Intelligent Systems with Applications 16 (2022) 200109

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.journals.elsevier.com/intelligent-systems-with-applications

Argumentation approaches for explanaible AI in medical informatics

Luciano Caroprese [*,a], Eugenio Vocaturo [a,b], Ester Zumpano [a,b]

[a] DIMES, University of Calabria, Italy
[b] CNR - Nanotec, Rende (CS), Italy

ARTICLE INFO

Keywords:
Argumentation
Explanable AI
Medical informatics

ABSTRACT

Artificial Intelligence algorithms are powerful in performing accurate predictions, but they are often considered *black boxes* as they do not provide any explanation about *how* outputs are derived from inputs or *why* a decision is taken. Therefore, urgent is the need of a completely transparent and eXplainable Artificial Intelligence (XAI) as also recognized by the explicit inclusion of *the right to explanation* in the General Data Protection Regulation (GDPR).

There has been much study on diagnosis, decision support, and interpretability, and there is significant interest in the development of Explainable AI in the realm of medicine. Interpretability in the medical field is not just an intectual curiosity, but a key factor. Medical choices impact the life of patients, and include risk and responsibility for the clinicians.

This proposal investigates the benefit of using logic approaches for eXplainable AI by evidencing how their natural characteristics of explainability and expressiveness help in the design of ethical, explainable and justified intelligent systems. More specifically, the paper focuses on a detailed topic related to the use of argumentation theory in Medical Informatics by overviewing existing approaches in the literature. The overview categorizes approaches on the basis of the specific purpose the argumentation is used for, into the following categories: *Argumentation for Medical Decision Making, Argumentation for Medical Explanations* and *Argumentation for Medical Dialogues*.

1. Introduction

Artificial Intelligence (AI) is everywhere. AI algorithms are powerful in performing accurate predictions as they are able to suggest correct decisions and reasoning in complex scenarios, but they are often considered *black boxes* because they do not provide any explanation about *how* outputs are derived from inputs, or differently said *why* a decision is taken. However, in important decisions that can impact a person's life, such as in diagnosing a disease, or in the security of a nation, such as in defense decision-making, it is critically important to fully understand the reason underlying an AI decision. To this end, the General Data Protection Regulation (GDPR), which mainly concerns the methods of data collection and management, also contains Article 22: *Automated individual decision-making, including profiling*[1]. The article refers to the use of automated processing and strongly requires the need *to obtain an explanation of the decision*[2] by evidencing the logic an automated system performs to obtain outputs from the given inputs. This topic is nicely investigated in Koshiyama et al. (2019) and the basic issue raised by the GDPR of *the right to explanation*, directly related to the use of machine learning (ML) applications, is discussed in Goodman and Flaxman (2017) together with the *right to non discrimination*.

Therefore, there is an urgent need for a completely transparent and understandable AI, eXplainable AI (XAI). An explainable and transparent AI system has to fulfill some general requirements: it has to provide decisions, suggestions, but it also has to be able to justify *how* and/or *why* the provided decisions, suggestions have been given.

A lot of research has been done in supporting diagnosis, decision support, and interpretability, and there is a growing interest in the field of medicine in the creation of XAI. Explainability is a crucial element in the medical profession, not only an intellectual curiosity. Patients' lives are impacted by medical decisions, and each diagnosis/decision includes risk, responsibility for the clinicians and provides lessons for

* Corresponding author.
E-mail address: l.caroprese@dimes.unical.it (L. Caroprese).

[1] https://gdpr-info.eu/art-22-gdpr

[2] *In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment and to challenge the decision.*

https://doi.org/10.1016/j.iswa.2022.200109
Received 21 April 2022; Received in revised form 20 July 2022; Accepted 8 August 2022
Available online 10 August 2022

future treatments. In this domain understanding why something fails, but also why things are going well and how to capitalize on them for future medical investigations, is of paramount importance.

Due to its inherent qualities of expressiveness and explainability, logic programming (LP) has a substantial impact on XAI. The use of LP leverages reasoning using various alternative semantics and allows different forms of inference, preference criteria, hard/soft constraint specification, handling of ambiguous information, modeling of incomplete and inconsistent knowledge, and declarative knowledge representation.

These characteristics ensure the satisfaction of the basic needs of XAI and candidate logic methods to serve as a key building block in the creation of explainable, justified, and ethic artificial intelligent systems.

Therefore, XAI methods can greatly benefit from logic approaches and the research community is working intensively on this relevant task. As for the specific application of argumentation this is also testified by competition involving computational models for argumentation Bistarelli et al. (2021).

Many different proposal in the recent literature reason on the concept of explainability in the medical domain Holzinger et al. (2019); Tonekaboni et al. (2019); Xie et al. (2019) and very interesting surveys have been written: see Miller (2019) for a general review on relevant papers that looks at XAI from different perspectives and Das and Rad (2020) for an overview of general challenges in XAI.

This paper is a survey on Argumentation Approaches for XAI in Medical Informatics and differently from existing surveys on the subject, that have a broad scope, pursues a different perspective: it looks at a detailed topic related to the use of argumentation framework in medical domain.

1.1. Main methods for explainable AI

In order to fulfill the requirement of a transparent and explainable AI, two main methods are pursued in eXplainable AI: *Transparent Design Methods* and *The Post-hoc explanation Methods*.

- *Transparent Design Methods*.
  Transparent Design Methods aim at making transparent *how* the solution has been achieved by looking at *how* the model works. These methods, in some sense, open the black box and reveal its inner implementation. The final achievement is to clarify any single detail that leads to the solution: the model, its structure, the value of the parameters, the training algorithms. Transparent Design Methods follow the perspective of the developer of the system and try to manage the delicate balance between the request of accuracy of the decision and the need of transparency. They answer the question: *How the solution has been obtained?*
- *The Post-hoc explanation Methods*.
  Post-hoc explanation Methods aim at providing a comprehensive explanations of the reasons for which a specific output decision has been suggested. Their effort is to give analytic insights on the motivations at the basis of the output solutions. They follow the perspective of the user of the system, do not interface with the black box implementation and answer to the question: *Why the solution has been obtained?*

1.2. Main contributions of the paper

In the following we report the main contributions of the paper.

- We give a thorough explanation of the many types of logic programming while emphasizing the inherent expressiveness and explainability of declarative approaches. These features ensure that the broad XAI criteria are met and also candidate logic methods to play a key role in the design of explainable, justified and ethical artificial intelligent systems.
- We overview the main features of argumentation as a process for producing explanations for a given claim using some basic premises and an argument that connects the premises and the claim. The paper describes the argumentation framework by Dung (1995); Dung and Son (1995) and its basic features. It also reports some discussions on the many different argumentation frameworks proposed in the literature and the most important tools of argumentation that allow developing argumentation systems for real life applications.
- We investigate a detailed topic related to the use of argumentation frameworks in the medical domain with a central focus on applications, tools and systems performing a specific task. In fact, as previously stated, the other surveys in the literature apply a huge perspective and therefore, the use of argumentation in medical informatics has not been surveyed in depth so far.
- We present an overview of ongoing research approaches connecting argumentation with medical Informatics and divide papers into three different categories on the basis of the specific purpose the argumentation is used for:
  Argumentation for Medical Decision Making, Argumentation for Medical Explanations and *Argumentation for Medical Dialogues*.
- We provide a summary of the main features of the reviewed literature papers, a discussion of the major opportunities in using argumentation approaches for XAI in medical informatics and a discussion on the major challenges that still remain to be faced.

1.3. Organization of the paper

The paper is organized as follows. Section 2 presents the state of the art of reviews related to the use of logic for explainable AI. For each survey the basic perspective and contribution will be reported. Section 3 describes the different branches of logic programming and evidences how the intrinsic characteristics of expressiveness and explainability of declarative approaches play a significant role in XAI. Section 4 recalls the basic features of argumentation frameworks, some discussion on the many different Argumentation Frameworks proposed in the literature and on the most important tools of argumentation for the development of argumentation systems for real life applications. Section 5 provides information on how relevant papers have been selected from the literature. Section 6 provides a review of the works in the literature using argumentation frameworks in medical informatics. Section 7 provides a discussion and finally, Section 8 traces the conclusions of this proposal.

2. Related surveys

Many different surveys on eXplainable AI have been proposed in the recent literature. See Miller (2019) for a general survey on relevant papers that looks at XAI from different perspectives and Das and Rad (2020) for an overview of general challenges in XAI. In this paper we are interested to reviews that investigate the connection between argumentation approaches and explainable AI. Therefore, the rest of this section reports a description of the surveys related to this relevant and challenging research topic.

