



HAL
open science

An Investigation of Argumentation Theory for the Prediction of Survival in Elderly Using Biomarkers

Lucas Rizzo, Ljiljana Majnaric, Pierpaolo Dondio, Luca Longo

► **To cite this version:**

Lucas Rizzo, Ljiljana Majnaric, Pierpaolo Dondio, Luca Longo. An Investigation of Argumentation Theory for the Prediction of Survival in Elderly Using Biomarkers. 14th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), May 2018, Rhodes, Greece. pp.385-397, 10.1007/978-3-319-92007-8_33 . hal-01821041

HAL Id: hal-01821041

<https://inria.hal.science/hal-01821041>

Submitted on 22 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

An investigation of argumentation theory for the prediction of survival in elderly using biomarkers

Lucas Rizzo¹, Ljiljana Majnaric², Pierpaolo Dondio¹, and Luca Longo¹

¹ School of Computing, Dublin Institute of Technology, Dublin, Ireland

² Department of Family Medicine, School of Medicine, University of Osijek, Croatia

luca.longo@dit.ie

Abstract. Research on the discovery, classification and validation of biological markers, or biomarkers, have grown extensively in the last decades. Newfound and correctly validated biomarkers have great potential as prognostic and diagnostic indicators, but present a complex relationship with pertinent endpoints such as survival or other diseases manifestations. This research proposes the use of computational argumentation theory as a starting point for the resolution of this problem for cases in which a large amount of data is unavailable. A knowledge-base containing 51 different biomarkers and their association with mortality risks in elderly was provided by a clinician. It was applied for the construction of several argument-based models capable of inferring survival or not. The prediction accuracy and sensitivity of these models were investigated, showing how these are in line with inductive classification using decision trees with limited data.

Keywords: Biomarkers, argumentation theory, defeasible reasoning

1 Introduction

In the medical domain, biomarkers can be objectively defined as a medical condition which can be measured precisely and reproducibly. It has to be observed from outside the patient, contrarily to medical symptoms which are an indication of health recognized by the patients themselves [27]. Simple examples include blood pressure, pulse and waist circumference. It can also range to several diseases diagnostics or more complex laboratory tests of blood and other tissues. Over the past 50 years the advances in biological sciences have generated more than 30,000 candidate biomarkers from which less than a thousand might have clinical value [23]. There is a vital issue in determining the relationship between biomarkers and relevant clinical endpoints, such as survival, stroke and myocardial infarction [27]. Data mining techniques incorporating machine learning algorithms have been used as a possible path for solving this problem [7, 28]. However, these are usually suited for large amounts of data which are not always available due to the complexity of measurement and quantity of biomarkers. In such cases, it is argued that defeasible reasoning might be considered a possible resolution technique.

In formal logics a defeasible concept is built upon a set of interactive pieces of evidence that can become defeated by additional reasons [21]. Efforts have been made within the field of Artificial Intelligence (AI) to perform and analyze the act of reasoning defeasibly. Argumentation Theory (AT) is a computational approach which has been widely employed for modelling defeasible and non-monotonic reasoning [3]. It has been broadly applied in the field of health care [5, 10] since data accounted in such problems is often uncertain, heterogeneous and incomplete. In this study the dataset employed follows this trend. It contains information of 93 patients and 51 different biomarkers collected in an European hospital from primary health care health records during the time span of five years. The survival of the patient is the dependent variable and indicates *survival* or *death*, while independent variables are both continuous and categorical. An initial investigation was performed and classification models such as decision trees could not provide accurate models, given the small amount of data. Thus, we proposed the investigation of computational argumentation theory for the prediction of survival in elderly. A knowledge-base constructed by a clinician was selected for examination. It adds dozens of rules and 7 different mortality risks. Its formalization was made by following the 5-layer modelling approach proposed by [16]. Finally, this schema was used for building a number of argument-based models with different parameters. Their accuracy and sensitivity (true positive rate) was subsequently computed for comparison purposes. The research question being investigated is: *to which extent can defeasible reasoning, implemented using formal argumentation theory, enhance the prediction of survival in elderly using information on biomarkers according to accuracy and sensitivity?*

The remainder of this paper is organized as follows. Section 2 presents related work on biomarkers and AT research, making the connection between the two areas. The design and methodologies of the study are detailed on section 3. Section 4 provides the comparison against decision trees and subsequent discussion. Finally, section 5 presents the conclusion and future work.

