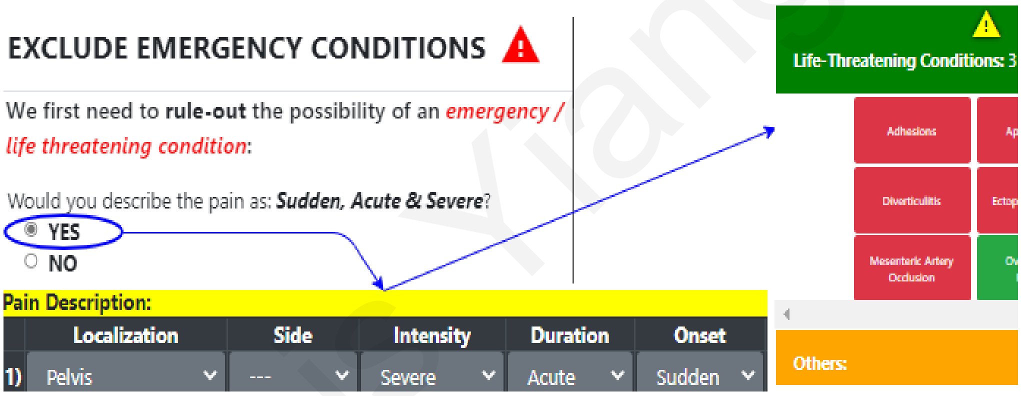


Figure 44. Dialog uncovers a Pelvic Pain Emergency condition

The above utilities help senior and experienced doctors remember information they may forgot about or they pass as trivial, as a result of cognitive fatigue or over-encountering common diseases and symptoms/information that made them forgot or ignore rarer cases. Another benefit is the enhancement of the learning process of junior and trainee doctors, as those prompting techniques guide them through a typical, natural diagnostic process and help them identify what is important, such as eliminating emergency conditions first, learning what to ask based on previous patient's answers and considering other, possible groups that are related to the information they collected latest.

4.8 Persistence and Cross-Form Consideration of Inputs

The Input forms of the system are programmed to be “persistent”. That means refreshing the current page, visiting another group for investigation or going back to previous pages will not cause any checkboxes or drop-down lists to forget the user’s selections, unless explicitly removing the “check” from the checkbox, or selecting an empty option in a drop-down list or clicking the “Reset” button. “Persistence” in a web page can be implemented by using “localStorage” [78], a special property of HTML Document Object Model, which allows the storage of data in a web browser. This is done to allow refreshing a page without losing previous inputs, as the system is programmed to identify changes in the inputs, then immediately execute a diagnostic process and automatically refresh the page to show the new results. This allows the system to act in almost real time, without the need to collect all information first and then click a “Submit” button. The doctor noted that this is especially useful for the user, since he/she can immediately observe and compare the changes in the colours of the buttons on the right side of the screen (Groups of Diseases or Diseases, accordingly), immediately after a change at the left side of the screen (i.e. when changing an input value).

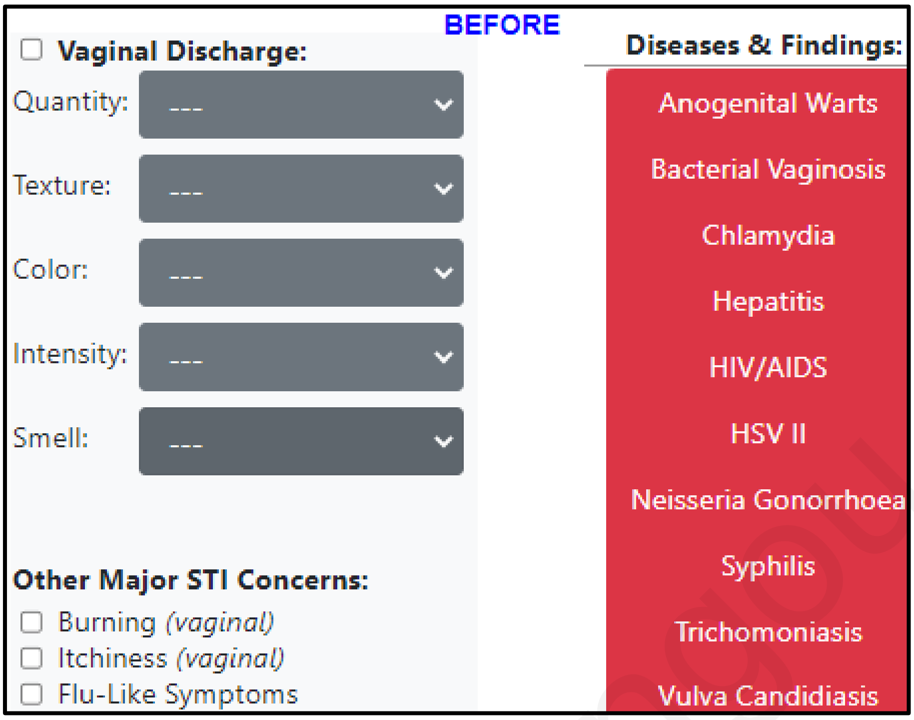


Figure 45. No inputs given leads to no possible diseases

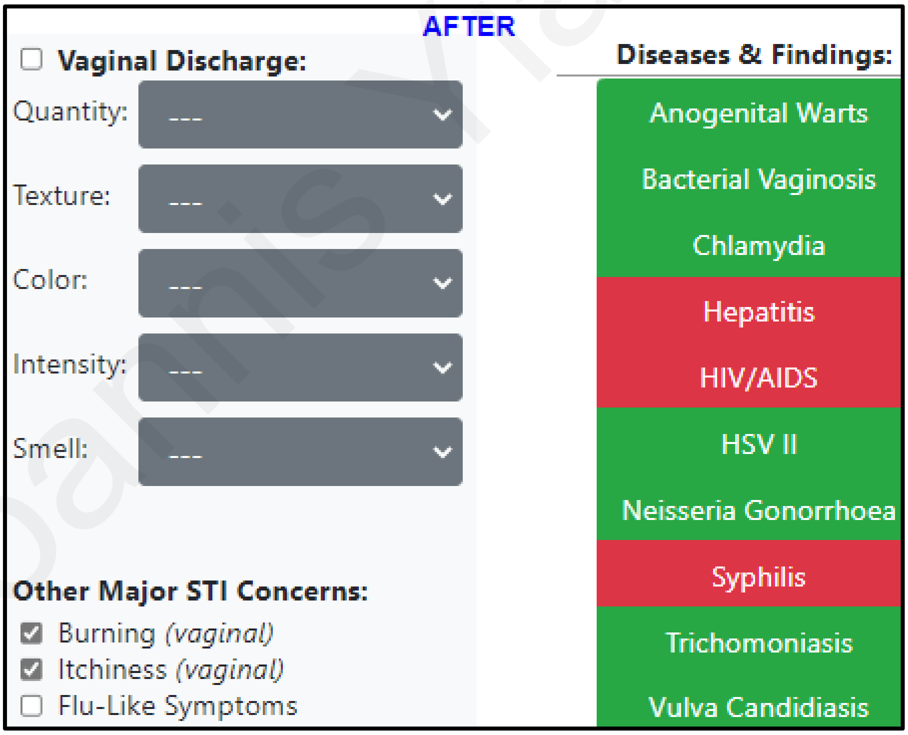


Figure 46. Result of a diagnostic process

The state presented in Figure 45 shows a case where no inputs have been given yet, therefore no diseases are possible so far. When the user inputs the information “Burning” and

"Itchiness", without the need to press submit, the system immediately runs the diagnostic process and refreshes the page to show the results, as presented in Figure 46 (notice the change in the colours of some buttons on the right of the screen, these are the possible diseases under given information). Notice that the checkboxes "Burning" and "Itching" remain checked, even after refreshing the page.

Another very important reason that the system uses persistent input forms is the achievement of cross-form consideration of inputs. Its significance lies in its potential to make previous inputs to be considered in any subsequent stage. If, for example, a user gives an input (clicking a checkbox, selecting an option from a drop-down list etc.) in the patient record screen, or the presenting complaints screen, or even inside a group of diseases, these inputs will still be applied to other screens too, since the investigation concerns a single patient, and his/her condition applies to the whole spectrum of his/her diagnostic process. This also helps the doctor remember what previous inputs he has selected, eliminating the chance of forgetting, ignoring or leaving important signs, symptoms or patient information out of the decision-making process.

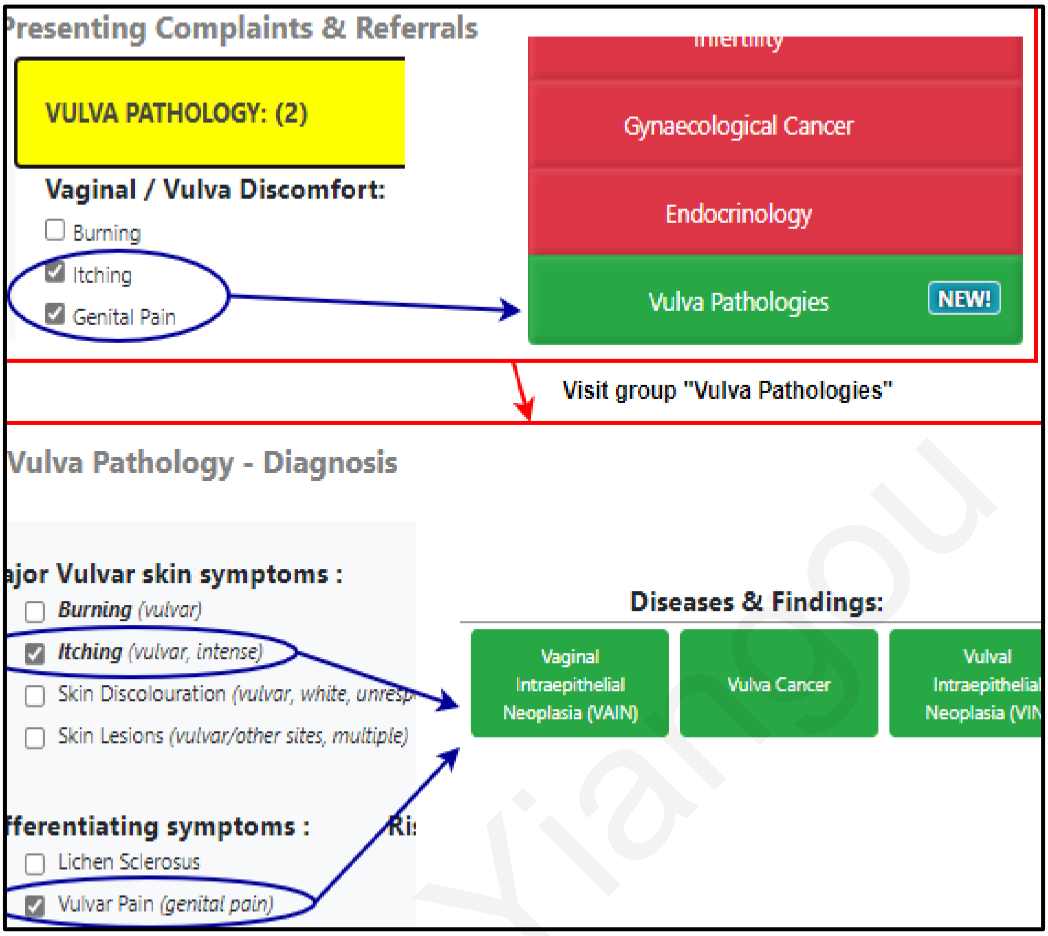


Figure 47. Input retention from Presenting Complaints Screen to a Group screen

Notice in Figure 47 the inputs that have been given in the screen of “Presenting Complaints & Referrals”. By visiting a group of diseases, the checkboxes of those inputs in the new screen are automatically clicked to remind the doctor that information from all previous stages of collecting information are still under effect and influence the decision-making process (notice that some diseases are found to be possible upon visiting the group).

The above also applies when leaving a group of diseases and entering another group. Any inputs given in the first group during investigation will be presented in the second group as well.

E.g. given the presenting complaint "Pelvic Pain (symptom)", the system recommends that we investigate two groups of diseases: "Pelvic Pain (group)" and "Gynecological Cancers". If the doctor visits "Pelvic Pain (group)" and during the investigation he learns that the patient also experiences "Vaginal Bleeding", by clicking the respective checkbox and visiting the group "Gynecological Cancers", he/she will notice that the checkbox for "Vaginal Bleeding" is already clicked, despite changing page

4.9 Query Optimization

In groups of diseases that have many scenario cases execution time increases dramatically. To mitigate this issue we needed to find a way to minimize the consulted knowledge to the most relevant to the current case.

Consider the following two assumptions:

1) Each time a user investigates a group it makes sense to currently only care for the results of this specific group
2) Knowledge is divided in multiple Gorgias files with each one focusing on the diagnosis of diseases of only one group (i.e. each Gorgias knowledge file is related to only one Group of Diseases)

Under the assumptions specified above, a form of Query Optimization is used to limit the number of knowledge we consult, to the most relevant to the inputs we collected each time. When the system is given a set of inputs, an algorithm is used in two consecutive steps, to identify the smallest set of Gorgias files that the system needs to consult in order to make a proper, but also optimized, diagnosis:

1) *1st step*: The Query Optimization algorithm identifies the Group of Diseases being currently investigated and only considers Gorgias files that are related to this group, for possible consultation

2) *2nd step*: Among the files that are found to be "possible" for consultation in the 1st step, the algorithm re-considers the inputs and rejects files that, although are relevant to the current Group of Diseases, they do not utilize any of the inputs given to generate arguments and extract conclusions.

The above steps are repeated each time a new piece of information becomes available, in order to reconsider which files are needed to be consulted to cover the new parameters.

To understand how the algorithm works, observe the following figures which contain pieces from the actual code:

```
 /**
  * BLEEDING GROUP
  */
 if (arg.compareTo("bleeding") == 0)
                ...
                ...
                ...
/**
 * VULVA PATHOLOGIES GROUP
 */
if (arg.compareTo("vulvaPathologies") == 0) {
                ...
                ...
                ...
```

1

```
// heavy periods
if (tempInfo.contains("heavyPeriods."))
    relevant.add("consult('GORGIAS/KNOWLEDGE/*/*/heavyPeriods.pl').");
```

2

```
// heavy periods && colon polyps
if (tempInfo.contains("heavyPeriods.")
        && (tempInfo.contains("colonPolyps.") || tempInfo.contains("history(colonPolyps)."))
    relevant.add("consult('GORGIAS/KNOWLEDGE/*/*/heavyPeriods_polyps.pl').");
```

3

```
// postcoitalBleeding
if (tempInfo.contains("postcoitalBleeding."))
    relevant.add("consult('GORGIAS/KNOWLEDGE/*/*/postcoitalBleeding.pl').");
```

Figure 48. Optimizing the diagnostic process

In Figure 48, we observe a piece of code from the 1st step of the optimization of the diagnostic process, where multiple "IF" statements are used to determine the current group of diseases, in order to only consult Gorgias knowledge files that are only related to it.

In Figure 4, we observe a piece of code from the 2nd step of the optimization of the diagnostic process in the group "Bleeding / Amenorrhea / Coagulation / Vascular Problems":

- The first piece of code (1) denotes that, if collected information so far, contains the symptom "Heavy Periods", then the system will consult the Gorgias file which deals with the scenarios that begin with the complaint "Heavy Periods".
- The second piece of code (2) denotes that, if collected information so far, contains both the symptom "Heavy Periods" and the information "Colon Polyps", then the system will consult the Gorgias file which deals with the scenarios that contain both "Heavy Periods" and "Colon Polyps".
- The third piece of code (3) denotes that, if collected information so far, contains the symptom "Post-Coital Bleeding", then the system will consult the Gorgias file which deals with the scenarios that build on the symptom "Post-Coital Bleeding".

Then, if a user reports "Heavy Periods" and "Post-Coital Bleeding", piece 1 and piece 2 of the above code will apply and the two files will be consulted, but not piece 3, since the user did not report "Colon Polyps", thus the system avoids broadening its search space with scenarios that are clearly not valid, under given information. The above approach is a sample of the Query Optimization code and is indicative of the rest of the code (i.e. the same philosophy applies to all groups that are affected by this algorithm).

The 1st step of the Query Optimization algorithm applies to all groups of diseases and is the main contributor to the reduction of execution time, as it lessens the number of consulted files considerably.