Cyras e al. in Cyras et al. (2021b) present an interesting overview on eXplainable AI models using techniques and methods deriving from the computational argumentation area, called *AF-based explanations*. The paper uses the term *model* in a very general perspective to refer to a very broad variety of categories, such as planning, tools for LP, decision support and recommender systems. The survey firstly overviews the different types of argumentative explanations and classifies them into: *intrinsic* and *post-hoc*. The post-hoc explanations have been further divided into *post-hoc approximate* and *post-hoc complete* explanations. The central focus of the survey is on what the model explains and on which argumentation framework is used in order to perform the task.

Kakas and Michael in Antonis and Loizos (2020) perform a very interesting and elegant discussion on abduction and argumentation, by pointing out their specific features and the basic role they have within

eXplainable AI. More in details, the paper first analyzes the connection between abduction and argumentation as the two main forms of reasoning, then expands the proposal and reviews the state of the art of the connection of these two forms of reasoning with machine learning.

Cocarasau and Toni in Cocarascu and Toni (2016) present an overview of the approaches using argumentation for ML and categorized them on the basis of the different types of ML methods they use, the argumentation framework adopted, the form of the arguments and the advantages they allow to achieve. The paper reviews approaches of Argumentation for Supervised Learning, Unsupervised Learning, Reinforcement Learning and finally provides a systematic comparison among them.

Toja and Guan in Tjoa and Guan (2021) investigate the issues of interpretability and explainability of ML algorithms and propose an interesting survey on the different forms of interoperability, from the concept of general interoperability, to a formal mathematically interoperability, passing through many other different concepts of interoperability also related to visualization. The same categories have been investigated in the medical domain in order to promote medical education in the ML domain and favor considerations from clinicians and practitioners to ML algorithms.

Vassiliades et al. in Vassiliades et al. (2021) perform a very broad and interesting survey on the topic of Argumentation and XAI. The focus of the proposal is on discussing how argumentation can enable XAI when solving decision making problems in which argumentation can help to achieve the correct decision, justify a claim or support an interactive dialogue. The survey overviews the use of argumentation in many different domains such as Security, Semantic Web, Medical Informatics, Law Informatics and Robotics. In addition, the paper nicely discusses the natural connection of ML with argumentation, evidencing how transparent can be a ML system thanks to argumentative reasoning.

Charwat et al. in Charwat et al. (2015) survey the different techniques proposed in the literature that implement the argumentation framework. The proposal divides these techniques into two different categories: (i) *reduction-based category* that refers to those approaches that transform an argumentation problem into a different problem, e.g. a satisfiability, ASP or constraint satisfaction problem. This strategy allows to use existing systems for the computation problems in an argumentation framework; (ii) *direct category* that refers to those approaches that implement argumentation from scratch.

Each of the above reported reviews has a wide scope and therefore a huge perspective and does not provide a detailed description of the different proposal in the literature of argumentation in medical domain. Differently from the above existing surveys on the subject, that have a broad scope, the present paper pursues a different perspective: it looks at a detailed topic related to the use of argumentation framework in medical informatics. The paper provides an overview of ongoing research approaches that takes advantages from an argumentative reasoning to obtain a specific task. Anyhow, it is important to stress that, as also pointed out in each paper that refers to logic to enhance XAI, this is a very broad problem and many challenges are still open.

3. Logic and explainable AI

LP is declined in different forms. Some of them, that mostly contribute to fit the basic requirements of XAI, are reported in the following[3]:

- **Abductive Logic Programming (ALP)**: Abduction is a well known and relevant form of nonmonotonic reasoning introduced by the American philosopher Charles Sander Peirce in 1865. Abduction can be informally described as a process for producing explanations for a given observation. Abductive Logic Programming (ALP) is a relevant subarea of abduction, in which the background theory is modeled by a logic program. This last, that often allows disjunction in the heads and negation in the bodies, is evaluated using one of the standard logic programming semantics Denecker and Kakas (2002). ALP enriches normal logic programming by allowing incompleteness in a subset of predicates, called *abducible predicates*. Given a problem, its solving strategy starts from the abducible predicates and uses an inference scheme to generate potential explanations of observations. Generally, in abductive reasoning there are many different abductive explanations which are not equally compelling. Therefore, the identification of a subclass, possibly narrow, of *preferred explanations* is a relevant task. Classical abductive problems can be *observations* for which an explanation is required, but also, as in normal logic programming, goals that need to be obtained. ALP can be naturally employed to solve different real life problems in diagnosis, social science, machine learning, planning and natural language.
- **Inductive Logic Programming (ILP)**: Inductive Logic Programming is a relevant area of symbolic artificial intelligence. From positive and negative examples and background knowledge, this technique automatically learns the induced declarative theory in the form of a logic program. This, called *hypothesis*, entails all the positive examples and does not entail any of the negative examples. The hypothesis constitutes an explanation of the given example in the background knowledge Muggleton and de Raedt (1994).
- **Fuzzy Logic (FL)**: Fuzzy logic is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is approximate rather than precise. Fuzzy logic F. Baldwin (1981); Zadeh (1965) attempts to quantify the degree of truth of propositions and consequences. Unlike standard first order logic (FOL) where either a proposition is *true* or *false*, in fuzzy logic, a proposition may have a truth value that ranges between 0 and 1, where 0 states for *false* and 1 states for *true*. Fuzzy logic easily represents, manipulates and interprets vague and uncertain information and therefore can be naturally employed in XAI. It allows to model human behaviors in thinking and making decisions under uncertain/imprecise conditions for solving real problems.
- **Probabilistic Logic (PL)**: Probabilistic Logic deals with uncertain scenarios by combining both logic and probability. Many different types and degrees of uncertainty can be modeled resulting in a variety of rich extensions of traditional logic with probabilistic features. PL can be effectively used in many different application areas and, combined with argumentation, is in charge of offering a flexible declarative framework to express uncertainty on claims, arguments and premises.

LP exhibits intrinsic characteristics of expressiveness and explainability. It allows to represent knowledge in a declarative and intuitive form, to provide different forms of inference, to specify preference criteria and hard and soft constraints, to deal with vague information, to model incomplete and inconsistent knowledge and to perform reasoning using different alternative semantics. These features naturally guarantee the satisfaction of the general requirements of XAI and candidate logic approaches to be a fundamental component in the design of explainable, justified and ethical artificial intelligent systems. The research literature in this context is really huge if we consider the connection between the different forms of logic and the explainability. Nevertheless, the goal of this paper is limited to investigate the use of argumentation theory in existing systems of Medical Informatics.

4. Argumentation frameworks and tools

The idea behind an *Argumentation Framework* (*AF*) Dung (1995) is that given a set of arguments, where some arguments attack others, we want to find the arguments that can ultimately be accepted. To determine whether an argument can be accepted or not, it is not sufficient to

[3] A formal presentation of the Argumentation Frameworks is reported in the next section

look at its defeaters because they could be defeated by other arguments.

More formally, an AF is a pair $\langle \mathscr{A}, attacks \rangle$ where:

- $\mathscr{A}$ is a set of elements, called *arguments*;
- $attacks \subseteq \mathscr{A} \times \mathscr{A}$ is a binary relation over $\mathscr{A}$. Given the arguments $x, y \in \mathscr{A}$, we say x *attacks* y if $(x,y) \in attacks$.

Given a set $X \subseteq \mathscr{A}$ of arguments and an argument $y \in \mathscr{A}$, X *attacks* y if there exists $x \in X$ s.t. $(x, y) \in attacks$, while X *defends* y if X attacks each argument $z \in \mathscr{A}$ s.t. $(z, y) \in attacks$. Given the sets $X, Y \subseteq \mathscr{A}$ of arguments, X *attacks* Y if there exists $x \in X$ and $y \in Y$ s.t. $(x,y) \in attacks$.

A basic question is whether *a given set of arguments can be collectively accepted* or, analogously, *what are the arguments that are not attacked by any other argument*.

Such a question, especially when considering AFs modeling real life complex scenarios, is far from being easy. As an example, look at the simple AF reported in Fig. 2.

We know that a patient with *retinopathy* often has *diabetes* (x). We also know that a patient most likely does not have diabetes if he/she has low blood *glucose* levels (y). The patient may have low blood glucose levels due to taking a particular *drug* whose side effect is to lower blood sugar levels (z). We observe that the patient has retinopathy and has low blood glucose levels due to a particular drug, whose side effect is to lower blood glucose levels.

We want to derive the acceptable arguments. In particular, we want to know if the patient is more likely to have diabetes or not.

In this case, at a first look, it seems that y is a reason against the argument x (the patient cannot have diabetes because of his/her low blood glucose level).