2 Related work

Research on prognostic information for mortality in older adults is of significant importance on clinical decision-making, for instance when deciding for more advance care planning for higher risks patients [12]. Non-communicable diseases, such as cardiovascular, are the main cause of mortality among elderly individuals [13]. Hence, the increase of biomarkers research to obtain early prognostics. Examples include [6, 1] which investigate possible biomarkers related to mortality. The predictive power of selected biomarkers is usually determined by a statistical analysis and so relies on data, not providing a complete explanation or a reasoning process. Recently, machine learning algorithms have also been used in this area of research [7, 28]. Nonetheless, the validation of biomarkers as possible features of endpoints classification is still a concern on medical research [27]. It falls on the same issues as other medical fields. In these, different pieces of information taken into account might be in contradiction with each other, thus a

method for resolving them is often necessary. Knowledge-based systems are approach to deal with uncertainty and potentially valid for overcoming such issues [25, 24]. These are well established within AI and have been used in several domains such as the pharmaceutical industry, clinical trials and care planning [26, 11, 14, 10, 18]. However, despite the similarity of the previous areas, it appears, from literature, that there is no research that attempts to employ defeasible reasoning in the biomarker domain for the prediction of survival. Reasoning is defeasible when a conclusion can be retracted in the light of new evidence. It is argued here that the knowledge required for modelling and assessing mortality risk using biomarkers can be seen as defeasible. It contains inconsistent pieces of evidence supporting different risk levels that are also retractable in the light of new information. Let us consider the following example with arguments:

- *Arg 1: Increased mean cell volume of red blood cells (MCV) might lead to macrocytosis. Older people with macrocytosis are more likely to have poorer cognitive functioning and increased mortality.*
- *Arg 2: Deficiency of vitamin B12 is an indicator of increased MCV. If there is no vitamin B12 deficiency then MCV can not be increased.*

On one hand, in the first argument, a clinician may argue that there is evidence to infer an increased mortality risk, due to the increased measure of MCV which can lead to macrocytosis. In case there is no other evidence, increased mortality risk might be a reasonable conclusion. On the other hand, the second argument argues that if there is no deficiency of vitamin B12 then increased measure of MCV can not be taken into account. In this case the inference of an increased mortality risk no longer holds. This example illustrates how the set of conclusions does not increase monotonically and can be retracted in the light of new information. The next section explains how such arguments and conflicts can be represented and possible approaches to reach justifiable conclusions.

3 Design and methodology

This section illustrates how a knowledge-base on mortality risk factors in elderly was translated into computational argument-based models following a 5-layer schema [15]. Due to space limitation the full knowledge-base is not shown here, but it can be found online³. It contains not only inference rules but also contradictions, preferences and a full description of all biomarkers utilized by the clinician, together with other 92 references used for the development of arguments (knowledge-base's rules).

3.1 Layer 1 - Definition of the structure of arguments

The first step on the argumentation process or the 5-layer schema is to define a set of *forecast* arguments. These can be represented like:

Forecast argument : premises \rightarrow conclusion

³ <http://dx.doi.org/10.13140/RG.2.2.20905.49764>

The objective is to determine a premise or set of premises from which a conclusion can be reasonably inferred (survival or not). Survival was deduced from mortality risks which were in turn based on natural language expressions utilized by the clinician. Expressions like “may affect survival”, “strong mortality risk factor” and “increased mortality” are a few examples. These expressions were separated in seven different mortality risks: *no risk* (r_1), *low risk* (r_2), *medium low risk* (r_3), *medium risk* (r_4), *medium high risk* (r_5), *high risk* (r_6) and *extremely high risk* (r_7). From the initial set of 51 biomarkers, 44 had a natural language description employable for the inference of some mortality risk. Let us consider a description for chronic obstructive pulmonary disease (COPD) and the possible argument associated to it:

- *Description: COPD is a major cause of mortality and also a cardiovascular risk factor. There may be a survival benefit for treatment with new inhalatory drugs, however, conclusive data are currently lacking.*
- *Arg: presence of COPD \rightarrow extremely high risk (r_7)*

From the knowledge-base’s descriptions it is possible to infer some mortality risk for some biomarkers, but the deduction from risks to survival or death is not known. Because of these two approaches are selected for investigation:

- *Cautious:* all risks are considered as potential predictors. r_{1-3} are predictors for survival while r_{4-7} are predictors for death.
- *Skeptical:* only r_{1-2} are associated with survival and r_{6-7} with death. Arguments that support other conclusions (r_{3-5}) are no longer part of the reasoning process. Since risks r_{3-5} are in the medium range it might be argued that they do not provide strong evidence to infer survival or not.

3.2 Layer 2 - Definition of the conflicts of arguments

At this layer the relationship between arguments is defined. While the first layer allows the definition of the monological structure of arguments, the second layer allows the creation of dialogical structures of knowledge. The objective is to support the examination of invalid arguments that have the earmarks of being legitimate. According to [20] these can be referred to as *mitigating arguments*. Their internal structure is defined by a set of premises and an undercutting inference \Rightarrow to an argument B (forecast or mitigating):

$$\textit{Mitigating argument} : \textit{premises} \Rightarrow B$$

A few classes of mitigating arguments can be found in [22]. However, only one of them, *undercutting attack*, is adopted in this research for the implementation of conflicts. An undercutting attack occurs when there is a special case, or a defeasible inference rule, that does not allow the application of the knowledge carried in some other argument. Examples include expressions such as:

- *Arg: If total cholesterol is high then hematocrite (HTC) is not low.*

Based on this description an undercutting attack to a forecast argument (or a mitigating argument) can be built, that assumes low HTC as one of its premises. For instance, consider a forecast argument F1 and an undercutting attack UC1:

- F1: *male AND low HTC* \rightarrow *medium low risk* (r_3)
- UC1: *high total cholesterol* \Rightarrow F1

Since it is known that HTC can not be low if total cholesterol is high, undercutting attack UC1 ensures that forecast rule F1 is not applied in this situation. A set of strict preferences among pairs of biomarkers for predicting survival are also provided by the clinician (expert) who was interviewed in this study. Such preferences can also be seen as undercutting attacks⁴. Consider the following preference P1, forecast arguments F2, F3 and an undercutting attack UC2:

- P1: *Hypertension > Age* - F2: *Age* \in [66, 70] \rightarrow *medium risk* (r_4)
- F3: *high Hypertension* \rightarrow *high risk* (r_6) - UC2: F3 \Rightarrow F2

In the above example P1 suggests that hypertension is more important for inferring survival than age, thus forecast argument F3 should be considered instead of F2. Let us point out that forecast arguments F1 and F2 would not be defined according to the skeptical approach (layer 1) because their risks are not in the survival set (r_{1-2}) or death set (r_{6-7}), consequently UC1 and UC2 would also not be defined in this approach. Once layers 1 and 2 are finalized, the set of arguments and attacks can be seen now as a graph, or *argumentation framework* (AF). Figure 1 depicts AFs for cautions and skeptical approaches. Node labels follow the same names used for arguments in the full knowledge-base⁵.

