On the Need of Interpretability for Biomedical Applications: Using Fuzzy Models for Lung Cancer Prediction with Liquid Biopsy

Nicolas Potie
Dept. of Computer Science and A.I.
University of Granada
Granada, Spain
npotie@ugr.es

Stavros Giannoukakos
Dept. of Genetics
University of Granada
Granada, Spain
sgiannoukakos@ugr.es

Michael Hackenberg
Dept. of Genetics
University of Granada
Granada, Spain
hackenberg@go.ugr.es

Alberto Fernandez
Dept. of Computer Science and A.I.
University of Granada
Granada, Spain
alberto@decsai.ugr.es

***Abstract*—In the latter years, we are witnessing a movement from the standard Data Mining towards a more profitable and challenging scenario known as Data Science. It can be defined as a set of quantitative and qualitative approaches that are applied to current relevant problems. In order to be able to "dig" to the deepest level considering the whole information available, the knowledge domain and the analysis of the data must have a strong synergy.**

There are many fields of application where it is necessary, if not essential, to give an explanation of the phenomenon under study. It is no longer enough to simply apply a Machine Learning model, but it must be comprehensible in order to provide a real decision support system. For this reason, a strong movement has emerged in favour of the eXplainable Artificial Intelligence that aims to respond to the "how" and "why" of the operation of automatic models.

In this work, our objective is to show the benefits of one of the learning paradigms of Computational Intelligence: Fuzzy Rule Based Systems and Evolutionary Fuzzy Systems. To this end, we focus on biomedical applications by presenting a case study based on lung cancer prediction from samples taken by liquid biopsy. Liquid biopsy enable us to study genomic alterations for each individual independently, a step towards personalised medicine. The results show the goodness of the solution based on Evolutionary Fuzzy Systems in terms of interpretability and comprehensibility, obtaining a low number of rules with less than 3 fuzzy linguistic labels per antecedent.

***Index Terms*—eXplainable Artificial Intelligence, Evolutionary Fuzzy Systems, Lung Cancer, Liquid Biopsy, Interpretability**

I. Introduction

Technological developments around Big Data and intelligent data analysis have recently given rise to the term Data Science. It is an emerging area of work as a natural evolution of the data mining field, that encompasses all the technologies related to Big Data to deal with the collection, preparation, analysis, visualization, management and conservation of large collections of information [1].

This work have been partially supported by the Spanish Ministry of Science and Technology under project TIN2015-68454-R, including European Regional Development Funds; and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions (grant agreement ELBA, 765492)

978-1-5386-1728-1/19/$31.00 ©2019 IEEE

It is known that data alone do produce information but not knowledge. The real value lies in the possibility of extracting useful information for decision making or the exploration and comprehension of the phenomenon that produced the data. Due to this, there is a need to propose methodologies for the intelligent analysis of data, which will allow us to obtain useful knowledge from them. To this end, data analysis methods, such as statistics and data mining, are involved in the Data Science process, mainly via Machine Learning (ML) techniques [2].

Generally speaking, an ML model is said to be useful when it effectively summarizes the underlying data, i.e. it provides good predictive performance. For this reason, we are witnessing a movement towards black-box ML approaches with an excellent ability to learn accurately from the input data, but that are not able to inform how they arrive at a certain decision [3]. However, there are several scenarios, such as medical diagnosis, where other criteria such as confidence, robustness, reliability or trust are necessary [4], [5].

Most of the aforementioned criteria often cannot be completely quantified, but if the system is **interpretable**, i.e. if it can **explain** its reasoning, it can be verified whether that reasoning is sound with respect to these auxiliary criteria. This can be compiled into a fundamental property of what was defined as eXplainble Artificial Intelligence (XAI) [3]. In this context, the fuzzy set theory might be regarded as a valuable tool [6]. Its advantages are clear, including the use of linguistic labels as a natural knowledge representation allowing the direct human semantic interaction [7]. In addition, from a learning perspective, translating the input features into fuzzy variables with fuzzy membership functions permits obtaining smoothed descriptive models that adapt well to data with a certain degree of uncertainty.

An ML paradigm that presents a strong synergy between accuracy and XAI are Evolutionary Fuzzy Systems (EFS) [8], which combine a good degree of understandability, comprehensibility, and explainability associated with Fuzzy Rule Based Systems (FRBS), and the potential of Evolutionary Algorithms (EAs) as the optimization technique for improving FRBSs.

In this work, our objective is to stress the importance of EFS to provide real XAI systems. To do so, we will make use of a biomedical case study based on lung cancer prediction. Specifically, we will consider the information extracted from a novel and very promising biotechnology known as liquid biopsy [9]. The underlying mechanism captures cell information from the blood stream of the patient, and translates it into gene expression data, leading to thousands of genes under study.