Anyhow y *is defeated by* z (the patient has a low blood glucose level because he/she has taken a drug that lowers blood glucose levels), that in turn has no counterarguments. Therefore, in this situation z *is accepted*, y *is defeated* and, as y is not anymore a counterargument for x, x *is accepted*. The conclusion is that the patient most likely has diabetes.

In general, more than one argument can be acceptable and a set of acceptable arguments is called *extension*. Given an AF $\langle \mathscr{A}, attacks \rangle$, a subset of arguments $X \subseteq \mathscr{A}$ is *conflict-free* if it does not contain any argument x that attacks another argument y in X. Referring to the seminal work in Dung (1995), an extension $X \subseteq \mathscr{A}$ is said to be:

- *admissible* iff X is conflict-free and X attacks each set Y of arguments that attacks it;
- *complete* iff X is admissible and all the arguments that X defends belong to X;
- *grounded* iff X is minimally complete (w.r.t. set inclusion);
- *preferred*, iff X is maximally admissible (w.r.t. set inclusion).

Many different Argumentation Frameworks have been proposed in the literature based on different visions of *arguments* and *semantics*. Many of them stem from the basic framework in Dung (1995) whose main concepts (*argument, extension, type of extensions*) have been reported above. This first category of approaches include: the *Bipolar Argumentation Framework (BAF)* Cayrol and Lagasquie-Schiex (2005), the *Label Based Argumentation Framework (LBAF)* Caminada (2008), the *Structured Argumentation Framework (SAF)* Dung (2016), the *Quantitative Bipolar Argumentation Framework (QBAF)*Baroni et al. (2018), and the *Probabilistic Bipolar Argumentation Framework (PBAF)* Baumeister et al. (2021); Fazzinga et al. (2018). Some other approaches propose a different concept for arguments that become *structured arguments* and are obtained using a deductive process Bondarenko et al. (1997); García et al. (2013); García and Simari (2004); Modgil and Caminada (2009). A new *labelling semantics*, i.e. a semantics assigning a truth value to each argument, for Weighted Argumentation Frameworks (WAFs) has been recently proposed in Bistarelli and Santini (2021); Bistarelli and Taticchi (2021) together with an interesting discussion on the definition of strong admissibility for WAFs. Additional interesting papers are: Caminada et al. (2015) that reasons on the equivalence between logic programming semantics and argumentation semantics, Cayrol and Lagasquie-Schiex (2020) that proposes a logical encoding of argumentation frameworks with higher-order attacks, and Alfano et al. (2020) that explores the relationships between argumentation based frameworks and Partial Stable Models.

An exhaustive treatment of these various frameworks is out of the scope of the paper. We refer the reader to the works Cyras et al. (2021b); Vassiliades et al. (2021) for a detailed overview of the Abductive Frameworks in the literature.

4.1. Argumentation tools for developing argumentation systems

This section reports the main tools of argumentation. They are used for a fast development of argumentation systems related to real life domains. These systems are in charge of solving decision problems in many different environments also ensuring the management of uncertain and incomplete information.

CaSAPI is a system presented in Gaertner and Toni (2007) that implements a generalization of the argumentation frameworks presented in Dung et al. (2006a, 2006b, 2006c). In the CaSAPI system different reasoning semantics are implemented: a brave (credulous) semantics and two different forms of cautious (skeptical) semantics. More specifically, the three different mechanisms for computing arguments are: the *AB-dispute derivation* for computing the brave admissible semantics in Dung et al. (2006a,b), the GB-dispute derivations for computing the brave grounded semantics in Dung et al. (2006b) and the *IB-dispute derivation* for computing the cautious ideal semantic Dung et al. (2006b, c). Therefore, given a claim, this can be credulously or sceptically entailed by the argumentation framework.

Gorgias (*http://www.cs.ucy.ac.cy/~nkd/gorgias/*) Kakas et al. (2019) is a preference based argumentation framework in which arguments are obtained using a basic argument scheme of Modus Ponens that connects the premises to the claim. In Gorgias an argument A consists of a set of argument rules, where an argument rule links a set of premises, that are generally facts, with the claim. Generally, it is said that an argument rule supports the claim. Given an argument A, the application of its constituent argument rules allow to obtain several claims. The final derived claim is said *supported by A*. In Gorgias, the syntax of argument rules is that of Extended Logic Programming and the conclusion of an argument rule can be either a positive atomic statement or a negative one. GorgiasB[4], is a tool, developed on the top of Gorgias, that allows domain expert to develop applications of argumentation in a systematic way, with really little knowledge of argumentation theory. The notable advantage of using Gorgias is the possibility to describe the application as a set of object level arguments and fix priority argument rules. Gorgias is open source and has been used to implement applications in many different domains including medical support systems, ambient intelligence, security and cognitive personal assistants[5].

DeLP (Defeasible Logic Programming) García and Simari (2004) is an argumentation formalism that combines Logic Programming with Defeasible Argumentation. DeLP declaratively represents information also using weak rules and proposes an argumentation inference mechanism that guarantees to treat inconsistent knowledge. An interpreter of DeLP[6] has been implemented in Prolog and an abstract machine, JAM (Justification Abstract Machine) Garcia and Alejandro (2000) has been designed for DeLP. DeLP allows to model real life applications presenting incomplete and inconsistent information and in addition the use of defeasible argumentation allows to build dynamic domains in which knowledge may change. An interesting real application for stock market has been presented in Garcia and Alejandro (2000).

[4] http://gorgiasb.tuc.gr
[5] http://gorgiasb.tuc.gr/Apps.html
[6] http://cs.uns.edu.ar/ajg/DeLP.html

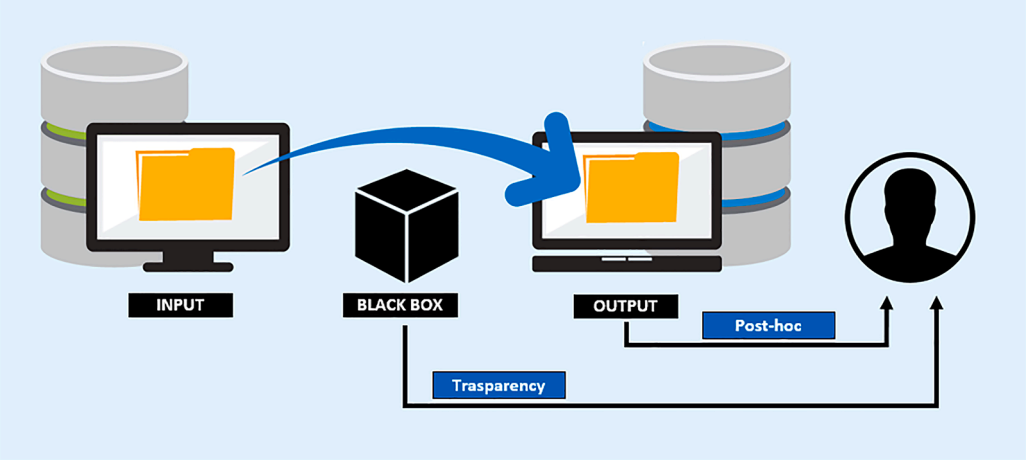

Fig. 1. Methods for XAI.

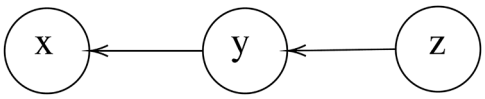

Fig. 2. A simple Argumentation Framework.

TOAST (The Online Argument Structures Tool) Snaith and Reed (2012) system is an implementation of the ASPIC$^+$ theory Prakken (2010) of structured argumentation. TOAST is implemented in Java, and the argument evaluation is performed by a web service Snaith et al. (2010). The TOAST system can be used through a web form[7] that allows to construct an argumentation system and theory by providing a knowledge base, a set of rules and a set of preferences and is deployed as a web service.

5. Methods and materials

We performed a systematic literature analysis to categorize relevant research. Our research and revising procedure for literature includes the following steps:

1. Defining the research questions to be addressed and the related queries to submit on most relevant academic article databases.
2. Establishing criteria for including publications from the screening process.
3. Revisioning the selected publications.
4. Establishing a taxonomy to assist in the classification of publications.

In the following we report the research questions at the basis of the proposal:

- *RQ1*: In the field of healthcare, which argumentation approaches have been used?
- *RQ2*: How can argumentation approaches be classified with respect to medical informatics?

To find the most targeted research and examine the quantity of potentially relevant papers, we created the following query:

```
((XAI AND logic) OR (XAI AND argumentation)) AND (healthcare OR medical)
```

The query has been submitted to PubMed, dblp, Google Scholar, Scopus, Web of Science, IEEE Xplore, ACM Digital Library, SpringerLink, and ScienceDirect. The papers obtained were the result of the aforementioned query, which was evaluated with respect to the title, the keywords and the abstract.