The 2nd step of the Query Optimization algorithm is not applied to all groups of diseases, since their knowledge base is very limited and their consultation does not cause much delay, compared to other groups. This technique is applied only where it will *significantly* improve the performance of the system, such as to the following larger groups of diseases:

- *Sexually Transmitted Infections*

- *Bleeding / Amenorrhea / Coauglation / Vascular Problems*
- *Gynaecological Cancers*
- *Endocrinology*

The above groups have many possible presenting complaints/referrals and symptoms that may have triggered their investigation and they also have many, large scenarios with many diseases and differentiations involved. *"Pelvic Pain" diseases, although a large group with many diseases and possible symptoms, is handled differently (see "Pelvic Pain Approach" section).*

E.g. investigation of "Sexually Transmitted Infections" may begin with reporting "Vaginal Burning", "Vaginal Itching", "Vaginal Discharge", "Flu-Like Symptoms / Fever" and/or "Unprotected Sexual Intercourse". Then, combinations of the above with important differentiating information such as the description of a vaginal discharge or "Inter-Menstrual Bleeding" and "Post-Coital Bleeding" yield a large number of scenarios. Consulting all this knowledge, when only given the input "Vaginal Discharge" is not needed. We only consult a file that utilizes "Vaginal Discharge" to create arguments, until further, new information becomes available, that requires the consultation of more Gorgias files

This will potentially limit the execution time, as the search space narrows-down significantly and the system does not try to check cases that have nothing to do with the inputs.

It is reasonable to say that giving more inputs will likely involve the consultation of more Gorgias files, depending on the case, thus increasing the knowledge that the system has to consider to reach a decision, and as a result execution time increases.

The algorithm is written in JAVA and communicates with the rest of the system via PHP service. It is executed shortly after printing the user inputs to the file, before any other process in Gorgias is executed, in order to regulate the number of Gorgias files that are consulted, before the Argumentation reasoning process.

4.10 Architecture

The system incorporates a variety of technologies, tools, methods and algorithms, as described in all previous sections of this chapter, and in conjunction with the large scale of the theoretical problem we are studying (as presented in Chapter 2: "Theoretical Model"), it was not impossible to create a system that, although works as expected, would have not been easy to expand, edit or have a part of it isolated and replaced. One of our goals was to avoid this pitfall, by implementing the system using a "Modular Architecture": a structure which will allows future works to be added to this project with minimal effort, all without interrupting the workflow or codes of sections that are not being edited or changed at the time.

Each component of the system is designed so that it can be changed without affecting the functionality of other modules. Below is an example of the architecture, in the level of tools & technologies:

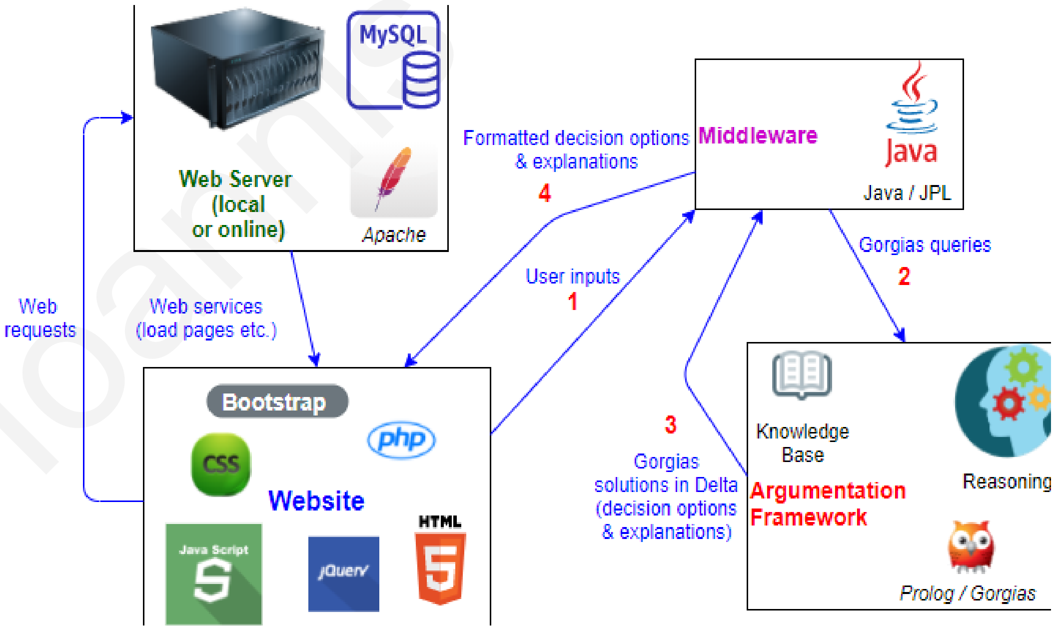


Figure 49. Modular Architecture

If, for example, in the near future, new medical knowledge is discovered about a disease, and we need to adapt the system to consider this new information, we only change the Argumentation framework by creating additional Gorgias rules to account for the new information that is related to the disease.

If a new disease is discovered, we also add new decision options and later we create new buttons in the user interface. All those happen, without touching the website source code or the middleware source code while implementing the changes in the Argumentation framework. If a new prompting technology is needed, we make changes to the code of the website, without the risk of changing Gorgias/Prolog codes, something that could have caused serious undesired technical or logical errors.

The point remains that, whenever a new change triggers the need for adaptation, the expansion of a part of the system will not affect the code, the technologies or processes in other modules, eventually avoiding the danger of causing errors and problems in other areas of the system. It goes without saying that some changes may be needed in other modules too, when changing a specific module. However this is minimized to the point that we are confident enough to avoid the need of multi-tasking in the form of changing various coding files at the same time, something that could mess-up the connections between the system components.

4.11 Operation & Applications in Real-Life Scenarios

In order to capture the whole scope of the system, APPENDIX B presents some important scenarios from all 8 groups of diseases we defined, as well as from different events that trigger the need for a different approach, such as continuing the investigation in another group. To avoid further congestion of the main chapters of this thesis, please refer to APPENDIX B for “Operation & Applications in Real-Life Scenarios”.

Chapter 5

Evaluation

This chapter is divided into two sections. The first one “General Study on AI Evaluation” is a review of the Evaluation process in AI systems, discussing topics such as the importance of this step in the development of such systems, the challenges that we have to overcome in order to ensure the quality of the evaluation phase as well as the related important methods and metrics that are widely used in this process. The second section “Evaluation Program for GAIA” is a complete description of the three-phased testing and evaluation program we planned to execute, in order to validate and evaluate the clinical use and user experience of our system, GAIA. The most important goal we had in mind when setting up this program is the development of a version of the system that can be used in real-life clinical practice, and a proper evaluation phase will identify the weaknesses and missing concepts that we need to overcome, in order to achieve this goal.

5.1 General Study on AI Evaluation

<u>The Importance of Evaluating AI systems</u>

The commercial and social pressure for speedy and wide adoption of Artificial Intelligence in every domain that could have had a meaningful application, favours the rushing implementation over the evidence quality. This could lead to undesired and unforeseen effects of those technologies that are capable of undermining the delivery of services, something that is extremely dangerous in domains of utmost importance, such as healthcare, where this could mean risking a patient’s life [49]. In order to identify and resolve such issues, we need to give great focus on the evaluation of the system, which is the most critical part of the development cycle in letting us know what components of the system are leading (or may lead to the future) to more disservice than good, in case the system will be put to actual practical use.

The importance of evaluation does not stop at finding out if the system works and at what degree. A study shows that doctors do not tend to use Diagnostic Decision Support systems (DDSS), at least widely, because of usability and integration issues [21]**.** Also, studies reveal that explanations returned by many such systems are not trusted, are not clear, are not transparent or are not understood by the patients [49], [81]. It may be possible to explain these observations, once an evaluation cycle is executed that uncovers the reason that the systems were not successful, not only in terms of performance, but also in terms of their acceptance and use.

Challenges in the Evaluation of AI systems

Evaluating an AI system comes with its own set of challenges. Firstly, when evaluating the performance of the system, we have to consider a number of metrics and their importance in this process. What do we want to measure and how? What is the importance of comparing the results? We need to answer those questions before using random metrics that serve no actual purpose. For the performance of the system, it is usually the following metrics that are used: discrimination, accuracy and precision [49].

In the second place, the system has to reflect the current state of the medical knowledge, which means following the latest trends, practices and studies in the field and incorporate them in the knowledge of the system. The challenge here is to evaluate whether or not a system does indeed adapt to the ever-changing knowledge of the field [49]. If we also consider that the amount of information in a domain becomes increasingly larger over the course of the time, it is certain that some parts of it may unwittingly be ignored or go unnoticed. This is not inherently wrong however an evaluation process has to be able to define if the magnitude of evidence in the system is sufficient, in order to perform as expected [49].

The next challenge refers to the evaluation of the impact of use of AI systems in the diagnostic process and how it compares to the diagnostic process without it. This means that the decisions of a DDSS have to be compared to a doctor's respective decisions for the same case, in terms of validity, correctness and usefulness [49]. An interesting observation here is whether or not the medical expert will change his/her mind, when assisted by the DSS. The challenge here is to ensure that the decisions are correct in order to encourage the doctor to adopt them without the fear of horrible consequences, especially since "automation bias" (i.e. the over-relying on computer decisions), which occurs very often, will magnify the problem [49].

Finally, trying to find reliable metrics that would enable the objective and quantified evaluation of the explanations of the decision support system is by itself an obstacle to overcome [81]. The question here is how to compare and measure sentences in Natural Language (i.e. explanations), in a way that is meaningful for the improvement of the system and the promotion of trust and transparency.

Evaluation Methods

There are three main methods of evaluating an AI decision support system and are all essential for a complete coverage of all the challenges discovered above.

The first method is the general evaluation of the system by reviewing relevant literature. This is done by gathering knowledge from various papers and other works that offer an insight on how a system should react and perform, based on given parameters and metrics. This evaluation method can be done without the presence of a medical expert or patient and mainly concerns the general performance of the system, compared to collected knowledge [49].

The second method is the specific evaluation of the system by using expert focus groups. This means that the decisions and performance of the system is compared to these of the doctor in order to address any issues that concern the correctness, validity and meaningfulness of the system results, as well as the level of discrepancy between two decisions for the same case (doctor versus the system) [21]**,** [79]. In this method, the doctor is the rater of the results [79] and is responsible for reporting the level at which the system is correct, the level at which it helped him execute his/her tasks and whether or not his decision changed because of discrepancy between their initial decision and that of the system [21].

The third and final method is the use of real patients in order to assess the performance of the system. Their resulting pathologies and patient information can be used to evaluate the accuracy and precision of the system [80], while their critical opinion can be used to rate how satisfactory are the explanations given by a Diagnostic Decision Support system, compared to those of the human doctor, in order to assess their levels of trust and transparency [81].

Evaluation Metrics

There are various metrics that can be used to evaluate a decision support system. The most common are “Accuracy” and “Precision” [49]. Accuracy refers to how close a system decision approached the expected decision for a given case, while Precision refers to how often the system returns the same, or similar decision, for the same case [82]**.**

Concerning this project, there is no meaningful reason to measure the Precision, since the system uses Knowledge Bases, where inference rules always return the same result when given the same inputs over and over again i.e. the system will diagnose the same diseases if given the same inputs.

What we seek is to match the predictions of the doctor and find out how accurate the system was in the process. What matters more is the measurement of Accuracy, which in our context refers to how close was the diagnosis made by the system, to the diagnosis made by the medical expert, if both parties were presented with the same set of patient information and complaints. In order to use the Accuracy metric in the evaluation phase, we need to define the following Classifications:

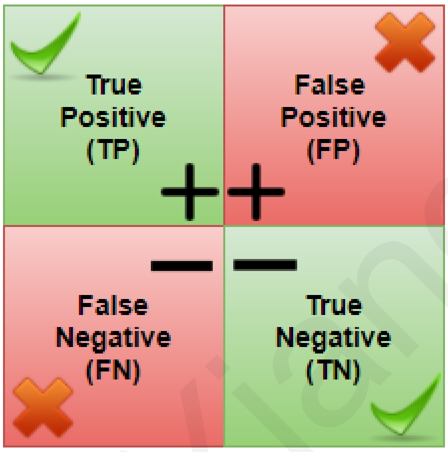


Figure 50. Classification Table

Observe Figure 50**.** Given a set of input patient information P, the following applies:

- *True Positive (TP):* number of diseases deemed "possible under P" by both the system and the doctor
    - *i.e. number of diseases, correctly found to be "possible"*
- *False Positive (FP):* number of diseases deemed "possible under P" by the system and "not possible under P" by the doctor
    - *i.e. number of diseases, incorrectly found to be "possible"*
- *False Negative (FN):* number of diseases deemed "not possible under P" by the system and "possible under P" by the doctor
    - *i.e. number of diseases, incorrectly found to be "not possible"*
- *True Negative (TN):* number of diseases deemed "not possible under P" by both the system and the doctor

- *i.e. number of diseases, correctly found to be "not possible"*

The above classifications of each disease diagnosed will help us determine the level of agreement and discrepancy we have with the doctor.

We can define the level of agreement with the doctor as the number of correct diagnoses of the system. Using the above terms, we mean the number of diseases the system correctly found to be "possible" (True-Positives), plus the number of diseases the system correctly found to be "not possible" (True-Negatives), out of all the diseases.

$$\text{Level of } Agreement = (TP + TN) / (TP + FP + FN + TN)$$

Additionally, we define the level of discrepancy as the number of diseases on which the decision of the system was different from that of the doctor. Using the above terms, we mean the number of diseases the system incorrectly found to be "possible", plus the number of the diseases the system incorrectly found to be "not possible", out of all the diseases.

$$Level\ of\ Discrepancy = (FP + FN) / (TP + FP + FN + TN)$$

The level of discrepancy can be used to identify how many (and which) cases have been diagnosed differently from the doctor. We are not hurried to label those cases as "wrong", but we follow a different approach with another idea in mind. Since the system is a Diagnostic Decision Support system, we want to evaluate another of its qualities, and that is the ability to remind the doctor about cases they forgot, ignored or did not notice, in the quest of reducing "over-diagnosis" of common diseases, out of comfort. This means that the results of discrepancy can be used to learn from the doctor about the cases we had different decision than him and he changes his mind in order to support the decision of our system, instead of his previous decision.

We can use the same metrics to calculate the levels of agreement and discrepancy for the cases where we argue on whole Groups of Diseases, instead of single diseases, such as when a patient gives a presenting complaint or referrals at the very start of the process, and when we are inside the groups for focused investigation and a new group of diseases becomes available along the way.

Finally, to measure the effectiveness level of prompting and explanations, we need to consider that both are more appropriate for subjective, than to objective, criticism. This happens because people have different tolerance thresholds concerning things like "alert fatigue" (i.e. the point from which an alert starts being annoying), trust in words and overall satisfaction.

We can use questionnaires to learn, from a scale of 0 to 10, what are the levels of satisfaction, transparency and trust of those features, as well as to what extent they convinced the doctor to change his decision, remain faithful to his initial one, check the newly-available groups, or has thought the same question as the one suggested by the dialog box to ask next etc.

5.2 Evaluation Program for GAIA

The importance of the evaluation phase and the challenges that accompany it will let us know what issues we have to face and how crucial it is to overcome them, while the methods and metrics discussed above helped us design and adopt a strategy for the testing and evaluation of the system proposed in this thesis.

The overall aim of the evaluation process of our GAIA system is to help us achieve our goal to provide diagnostic decision support for medical gynecologists, to discourage the over-diagnosing of common diseases and to avoid missing out emergency cases. We approach this evaluation process in three-phases, as follows:

- *Phase 1:* A senior consultant generates test cases that are based on real-life patient cases
    - *Main task A:* evaluation of the accuracy of diagnosis of the system
    - *Main task B:* definition of any knowledge or relevant information that are missing from the system
- *Phase 2:* A junior doctor will use the system, in parallel to a senior consultant's diagnostic practice, in order to compare their results
- *Phase 3:* A focus group of practicing gynecologists will use the system in their daily practice

The results of this three-phase evaluation process will give us a better idea on how the GAIA system can be utilized by doctors of different experience levels (e.g. senior / junior) or even individuals from different interest groups, such as nurses and patients. This would help us proceed to the creation of specialized versions for each one of them.

For Phase 1, the aim is to receive feedback regarding the accuracy and completeness of coverage of cases that may occur in real-life, with the intention to improve the system to the point that it is considered ready for Phase2 of the evaluation process. Collection of information for Phase 1 will follow the usual practice flow, as described in the theoretical model of the system:

1) Fill the patient record

2) Give a hypothetical presenting complaint or referral, along with the expected Groups of Diseases that would have been activated by this decision

3) Continue the testing phase with investigations inside the Groups of Diseases. The investigation on each group follows the following cycle (rounds 2 and 3 are optional)

- Round 1: Give a further description of the presenting complaint or collect other patient information, along with the expected results of this process:

which diseases are possible and which groups of diseases are expected to be suggested by the system

- *Round 2*: Give additional information that would further differentiate between diseases, along with the expected result of this process: which diseases are possible and which groups of diseases are expected to be suggested by the system
- *Round 3*: Give a last set of additional, more focused information, along with the expected result of this process: which diseases are possible and which groups of diseases are expected to be suggested by the system
- *Comments*: We give the chance to the doctor to give additional comments

All phases will also include a task of evaluating how successful is the system at guiding the process of symptomatology, in the quest of assisting training doctors to collect relevant information, to focus on specific relevant diseases and to consider the existence of rare, but still possible, diseases and conditions. A great focus will be given on evaluating the prompting dialogs of the system and the explanations provided for each disease.