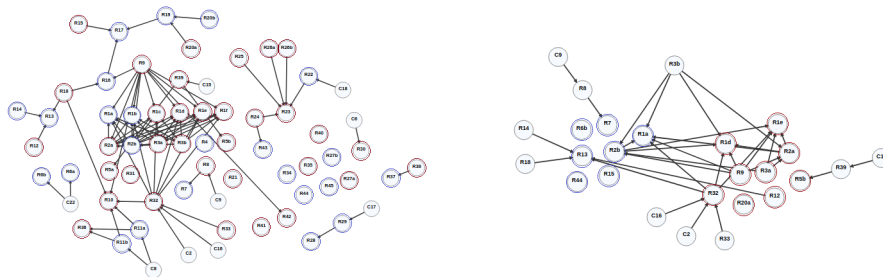


Fig. 1. Argumentation framework: graphical representation of cautious (left) and skeptical (right) knowledge-bases. Double circles represent forecast arguments supporting survival (blue) or death (red). Other nodes represent premises of undercutting attacks. Edges are directed and represent attacks from an argument to another argument, or from a forecast argument to another forecast argument.

⁴ Note that preference among biomarkers is not the same as preferentiality of arguments, a notion mentioned on layer 3.

⁵ <http://dx.doi.org/10.13140/RG.2.2.20905.49764>

3.3 Layer 3 - Evaluation of the conflicts of arguments

Once an AF is constructed, it can be elicited with data⁶. Some of arguments will be activated and kept, while others will not be activated and discarded. For instance, F1 premises do not evaluate true for female individuals or individuals that do not have low HTC, hence it is not activated and is discarded. Activated arguments might be in conflict, according to attacks defined in layer 2. These can be evaluated employing different strategies such as the strength of argument, preferentiality and strength of attack relations [9]. Here attacks only present a form of a binary relation. In other words, once the arguments of an attack are activated, the attack is automatically considered efficacious. Activated arguments together with valid attacks form a *sub-argumentation framework* (sub-AF), that can now be evaluated against inconsistencies. Figure 2 depicts a sub-AF.



Fig. 2. Argumentation framework: an example of activated arguments (blue nodes inside) given one record of the dataset. Light nodes and edges are still part of the knowledge-base but are not considered for the next layers.

3.4 Layer 4 - Definition of the dialectical status of arguments

In order to accept or reject arguments, acceptability semantics are applied on top of each sub-AF. Note that each record in the dataset produces a different sub-AF, so semantics should be applied for each separated case. Most well known semantics, such as *grounded* and *preferred*, can be found in [8]. The goal in this layer is to evaluate not only if an argument is defeated but if the defeaters are defeated themselves. It is said that an argument A *defeats* argument B if and only if there is a valid attack from A to B. Dung's acceptability semantics [8] will generate one or more sets of *extensions* (conflict free sets of arguments). Each extension can be seen as a different point of view that can be used in a decision making process. At this stage the internal structure of arguments is not considered and in the literature this is known as an *abstract argumentation framework* (AAF). It is a pair $\langle Arg, attacks \rangle$ where: Arg is a finite set of (abstract) arguments, $attacks \subseteq Arg \times Arg$ is binary relation over Arg . Given sets $X, Y \subseteq Arg$ of arguments, $X attacks Y$ if and only if there exists $x \in X$ and $y \in Y$ such that $(x, y) \in attacks$. A set $X \subseteq Arg$ of argument is

- *admissible* iff X does not attack itself and X attacks every set of arguments Y such that $Y attacks X$;
- *complete* iff X is admissible and X contains all arguments it *defends*, where $X defends x$ if and only if $X attacks all attacks against x$;
- *grounded* iff X is minimally complete (with respect to \subseteq);
- *preferred* iff X is maximally admissible (respect to \subseteq)