We used the following inclusion criteria to choose which publication to include in our review:

- Articles in which argumentation has been used in the healthcare field.
- Articles written in English.
- Articles published in relevant journals and conferences.

Next section provides a review of the selected papers.

6. Argumentation for medical informatics

The use of argumentation in the healthcare domain has been proved to be appealing and useful to manage inconsistent, incomplete, complex and heterogeneous information.

Example 1. Jane is a 22-year-old female who had just suffered of an ischemia. She is alcoholic and depressed and sufferes of hypertension. Several treatment options can be consider, and the choice is dependent on the priorities of the clinician and patient, which may not be aligned. Jane could prefer drugs over a change of her lifestyle or viceversa, but the clinician could suggest firstly a change in lifestyle over drug prescription. It could be the case that besides preferences, constraints should be managed: suppose the patient has co-morbidities that forbid to assume some specific drugs.

In the described scenario, argumentation can be used to:

- *recommend* a possible treatment by considering patient and clinicians preferences and constraints (*Medical Decision Making*);
- *reason* on a specific prescription in order to justify/strengthen it (*Medical Explanations*);
- *support* an argumentative conversation among parties in order to debate on the validity of a statement (*Medical Dialogues*).

The rest of this section describes different works in the literature that use argumentation in Medical Informatics to support medical decision making, to explain a claim/decision or to support an argumentation dialogue. In more details works in the literature are divided in three different categories, based on the different specific goal for which the argumentation theory is used:

- *Argumentation for Medical Decision Making*,
- *Argumentation for Medical Explanations*,
- *Argumentation for Medical Dialogues*.

[7] http://toast.arg-tech.org/

6.1. Argumentation for medical decision making

Achilleo et al. in Achilleos et al. (2020) use an abductive framework to aid in the diagnosis of Alzheimer's disease, a form of dementia that involves memory, thinking and behavior and affects millions of people around the world. The paper contributes to the assessment of the disease and evaluates changes in brain structure as evidenced in Magnetic Resonance Imaging (MRI). The Hippocampus Volume is the most important feature in the classification of the disease. The brain MRI images are acquired by the Alzheimer's Disease Neuroimaging Initiative (ADNI)[8] and the final dataset consists of 144 MRI images of normal cases and 69 MRI images of subjects presenting the pathology. In the proposal, decision trees (DT) and random forests (RF) algorithms are used to extract a set of rules starting from the discretized MRI images. The output of the DT and RF models constitutes the input of the argumentation module which has the final objective to propose the final learned model. This has been implemented by using the *preference based argumentation component GorgiasB*. The proposal, in its first step, classifies individuals in *healthy*(P) and *diagnosedAlz*(P). In the second step, option predicates, that essentially correspond to the features, are defined. As for the third step it consists in the authoring of the arguments (rules); more specifically, each argument consists of a set of option predicates, restriction on variables and the claim of *healthy*(P) and *diagnosedAlz*(P). The final step is performed manually and consists in specifying the rules that have to be included in the argumentation theory ad their associate priority. Priority of a rule over another can be set either under some specific conditions or unconditionally. The proposed argumentation model showed improvements compared with models of DT and RF and achieved an average accuracy of 91%.

Kokciyan et al. in Kökciyan et al. (2021) propose an interesting metalevel argumentation framework (MAF) - called CONSULT - as a support for medical decision processes. The CONSULT system is developed with the final aim of helping stroke patients in taking care of their conditions and adhere to treatment plans. CONSULT manages both static knowledge, consisting of a set of facts and a set of rules stored in a knowledge base, and dynamic data obtained from wellness sensors and reporting health parameters on vital signs of the patients, such as blood pressure and heart rate. The system is therefore able to integrate and correctly manage information provided by heterogeneous information sources. Information stored in the knowledge base (KB), derived from patient medical records, is modeled using first order logic whose expressive power is sufficient to satisfy recommendations provided by the hypertension guidelines. More specifically, arguments, attacks and explanations schemes are represented as sentences (including variables) instantiated over the KB. Scheme rules allow to automatically construct object-level arguments and a meta engine is in charge of translating object-level arguments and attacks into a metalevel argumentation frameworks (MAF). In addition, CONSULT also allows to specify preferences that can be embedded in the KB if a priori knowledge on priority statements exists or can be interactively specified to the system. More in details, different preferences over possible treatments can be specified: some of them can be specified by the specialist, whereas some others can be specified by the patient himself and it is possible to set a priority among them Kokciyan et al. (2018a). The obtained argumentation framework can be used with an answer set programming (ASP) solver, such as DLV[9], in order to retrieve the justified arguments of the MAF. These arguments are processed by an *Explanation Generation* component that *translates* each justified argument in the form of a textual explanation that can be provided to the patient. The CONSULT system has been implemented as a mobile application running on Android, available in two different versions: i) *dashboard* or ii) *dashboard plus a chatbot*. A pilot study involving six healthy volunteers for seven days demonstrated the usability and acceptability of the system. The CONSULT system has been extended in Chapman et al. (2019); Kokciyan et al. (2019) to deal with chronic disease.

Longo et al. in Longo et al. (2012) apply argumentation theory to the breast cancer recurrence problem, i.e. the recurrence of the disease after breast cancer surgery. The problem is of paramount importance and is related to the ability of examining unstructured information related to relevant features, such as the dimension of the tumor, the menopausal state of the patient, the type of the tumor, and the past experiences of similar cases in order to provide a prediction. By considering specific features, the expertise of the clinicians and the past similar cases it is possible to model clinical evidences (e.g. in the presence of a small-sized cancer, the possibility of a recurrence of the cancer is low, or in the presence of a significant number of involved nodes in the breast cancer area, the possibility of a recurrence of the cancer is high). The paper proposes an argumentation framework that models clinical evidence as knowledge-based arguments, represents defeat relations among arguments and finally applies argumentation semantics in order to predict the recurrence of the disease and justify the outcome. An interesting feature of the proposed framework is the idea of performing the reasoning task by modeling uncertain and vague knowledge by considering fuzzy sets and membership functions. The paper uses the Ljubljana breast cancer dataset UKWorkingGrouponthePrimaryPreventionofBreastCancer (2005), consisting of 286 instances of women that experienced breast cancer surgery. As for the semantics, in order to enhance the medical decision process, the strong suggestion is to reason using a skeptical (or cautious) semantics and adopt a brave reasoning, combined with some preference strategy, only in the case the cautious semantics cannot be used. Results obtained by the proposed argumentation framework for breast cancer recurrence have been defined *encouraging* by the authors of the papers as the predictive capacity of the proposed framework is similar to the one obtained using well known machine learning tools.

Williams and Williamson in Williams and Williamson (2006) model the breast cancer prognosis in a framework combining logic and probability. More specifically, two different components interact in the proposal: a standard argumentation theory is used to model knowledge related to the disease and a standard probabilistic formalism, in the form of a Bayesian network, is used to capture the probabilistic relations among arguments. The Bayesian network component is used in the proposal to provide a prognosis, whereas the argumentation component is in charge of helping in justifying and explaining the prognosis itself.

Spanoudakis et al. in Spanoudakis et al. (2017) investigate a relevant problem in the health domain related to the medical data access, referring to the regulations established by the European Union (EU) and those of the Cyprus member state. An argumentation framework is used in order to model such a set of regulations. This interesting and important real life application of compliance of access, with respect to specific EU and Cyprus rules, is modeled as a decision problem. In the presence of different arguments, supporting different levels of access, grant access is allowed by the stronger arguments. The legislation for medical data access is modeled by using the GorgiasB tool that, once the argumentation theory is ready, allows to automatically generate the Gorgias Prolog source code. In the specific context of data sharing in the medical field, the proposed argumentation framework is a valuable solutions as it allows an high declarative representation of the regulation policies. It does not require an analysis of the policies and a strategy to solve and fix conflicts and, in addition, the system can be easily modified if changes occur in regulation policies.

Glasspool et al. in Glasspool et al. (2006) describe a software application, based on argumentation logic, called REACT (Risks, Events, Actions and their Consequences over Time) which supports clinicians and patients involved in medical planning. REACT allows to operate on a single care plan and provides a summary of the alternatives (*what if*) when the user inserts or removes an event from the planning chart. Therefore, the tool allows a real time feedback and helps in deciding the

[8] http://adni.loni.usc.edu/
[9] https://www.dlvsystem.it/dlvsite/

right care plan.