This thesis presents the results of the 1st round of Phase 1, which along with the 2nd round of Phase 1, is currently guiding us to address several weaknesses in the system, mainly from the point of view of its clinical use. Other phases will be carried out later on:

- Phase 2 is expected to be carried out over the following July, August and September
- Phase 3 is expected to be carried out over the following Autumn and Winter

The "Future Work" section of the thesis includes a brief reference to the execution of Phases 2 and 3.

This extensive evaluation program is the first essential step towards developing the system into clinical real-life use, as each phase will detect even more information that the system could utilize to improve the quality of the decision-making process and the user's experience and interaction with the system.

Material:

The senior doctor that co-supervises this thesis prepared sixteen test cases independently, using a Word file template as shown in APPENDIX C. Each test case presents the whole process of consultation of the patient with the doctor (as described earlier in this chapter) up to the point where a clinical examination and/or laboratory test will be requested for further investigation.

Method:

The evaluation testing is carried out in two parts.

The first part calls for testing each test case in the system and keeping a record of the outputs (possible groups or specific diseases), with the purpose of comparing them with those given by the doctor, using various metrics described below. If a piece of information is missing from the system, but is included in a test case, we record it as "missing information" to implement it in a later stage.

The second part calls for identifying any discrepancies between the doctor's suspicions in a test case and those returned by the system. This includes the case of a group or disease that the system presents as "possible" but the doctor does not (for the same test case), or the opposite, the case of a group or disease that the doctor presents as "possible" but the system does not (also for the same test case). For each such case, the doctor is shown the explanations of the system for its own conclusions and then they reconsider and decide whether the discrepancy is:

1) *Valid*: this means the doctor changed his mind and agrees with the system

2) *Possible*: this means the doctor finds the conclusions of the system relevant, although remotely

3) *Invalid*: this means the doctor considers this discrepancy as a system error that needs to be fixed

We then can categorize invalid discrepancies (i.e. system errors) into two categories that describe their error prediction:

1) *Incorrectness*
2) *Incompleteness*

The most important findings of this step, that are given more weight, are those of test cases where the doctor did not record a possible disease that has been predicted by the system. Those will help us find out if the system really misses important information, or if it really achieved its goal of assisting senior doctors in remembering pieces of information they forget or rarely consider.

Comparison Metrics:

The metrics that can be used to measure the weighted closeness of the doctor's prediction and the prediction of the system, for each test case, are the following:

$$\text{Level of Agreement} = (TP + TN) / (TP + FP + FN + TN)$$

$$\text{Level of Discrepancy} = (FP + FN) / (TP + FP + FN + TN)$$

More information about the above metrics is included in the "***Study on AI Evaluation: Evaluation Metrics***" section of this chapter. These metrics will be used for calculating, both the closeness of the prediction of diseases, and the closeness of the prediction of the groups of diseases, for each test case.

Results, Observations and Conclusions for Phase 1:

By studying the test cases and trying to simulate them in the system, we were able to identify and address the issues that are needed to be resolved, before even considering the real-life use of the system.

Firstly, we identified several vocabulary discrepancies between the doctor's expressions in the test case forms, something which we encouraged in order to create a list of synonym terms that can be implemented to help the system understand and accept different words and terms for the same concept. Some examples are the following:

SYNONYM TERMS:		
Attribute:	**Term expressed by medical expert:**	**Term provided by the system:**
Symptom	Increased Vaginal Discharge	Profuse Vaginal Discharge
Symptom	Intense Bleeding	Heavy/Profuse Bleeding

Table 30. Sample of the collection of Synonym terms

To add to the above, some of the terms expressed by the doctor in the test case forms were completely unknown to the system. Also, the doctor described some concepts that are currently not implemented fully in this system, such as the use of "blocking arguments" in the diagnostic process. A great example of this concept, which was given by the medical expert that co-supervises this thesis, is an Ultra-Sound image which does not show any Ovarian Cysts. Although the system provides the ability to input the information "Ultra-Sound Scan shows Ovarian Cyst Rupture / Torsion/ Hemorrhage etc.", it did not account for the case of not showing any Ovarian Cysts, something that would have eliminated the possibility of a disease / condition related to Ovarian Cyst accidents. We noted these terms in order to observe the level at which the system is incomplete, with the goal to implement them in the near future. Some examples of unknown (to the system) terms are the following:

UNKNOWN / MISSING TERMS:		
Attribute:	**Term expressed by medical expert:**	**Comments:**
Symptoms / Inputs	**NO** Urination Urgency	*Blocking argument for Urological / Urinary diseases & conditions*

Patient Record: Imaging test result	US shows **NO** ovarian cyst	*Blocking argument for Ovarian Cyst diseases & conditions*
Patient Record: Laboratory test result	BetaHCG **Negative**	*Blocking argument for Pregnancy*
Disease / Finding / Output	Ureoplasma	*Did not exist in scenario tables*
Disease / Finding / Output	Endometritis	*Did not exist in scenario tables*

Table 31. Sample of the collection of Unknown/ Missing terms

Phase 1 of the Evaluation program confirmed some issues we already had in mind that would occur, something which is very natural when working with a real-life problem of this magnitude. Most notably, we observed the increased need for flexibility in the vocabulary and syntax of information, in order to recognize the references to a specific concept, even if the concept is described with different, but synonym, terms. There is also the need to separate the patient record information into different intervals, as we discovered from the medical expert that some of the fields are very standard and are being completed in the very beginning of the diagnostic process, while other fields might be less required and are completed during the dialog with the patient, and only when needed, or when explicitly reported by the patient. In addition to the above, another very important issue is the incompleteness in the system to recognize some overlapping diseases that occur in more than one groups of diseases, and that is mainly attributed to the fact that each group becomes available by giving a specific piece of information *(e.g. "Dysuria" leads to the recommendation of "Urogynecology & Prolapses").* However this does not account that a disease might be possible, even in the absence of a specific related symptom, mainly because several other identifying symptoms are presented:

e.g. (the below is given as an example by the medical expert that co-supervises this thesis)

- *"Ovarian Cyst Hemorrhage" is a decision option in the group "Pelvic Pain"*
- *In order to "activate" the group of "Pelvic Pain", a patient must present Pelvic Pain, Lower-Abdominal Pain or other pain of gynecological nature in the pelvic girdle or abdominal area*
- *However, "Ovarian Cyst Hemorrhage" can be possible even without pain, mainly because other symptoms are present such as uterine/vaginal bleeding*
- *In this case, we would have to account that "Ovarian Cyst Hemorrhage" belongs to other groups of diseases at the same time (overlapping), therefore it could be possible, given any symptom that is related to it (not just with Pelvic Pain, but with Bleeding too)*

Finally, this evaluation process confirmed our suspicion that a better prompting method for acquiring explicit negative information is needed, in order to fully cover the concept of "blocking arguments", as we observed that, just assuming a piece of information as "missing" does not help to exclude diseases as much as if we had known that it is either positive or negative.

As we already mentioned earlier, the intention of this evaluation program was to address these issues and adopt strategies that would help us deal with them over the near future. Further actions will be taken in the initial stages of the future work on this project, as presented and discussed in the respective, final section of the thesis.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

From a general aspect, this thesis provided detailed guidelines and proposed a methodology for approaching large-scale, real-life problems, such as the diagnostic process in the field of Gynecology, and the modelling of the complete medical expertise, using Argumentation frameworks in order to enable Cognitive AI Systems to conceive the way human experts find acceptable solutions in difficult, complex and high-risk problems, all while being able to justify those decisions with valid, indisputable arguments. The theoretical model and development methods provided in this thesis can be utilized for solving a variety of problems inspired by situations and dilemmas, us, humans face in real life, with great success.

From a more specific aspect, the result of this thesis work is an AI Diagnostic Decision-Support System based on Argumentation, for the medical field of Gynecology. GAIA can provide support to Gynecologists in the form of cognitive assistance, during the diagnostic decision-making process of their clinical practice, benefiting doctors of various experience levels, as senior and consultant gynecologists refresh their memory and hone their skills, while junior and training doctors could experience an enhanced training and learning process in gynecology. As an additional benefit, the system is able to lighten the cognitive load of the user, since a large part of the required processes, such as logical reasoning, the consideration of large amounts of knowledge, the handling of contradicting information and the decision-making, are all executed by the system.

Our goals for this project are reflected in the requirements and specifications described earlier, and most of them have been achieved. These include the hosting of a complete knowledge of medical expertise that will stay updated as it evolves over time and its formalization in Argumentation logic, thus enabling us to integrate the concept of “Explainable Artificial Intelligence”. Another important goal was to enhance the accuracy and speed of the diagnostic process. This was achieved by alerting the user about various events such as when new groups of diseases become available, when a piece of information is required to be acquired and when an emergency condition is possible. Also, by showing all the diseases that are available and possible, at all times, and by encouraging broader diagnoses, the system accounts for “Differential Diagnoses” while eliminating the phenomenon of “overdiagnosing” common diseases. Furthermore, the dynamic revision of previous conclusions, when new information becomes known, triggers the immediate, automatic update of the current state of the system, to match the new context.

In addition to the above goals, we also tried to make the system more user-friendly, by providing a graphical interface in the form of a website, which is easy to become familiarized with, and to simulate a natural human-system interaction, with our ideal model being the typical diagnostic process as it happens in real life, and that includes dialogs, alerts and other forms of interaction between the two parties.

Finally, a three-phased evaluation program was devised to test and evaluate the system, as well as to capture any piece of knowledge that is missing, or is incomplete, or is described using different vocabulary in order to implement them at a later stage, in the near future, a task that is much easier since we follow a modular architecture. While Phases 2 and 3 are programmed for a later stage, Phase 1 of this process already gave us a good insight on what we need to do, and how we need to act, in order to achieve the ultimate goal we had in mind while working on the system, and that is the development of a version of the system that can be used in real-life clinical practice.

6.2 Future Work

The Cognitive Gynecological Assistant satisfies the most part of the initially defined requirements, making it an acceptable tool for use in the diagnostic decision-making process, as source of a reliable, second opinion. However, the system certainly has room for improvements since the great magnitude of the demanded task, coupled with the need to be completed within the time frames of the dissertation made it difficult to implement potential enhancements we had in mind.

A number of possible future works are described in this chapter, with some of them being attempts to improve the solution of the problem and expand its domain to a larger spectrum, while others are more interface-oriented by attempting to enhance the communication and interaction between the user and the system. These resulted from various discussions with both supervisors (medical expert and computer expert), in order to guarantee that any potential improvements, expansions and recommendations are acceptable, relevant and meaningful as goals.

Finally, by explaining "why" it is important to implement each suggestion, we provide the idea and motivation. By explaining "how" to achieve this, we provide the methodology and tools. Both are necessary in order to pave the way for further, constructive expansion of this project and as an extension, for further contribution to both Medical (Gynecology/Obstetrics) and Computer Sciences.

- **Execution of the remaining two steps of the three-phase Evaluation program (Phase 2 and Phase 3), that is described in the "Evaluation" chapter of this thesis**

Importance: By completing all testing and evaluation phases that are scheduled in the three-phase evaluation process, we ensure that the system does not miss any critical information and that it works as expected, in order to meet all the technical requirements and specifications that we introduced earlier in this thesis

How: Phase 2 will help us acquire a clear image on how useful the system is, for junior, and training doctors and what is needed to change in order to further improve the learning quality and experience of this focus group. The successful completion of Phase 3 will be a big step towards the acceptance of the system by the expert focus group of gynecologists and the application of the system in real-life. This would mean that the system will be ready to offer actual, valuable assistance that would aid to the improvement of the healthcare services in this domain.

- **Use of Negative Information as Blocking Arguments**

This refers to a future potential of the system to consider, not only the presence of a piece of information, but also its absence, during the decision-making process. An example is the absence of an Ovarian Cyst as shown in an Ultra-Sound scan, something that the system does not account for in the current stage (it only accounts for the presence of a condition, and in case such information is missing, the system does not assume the opposite option)

Importance: This kind of information can be used as general strong arguments against specific diseases, under a specific context. This means that the system will be able to exclude diseases, based on both information that is present and information that is missing

How: To implement the complementary relation between the two states of an information (present or absent), we can construct predicates for negative information in Logic Programming (Prolog), as negations of existing information. An example is the presence of vaginal discharge (vaginalDischarge) and the absence of vaginal discharge (not(vaginalDischarge)). These predicates can be used by Gorgias to construct the appropriate rules for argumentative reasoning using this concept. This way, the system will be able to argue upon the absence of a piece of information too. To collect such information, an extension of the interface of the system is suggested, in order to seek negative information, using chat boxes to ask the user with Yes/No options for a particular condition, symptom or any other piece of information, which its absence has value for the decision-making process.

- **Implementation of a Medical Image Processing module for supplying the system with Imaging Tests results**

Importance: Provides higher accuracy and more efficiency since it will not be required for the doctor to examine the image in order to know what patient information to input to the system. This module will do it automatically.

How: The Medical Image Processing module would be able to analyze, process and extract conclusions from real imaging examinations (X-Ray, Ultra-Sound Scan, Magnetic Resonance Imaging etc.) using Machine Learning algorithms. Then, the results (accuracy, variance etc.) will be used by a Common Sense module, consisted of Logic Programming inference rules, to infer high-level concepts (i.e. interpretable results, *e.g. "MRI shows Adenomyosis")* to the Knowledge.

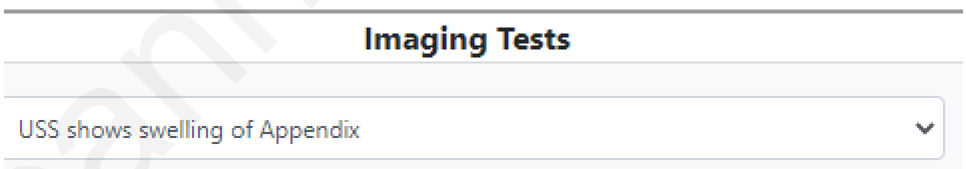


Figure 51. Imaging Test input

e.g. instead of searching a drop-down list for "USS shows swelling of Appendix" (like in the figure above), if this module identified a swelled appendix in an Ultra-Sound Scan image, then it would input this finding, instead of the doctor

- **Extend the theoretical model to account for a wider range of clinical examinations and medical tests**

Importance: this would provide higher diagnostic accuracy, as such piece of information can help to further focus on possible diseases

How: by extending the patient record to consider more examination and test types and possible results, as well as different related concepts too (e.g. "Normal levels" of a test marker may exclude some options, or "Ultra-Sound shows NO Ovarian Cyst" indicates a strong argument based on the absence of a specific condition)

- **Implementation of a module to translate Laboratory test results to High-Level concepts**

Importance: provides higher accuracy and more efficiency since it will not be required for the doctor to find out what the numbers mean in the results. This module will make conclusions, based on those numbers, which will be interpretable by humans. This will also help users with no medical background or knowledge, to understand their test results.

How: the results will be used by a Common Sense module, consisted of Logic Programming inference rules, to infer high-level concepts to the Knowledge

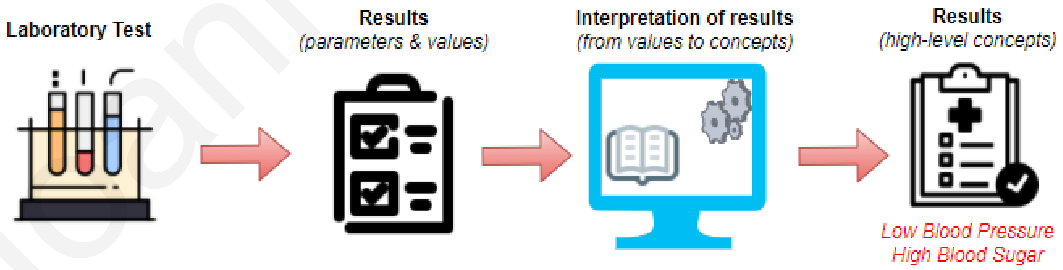


Figure 52. Laboratory Test High-Level Interpretation

To add to the above suggestions, a service which lets the doctor order various tests (laboratory or imaging) could also be very convenient and practical.

- **Make targeted suggestions to doctors on what piece of information to seek next (e.g. ask for a specific symptom) for all groups of diseases.**

<u>*Importance*</u>: This would enhance the training and learning process of junior / trainee doctors by showing them a possible sequence that a diagnosis of a specific condition may have. Also, it would remind the experienced doctor for cases he/she forgot to investigate due to stress or cognitive fatigue.

Part of this future work has already been implemented for the group of "Sexually Transmitted Infections (STIs)" and here is suggested the further expansion of this work to other groups too

<u>*How*</u>: by considering the inputs given previously, the system will be able to identify refinements of the current scenario that would occur by assuming a piece of missing information. The suggestion to the doctor will consist of pieces of these missing information (i.e. symptoms and other patient information) that are assumed to be the next relevant thing to ask for, in order to achieve refinement of the current scenario. This could be possible by using the "abducible" predicate in Gorgias, which lets us program assumed information and scenarios that occur under assumptions [83].