⁶ <http://dx.doi.org/10.13140/RG.2.2.17130.62402>

It is important to highlight that arguments with different conclusions (survival or death in this study), are not necessarily conflicting (given the expert’s knowledge-base) and might be part of the same extension. This is because they are originally defined according to the set of mortality risks (r_{1-7}) and not survival, thus, unless some inconsistency is explicitly defined, they can coexist. Nonetheless, in order to minimize this inconsistency, ranking-based semantics were also applied. This class of semantics is capable of distinguishing arguments not only as accepted or rejected, but it also provides a mechanism to fully rank them, therefore defining a form of importance. Here, the *categorizer* semantic is employed [2]. It takes into consideration the number of direct attackers to compute the strength of an argument. In this sense, attacks from non-attacked arguments are stronger and are more impactful than attacks from arguments attacked several times. Given an $AAF = \langle Arg, attacks \rangle$, and the set of direct attackers $att(a) \forall a \in Arg$, the categorizer function $Cat : Arg \rightarrow]0, 1]$ is given by:

$$Cat(a) = \begin{cases} 1 & \text{if } att(a) = \emptyset \\ \frac{1}{1 + \sum_{c \in att(a)} Cat(c)} & \text{otherwise} \end{cases} \quad (1)$$

Finally, a ranking is given according to the value computed by Cat for each argument. Forecast arguments with different conclusions still might have the same ranking, for instance, if they are both not attacked. In this case another layer is required to accrue arguments and produce a final inference. Figure 3 illustrates *grounded* and *categorizer* semantics computed for the activated graph on figure 2. Note that forecast arguments attacked only by rejected arguments can still be rejected under the categorizer semantics.

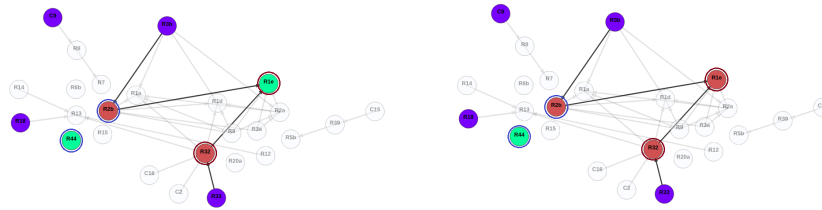


Fig. 3. Argumentation framework: acceptable arguments computed by the grounded semantics (left) and categorizer semantics (right). Blue nodes inside are activated but do not support a conclusion, so are not accepted neither rejected. Red and green nodes inside are forecast arguments rejected and accepted respectively. Double circles indicate whether forecast arguments support death (red) or survival (blue).

3.5 Layer 5 - Accrual of acceptable arguments

The final step on the argumentation process of this study is to select one of the possible two outcomes: survival or death. As defined in layer 2 (section 3.2), two types of arguments can be part of an extension or have the same ranking:

forecast and mitigating. Mitigating arguments, by definition, do not support any conclusion and have already finalized their role attacking other arguments and contributing to the definition of the set of acceptable arguments. Thus, the final inference is given only by the forecast arguments. Three situations are possible:

1. If the set of acceptable forecast arguments support the same conclusion, then final inference coincides to that conclusion.
2. If both conclusions are supported we argue that the one with the highest number of supporters should be chosen. Preferences have already been taken into account, hence the amount of evidence, or number of arguments supporting a conclusion, may be seen as a possible reason for choosing one.
3. If there is still a tie (survival and death being supported by an equal number of acceptable forecast arguments), then logically there is no reason for the system to take any conclusion. This might reflect the uncertainty of the domain, in which the knowledge-base, as coded from the expert in this study, does not contain information for dealing with this type of situation. Nonetheless, because one conclusion is eventually required for comparison purposes, the preference goes to death making a skeptical inference.

Table 1 lists the constructed models, summarising their settings for each layer.

Table 1. Set up of each model investigated. For layers 1, 2 and 4 ($L1$, $L2$, $L4$) a parameter is set. Layers 3 and 5 are equal for all cases.