Cyras et al. in Cyras et al. (2018); Cyras and Oliveira (2019); Cyras et al. (2021a) present an interesting work that uses an argumentation formalism to reason in the presence of conflicting clinical guidelines. The proposal is implemented into a system that provides recommendations by also examining specific information and preferences. The proposal allows to obtain personalized recommendations by combining patient's information and clinical guidelines. Conflicts within guidelines, if present, are managed using computational argumentation techniques ensuring the satisfaction of specified preferences. The results, provided by the system, are explainable and conflict-free recommendations. ABA$^+$Cyras and Toni (2016) is the structured argumentation formalism used in the proposal.

6.2. Argumentation for medical explanations

Shankar et al. in Shankar et al. (2006) propose an explanatory framework, called WOZ, that provides justification to the outputs of a clinical decision-support system. More in details, WOZ is a multi-client framework that is part of the knowledge base system EON Musen et al. (1996), a collection of software components and clinical models that may be used to support clinical decision-making in the implementation of guidelines-based care. The WOZ explanation framework is used on the ATHENA DSS, a decision-support system for managing primary hypertension Goldstein et al. (2000) based on the EON architecture. ATHENA DSS analyzes the hypertension guideline model using individual patient data and generates patient-specific clinical care recommendations, such as adding a specific medication to the patient's regimen. The WOZ explanatory framework on top of ATHENA DSS provides arguments to support and justify ATHENA DSS conclusions to the user. The problem-solving components of WOZ, like in EON, use explicit models of clinical procedures and medical domain knowledge to abstract the strategy used for explanations. It defines the information that WOZ uses to justify a claim and follows the Toulmin's argument structure Toulmin (1958) in order to organize medical evidences supporting a diagnosis or a treatment plan. The strategy used to derive an explanation in WOZ is therefore conceived in order to include more information than just the guidelines that led to the claim (e.g. the links to additional resources). In addition, the strategy is arranged and presented in the form of an argument structure whose different levels can be easily navigated by the user so that understanding relevant patient data and drug recommendations, a summary of the main reasons supporting the claim and all the links to relevant resources that help in providing a justification.

In Grando et al. (2013), Grando et al. propose an extension of EIRA (Explaining, Inferencing, and Reasoning about Anomalies) Moss (2010), an existing knowledge-based system that has been successful in detecting anomalous patient responses to treatments in the Intensive Care Unit (ICU), but is unable to explain to clinicians the rationales behind the anomalous detections. While EIRA has proven to be extremely accurate, it lacks a justification system that could make explicit, in a user-friendly manner, the complex reasoning behind the algorithms. Then they propose arguEIRA, an extension of EIRA that replaces the most complex algorithms, used to generate potential explanations for anomalous patient responses to ICU treatments, with an argumentation-based decision system.

Tobias Mayer, in Mayer (2020) concentrates on the definition of an argument mining system for extracting arguments from clinical trials, addressing the following issues: (i) How it's possible to discriminate, in natural language clinical trials, between argumentative and non-argumentative components? (ii) How should the identified argumentation components be classified as *evidence* or *claims*? Clinical trials are documents written in natural language that compare the relative advantages of different therapies. As a result, there is a compelling need to research approaches for extracting structured data from unstructured text in order to enable argument-based decision-making frameworks. The authors provide an automated method for extracting argumentation information from Randomized Clinical Trials and a claim detection technique. The proposed argument mining system, in particular, identifies argumentative segments, and then classifies the detected segments into *evidences* (i.e., the argument's premises) and *claims* (i.e., the conclusions of the argument). The authors rely on the pre-existing system MARGOT Schuster and Manning (2016) and propose an enhanced version of MARGOT that recognizes evidences and claims in clinical data. The proposal uses PubMed Randomized Clinical Trials abstracts related to four distinct diseases, including glaucoma, diabetes, hepatitis B, and hypertension. Starting from these data, an annotated dataset containing 976 argument components, classified into 697 evidences and 279 claims has been extracted. Two different datasets were used to train two binary classifiers: one for claim detection and the other one for evidence detection. Experiments were performed with three different classifiers: (i) SSTK exploiting constituency parse trees (ii) SVM with BoW features weighted by TF-IDF (iii) a kernel machine combining the two approaches. The evaluation of the models provided different performance of the outcomes for evidence and claim detection tasks and show that the results are promising.

Zeng et al. in Zeng et al. (2018) present an argumentation-based approach for making context-based and explainable judgments. The context in which a decision has to be made, is an important piece of information to consider in order to make the best possible decision. Contexts provide a layer of complexity and dynamism to decision making by allowing a decision to be a good choice in one context, but a bad choice in another one. Incorporating context into a formal representation of a decision issue is the first step towards incorporating context into decision making. The authors use Decision Graphs with Context (DGC) to address this issue. DGCs provide expressiveness and flexibility in modeling decision issues by being able to represent the many links between decisions and objectives in various contexts. The goal of selecting a good decision is pursued by mapping DCGs into Assumption-based Argumentation (ABA) frameworks, and transforming the process of making context-based judgements in DGCs into the selection of admissible argument in ABA frameworks.

Donadello et al. in Donadello et al. (2019) propose a XAI system built on logical reasoning to enable users behavior monitoring and persuade them to conduct healthy lifestyles. The system implements a Persuasive Explanation of Reasoning Inferences that pursues the goal of supporting users to adhere to an healthy diet. The Tbox of the HeLiS ontology Dragoni et al. (2018a) specifies healthy behavior principles and restrictions. After receiving the intake food input, the system utilizes logic to assess if the user does not adhere to healthy living standards. In this case a reasoner module utilizes knowledge and user data (Tbox and Abox) to infer user behavior and cause inconsistencies. As a consequence the system sends the user a natural language message outlining the erroneous behavior and its consequences. As part of the Key To Health pilot project Dragoni et al. (2018b), the recommended technique was deployed into the HORUS.AI platform and evaluated using a mobile application. When compared to simple warnings of abnormalities, the results show that convincing explanations can help users to improve their adherence to dietary rules.

Prentzas et al. in Prentzas et al. (2019), present a method for building explainable AI models using argumentation on top of machine learning. The proposed methodology was experimented using Random Forest in a classification challenge for stroke prediction vs different machine learning algorithms. The basic idea can be applied on top of any learning model that can be used with a rule-generation method, and can be used for both classification and regression tasks. The application generates an argumentation theory in Gorgias, as a Prolog executable file. An explanation of the prediction consists of a set of arguments and decision-making guidelines. The results are encouraging as the proposed explainable model generates human-interpretable explanations and outperforms the machine learning model (random forest accuracy: 69%) and the SVM classifier (accuracy: 74%) in terms of accuracy, correctly

predicting 77% of the test scenarios. The performance of probabilistic neural networks was likewise lower, at 64%.

6.3. Argumentation for medical dialogues

Costa et al. in Costa et al. (2017) suggest the use of persuasive techniques to guide users mainly belonging to the elderly community towards interacting with the iGenda Ambient Assisted Living (AAL) framework Costa et al. (2016). iGenda AAL framework is a cognitive assistant that handles active daily living tasks, keeps track of the user's health, and connects people via mobile devices in a social network.

The used persuasive architecture is based on argumentation schemes and aims at offering users with recommendations that are suited to their profile and interests. The focus is on merging iGenda with a persuasive module that comprises a collection of argumentation schemes that map physicians' and caregivers' reasoning approaches for promoting activities to patients. These approaches are used to build arguments in support of proposed actions or to critique potential alternatives. The authors' objective is to boost the system's persuasion power in order to raise user approval of the activities. When iGenda requests that the module recommend activities, the system tries to come up with one (or more) justifications to support each activity and identify which one the user likes. The next step is to apply an internal reasoning technique to assess which action is best supported by the arguments. This framework takes into account the values that the arguments promote (the users' preferences over the activities' motion characteristics, location, social requirements, environmental conditions, or health conditions), the users' preference relations, and the dependency relations between agents to evaluate arguments and determine which ones defeat others. The argumentation process is an internal iGenda mechanism for assessing which activities are best justified.

Sassoon et al. in Sassoon et al. (2019), argued about how computational argumentation and argumentation-based discourse can be used in the field of clinical consulting, with a focus on chronic health self-management. Chronic diseases are characterized by regular and frequent monitoring of a variety of biometric indicators, as well as adherence to a certain diet, exercise and pharmacological regimen. A patient managing chronic illnesses needs his/her healthcare provider(s) to provide information, recommendation, and explanation.