- **Provide various types of Explanations**

<u>*Importance:*</u> This would help various interest groups to seek the explanations they need for things they matter to them. The various types of explanations have been discussed in Chapter 2 - Background, under the section "Explanation Types"

  - *e.g. a computer programmer might want to replicate a part of the diagnostic module or fix a bug therefore he/she is interested in Trace-based Explanations which offer details of processes used by the system to derive the decision, step-by-step* [34]

- *e.g. a statistical analyst which studies cancer statistics might be benefitted by a Statistical Explanation, one which utilizes numerical evidence, probabilities and percentages to support decisions* [34]

How: Since the admissible arguments are all collected in a place (i.e. the admissible set), it is only a matter of tailoring the explanation of each one of them, using string-manipulation techniques, in a different structure and vocabulary than the current (which is more focused on "justifying" the decision). Any other required components such as statistics and diagrams may be produced by using Machine Learning algorithms. We can give the option to the users to select a type of explanations they desire from the system.

- **Provide Decision Support for choosing Medication & Treatments (Gynecology)**

Expand the implementation to also assist the doctor in the decision-making process regarding the selection of the best Medication & Treatment solutions for the diseases that were deemed possible in the diagnostic step.

Importance: Increases the efficacy and accuracy in a larger part of the doctor's job, compared to only offering diagnostic decision-making support

How: We can use the same methodology proposed in this thesis (Scenario Tables, Argumentation, Explanations, Prompting, Theoretical Model etc.) and adjust the context accordingly. The approach here is an Argumentation framework with arguments constructed by considering the following:

- various drug interactions (drug-drug, drug-food, drug-condition) [84]
- possibilities of success / failure / necessity for medications and treatments
- notable side-effects that accompany a solution
- risk factors involved etc.

- **Provide Diagnostic Decision Support for the field of Obstetrics**

Expand the implementation to also assist obstetricians and gynecologists in the diagnostic decision-making process regarding Obstetrics cases

Importance: Will provide Obstetricians and Obstetrician-Gynecologists with full diagnostic support covering both the gynecologic part with this thesis work, and the obstetric part with this future work recommendation

How: We can use the same methodology proposed in this thesis (Scenario Tables, Argumentation, Explanations, Prompting, Theoretical Model etc.) and adjust the context accordingly.

- **Implementation of a proper Database for Historical Information of Patient**

Importance: The system mainly relies on text files to handle patient information *(considering that the system only handles hypothetical cases at this stage, not real-life ones yet).* A proper database would provide a more structured and efficient way of storing and handling information, especially when the number of patients is dramatically increased. It would also offer a more secure and standard way of protecting sensitive information once the system starts handling information from real patients.

How: The implementation of a patient record based on established guidelines in order to comply to the national and EU standards will ensure the achievement of this future goal

- **Implementation of a Search Engine for medical terms**

Importance: Since the field of medicine mainly uses Natural Language as mean for expression and communication, there are various terms or phrases that are referring to the very same condition and are used interchangeably (e.g. "Spotting" is the same as "Light Inter-Menstrual Vaginal Bleeding" [85]). The current state of the system does not support multiple descriptions of the same concept, however if it did, it would make doctors feel more confident on describing what they have in mind more accurately. This would also help users with no medical knowledge (e.g. patients) to translate medical terms to simpler ones, in order to describe a symptom or condition

A possible solution is to implement a search engine that accepts a user's desired input and maps it to the way it is named in the system, then highlighting the corresponding input in the interface, or even completing the field directly.

<u>How</u>: A computer system cannot understand those terms without explicitly stating the relations between different terms and phrases (synonyms, antonyms, unrelated etc.). Therefore, Logic Programming inference rules can be used to relate terms in the head of the rule, with the terms given by the doctor in the body of the rule

e.g. light_intermenstrual_vaginal_bleeding :- search_term('spotting').

The search engine will accept semi-structured free-text, as in the case of keyword searching. The system will break the input string to its consisting terms/keywords and will consult a thesaurus of terms and the term relation rules to relate the keywords to inputs, as presented in the interface. Those inputs will be highlighted to be noticeable. An interesting alternative input method would be voice recognition, something that would require the collaboration of computer experts from different domains.

- **Make steps to promote the implementation of a commercial version of the project**

<u>Importance</u>: The commercialization of this work will put the theory into action, by offering solutions to real-life cases and by generally contributing to the improvement of the gynecological (and generally, the medical) diagnostic practice.

<u>How</u>:

- Use data from real patient records and store them on cloud, under strict security and permission levels (e.g. nurse, personal doctor, other doctor)

- Export complete medical reports that are accepted by standard guidelines from the field of medicine, and are comprehensible by both health service providers and patients. So far, the system was intended solely for diagnostic support, however this feature will assist doctors in tasks beyond diagnostic process
- Introduce formal, medical coding standards (e.g. ICD [86] and SNOMED CT [87]) in order to promote the adoption of "Interoperability" [56]

- **Create a minimal version for users with no medical knowledge**

Importance: This version will let users have a first glance at what they might be suffering from, and asses the seriousness of the condition in order to arrange an appointment with a gynecologist, all from the comfort of their own home. All of the above, in conjunction with giving simple advice or at-home treatment methods for non emergency conditions will lighten the work load of the doctor, as patients with less serious and easily manageable conditions will not need to visit him/her.

How: We can use the same methodology proposed in this thesis (Scenario Tables, Argumentation, Explanations, Prompting, Theoretical Model etc.) and adjust the system to contain more user-friendly terms and concepts, as well as to simplify highly technical medical terms, while still retaining their meaning. The proposed system will guide the user into accurately describing his/her condition by answering simple and comprehensible questions that are based on previous inputs given, in order to be more relevant and natural. Finally, for a more wide adoption of the system, it could be implemented as a mobile application, which has the added benefit of enabling users to use it at any time and place.

Bibliography

[1] R. T. Sutton, D. Pincock, D. C. Baumgart, D. C. Sadowski, R. N. Fedorak, and K. I. Kroeker, “An overview of clinical decision support systems: benefits, risks, and strategies for success,” *NPJ Digit. Med.*, vol. 3, p. 17, 2020.

[2] J. Klein, “Differential diagnosis: Definition, examples, and more,” *Medicalnewstoday.com*, 15-Jul-2020. [Online]. Available: https://www.medicalnewstoday.com/articles/differential-diagnosis. [Accessed: 17-May-2021].

[3] MedlinePlus, “Recognizing medical emergencies,” *Medlineplus.gov*. [Online]. Available: https://medlineplus.gov/ency/article/001927.htm. [Accessed: 17-May-2021].

[4] Á. Carrera and C. A. Iglesias, “A systematic review of argumentation techniques for multi-agent systems research,” *Artificial Intelligence Review*, vol. 44, no. 4, pp. 509–535, 2015.

[5] A. Kakas, L. Michael, “Cognitive Systems: Argument and Cognition”, *IEEE Intelligent Informatics Bulletin*, vol. 17, no. 1, pp. 14-20, December, 2016

[6] C. E. Alchourrón, P. Gärdenfors, and D. Makinson, “On the logic of theory change: Partial meet contraction and revision functions,” in *Readings in Formal Epistemology*, Cham: Springer International Publishing, 2016, pp. 195–217. [Online]. Available doi: 10.1007/978-3-319-20451-2_13

[7] A. Etzioni and O. Etzioni, "AI assisted ethics," *Ethics Inf. Technol.*, vol. 18, no. 2, pp. 149–156, 2016.

[8] P. Lin, "The ethics of autonomous cars," *Atlantic monthly (Boston, Mass.: 1993)*, 08-Oct-2013, [Online]. Available: https://www.theatlantic.com/technology/archive/2013/10/the-ethics-of-autonomous-cars/280360/

[9] C. Zednik, "Solving the black box problem: A normative framework for explainable artificial intelligence," *Philos. Technol.*, 2019.

[10] J. Gerlings, A. Shollo, and I. Constantiou, "Reviewing the need for explainable artificial intelligence (xAI)," in *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2021, p. 1284, [Online]. Available: http://hdl.handle.net/10125/70768

[11] R. Schmelzer, "Understanding Explainable AI," *Forbes Magazine*, 23-Jul-2019., [Online].Available:https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/?sh=45cbfa9e7c9e

[12] G. Rabindranath, "What are explainable AI principles?," *Towards Data Science*, 17-Mar-2021. [Online]. Available: https://towardsdatascience.com/what-are-explainable-ai-principles-53aa52879cbb. [Accessed: 02-May-2021].

[13] WebMD, "What is a gynecologist?," *Webmd.com*. [Online]. Available: https://www.webmd.com/women/what-is-gynecologist. [Accessed: 07-May-2021].

[14] E. I. Emin, E. Emin, A. Papalois, F. Willmott, S. Clarke, and M. Sideris, "Artificial intelligence in obstetrics and gynaecology: Is this the way forward?," *In Vivo*, vol. 33, no. 5, pp. 1547–1551, 2019.

[15] F. Walsh, "AI 'outperforms' doctors diagnosing breast cancer," *BBC*, 02-Jan-2020. [Online]. Available: https://www.bbc.com/news/health-50857759.

[16] O. Sanchez Graillet and P. Cimiano, "Argumentation Schemes for Clinical Interventions Towards an Evidence-Aggregation System for Medical Recommendations," in *HEALTHINFO 2019 : The Fourth International Conference on Informatics and Assistive Technologies for Health-Care, Medical Support and Wellbeing*, 2019.

[17] M. E. Buron Brarda, L. H. Tamargo, and A. J. García, "An approach to enhance argument-based multi-criteria decision systems with conditional preferences and explainable answers," *Expert Syst. Appl.*, vol. 126, pp. 171–186, 2019.

[18] A. Kakas, L. Michael, "Cognitive Systems: Argument and Cognition", *IEEE Intelligent Informatics Bulletin*, vol. 17, no. 1, pp. 14-20, December, 2016

[19] C. E. Alchourrón, P. Gärdenfors, and D. Makinson, "On the logic of theory change: Partial meet contraction and revision functions," in *Readings in Formal Epistemology*, Cham: Springer International Publishing, 2016, pp. 195–217. [Online]. Available doi: 10.1007/978-3-319-20451-2_13

[20] ForeSee Medical, Inc, "Clinical Decision Support Systems (CDSS)," *Foreseemed.com*. [Online]. Available: https://www.foreseemed.com/clinical-decision-support. [Accessed: 07-May-2021].

[21] T. Porat, B. Delaney, and O. Kostopoulou, “The impact of a diagnostic decision support system on the consultation: perceptions of GPs and patients,” *BMC Med. Inform. Decis. Mak.*, vol. 17, no. 1, 2017. [Online]. Available doi: https://doi.org/10.1186/s12911-017-0477-6

[22] “GDPR Article 1: Subject-matter and objectives,” *Advisera.com*. [Online]. Available: https://advisera.com/eugdpracademy/gdpr/subject-matter-and-objectives/. [Accessed: 03-May-2021].

[23] “GDPR Article 21: Right to object,” *Advisera.com*. [Online]. Available: https://advisera.com/eugdpracademy/gdpr/right-to-object/. [Accessed: 03-May-2021].

[24] “GDPR Article 22: Automated individual decision-making, including profiling,” *Advisera.com*. [Online]. Available: https://advisera.com/eugdpracademy/gdpr/automated-individual-decision-making-including-profiling/. [Accessed: 03-May-2021].

[25] IBM Cloud Education, “What is Artificial Intelligence (AI)?,” *Ibm.com*. [Online]. Available: https://www.ibm.com/cloud/learn/what-is-artificial-intelligence. [Accessed: 30-Apr-2021].

[26] Y. C. Shin and C. Xu, *Intelligent systems: Modeling, optimization, and control*. Boca Raton, FL: CRC Press, 2008, p. xiii.

[27] Stanford University, “Ancient myths reveal early fantasies about artificial life,” *Stanford.edu,* 28-Feb-2019. [Online]. Available:

https://news.stanford.edu/2019/02/28/ancient-myths-reveal-early-fantasies-artificial-life/. [Accessed: 30-Apr-2021].

[28] Apple, "Siri," *Apple.com*. [Online]. Available: https://www.apple.com/siri/. [Accessed: 30-Apr-2021].

[29] B. Bosker, "SIRI RISING: The inside story of siri's origins -- and why she could overshadow the iPhone," *HuffPost UK*, 22-Jan-2013. [Online]. Available: https://www.huffpost.com/entry/siri-do-engine-apple-iphone_n_2499165. [Accessed: 30-Apr-2021].

[30] Microsoft, "What is Cortana?," *Microsoft.com*. [Online]. Available: https://support.microsoft.com/en-us/topic/what-is-cortana-953e648d-5668-e017-1341-7f26f7d0f825. [Accessed: 30-Apr-2021].

[31] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, Boston, MA: Springer US, 2011, pp. 1–35.

[32] F. Davidoff, "Understanding contexts: how explanatory theories can help," *Implement. Sci.*, vol. 14, no. 1, p. 23, 2019.

[33] K. Cyras *et al.*, "Machine Reasoning Explainability," *arXiv [cs.AI]*, 2020. [Online]. Available: https://arxiv.org/abs/2009.00418.

[34] S. Chari, D. M. Gruen, O. Seneviratne, and D. L. McGuinness, "Directions for explainable knowledge-enabled systems," *arXiv [cs.AI]*, 2020., [Online]. Available: https://arxiv.org/abs/2003.07523.

[35] DeepAI, "Logic Programming," *Deepai.org*, 17-May-2019. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/logic-programming. [Accessed: 04-May-2021].

[36] K. R. Apt, "Declarative Programming in Prolog," *Cwi.nl*. [Online]. Available: https://homepages.cwi.nl/~apt/ps/a-mit.pdf. [Accessed: 04-May-2021].

[37] X. Fan and F. Toni, "On Computing Explanations in Argumentation," *Core.ac.uk*. [Online]. Available: https://core.ac.uk/download/pdf/77001864.pdf. [Accessed: 05-May-2021].

[38] Y. Dimopoulos and A. Kakas, "Logic Programming without Negation as Failure," in *Logic Programming: The 1995 International Symposium*, J. W. Lloyd, London, England: MIT Press, 1995, pp. 369–383

[39] P. Baroni, M. Caminada, and M. Giacomin, "An introduction to argumentation semantics," *Knowl. Eng. Rev.*, vol. 26, no. 4, pp. 365–410, 2011.

[40] C. E. Briguez, M. C. D. Budán, C. A. D. Deagustini, A. G. Maguitman, M. Capobianco, and G. R. Simari, "Argument-based mixed recommenders and their application to movie suggestion," *Expert Syst. Appl.*, vol. 41, no. 14, pp. 6467–6482, 2014.
[Online]. Available doi: https://doi.org/10.1016/j.eswa.2014.03.046

[41] J. Liebowitz, "Expert systems: A short introduction," *Engineering Fracture Mechanics.*, vol. 50, no. 5–6, pp. 601–607, 1995. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0013794494E0047K

[42] T. J. M. Bench-Capon and P. E. Dunne, "Argumentation in AI and law: Editors' introduction," *Artificial Intelligence and Law*, vol. 13, no. 1, pp. 1–8, 2005. [Online]. Available doi: 10.1007/s10506-006-9007-z

[43] L. Longo, B. Kane, and L. Hederman, "Argumentation theory in health care," in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, 2012. [Online]. Available doi: 10.1109/CBMS.2012.6266323

[44] D. N. Walton, "Reasoned use of expertise in argumentation," *Argumentation*, vol. 3, no. 1, pp. 59–73, 1989. [Online]. Available doi: 10.1007/BF00116417

[45] A. C. Kakas, P. Moraitis, and N. I. Spanoudakis, "GORGIAS: Applying argumentation," Argument & Computation., vol. 10, no. 1, pp. 55–81, 2018. [Online]. Available doi: 10.3233/AAC-181006

[46] Gorgias-B, "GorgiasB tool," *Tuc.gr*. [Online]. Available: http://gorgiasb.tuc.gr/Tutorial3.html. [Accessed: 06-May-2021].

[47] S. Oumaima, K. Soulaimane, and B. Omar, "Latest Trends in Recommender Systems applied in the medical domain: A Systematic Review," in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, 2020. [Online]. Available doi: 10.1145/3386723.3387860

[48] D. Zikos, "A framework to design successful clinical decision support systems," in *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, 2017. [Online]. Available doi: 10.1145/3056540.3064960

[49] F. Magrabi *et al.*, “Artificial intelligence in clinical decision support: Challenges for evaluating AI and practical implications: A position paper from the IMIA Technology Assessment & Quality Development in Health Informatics Working Group and the EFMI Working Group for Assessment of Health Information Systems,” *Yearb. Med. Inform.*, vol. 28, no. 1, pp. 128–134, 2019. [Online]. Available doi: 10.1055/s-0039-1677903

[50] Duodecim, “EBMEDS Clinical Decision Support System”, *Ebmeds.org*. [Online].Available:https://www.ebmeds.org/wp-content/uploads/sites/16/2019/06/EBMEDS_brochure_2019-2.pdf. [Accessed: 07-May-2021].

[51] Symptoma, “About Symptoma,” *Symptoma.com*. [Online]. Available: https://www.symptoma.com/en/about. [Accessed: 07-May-2021].