Model	Survival ($L1$)	AF ($L2$)	Semantics ($L4$)
M1	Cautious	figure 1 (left)	Grounded
M2	Cautious	figure 1 (left)	Categorizer
M3	Skeptical	figure 1 (right)	Grounded
M4	Skeptical	figure 1 (right)	Categorizer

4 Results

Collected data of 93 patients and 51 different biomarkers obtained from an European hospital from primary health care health records during the time span of five years⁷ was used to instantiate the designed argumentation-based models (table 1). In order to gauge the quality of the solutions an analysis was also performed using classification trees (CT) on the same dataset. Data was normalized and stratified 10-fold cross validation was applied for building CT models. The percentage of death and survival records is 39% and 61% respectively. The evaluation metrics selected were accuracy, obtained from each respective confusion matrix, and sensitivity or true positive rate. It is important to highlight that the comparison between argument-based inference (conducted case-by-case) and a

⁷ <http://dx.doi.org/10.13140/RG.2.2.17130.62402>

learning technique (conducted on the whole dataset with CT) is not straightforward. In fact, argument-based models do not rely on data and consequently are not built on training sets. To be as fair as possible, oversampling could not be applied on the classification tree considering that synthetic data might be outside the knowledge-base capacity of understanding. The accuracy and sensitivity of each fold was computed across the argument-based models and the CT models (figures 4 and 5).

4.1 Accuracy

The accuracy of CT was higher than argument-based models in 4 folds, lower in 4 folds and equal in 2 folds. Given the size of the dataset, there is approximately 9 records on each fold, and because of that there is a large fluctuation in the accuracy percentage. Nonetheless, it is reasonable to say that, overall, tested approaches are nearly equivalent according to their accuracies (figure 4). Among the argument-based models, also no significant difference can be observed. This demonstrates a slight advantage of models M3-4 over M1-2, since the skeptical approach contain less information and could achieve similar results. Cautious models (M1-2) and skeptical models (M3-4) did not present any significant difference between them, suggesting that, given the topology of the argumentation frameworks (figure 1), grounded and categorizer semantics have no meaningful impact on inferences.

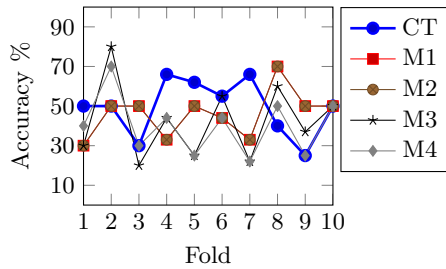


Fig. 4. Accuracy of classification tree (CT) and each model (M) by fold.

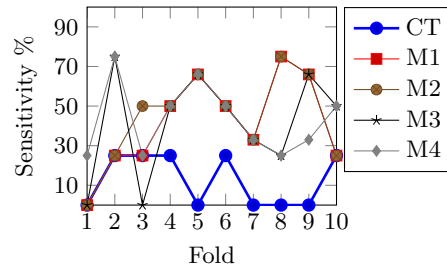


Fig. 5. Sensitivity of classification tree (CT) and each model (M) by fold.

4.2 Sensitivity

Argument-based models could outperform CT in all folds. Only model M3 had a lower sensitivity in one fold. This shows how the techniques are equivalent based on their accuracies but argument-based models have a better prediction capacity for death, which is highly important in this domain. Among argument-based models, folds 4 to 7 had the same sensitivity. Other folds presented a

mixed prediction, sometimes having better results for models M1-2 and other times for models M3-4. Again, this suggests how the skeptical approach might be as predictive as the cautious one, making the use of medium risk biomarkers (by models M1-2) questionable for the prediction of survival.

4.3 Discussion

The analysis of accuracy of the investigated models demonstrates that the prediction capacity of survival by argument-based models is in line with the classification tree approach. Sensitivity, however, suggests that more precise inferences of death can be done by argument-based models. One can argue that learning algorithms might not be suited for the proportion of biomarkers by records in this problem (51 biomarkers and 93 records), thus another knowledge-base approach should be employed for comparison purposes. However, to the best of our knowledge, no other knowledge-base approach has been applied in the biomarker domain. In fact, the use of other knowledge-base techniques such as expert systems and fuzzy logic systems are mentioned as future work.