The last key feature, in particular, requires the patient to request clarification or explanation on a response provided by the system, resulting in an iterative process in which several questions and answers are exchanged until the patient understands the information and/or the system recommendation. The recommended method for allowing explanation functionality as part of a wellness consultation between a human and an agent includes three phases: (i) establishing a specific argumentation structure for the provision of health-related treatments or acts; (ii) finding, and maybe defining, existing argumentation-based frameworks; (iii) illustrating how the agent may employ these strategies in conjunction with argumentation-based interactions to provide patients with explanations. The authors employ the Argumentation Scheme for Proposed Treatment (ASPT) Kokciyan et al. (2018b), which is a subset of the practical reasoning argumentation scheme (ASPR). ASPT is the foundation for a wellness conversation between a patient and an agent to discuss possible actions or treatments. Critical questions (CQs) are posed to ASPT, and they can be used to generate extra or counter-arguments to the arguments generated by this approach. To enable evidence-backed argumentation based decision assistance, the prototype system incorporates data from wellness sensors, electronic health records, and appropriate recommendations. The patient accesses the system using a mobile app that includes a dashboard and a chatbot.

Toniolo et al. in Toniolo et al. (2020) address the issue of justifying the output of Satisfiability Module Theories (SMT) solvers for chronic conditions, i.e. the need to explain suggested optimum plans in a human-friendly manner. Treatment strategies for particular chronic diseases may be thought of as graphs, and in the case of multimorbidities, it result to be helpful to search for the best path across numerous graphs that minimizes adverse drug reactions. The proposed approach uses an argumentation framework on top of the SMT solver to explain decisions. Explanation is performed interactively through argumentation-based dialogues.

In Snaith et al. (2021), Snaith et al. propose an interesting discussion on the two primary issues for Responsible Research and Innovation (RRI) that are extremely relevant to the use of conversation and argumentation in the implementation of e-health systems for advise providing. Collecting and managing health data, as well as proper trust, are relevant challenges. Each of these challenges can be further broken down into additional problems: privacy, informed consent, and addressing dispute, as well as fairness and explanation. The paper evidences the relationship between the main challenges and issues and outlines what result to be relevant to argumentation and dialogue. The proposal also evidences that dialogue plays a vital role in e-health systems, and computational models of formalized dialogue games can help to support this.

7. Discussion

7.1. Summary of the main features of the literature papers

This section summarizes the main features of the reviewed papers. We used the categorization reported in the previous sections, i.e. *Argumentation for Medical Decision Making, Argumentation for Medical Explanations* and *Argumentation for Medical Dialogues*. Table 1 reports the category, a set of references and a brief idea of each proposal. A lot of research has been done in supporting diagnosis, decision support, and interpretability, and there is growing interest in the field of medicine in the creation of Explainable AI also through Argumentation Frameworks.

The most significant contributions of the works in the category *Argumentation for Medical Decision Making* include *support* or *opposition* to a *choice, reasoning* for a *decision, handling KBs with uncertainty*, and *recommendations*. Among studies that mix argumentation with decision-making, the issue of choosing the best option from a range of options is perhaps the most common. Many of the proposals pursued a very specific task: Kokciyan et al. in Kökciyan et al. (2021) propose the CONSULT system for helping stroke patients in taking care of their conditions and adhere to treatment plans; Achilleos et al. in Achilleos et al. (2020) use an abductive framework to help in Alzheimer disease diagnosis (AD); Longo et al. in Longo et al. (2012) apply argumentation theory to the breast cancer recurrence problem; Williams and Williamson in Williams and Williamson (2006) model the breast cancer prognosis in a framework combining logic and probability. Some other proposals have a more general perspective: Glasspool et al. in Glasspool et al. (2006) present the system REACT to support clinicians and patients involved in medical planning; Spanoudakis et al. in Spanoudakis et al. (2017) aims at determining the level of access to a patient's medical record; Cyras et al. in Cyras et al. (2018); Cyras and Oliveira (2019); Cyras et al. (2021a) reason in the presence of conflicting clinical guidelines.

As for the approaches in the category *Argumentation for Medical Explanations* the aim is providing a strategy for reasoning on an argument in order to justify/strengthen it and persuade the other party by using background information, AF defenses, and external knowledge. the work in Shankar et al. (2006) proposes an explanatory framework, called WOZ, that provides justification to the outputs of a clinical decision-support system. Grando et al. in Grando et al. (2013) propose a system for detecting anomalous patient responses to treatments in the Intensive Care Unit (ICU). Mayer, in Mayer (2020) extracts arguments from clinical trials; Zeng et al. in Zeng et al. (2018) present a unique argumentation-based approach for making context-based and explainable judgments. Donadello et al. (2019) in Donadello et al. (2019) reason on the dietary practices that a patient with a chronic illness should follow, whereas Prentzas et al., in Prentzas et al. (2019), use an

Table 1
Summary of Argumentation Methods for Medical Informatics.

Category	Ref.	Proposals' overview
Decision Making	Achilleos et al. (2020) Kökciyan et al. (2021) Longo et al. (2012) Williams and Williamson (2006) Spanoudakis et al. (2017) Glasspool et al. (2006) Cyras et al. (2018) Cyras and Oliveira (2019) Cyras et al. (2021a)	Achilleos et al. in Achilleos et al. (2020) use an abductive framework to help in Alzheimer disease diagnosis (AD) a form of dementia that involves memory, thinking and behavior and affects million people around the world. Kokciyan et al. in Kökciyan et al. (2021) propose the CONSULT system developed with the final aim of helping stroke patients in taking care of their conditions and adhere to treatment plans. Longo et al. in Longo et al. (2012) apply argumentation theory to the breast cancer recurrence problem, i.e. the recurrence of the disease after breast cancer surgery. Williams and Williamson in Williams and Williamson (2006) model the breast cancer prognosis in a framework combining logic and probability. Glasspool et al. in Glasspool et al. (2006) present the system REACT (Risks, Events, Actions and their Consequences over Time) which supports clinicians and patients involved in medical planning. Spanoudakis et al. in Spanoudakis et al. (2017) aim at determining the level of access to a patient's medical record. Cyras et al. in Cyras et al. (2018); Cyras and Oliveira (2019); Cyras et al. (2021a) present a interesting abductive framework to reason in the presence of conflicting clinical guidelines.
Explanations	Shankar et al. (2006) Grando et al. (2013) Mayer (2020) Zeng et al. (2018) Donadello et al. (2019) Prentzas et al. (2019)	Shankar et al. in Shankar et al. (2006) present WOZ, an explanatory framework that verifies the outputs of a clinical decision-support system. Grando et al. in Grando et al. (2013) propose a system for detecting anomalous patient responses to treatments in the Intensive Care Unit (ICU). Mayer, in Mayer (2020) concentrates on the definition of an argument mining system for extracting arguments from clinical trials. Zeng et al. in Zeng et al. (2018) present a unique argumentation-based approach for making context-based and explainable judgments. Donadello et al. (2019) in Donadello et al. (2019) employ an OWL ontology that has details on the dietary practices that a patient with a chronic illness should follow. Prentzas et al., in Prentzas et al. (2019), offer a method for building explainable AI models using argumentation to evaluate the suggested approach in a classification challenge for stroke prediction vs different machine learning algorithms.
Dialogues	Costa et al. (2017) Sassoon et al. (2019) Toniolo et al. (2020) Snaith et al. (2021)	Costa et al. in Costa et al. (2017) suggest the use of persuasive techniques to persuade users towards interacting with the iGenda AAL framework. Sassoon et al. in Sassoon et al. (2019) argued about how computational argumentation and argumentation-based discourse can be used in the field of clinical consulting, with a focus on chronic health self-management. Toniolo et al. in Toniolo et al. (2020) address the issue of Satisfiability Module Theories (SMT) solvers, which emphasize the necessity to explain optimum plans in a human-friendly manner. Snaith et al., in Snaith et al. (2021) discuss the primary issues to the use of conversation and argumentation in the implementation of e-health systems for advise providing.

argumentation framework to reason on stroke prediction.

The approaches in the third category, i.e. *Argumentation for Medical Dialogues*, establish an argumentative conversation to explain a position. These discussions take place between two parties who debate the validity of one or more statements or arguments, each seeking to persuade the other to agree with his/her viewpoint. These conversations are also known as persuasion dialogues. Sassoon et al. in Sassoon et al. (2019) propose the use of argumentation-based discourse in the field of chronic health self-management; Costa et al., in Costa et al. (2017), suggest the use of persuasive techniques to persuade users towards interacting with the iGenda AAL framework. Toniolo et al. in Toniolo et al. (2020) emphasize the need to explain optimum plans in a human-friendly manner; Snaith et al. in Snaith et al. (2021) discuss the primary issues to the use of conversation and argumentation in the implementation of e-health systems for advise providing.