[52] Babylon Health, “Babylon health UK - the online doctor and prescription services app,” *Babylonhealth.com*, 26-Feb-2018. [Online]. Available: https://www.babylonhealth.com/. [Accessed: 07-May-2021].

[53] R. Kunhimangalam, S. Ovallath, and P. K. Joseph, “A clinical decision support system with an integrated EMR for diagnosis of peripheral neuropathy,” *J. Med. Syst.*, vol. 38, no. 4, p. 38, 2014. [Online]. Available doi: 10.1007/s10916-014-0038-9

[54] A. I. Martinez-Franco *et al.*, “Diagnostic accuracy in Family Medicine residents using a clinical decision support system (DXplain): a randomized-controlled trial,” *Diagnosis (Berl)*, vol. 5, no. 2, pp. 71–76, 2018. [Online]. Available doi: https://doi.org/10.1515/dx-2017-0045

[55] H. Singh, A. N. D. Meyer, and E. J. Thomas, "The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations," *BMJ Qual. Saf.*, vol. 23, no. 9, pp. 727–731, 2014.

[56] European Commission, "eHealth Action Plan 2012-2020: Innovative healthcare for the 21st century," *Europa.eu*, 07-Dec-2012. [Online]. Available: https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century. [Accessed: 08-May-2021].

[57] Massachusetts General Hospital Laboratory of Computer Science, "Using Decision Support to Help Explain Clinical Manifestations of Disease," *Mghlcs.org*. [Online]. Available: http://www.mghlcs.org/projects/dxplain. [Accessed: 08-May-2021].

[58] Babylon Health, "Product," *Babylonhealth.com*, 30-Apr-2018. [Online]. Available: https://www.babylonhealth.com/product. [Accessed: 08-May-2021].

[59] Symptoma, "Digital Health Assistant & Symptom Checker," *Symptoma.com*. [Online]. Available: https://www.symptoma.com. [Accessed: 08-May-2021].

[60] A. Pietrangelo, "Endocervix culture: Preparation, purpose, and procedure," *Healthline.com*, 31-Mar-2017. [Online]. Available: https://www.healthline.com/health/culture-endocervix. [Accessed: 08-May-2021].

[61] MedicineNet, "Definition of Risk factor," *Medicinenet.com*. [Online]. Available: https://www.medicinenet.com/risk_factor/definition.htm. [Accessed: 17-May-2021].

[62] Centers for Disease Control and Prevention, “What are the risk factors for lung cancer?,” *Cdc.gov*, 23-Sep-2020. [Online]. Available: https://www.cdc.gov/cancer/lung/basic_info/risk_factors.htm. [Accessed: 17-May-2021].

[63] NHS website, “Your health records,” *Nhs.uk*. [Online]. Available: https://www.nhs.uk/using-the-nhs/about-the-nhs/your-health-records/. [Accessed: 12-May-2021].

[64] MedicineNet, “Definition of Pathognomonic,” *Medicinenet.com*. [Online]. Available: https://www.medicinenet.com/pathognomonic/definition.htm. [Accessed: 17-May-2021].

[65] “NCI dictionary of Cancer Terms,” *Cancer.gov*, 02-Feb-2011. [Online]. Available: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/diagnosis. [Accessed: 17-May-2021].

[66] A. Santos-Longhurst, “Types of pain: Classifications and examples to help describe your pain,” *Healthline.com*, 29-Nov-2018. [Online]. Available: https://www.healthline.com/health/types-of-pain. [Accessed: 17-May-2021].

[67] Mayo Clinic, “Pelvic pain,” *Mayoclinic.org*, 10-May-2019. [Online]. Available: https://www.mayoclinic.org/symptoms/pelvic-pain/basics/causes/sym-20050898. [Accessed: 17-May-2021].

[68] T. Malmström, O. Huuskonen, P. Torkki, and R. Malmström, “Structured classification for ED presenting complaints - from free text field-based approach to

ICPC-2 ED application," *Scand. J. Trauma Resusc. Emerg. Med.*, vol. 20, no. 1, p. 76, 2012.

[69] "Definition of Referral," *Medicinenet.com*. [Online]. Available: https://www.medicinenet.com/referral/definition.htm. [Accessed: 17-May-2021].

[70] A. Kahn, "Bleeding Disorders," *Healthline.com*, 26-Feb-2018. [Online]. Available: https://www.healthline.com/health/bleeding-disorders. [Accessed: 17-May-2021].

[71] P. W. Grigsby, "Vaginal Cancer," in *Gynecologic Cancer*, Elsevier, 2004, pp. 113–118.

[72] "Vaginal Discharge: What's Abnormal?," *Webmd.com*. [Online]. Available: https://www.webmd.com/women/guide/vaginal-discharge-whats-abnormal. [Accessed: 17-May-2021].

[73] S. Y. Morris, "Bleeding after sex: Causes, risk factors, and more," *Healthline.com*, 06-Nov-2020. [Online]. Available: https://www.healthline.com/health/womens-health/bleeding-after-sex. [Accessed: 17-May-2021].

[74] S. Watson, "Childbearing age: What's ideal and what are the risks?," *Healthline.com*, 06-Jun-2018. [Online]. Available: https://www.healthline.com/health/womens-health/childbearing-age. [Accessed: 17-May-2021].

[75] National Cancer Institute, "Risk Factors: Age," *Cancer.gov*, 2015. [Online]. Available: https://www.cancer.gov/about-cancer/causes-prevention/risk/age. [Accessed: 17-May-2021].

[76] “Defining adult overweight & obesity,” *Cdc.gov*, 28-Apr-2021. [Online]. Available: https://www.cdc.gov/obesity/adult/defining.html. [Accessed: 15-May-2021].

[77] American Cancer Society, “Ovarian Cancer Risk Factors,” *Cancer.org*.[Online]. Available:https://www.cancer.org/cancer/ovarian-cancer/causes-risks-prevention/risk-factors.html.
[Accessed: 17-May-2021].

[78] W3Schools, “Window localStorage Property,” *W3schools.com*. [Online]. Available: https://www.w3schools.com/jsref/prop_win_localstorage.asp. [Accessed: 20-May-2021].

[79] J. Schaaf, M. Sedlmayr, B. Sedlmayr, H.-U. Prokosch, and H. Storf, “Evaluation of a clinical decision support system for rare diseases: a qualitative study,” *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, p. 65, 2021.

[80] Z.-G. Zhou *et al.*, “An evidential reasoning based model for diagnosis of lymph node metastasis in gastric cancer,” *BMC Med. Inform. Decis. Mak.*, vol. 13, no. 1, p. 123, 2013.

[81] A. Rosenfeld, “Better Metrics for Evaluating Explainable Artificial Intelligence: Blue Sky Ideas Track,” in *Proc. of the 21th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021)*, 2021.

[82] Forecast, “What is the Difference Between Accuracy and Precision?,” *Forecast.app*. [Online]. Available: https://www.forecast.app/faqs/what-is-the-difference-between-accuracy-and-precision. [Accessed: 19-May-2021].

[83] University of Cyprus Department of Computer Science, "Gorgias -- Tutorial," *Ucy.ac.cy*. [Online]. Available: http://www.cs.ucy.ac.cy/~nkd/gorgias/tutorial.html. [Accessed: 18-May-2021].

[84] Center for Drug Evaluation and Research, "Drug interactions: What you should know," *Fda.gov*, 2020. [Online]. Available: https://www.fda.gov/drugs/resources-you-drugs/drug-interactions-what-you-should-know. [Accessed: 18-May-2021].

[85] J. Marks, "Spotting before periods: 17 causes and when to seek help," *Healthline.com*, 08-May-2018. [Online]. Available: https://www.healthline.com/health/womens-health/spotting-before-periods. [Accessed: 18-May-2021].

[86] World Health Organization, "Classification of Diseases (ICD)," *Who.int*. [Online]. Available: https://www.who.int/standards/classifications/classification-of-diseases. [Accessed: 18-May-2021].

[87] T. Benson and G. Grieve, "SNOMED CT," in *Principles of Health Interoperability*, Cham: Springer International Publishing, 2016, pp. 155–172.

[88] X. Ma *et al.*, "Understanding adversarial attacks on deep learning based medical image analysis systems," *Pattern Recognit.*, vol. 110, no. 107332, p. 107332, 2021.

[89] N. Peiffer-Smadja *et al.*, "Machine learning for clinical decision support in infectious diseases: a narrative review of current applications," *Clin. Microbiol. Infect.*, vol. 26, no. 5, pp. 584–595, 2020.

[90] G. H. Choi *et al.*, “Development of machine learning-based clinical decision support system for hepatocellular carcinoma,” *Sci. Rep.*, vol. 10, no. 1, p. 14855, 2020.

[91] P. Khumrin, A. Ryan, T. Judd, and K. Verspoor, “Diagnostic machine learning models for acute abdominal pain: Towards an e-learning tool for medical students,” *Stud. Health Technol. Inform.*, vol. 245, pp. 447–451, 2017.

[92] S. Asgari, F. Scalzo, and M. Kasprowicz, “Pattern recognition in medical decision support,” *Biomed Res. Int.*, vol. 2019, pp. 1–2, 2019.

[93] C. Gerlinger, J. Wessel, G. Kallischnigg, and J. Endrikat, “Pattern recognition in menstrual bleeding diaries by statistical cluster analysis,” *BMC Womens. Health*, vol. 9, no. 1, 2009.

Appendix A [Important Table Structures]

APPENDIX A is related to table structures that are used in this project.

<u>Section 1:</u>

Mapping of presenting complaints and referrals to their relevant Groups of Diseases, using a table:

Bleeding / Amenorrhea / Coagulation / Vascular Problems
Bleeding Gums
Brown, Watery Vaginal Discharge
Haematuria
Heavy Periods
Heavy, Profuse Bleeding
Hematochezia
Inter-Menstrual Bleeding
Leg Burning
Leg Cramping
Leg fullness
Leg Pain
Leg Swelling
Signs of Necrosis
Nose Bleeding
Oligomenorrhea
Pain, possibly related to Deep Vein Thrombosis (DVT)
Post-Coital Bleeding
Post-Operative Bleeding
Primary Amenorrhea
Prolonged Bleeding
Secondary Amenorrhea
Skin Redness possibly related to Deep vein thrombosis (DVT)
Skin Warmth possibly related to Deep vein thrombosis (DVT)
Swelling possibly related to Deep vein thrombosis (DVT)
Vaginal Bleeding

Urogynecology & Prolapses
Dark / Cloudy Urine
Dysuria
Feeling like rectum is not emptied after defecation
Feeling like sitting on a small ball
Feeling like something is falling out of the vagina
Frequent Urination
Haematuria
Heaviness / Pulling Sensation in Vagina
Lower-Abdominal Aches
Lower-Abdominal Fullness
Lower-Abdominal Heaviness
Lower-Abdominal Pressure
Lower-Back Pain
Nocturia
Pelvic Aches
Pelvic Fullness
Pelvic Heaviness
Pelvic Pressure
Pelvic Pulling
Rectum Fullness
Rectum Pressure
Smelly urine
Tissue Protruding from Vagina
Tissue Protruding from Anterior Vaginal Wall
Tissue Protruding from Posterior Vaginal Wall
Urinary Stress Incontinence
Urinary Urgency
Urination discomfort
Urination in small amounts
Vaginal Bulging

Table 32. Mapping presenting complaints and referrals to Groups of Diseases - 1

Sexually Transmitted Infections

Anogenital Warts
Vaginal Burning
Anogenital Warts
Ectopic Pregnancy
PID
Fever
Flu-Like Symptoms
Vaginal Genital Pain
History: Anogenital Warts
History:Ectopic Pregnancy
History: PID
Vaginal Itching
PCR Test: Anogenital Warts
Anogenital Warts
Unprotected Sexual Intercourse
Vaginal Discharge

Vulva Pathologies

Anogenital Warts
Vaginal Burning
Anogenital Warts
Vaginal Genital Pain
History: Anogenital Warts
Vaginal Itching
PCR Test: Anogenital Warts
Lichen Sclerosus
Anogenital Warts
Skin Discolouration
Skin Lesions
White skin discolouration

Infertility

Conception Failure
Delayed Period

Endocrinology

Abnormal Growth in Puberty
Abnormal Growth in Teenager
Acne
Conception Failure
Galactorrhea
Goiter
Hirsutism
Oligomenorrhea

Pelvic Pain

Abdominal Pain
Dysmenorrhea
Dyspareunia
Lower-Abdominal Pain
PID
Pelvic Pain
Pregnant + Any of the above

Gynaecological Cancer

Abdominal Bloating
Abdominal Mass
Abdominal Pain
Abdominal Swelling
Abdominal Tumor
Anogenital Warts
Breast Skin Abnormal Changes
Breast Skin Increased Density
Breast Skin Irritation
Breast Skin Peeling
Breast Skin Pitting
Breast Skin Redness
Breast Skin Scaling
Breast Tumor
Brown, Watery Vaginal Discharge
Cervical Smear Test: Abnormal (Cytology)
Cervical Smear Test: Abnormal (Dyskaryosis)
Dyspareunia
Galactorrhea
Gestational Trophoblastic Disease (GTD)
Heavy, Profuse Bleeding
High CA125 levels in Laboratory Test
High CEA levels in Laboratory Test
History: Anogenital Warts
Inter-Menstrual Bleeding
Lower-Abdominal Pain
Newly Invented Nipple
Pain on breast area during palpation exam
Palpation Exam shows lump in breast area
Palpation Exam shows thickened breast area
PCR Test: Anogenital Warts
Pelvic Mass
Pelvic Pain
Post-Coital Bleeding
USS: Thickened Endometrium
Vaginal Bleeding
Pregnant + Any of the above

Table 33. Mapping presenting complaints and referrals to Groups of Diseases - 2

Section 2:

Below is an example of algorithmically translating a whole Scenario Table from the group of "Sexually Transmitted Infections (STIs)" to Gorgias code, in order to understand the concept and the importance of Scenario Tables, as described in *Chapter 2 – Theoretical Model* (section "*Scenario-Based Preferences*") and in *Chapter 3 – Requirements and Implementation* (section "*Modeling the Scenario Tables in Gorgias*").

Scenarios	**Bacterial Vaginosis**	**Trichomoniasis**	**Vulva Candidiasis**	**Neisseria Gonorrhoea**	**Chlamydia**	**HSV II**	**Anogenital Warts**
Burning / Itching	✔	✔	✔	✔	✔	✔	✔
(Burning / Itching) ++ (Inter-menstrual bleeding / Post-coital bleeding)			✔	✔	✔		✔
(Burning / Itching) ++ (Inter-menstrual bleeding / Post-coital			✔	✔	✔		

bleeding) ++ Vaginal Discharge							

Table 34. Scenario Table translation

ROW 1:

```
burningANDORitching(L):- findall(X, burningANDORitching_instance(X), L),
not(listIsEmpty(L)).
burningANDORitching_instance(contextARG_burning):-burning.
burningANDORitching_instance(contextARG_itching):-itching.

rule(rName1(D,E), prefer(poss(D), notPoss(D)), []):-burningANDORitching(L), E= [L],
(
        D=bacterialVaginosis;
        D=trichomoniasis;
        D=vulvaCandidiasis;
        D=neisseriaGonorrhoeae;
        D=chlamydia;
        D=hsv;
        D=anogenitalWarts
).

rule(rName1 (D,E), prefer(notPoss(D), poss(D)),[]):-burningANDORitching(L), E= [L],
not(
        D=bacterialVaginosis;
        D=trichomoniasis;
        D=vulvaCandidiasis;
```

```
        D=neisseriaGonorrhoeae;
        D=chlamydia;
        D=hsv;
        D=anogenitalWarts
).
```

ROW 2:

```
intermenstrualBleedingANDORpostcoitalBleeding(L)
        :-findall(X, intermenstrualBleedingANDORpostcoitalBleeding_instance(X), L),
        not(listIsEmpty(L)).

intermenstrualBleedingANDORpostcoitalBleeding_instance(supportingARG_intermenstrual
Bleeding)
        :-intermenstrualBleeding.
intermenstrualBleedingANDORpostcoitalBleeding_instance(supportingARG_postcoitalBleed
ing)
        :-postcoitalBleeding.

rule(rName2(D,E), prefer(poss(D), notPoss(D)), [])
        :-burningANDORitching(L1),
        intermenstrualBleedingANDORpostcoitalBleeding(L2), E= [L1|L2],
(
        D=vulvaCandidiasis;
        D=neisseriaGonorrhoeae;
        D=chlamydia;
        D=anogenitalWarts
).
```

```
rule(rName2(D,E), prefer(notPoss(D), poss(D)),  [])
        :-burningANDORitching(L1),
        intermenstrualBleedingANDORpostcoitalBleeding(L2), E= [L1|L2],
not(
        D=vulvaCandidiasis;
        D=neisseriaGonorrhoeae;
        D=chlamydia;
        D=anogenitalWarts
).
```

%prefer this over the scenario in the previous row of the Scenario Table

```
rule(cName1(D,E), prefer(rName2(D,E), rName1(_,_)), []).
```

ROW 3:

```
rule(rName3(D,E), prefer(poss(D), notPoss(D)),  [])
:-burningANDORitching(L1), intermenstrualBleedingANDORpostcoitalBleeding(L2),
vaginalDischarge, E= [L1| [L2, strengtheningARG_vaginalDischarge]],
(
        D=vulvaCandidiasis;
        D=neisseriaGonorrhoeae;
        D=chlamydia
).

rule(rName3(D,E), prefer(notPoss(D), poss(D)), [])
:-burningANDORitching(L1), intermenstrualBleedingANDORpostcoitalBleeding(L2),
vaginalDischarge, E=[L1|[L2, strengtheningARG_vaginalDischarge]],
not(
        D=vulvaCandidiasis;
```

```
        D=neisseriaGonorrhoeae;
        D=chlamydia
).
```

%prefer this over the scenarios in all previous rows of the Scenario Table

```
rule(cName2_1(D,E), prefer(rName3(D,E), rName1(_,_)),[]).
rule(cName2_2(D,E), prefer(rName3(D,E), rName2(_,_)),[]).
```

Appendix B [Scenarios of Real-Life Applications]

APPENDIX B is related to sample runs of the system.