In line with this, it is important to mention the interest in pursuing explanation systems [4] in the health-care sector opposed to learning techniques where outcomes are not fully comprehended by humans. Defeasible reasoning and argument-based models allow a clearer reasoning process under uncertainty and incomplete data, are not based upon statistic or probability and allow the comparison of different knowledge-bases [19]. Efforts have already been made in order to apply argument-based systems in the medical field [17], so it is expected that this integration might be enhanced by this research. For instance, take the knowledge-base applied here. Although it presents a large set of biomarkers it is not completely clear how they interact with each other. It is also not clear how the deduction from mortality risk to survival should be made. However, these are questions that are possibly not known even by domain experts. Nonetheless, two models or interpretations of this knowledge-base (cautious and skeptical) could be investigated and hopefully shed some light on these questions. For example, skeptical and cautious models presented similar performance, demonstrating that medium risk biomarkers did not have a significant contribution in the prediction of survival. Furthermore, each record analyzed has its own sub-argumentation framework, its own set of acceptable arguments and its own set of successful attacks. Such information can also be used for additional reasoning and improvement of biomarkers understanding towards survival.

5 Conclusion and future work

This research investigated the use of defeasible reasoning, formally implemented via computational argumentation theory, for the prediction of survival in elderly built upon evidence on biomarkers. The motivation for applying this technique comes from the lack of other studies employing knowledge driven approaches in the biomarker domain and also from the small amount of available data. A 5-layer schema was used to translate an extensive expert's knowledge-base

into arguments. These were used to infer death or survival of elderly patients. The main disadvantage of building argument-based models is the time required to translate a knowledge-base generally expressed in natural language terms, into formal rules. Findings suggest that argument-based models are a promising avenue for the investigation of the relationship between biomarkers and survival in elderly. A comparison between argument-based models, built upon this 5-layer schema, and a classification tree algorithm showed that the formers can achieve a similar accuracy than the latter but a better sensitivity. In addition, argument-based models are not based on statistic or probability and can reason with unclear and incomplete data. Future work will be focused on the application of different knowledge driven approaches, including expert systems and fuzzy logic. Moreover, given the uncertainty of the application (inference of survival of elderly using biomarkers) the same methodology of this study (5-layer schema) is going to be employed with different knowledge-bases designed by distinct experts.

Acknowledgments

Lucas Middeldorf Rizzo would like to thank CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) for his Science Without Borders scholarship, proc n. 232822/2014-0.

References

1. Barron, E., Lara, J., White, M., Mathers, J.C.: Blood-borne biomarkers of mortality risk: systematic review of cohort studies. *PloS one* 10(6), e0127550 (2015)
2. Besnard, P., Hunter, A.: A logic-based theory of deductive arguments. *Artificial Intelligence* 128(1-2), 203–235 (2001)
3. Bryant, D., Krause, P.: A review of current defeasible reasoning implementations. *The Knowledge Engineering Review* 23(3), 227–260 (2008)
4. Core, M.G., Lane, H.C., Van Lent, M., Gomboc, D., Solomon, S., Rosenberg, M.: Building explainable artificial intelligence systems. In: *AAAI*. pp. 1766–1773 (2006)
5. Craven, R., Toni, F., Cadar, C., Hadad, A., Williams, M.: Efficient argumentation for medical decision-making. In: *KR* (2012)
6. De Ruijter, W., Westendorp, R.G., Assendelft, W.J., den Elzen, W.P., de Craen, A.J., le Cessie, S., Gussekloo, J.: Use of framingham risk score and new biomarkers to predict cardiovascular mortality in older people: population based observational cohort study. *Bmj* 338, a3083 (2009)
7. Dipnall, J.F., Pasco, J.A., Berk, M., Williams, L.J., Dodd, S., Jacka, F.N., Meyer, D.: Fusing data mining, machine learning and traditional statistics to detect biomarkers associated with depression. *PloS one* 11(2), e0148195 (2016)
8. Dung, P.M.: On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial intelligence* 77(2), 321–358 (1995)
9. García, D., Simari, G.: Strong and weak forms of abstract argument defense. *Computational Models of Argument: Proceedings of COMMA 2008* 172, 216 (2008)
10. Glasspool, D., Fox, J., Oettinger, A., Smith-Spark, J.: Argumentation in decision support for medical care planning for patients and clinicians. In: *AAAI Spring Symposium: Argumentation for Consumers of Healthcare*. pp. 58–63 (2006)