7.2. *Major opportunities in using argumentation approaches for explainable AI in medical informatics*

A lot of research has been done in supporting diagnosis, decision support, and interpretability, and there is a growing interest in the field of medicine in the creation of Explainable AI. Interpretability in the medical field is not just an intellectual curiosity, but a key factor. The major opportunities in using argumentation approaches for eXplainable AI in Medical Informatics are reported below.

- **Natural Explanatory Capacity**: Medical Informatics advantages of the natural explanatory capacity of argumentation. The translation of the knowledge base into interactive arguments is intuitive for clinicians since it adheres to a modular procedure based on clinically relevant natural language words. An argumentation based framework performs a gradual, modular approach that uses evidence to construct an explanatory reasoning and allows to select a specific semantic through which calculating the justification so that ensuring a more intuitive interpretation of the results. Additionally, final results are conflict-free sets of the same input arguments, i.e. the output of an argumentation based framework are the arguments that are susceptible to be accepted under a given semantics. Therefore, clinicians can examine each argument within the final results in order to comprehend and defend the suggested plan of action.

- **Partial, Inconsistent, Uncertain and Vagueness knowledge**: An argumentation based framework results to be an effective approach in the case of partial or inconsistent knowledge, that results to be a frequent scenario in the medical domain. Reasoning is applied on available data, by simply discarding missing data and arguments that results to be not useful for the argumentative process. Moreover, an argumentation approach results particularly suited to express uncertain and vagueness knowledge, often present in medical domain, using

natural language assertions or propositions.

- **Extensibility**: Argumentation Theory is an extensible framework, that simply allows to insert/remove an argument as soon as a new evidence emerges.

7.3. Major challenges in argumentation approaches for explainable AI in medical informatics

Argumentation has been proved to be a viable effective approach in the development of Explainable AI in the realm of medicine, however, many are still the challenges that remain to be addressed.

- **Computational Complexity**: The efficient computation of argumentation-based explanations is required to enable XAI solutions. This focuses on effective systems for the key reasoning problems and a thorough comprehension of their computational complexity.

- **Translation of the knowledge base**: The first step related to the conversion of a knowledge-base into an interactive set of arguments may be a difficult and time consuming task, particularly if there are many pieces of evidence. This conversion is a key piece for ensuring explainability as it adheres to a modular procedure based on clinically relevant natural language words. Effort should be pursued by the research community in supporting and making simple this specific task, while ensuring the production of an intuitive knowledge base by embedding argumentation reasoning in more general settings.

- **Lack of Learning**: An Argumentation based framework is not a learning-based paradigm. Therefore, specific actions on the knowledge base are not automatically performed, but require a specific management. Also in this specific case, it is required effort by the scientific community in order to combine the benefit of argumentation frameworks with machine learning techniques.

8. Conclusions

This paper describes the benefit of logic approaches for XAI evidencing how the intrinsic characteristics of expressiveness and explainability guarantee the satisfaction of the general requirements of XAI and candidate declarative approaches to be a fundamental component in the design of explainable, justified and ethical artificial intelligent systems. More specifically, the present proposal focuses on a detailed topic related to the use of argumentation approaches in Medical Informatics by overviewing existing approaches in the literature.

Declaration of Competing Interest

No conflict of interest exists.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Achilleos, K. G., Leandrou, S., Prentzas, N., Kyriacou, P. A., Kakas, A. C., & Pattichis, C. S. (2020). Extracting explainable assessments of alzheimer's disease via machine learning on brain MRI imaging data. *IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, 1036–1041.

Alfano, G., Greco, S., Parisi, F., & Trubitsyna, I. (2020). On the semantics of abstract argumentation frameworks: A logic programming approach. *Theory and Practice of Logic Programming, 20*(5), 703–718.

Antonis, K., & Loizos, M. (2020). Abduction and argumentation for explainable machine learning: A position survey. CoRR abs/2010.12896.

Baroni, P., Rago, A., & Toni, F. (2018). How many properties do we need for gradual argumentation? *Proceedings of the AAAI Conference on Artificial Intelligence, 32*(1).

Baumeister, D., Järvisalo, M., Neugebauer, D., Niskanen, A., & Rothe, J. (2021). Acceptance in incomplete argumentation frameworks. *Artificial Intelligence, 295*, 103470.

Bistarelli, S., Kotthoff, L., Santini, F., & Taticchi, C. (2021). Summary report for the third international competition on computational models of argumentation. *AI Magazine, 42*(3), 70–73.

Bistarelli, S., & Santini, F. (2021). Weighted argumentation. *FLAP, 8*(6), 1589–1622.

Bistarelli, S., & Taticchi, C. (2021). A labelling semantics and strong admissibility for weighted argumentation frameworks. *Journal of Logic and Computation, 32*(2), 281–306.

Bondarenko, A., Dung, P. M., Kowalski, R. A., & Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence, 93*, 63–101.

Caminada, M. (2008). A gentle introduction to argumentation semantics. *Lecture Material, Summer*.

Caminada, M., Sá, S., Alcântara, J. F. L., & Dvorák, W. (2015). On the equivalence between logic programming semantics and argumentation semantics. *International Journal of Approximate Reasoning : Official Publication of the North American Fuzzy Information Processing Society, 58*, 87–111.

Cayrol, C., & Lagasquie-Schiex, M. C. (2005). On the acceptability of arguments in bipolar argumentation frameworks. *In european conference on symbolic and quantitative approaches to reasoning and uncertainty* (pp. 378–389). Springer.

Cayrol, C., & Lagasquie-Schiex, M. C. (2020). Logical encoding of argumentation frameworks with higher-order attacks and evidential supports. *International Journal on Artificial Intelligence Tools*.29(03n04): 2060003:1-2060003:50

Chapman, M., Balatsoukas, P., Ashworth, M., Curcin, V., Kökciyan, N., Essers, K., Sassoon, I., Modgil, S., Parsons, S., & Sklar, E. (2019). *Computational argumentation-based clinical decision support. In proceedings of the 18th international conference on autonomous agents and multiagent systems, 2345–2347*. International Foundation for Autonomous Agents and Multiagent Systems.

Charwat, G., Dvorak, W., Gaggl, S. A., Wallner, J. P., & Woltran, S. (2015). Methods for solving reasoning problems in abstract argumentation: A survey. *Artificial Intelligence, 220*, 28–63.

Cocarascu, O., & Toni, F. (2016). Argumentation for machine learning: A survey. *COMMA*, 219–230.

Costa, A., Heras, S., Palanca, J., Jordán, J., Novais, P., & Julián, V. (2017). Argumentation schemes for events suggestion in an e-health platform. *International Conference on Persuasive Technology*, 17–30.

Costa, A., Heras, S., Palanca, J., Novais, P., & Julián, V. (2016). A persuasive cognitive assistant system. *International Symposium on Ambient Intelligence*, 151–160.

Cyras, K., Delaney, B., Prociuk, D., Toni, F., Chapman, M., Domínguez, J., & Curcin, V. (2018). Argumentation for explainable reasoning with conflicting medical recommendations. *MedRACER+WOMoCoE@KR*, 14–22.

Cyras, K., & Oliveira, T. (2019). Resolving conflicts in clinical guidelines using argumentation. *AAMAS*, 1731–1739.

Cyras, K., Oliveira, T., Karamlou, A., & Toni, F. (2021a). Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. *Argument & Computation, 12*(2), 149–189.

Cyras, K., Rago, A., Albini, E., Baroni, P., & Toni, F. (2021b). Argumentative XAI: A survey. CoRR abs/2105.11266.

Cyras, K., & Toni, F. (2016). ABA+: Assumption-based argumentation with preferences. In C. Baral, J. P. Delgrande, & F. Wolter (Eds.), *In principles of knowledge representation and reasoning, 15th international conference* (pp. 553–556). AAAI Press, Cape Town.

Das, A., & Rad, P. (2020). Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371*.

Denecker, M., & Kakas, A. C. (2002). Abduction in logic programming. *In Computational Logic: Logic Programming and Beyond*, 402–436.

Donadello, I., Dragoni, M., & Eccher, C. (2019). Persuasive explanation of reasoning inferences on dietary data. *PROFILES/SEMEX@ISWC*, 46–61.

Dragoni, M., Bailoni, T., Maimone, R., & Eccher, C. (2018a). Helis: An ontology for supporting healthy lifestyles. *ISWC*, (2), 53–69.

Dragoni, M., Bailoni, T., Maimone, R., Marchesoni, M., & Eccher, C. (2018b). HORUS.AI - A knowledge-based solution supporting health persuasive self-monitoring. *ISWC (P&D/Industry/BlueSky)*.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning. logic programming and n-person games. *Artificial Intelligence, 77*(2), 321–358.