Appendix B contains some sample scenarios of operation and application of our GAIA system in Real-Life cases, from all 8 groups of diseases.

<u>Sexually Transmitted Infections (STIs)</u>

<u>*Scenario*</u>*: "The patient presents abnormal vaginal discharge. The doctor asks to find out more about the description of the vaginal discharge, and any other symptoms that may accompany it"*

1. *Presenting Complaints & recommended Groups of Diseases:*

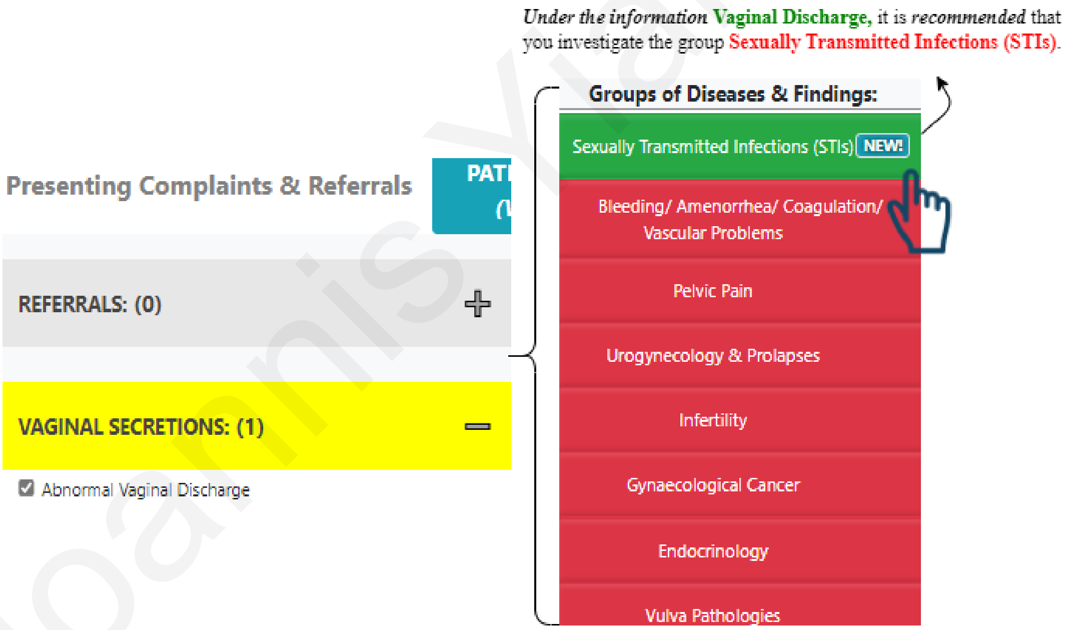


Figure 53. Scenario 1 – Sexually Transmitted Infections

2. *Prompting dialog box for describing the Vaginal Discharge*

Figure 54. Prompting for Vaginal Discharge

3. *Diagnosis based on description of vaginal discharge*

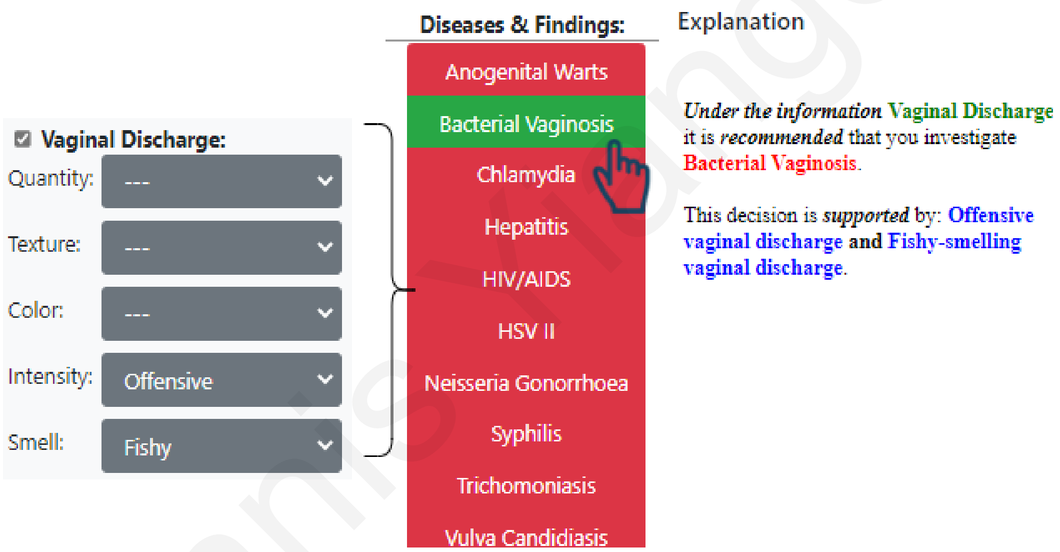


Figure 55. Diagnosis based on description of vaginal discharge

4. *Asking for further symptoms that would revise the result of the diagnostic process and alert the user about newly available groups of diseases for investigation*

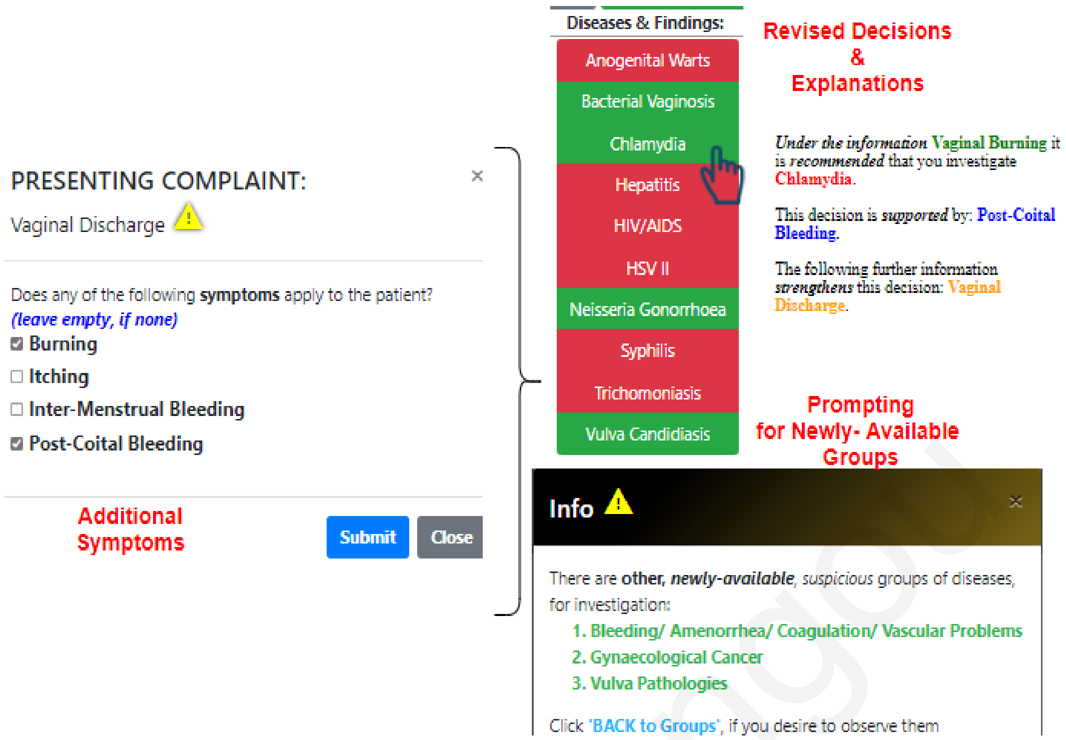


Figure 56. Newly-Available Groups prompt

<u>Bleeding / Amenorrhea / Coagulation / Vascular problems</u>

<u>Scenario</u>: *"The cases of two patients that have similar symptoms and conditions, but their different patient record information affects their individual diagnoses."*

- *Patient #1 presents Heavy Periods, Infertility and Dysmenorrhea, and is 33 years old.*
- *Patient #2 also presents Heavy Periods, Infertility and Dysmenorrhea, and is also 33 years old, but is also a Tamoxifen user (Breast Cancer medication)".*

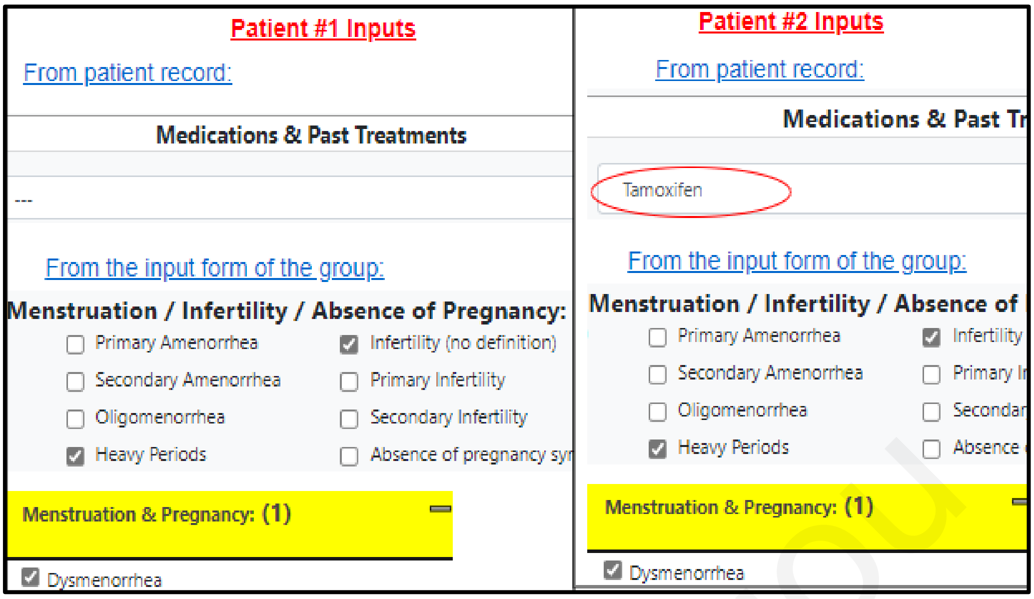


Figure 57. Different patient profiles

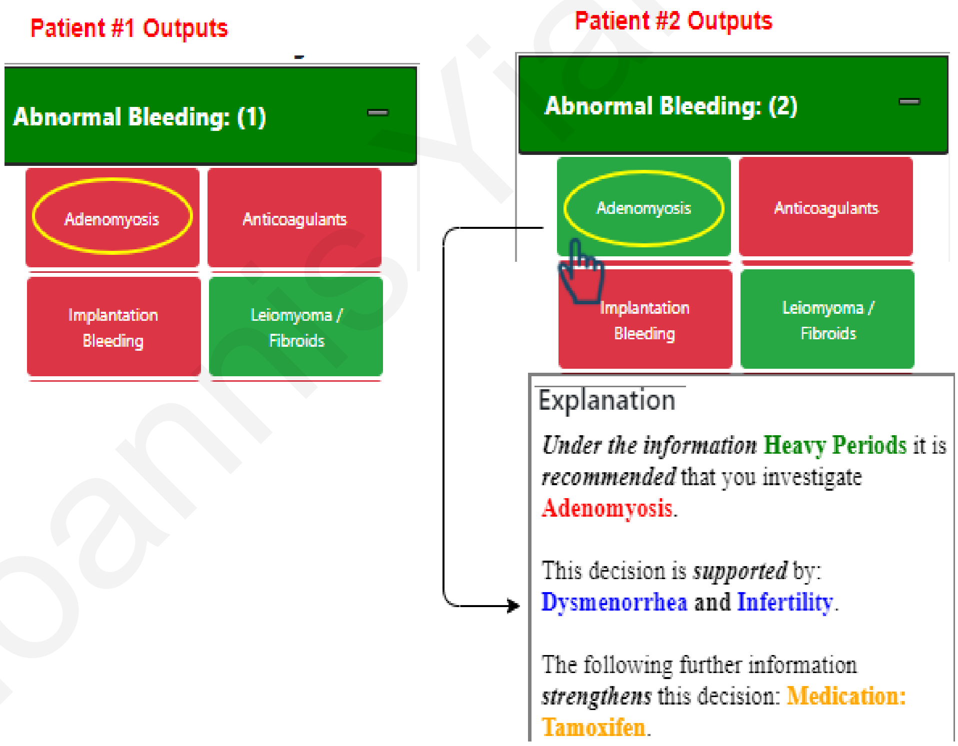


Figure 58. Different results for two different pages with same complaints

Pelvic Pain

Scenario: *"The patient describes an emergency pain. The doctor identifies additional symptoms and concludes that it is either Adhesions, Appendicitis or Diverticulitis"*

1) *After giving the presenting complaint "Pelvic Pain", the patient is asked to identify if the pain follows a specific description which corresponds to emergency cases. The patient answers yes.*

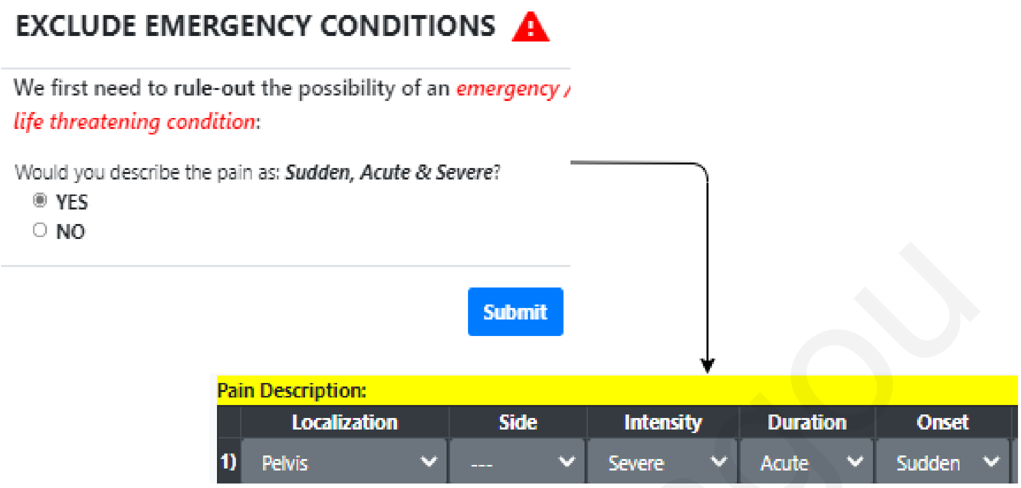


Figure 59. Emergency Pain profile

2) *The patient is asked to give further description, reporting that the pain is on the right side*

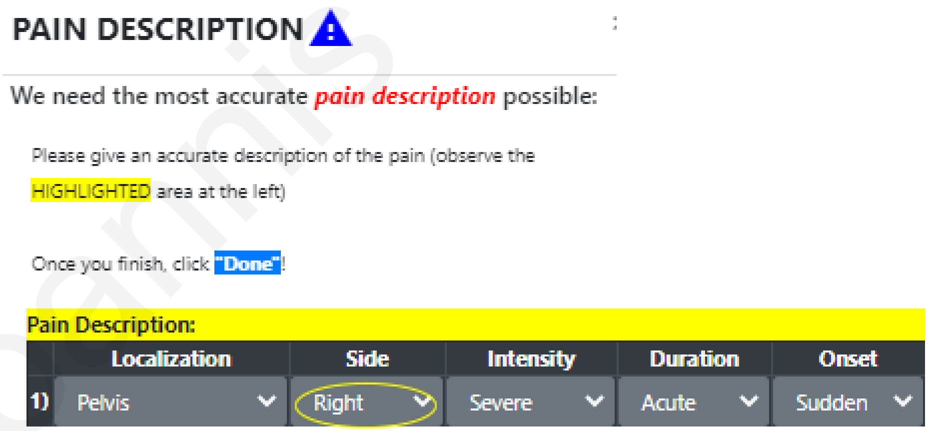


Figure 60. Further specification of Emergency pain

3) *Finally, the patient is asked to give further symptoms (other than pain) and she reports "Constipation", revealing a possible gastro-intestinal issue*