11. Hunter, A., Williams, M.: Argumentation for aggregating clinical evidence. In: Tools with Artificial Intelligence (ICTAI), 2010 22nd IEEE International Conference on. vol. 1, pp. 361–368. IEEE (2010)
12. Lee, S.J., Lindquist, K., Segal, M.R., Covinsky, K.E.: Development and validation of a prognostic index for 4-year mortality in older adults. *Jama* 295(7), 801–808 (2006)
13. Lloyd-Jones, D., Adams, R., Carnethon, M., De Simone, G., Ferguson, T.B., Flegal, K., Ford, E., Furie, K., Go, A., Greenlund, K., et al.: Heart disease and stroke statistics2009 update: a report from the american heart association statistics committee and stroke statistics subcommittee. *Circulation* 119(3), e21–e181 (2009)
14. Longo, L.: Formalising human mental workload as non-monotonic concept for adaptive and personalised web-design. In: International Conference on User Modeling, Adaptation, and Personalization. pp. 369–373. Springer (2012)
15. Longo, L.: A defeasible reasoning framework for human mental workload representation and assessment. *Behaviour and Information Technology* 34(8), 758–786 (2015)
16. Longo, L.: Argumentation for knowledge representation, conflict resolution, defeasible inference and its integration with machine learning. In: Machine Learning for Health Informatics, pp. 183–208. Springer (2016)
17. Longo, L., Dondio, P.: Defeasible reasoning and argument-based systems in medical fields: An informal overview. In: Computer-Based Medical Systems (CBMS), 2014 IEEE 27th International Symposium on. pp. 376–381. IEEE (2014)
18. Longo, L., Hederman, L.: Argumentation theory for decision support in healthcare: a comparison with machine learning. In: International Conference on Brain and Health Informatics. pp. 168–180. Springer (2013)
19. Longo, L., Kane, B., Hederman, L.: Argumentation theory in health care. In: Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on. pp. 1–6. IEEE (2012)
20. Matt, P.A., Morgem, M., Toni, F.: Combining statistics and arguments to compute trust. In: 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, Canada. vol. 1, pp. 209–216. ACM (May 2010)
21. Pollock, J.L.: Defeasible reasoning. *Cognitive science* 11(4), 481–518 (1987)
22. Prakken, H.: An abstract framework for argumentation with structured arguments. *Argument and Computation* 1(2), 93–124 (2010)
23. Pritzker, K.P., Pritzker, L.B.: Bioinformatics advances for clinical biomarker development. *Expert opinion on medical diagnostics* 6(1), 39–48 (2012)
24. Rizzo, L., Dondio, P., Delany, S.J., Longo, L.: Modeling Mental Workload Via Rule-Based Expert System: A Comparison with NASA-TLX and Workload Profile, pp. 215–229. Springer International Publishing, Cham (2016)
25. Rizzo, L., Longo, L.: Representing and inferring mental workload via defeasible reasoning: a comparison with the nasa task load index and the workload profile. In: 1st Workshop on Advances In Argumentation In Artificial Intelligence. pp. 126–140 (2017)
26. Slater, T., Bouton, C., Huang, E.S.: Beyond data integration. *Drug discovery today* 13(13), 584–589 (2008)
27. Strimbu, K., Tavel, J.A.: What are biomarkers? *Current Opinion in HIV and AIDS* 5(6), 463 (2010)
28. Swan, A.L., Mobasheri, A., Allaway, D., Liddell, S., Bacardit, J.: Application of machine learning to proteomics data: classification and biomarker identification in postgenomics biology. *Omics: a journal of integrative biology* 17(12), 595–610 (2013)