Dung, P. M. (2016). An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence, 231*, 107–150.

Dung, P. M., Kowalski, R. A., & Toni, F. (2006a). Dialectic proof procedures for assumption-based, admissible argumentation. *Artificial Intelligence, 170*.

Dung, P. M., Mancarella, P., & Toni, F. (2006b). *Computing ideal sceptical argumentation. Technical report*. Imperial College London.

Dung, P. M., Mancarella, P., & Toni, F. (2006c). A dialectic procedure for sceptical, assumption-based argumentation. *COMMA*, 145–156.

Dung, P. M., & Son, T. C. (1995). Nonmonotonic inheritance, argumentation and logic programming. *LPNMR*, 316–329.

F. Baldwin, J. (1981). Fuzzy logic and fuzzy reasoning. In E. H. Mamdani, & B. R. Gaines (Eds.), *In fuzzy reasoningand its applications*. London Academic Press.

Fazzinga, B., Flesca, S., & Furfaro, F. (2018). Probabilistic bipolar abstract argumentation frameworks: Complexity results. *In IJCAI*, 1803–1809.

Gaertner, D., & Toni, F. (2007). Computing arguments and attacks in assumption-based argumentation. *IEEE Intelligent Systems, 22*(6), 24–33.

Garcia, A., & Alejandro, J. (2000). *Defeasible logic programming: Definition, operational semantics and parallelism. Ph.d. thesis*. Computer Science Department, Universidad Nacional del Sur, Bahia Blanca, Argentina.

García, A. J., Chesñevar, C. I., Rotstein, N. D., & Simari, G. R. (2013). Formalizing dialectical explanation support for argument based reasoning in knowledge based systems. *Expert Systems With Applications, 40*(8), 3233–3247.

García, A. J., & Simari, G. R. (2004). Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming, 4*(2), 95–138.ISSN 1471-0684

Glasspool, D., Fox, J., Oettinger, A., & Smith-Spark, J. H. (2006). Argumentation in decision support for medical care planning for patients and clinicians. *AAAI Spring Symposium: Argumentation for Consumers of Healthcare*.

Goldstein, M. K., Hoffman, B. B., Coleman, R. W., Musen, M. A., Tu, S. W., Advani, A., Shankar, R., & O'Connor, M. (2000). Implementing clinical practice guidelines while taking account of changing evidence: ATHENA DSS, an easily modifiable decision-support system for managing hypertension in primary care. *Proceedings AMIA Symposium*, 300–304.

Goodman, B., & Flaxman, S. (2017). European union regulations on algorithmic decision making and a right to explanation. *AI Magazine, 38*(3), 50–57.

Grando, M. A., Moss, L., Sleeman, D. H., & Kinsella, J. (2013). Argumentation-logic for creating and explaining medical hypotheses. *Artificial Intelligence Medicine, 58*(1), 1–13.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining Knowledge Discovery, 9*(4).E1312

Kakas, A. C., Moraitis, P., & Spanoudakis, N. I. (2019). GORGIAS: Applying argumentation. *Argument & Computation, 10*(1), 55–81.

Kokciyan, N., Chapman, M., Balatsoukas, P., Sassoon, I., Essers, K., Ashworth, M., Curcin, V., Modgil, S., Parsons, S., & Sklar, E. I. (2019). A collaborative decision support tool for managing chronic conditions. *Studies in Health Technology and Informatics*, 644–648.Aug 21;264

Kökciyan, N., Sassoon, I., Sklar, E., Modgil, S., & Parsons, S. (2021). Applying metalevel argumentation frameworks to support medical decision making. *IEEE Intelligent Systems, 36*(2), 64–71.1 March-April

Kokciyan, N., Sassoon, I., Young, A. P., Chapman, M., Porat, T., Ashworth, M., Curcin, V., Modgil, S., Parsons, S., & Sklar, E. (2018a). Towards an argumentation system for supporting patients in self-managing their chronic conditions. *AAAI Workshops*, 455–462.

Kokciyan, N., Sassoon, I., Young, A. P., Chapman, M., Porat, T., Ashworth, M., Curcin, V., Modgil, S., Sanjay, S., Parsons, S., & Sklar, E. (2018b). Towards an argumentation system for supporting patients in self-managing their chronic conditions. *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*.

Koshiyama, A., Kazim, E., & Engin, Z. (2019). Xai: Digital ethics. *In HeXAI Workshop*.

Longo, L., Kane, B., & Hederman, L. (2012). Argumentation theory in health care. *25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, 1–6.

Mayer, T. (2020). *Argument mining on clinical trials. (fouille d'arguments à partir des essais cliniques)*. France: Université Côte d'Azur.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence, 267*, 1–38.

Modgil, S., & Caminada, M. (2009). Proof theories and algorithms for abstract argumentation frameworks. *Argumentation in Artificial Intelligence, 2009*, 105–129.

Moss, L. E. (2010). *Explaining anomalies : An approach to anomaly-driven revision of a theory*. University of Aberdeen, UK.

Muggleton, S., & de Raedt, L. (1994). Inductive logic programming: Theory and methods. *The Journal of Logic Programming, 19–20*(2), 629–679.

Musen, M. A., Tu, S. W., Das, A. K., & Shahar, Y. (1996). EON: A component-based approach to automation of protocol-directed therapy. *Journal of the American Medical Informatics Association: JAMIA, 3*(6), 367–388.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation, 1*(2), 93–124.

Prentzas, N., Nicolaides, A., Kyriacou, E., Kakas, A. C., & Pattichis, C. S. (2019). Integrating machine learning with symbolic reasoning to build an explainable AI model for stroke prediction. *BIBE*, 817–821.

Sassoon, I., Kökciyan, N., Sklar, E., & Parsons, S. (2019). Explainable argumentation for wellness consultation international workshop on explainable. *Transparent Autonomous Agents and Multi-Agent Systems*, 186–202.

Schuster, S., & Manning, C. D. (2016). Enhanced english universal dependencies: An improved representation for natural language understanding tasks. *LREC*.

Shankar, R. D., Tu, S. W., & Musen, M. A. (2006). Medical arguments in an automated health care system. *AAAI Spring Symposium: Argumentation for Consumers of Healthcare, 2006*, 96–104.

Snaith, M., Devereux, J., Lawrence, J., & Reed, C. (2010). Pipelining argumentation technologies. proceedings of the third international conference on computational models of argument (COMMA 2010).

Snaith, M., Nielsen, R., Kotnis, O., & Ramchandra, S. (2021). Pease, alison ethical challenges in argumentation and dialogue in a healthcare context. *Argument & Computation, 12*(2), 249–264.

Snaith, M., & Reed, C. (2012). TOAST: Online ASPIC+ implementation. computational models of argument – proceedings of COMMA 2012. *Vienna, Austria, September 10–12, 2012*, 509–510.

Spanoudakis, N. I., Constantinou, E., Koumi, A., & Kakas, A. C. (2017). Modeling data access legislation with gorgias. *30th international conference on industrial, engineering, and other applications of applied intelligent systems*. IEA/AIE.

Tjoa, E., & Guan, C. (2021). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE trans. *Neural Networks of Learning System, 32*(11), 4793–4813.

Tonekaboni, S., Joshi, S., McCradden, M. D., & Goldenberg, A. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. *MLHC, 2019*, 359–380.

Toniolo, A., Bowles, J., & Juliana, K. F. (2020). Others, dialogue games for explaining medication choices. *International Joint Conference on Rules and Reasoning*, 97–111.

Toulmin, S. (1958). *The uses of argument*. Cambridge MA.: Cambridge University Press.

UK Working Group on the Primary Prevention of Breast Cancer. (2005). *Breast cancer: An environmental disease - the case for primary prevention. Technical report*. The UK Working Group on the Primary Prevention of Breast Cancer.

Vassiliades, A., Bassiliades, N., & Patkos, T. (2021). Argumentation and explainable artificial intelligence: A survey. *The Knowledge Engineering Review, 36*.E5

Williams, M., & Williamson, J. (2006). Combining argumentation and bayesian nets for breast cancer prognosis. *Journal of Logic, Language and Information, 15*(1–2), 155–178.

Xie, Y., Gao, G., & Chen, X. A. (2019). Outlining the design space of explainable intelligent systems for medical diagnosis. *CoRR, vol. abs/1902.06019*, 1–5.

Zadeh, L. A. (1965). Fuzzy sets. *Journal of Information and Control, 8*, 338–353.

Zeng, Z., Fan, X., Miao, C., Leung, C., Chin, J. J., & Ong, Y. S. (2018). Context-based and explainable decision making with argumentation. *AAMAS*, 1114–1122.