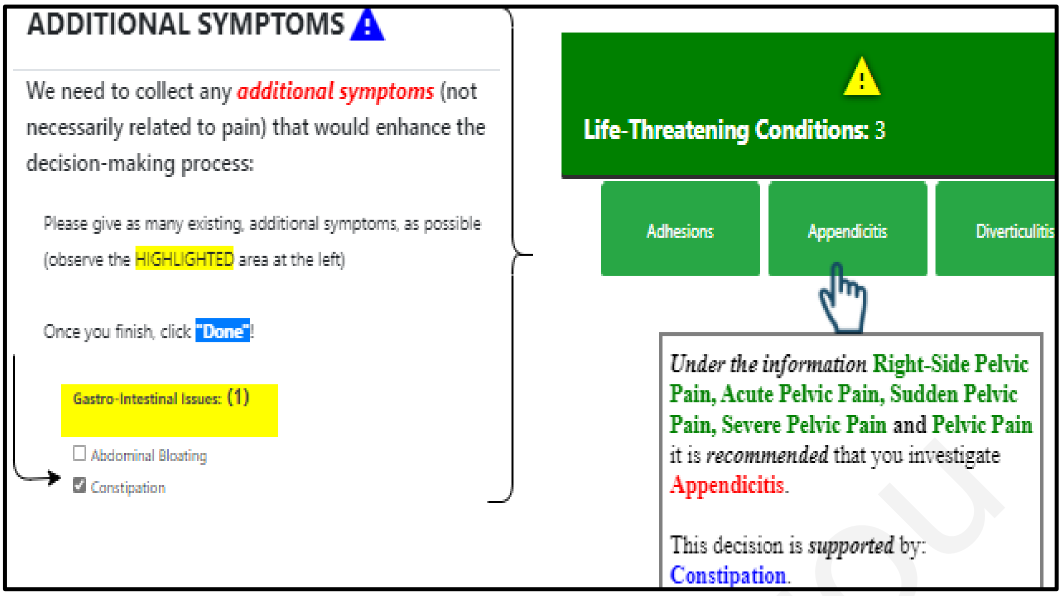


Figure 61. Additional symptoms along with Emergency Pelvic Pain

<u>**Gynaecological Cancer**</u>

<u>*Scenario*</u>*: "The cases of two patients, which both developed the same type of cancer, in a different way"*

- *Patient #1 inherited family genes that are highly risky for cancer development (familial cancer)*
- *Patient #2 developed cancer without genetic predisposition (unlike Patient #1)*

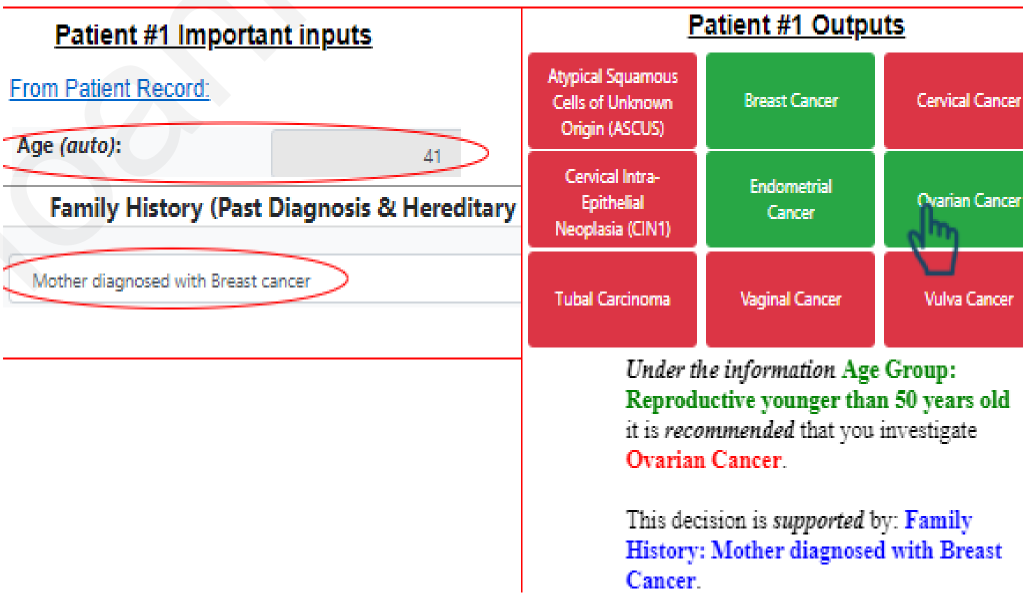


Figure 62. Familial Cancer

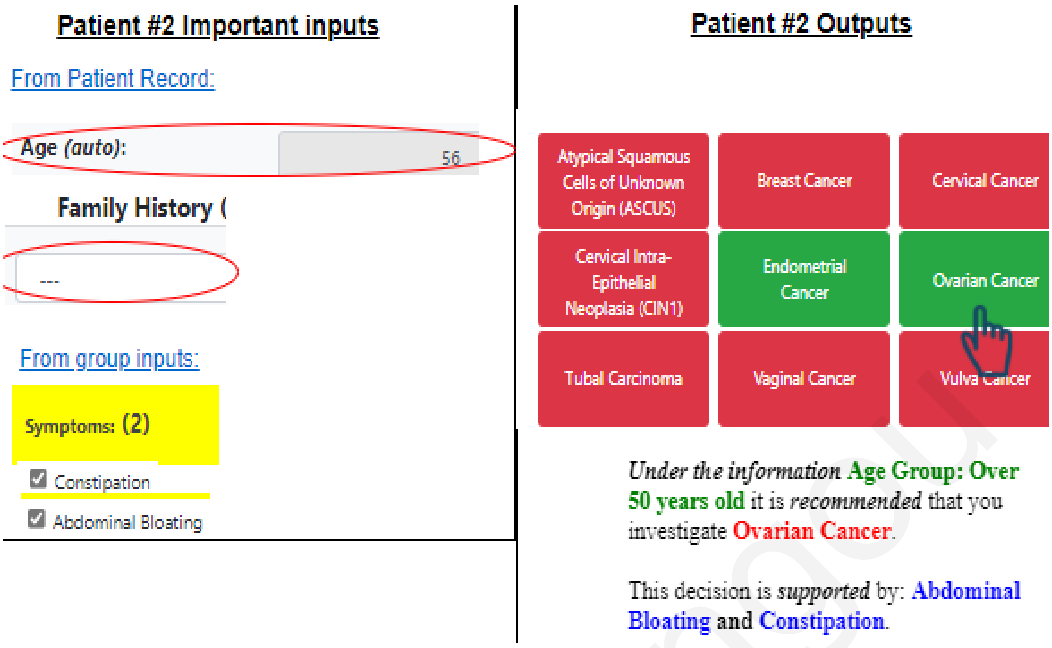


Figure 63. Cancer not related to family history

Endocrinology

Scenario: *"Hirsutism in different ages leads to different diagnoses"*

- *Patient #1 is a woman on Menopause (57 years old)*

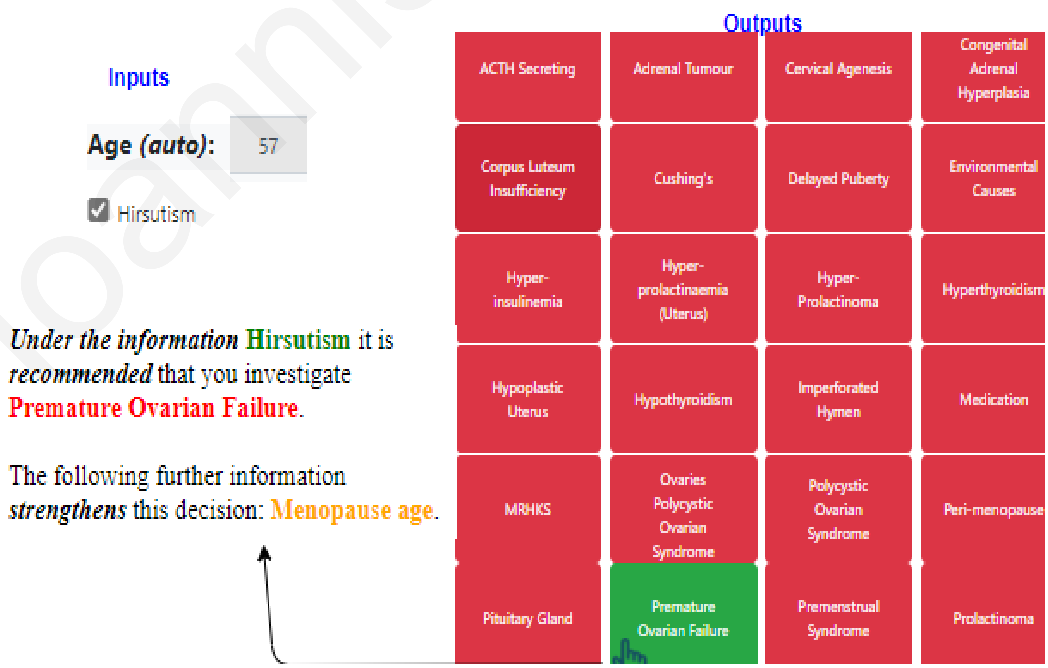


Figure 64. Hirsutism in a Menopausal woman

- *Patient #2 is a child (7 years old)*

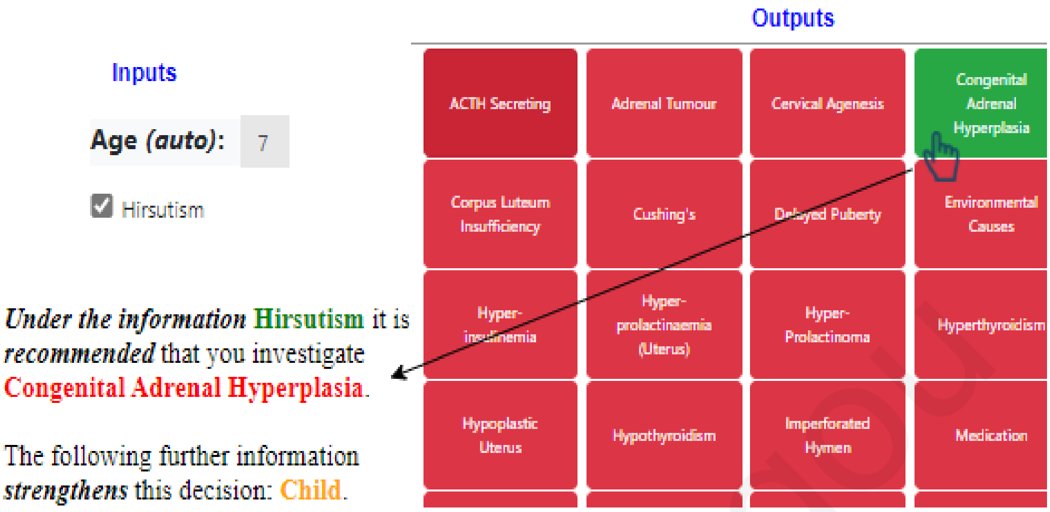


Figure 65. Hirsutism in a child

- *Patient #3 is a young adult (22 years old)*

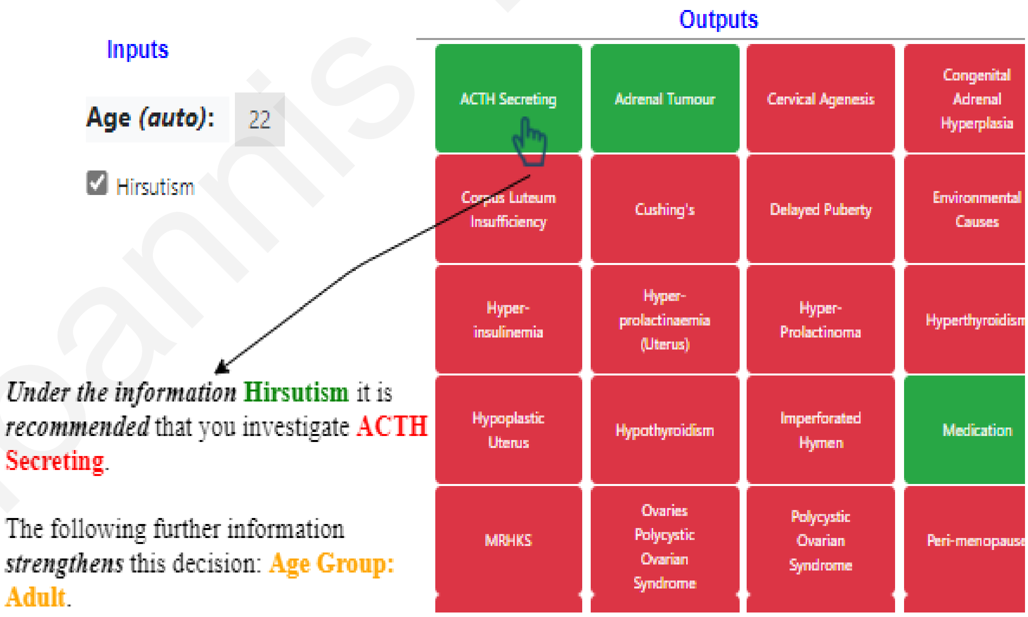


Figure 66. Hirsutism in an adult

<u>Infertility</u>

<u>Scenario</u>: *"Two women that cannot conceive child, but their infertility type differs"*

- *Patient #1 cannot conceive her second child (already given birth, at least once)*

Figure 67. Secondary Infertility scenario

- *Patient #2 never conceived a child, and neither can conceive now (never given birth before)*

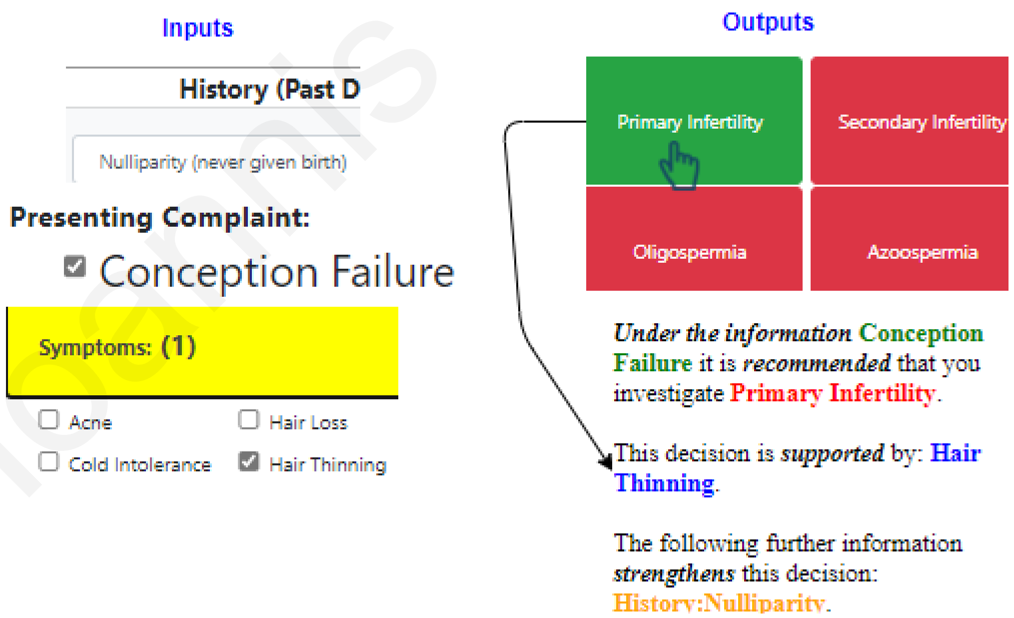


Figure 68. Primary Infertility scenario

<u>Vulva Pathologies</u>

<u>Scenario</u>: *"Itchiness in the vulvar area might lead to different diagnoses, depending on other symptoms that accompany it"*

- *Itchiness and white skin discoloration could lead to both "Vaginal Intraepithelial Neoplasia (VAIN)" and "Vulval Intraepithelial Neoplasia (VIN)"*

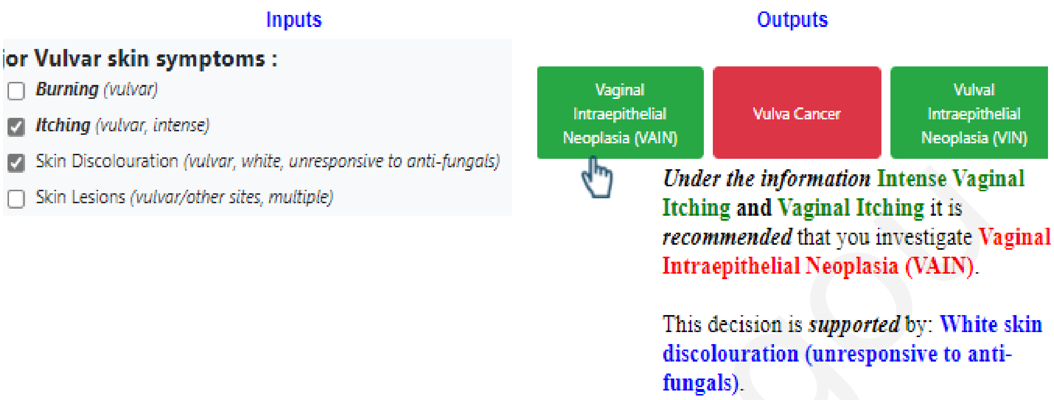


Figure 69. Vulva Pathologies and Skin Discolouration

- *Itchiness and multiple skin lesions lead to "Vulval Intraepithelial Neoplasia (VIN)" only, in women younger than 35 years old, while for women that are older than 50 years old, we only have "Vaginal Intraepithelial Neoplasia (VAIN)" (even without skin lesions)*

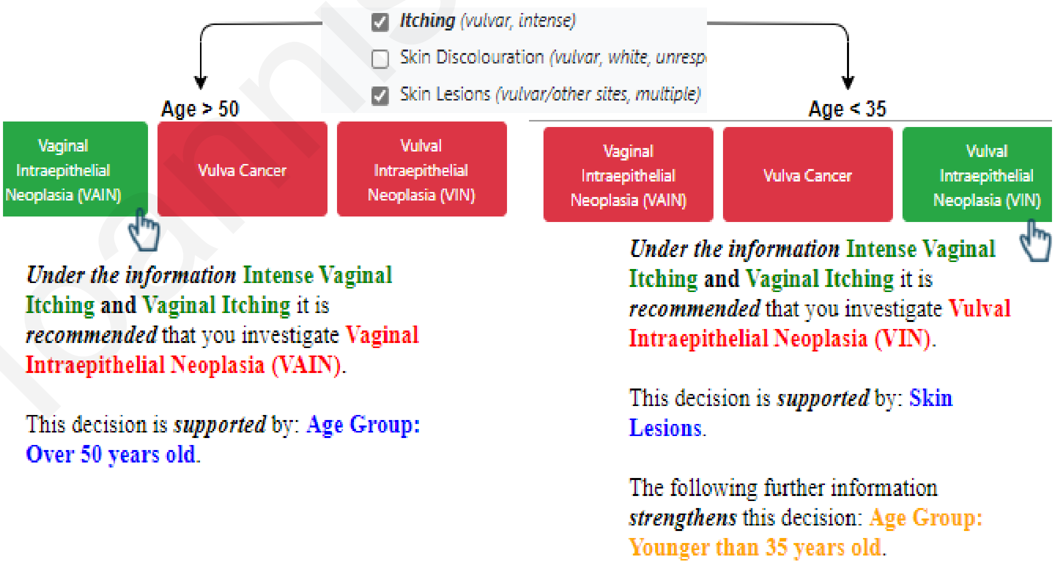


Figure 70. Vulva Pathologies, Skin Lesions and the age factor

<u>**Urogynecology & Prolapses**</u>

This group covers urological & urinary conditions which affect parts of the reproductive system too, as well as various prolapses of female reproductive organs

- <u>Scenario 1:</u> *"Patient initially reports Dysuria, but the doctor later discovers that Fever is also present, thus limiting the number of possible diseases"*

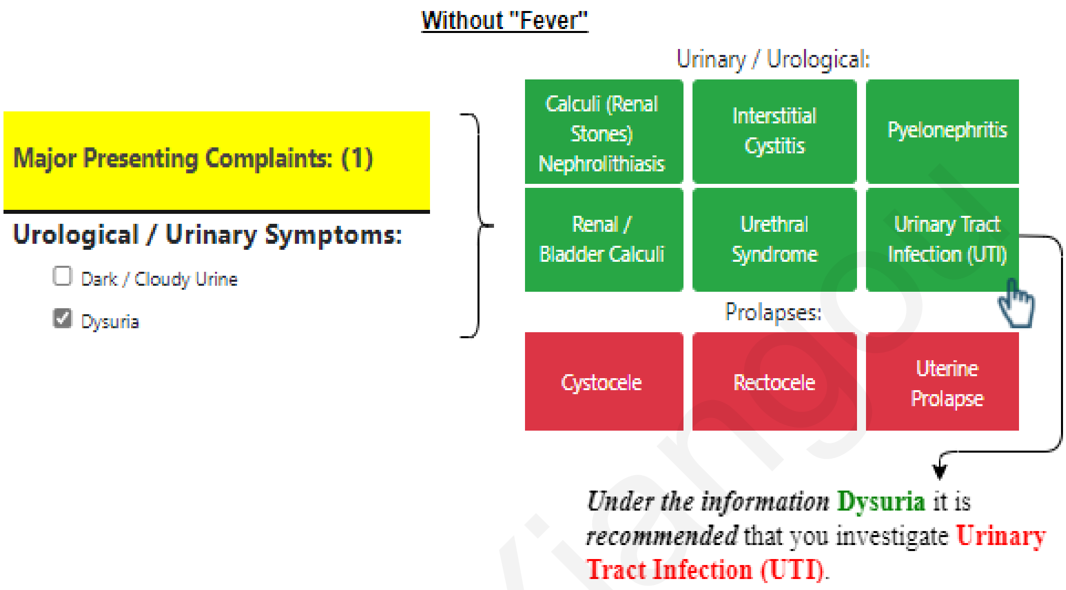


Figure 71. Urological / Urinary problem

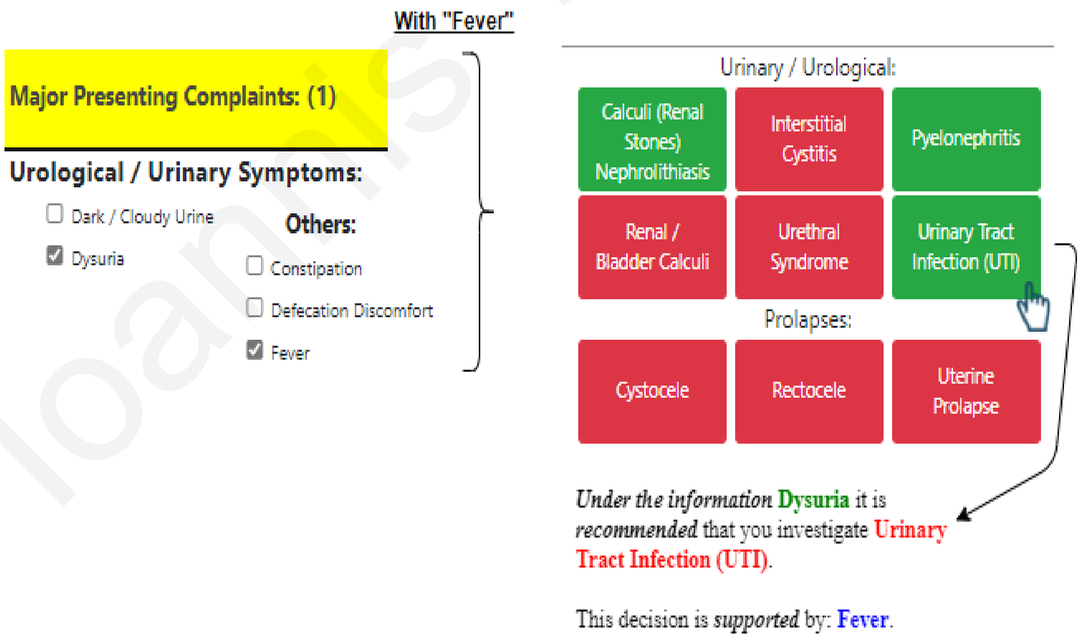


Figure 72. Urological / Urinary problem and additional symptoms

- <u>Scenario 2</u>: *"Tissue protruding from anterior vaginal wall is related to Cystocele, while tissue protruding from posterior vaginal wall is related to Rectocele"*

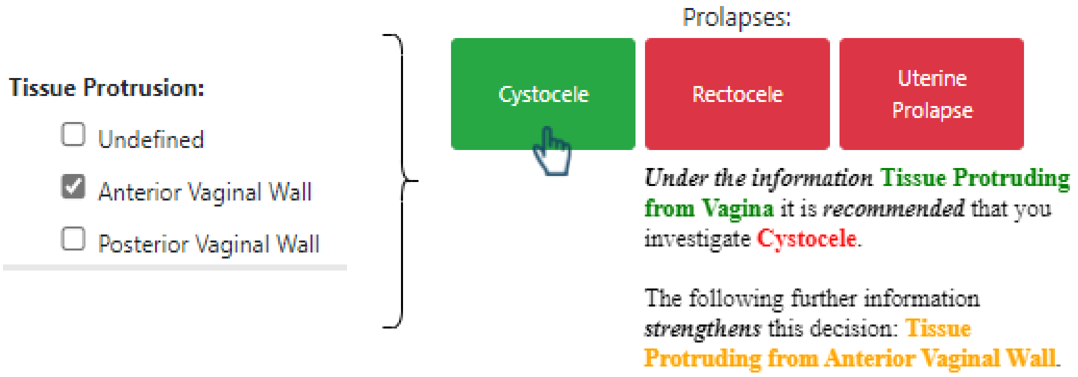


Figure 73. Prolapses - Cystocele

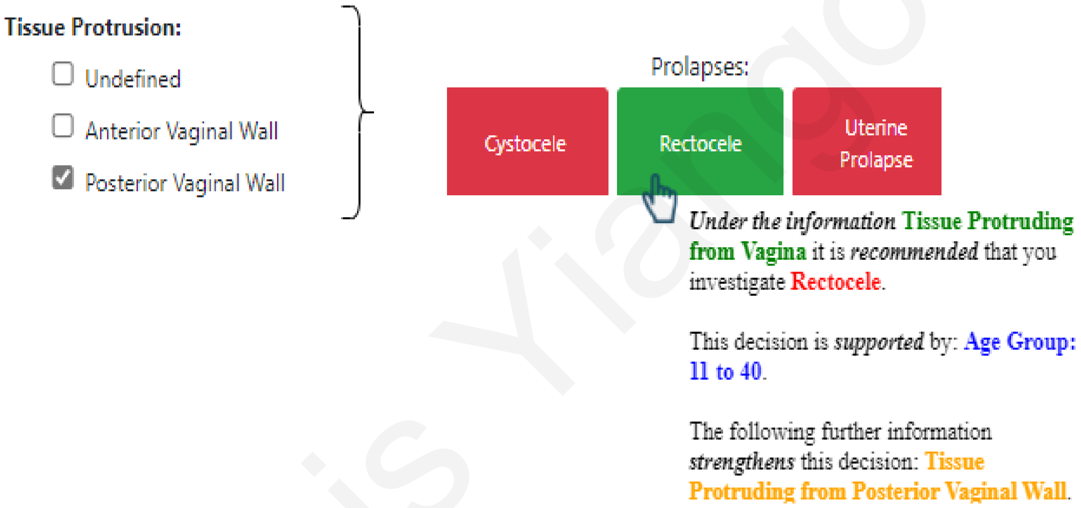


Figure 74. Prolapses - Rectocele

Appendix C [Test Case Form Structure]

APPENDIX C is related to the testing and evaluation phase of the system.

Appendix C contains file templates of test cases and other useful information related to the "Evaluation" phase of the project.

- ***File template of Test Case Form**: for Phase 1 of the three-phase Evaluation Process:*

Test ID: [IDNumber]

STEP 1.1: Known Historical Information:

Known Historical Information	
Age	
Periods (IR-Regular, Early, Delayed, Prolonged)	
Menopause Age	
Menopause Stage (pre-meno, post-meno**)**	
Medication	
OTHER Historical Information	
OTHER Historical Information	
OTHER Historical Information	

STEP 1.2: Presenting Complaint(s):

Presenting Complaint(s):
1.
2.
3.

RESPONSE 1 –Add a tick across each suspected group

GROUPS	SUSPECTED
Sexually Transmitted Infections (STIs)	
Bleeding/Coagulation/Vascular Problems	
Pelvic Pain	
Urogynecology & Prolapses	
Infertility	
Gynaecological Cancer	
Endocrinology	
Vulva Pathologies	

------------------------------------- END OF PHASE 1 --------------------------------

FOR EACH SUSPECTED GROUP WE GO THROUGH PHASE 2

Sexually Transmitted Infections (STIs)

GROUP: Sexually Transmitted Infections (STIs) ROUND 1: Description of Presenting Complaint(s) this Group		

GROUP: Sexually Transmitted Infections (STIs) ROUND 1: Suspected Possible Diseases IN THIS GROUP	

GROUP: Sexually Transmitted Infections (STIs) ROUND 2: Additional Information for this Group		
GROUP: Sexually Transmitted Infections (STIs)		

ROUND 2: Suspected Possible Diseases IN THIS GROUP	

GROUP: Sexually Transmitted Infections (STIs) ROUND 3: Additional Information for this Group		

GROUP: Sexually Transmitted Infections (STIs) ROUND 3: Suspected Possible Diseases IN THIS GROUP	

SUSPECTED NEW GROUP(S) ???	

Bleeding/Coagulation/Vascular Problems

GROUP: Bleeding/Coagulation/Vascular Problems ROUND 1: Description of Presenting Complaint(s) this Group		

GROUP: Bleeding/Coagulation/Vascular Problems ROUND 1: Suspected Possible Diseases IN THIS GROUP	

GROUP: Bleeding/Coagulation/Vascular Problems ROUND 2: Additional Information for this Group		

GROUP: Bleeding/Coagulation/Vascular Problems ROUND 2: Suspected Possible Diseases IN THIS GROUP	

GROUP: Bleeding/Coagulation/Vascular Problems ROUND 3: Additional Information for this Group		

GROUP: Bleeding/Coagulation/Vascular Problems ROUND 3: Suspected Possible Diseases IN THIS GROUP	

SUSPECTED NEW GROUP(S) if ANY	

Pelvic Pain

GROUP: Pelvic Pain ROUND 1: Description of Presenting Complaint(s) this Group		
ROUND 1: Suspected Possible Diseases IN THIS GROUP		

GROUP: Pelvic Pain ROUND 2: Additional Information for this Group		
ROUND 2: Suspected Possible Diseases IN THIS GROUP		

GROUP: Pelvic Pain ROUND 3: Additional Information for this Group		
ROUND 3: Suspected Possible Diseases IN THIS GROUP		

SUSPECTED NEW GROUP(S) if ANY	

Urogynecology & Prolapses

GROUP: Urogynecology & Prolapses ROUND 1: Description of Presenting Complaint(s) this Group		

ROUND 1: Suspected Possible Diseases IN THIS GROUP		

GROUP: Urogynecology & Prolapses **ROUND 2: Additional Information for this Group**		
ROUND 2: Suspected Possible Diseases IN THIS GROUP		

GROUP: Urogynecology & Prolapses **ROUND 3: Additional Information for this Group**		
ROUND 3: Suspected Possible Diseases IN THIS GROUP		

SUSPECTED NEW GROUP(S) if ANY	

Infertility

GROUP: Infertility

ROUND 1: Description of Presenting Complaint(s) this Group		

ROUND 1: Suspected Possible Diseases IN THIS GROUP	

GROUP: Infertility ROUND 2: Additional Information for this Group		

ROUND 2: Suspected Possible Diseases IN THIS GROUP	

GROUP: Infertility ROUND 3: Additional Information for this Group		

ROUND 3: Suspected Possible Diseases IN THIS GROUP	

SUSPECTED NEW GROUP(S) if ANY	

Gynaecological Cancer

GROUP: Gynaecological Cancer ROUND 1: Description of Presenting Complaint(s) this Group		

ROUND 1: Suspected Possible Diseases IN THIS GROUP	

GROUP: Gynaecological Cancer ROUND 2: Additional Information for this Group		

ROUND 2: Suspected Possible Diseases IN THIS GROUP	

GROUP: Gynaecological Cancer ROUND 3: Additional Information for this Group		

ROUND 3: Suspected Possible Diseases IN THIS GROUP	

SUSPECTED NEW GROUP(S) if ANY	

Endocrinology

GROUP: Endocrinology ROUND 1: Description of Presenting Complaint(s) this Group		
ROUND 1: Suspected Possible Diseases IN THIS GROUP		

GROUP: Endocrinology ROUND 2: Additional Information for this Group		
ROUND 2: Suspected Possible Diseases IN THIS GROUP		

GROUP: Endocrinology ROUND 3: Additional Information for this Group		
ROUND 3: Suspected Possible Diseases IN THIS GROUP		

SUSPECTED NEW GROUP(S) if ANY	

Vulva Pathologies

GROUP: Vulva Pathologies ROUND 1: Description of Presenting Complaint(s) this Group		
ROUND 1: Suspected Possible Diseases IN THIS GROUP		

GROUP: Vulva Pathologies ROUND 2: Additional Information for this Group		
ROUND 2: Suspected Possible Diseases IN THIS GROUP		

GROUP: Vulva Pathologies ROUND 3: Additional Information for this Group		
ROUND 3: Suspected Possible Diseases IN THIS GROUP		

SUSPECTED NEW GROUP(S) if ANY	

--------------------------------FINAL PHASE --------------------------------

FINAL SET OF SUSPECTED DISEASES:

Collect here all the possible diseases indicated above together with a short explanation, in free text, of why you suspect this.

FULL SET of SUSPECTED DISEASES in all GROUPS	EXPLANATION