Undoubtedly, medical doctors need a compact ML system for it to be truly useful. This compactness is given in both in terms of rules and gene attributes that explain the prediction. To this end, in our work we will stress the quality of EFS for achieving interpretable and comprehensible decision support systems in biomedical applications, namely via the Fuzzy Associative Rule-based Classifier for High Dimensional (FARC-HD) algorithm [10]. In the experimental study, we will confirm the superior capabilities of FARC-HD versus state-of-the-art rule learning models such as Decision Trees (DT) [11] and Random Forest [12].

The rest of this contribution is arranged as follows. In Section II we will introduce the paradigm of EFS and its relationship with the achievement of XAI in ML. Then, Section III will describe the characteristics of the biomedical case study that will serve us to show the good behavior of EFSs. The experimental study with the results on accuracy and interpretability will be shown in Section IV. Finally, Section V will summarize and conclude the paper.

II. Evolutionary Fuzzy Systems for Rule Learning: The Quest for eXplainable Artificial Intelligence

In the introduction of this contribution, we stressed that both accuracy and interpretability/explainability should be considered when determining what kind of ML techniques will be used to address a specific problem. In this case, the hitch is that usually both are in conflict, leading to a complex task. [6].

A compromise solution that is often considered is to take advantage of rule-based systems. This type of paradigm allows auditing the extracted knowledge with a double objective. On the one hand, obtaining a more direct explanation of the cognition process carried out by the system. On the other hand, being able to trust in the description of the rules and their relationship with the problem that is aimed to be solved.

FRBSs are composed of fuzzy IF-THEN rules where both antecedents and consequents usually contain fuzzy sets. The main components of any FRBS are the knowledge base (KB) and the inference engine module. The KB comprises all the fuzzy rules within a rule base (RB), and the definition of the fuzzy sets in the data base (DB). The inference engine includes a fuzzification interface, an inference system, and a defuzzification interface.

To boost the behavior and predictive abilities of FRBS, EFS are developed on top of them. The goal of this type of approaches is to learn or to tune the components of the FRBS using an evolutionary process commonly taken from available data. In this sense, rule sets, membership functions, and many other features of an FRBS can be easily optimized via EAs [8].

The goodness of EFS is mainly based on two characteristics. On the one hand, the inherent interpretability of the system, namely the comprehensibility associated with the use of a simple description mechanism in the form of fuzzy linguistic rules, as well as the understandability of the rule-based system and the inference procedure. On the other hand, the robustness of modeling scenarios that are difficult to represent with other types of paradigms, especially when users must deal with the lack of data or uncertainty in the definition of the input data.

However, it must be emphasized that FRBSs must remain simple and understandable since they are not interpretable *per se* [13], [14]. It is important to take into account different issues in order to obtain the FRBSs that represent knowledge that can be easily understandable by humans. Among others, the rule base compactness or the semantic comprehensibility of the fuzzy partitions must be stressed. Moreover, the EFSs must be properly designed to obtain the desired trade-off between accuracy and explainability for the problem at hand.

III. A Case Study on Lung Cancer Prediction via Liquid Biopsy

Cancer is a term that englobes a wide range of heterogeneous diseases that share two main properties: i) uncontrolled abnormal cell growth and ii) capacity to invade other parts of the body. The molecular cause are mutations in important genes like those involved in repair mechanisms, cell growth or cell cycle control. Both the DNA mutations in important driver genes and the changes in expression levels are used frequently as attributes in ML approaches to detect the presence of malignant cells or to predict the outcome of a certain treatment.

In this Section, our objective is to present the main features of lung cancer diagnosis (Section III-A). Then, we focus on the description of liquid biopsy as a novel technique to detect cancer from blood samples (Section III-B). Afterwards, we will introduce the dataset employed as case study, together with its main characteristics (Section III-C). Finally, we provide some comments on the bioinformatics methodology followed to prepare the dataset prior to the applications of the ML models (Section III-D).

A. Diagnosis of lung cancer: past and present

Lung cancer is a type of cancer located in the respiratory system. It was the leading cause of death worldwide in 2018 amongst all cancers and the third cause of death in the USA [15]. Cancer genome alterations have the potential to serve as a powerful cancer diagnostic tool. Nevertheless, such knowledge is only being used to better understand the biology of it and choose an appropriate treatment. This is primary because the sequencing of the genome (or the part of it) is traditionally based on tissues biopsy. Thus, the cancer status of the individual as well as its location must be known.

In the case of lung cancer, unfortunately individuals do not exhibit any symptoms at the initial stages of the disease, implying a difficulty of an early detection. Imaging diagnostic is a preferred choice nowadays. To this end, doctors use the newest version of computational tomography scan called low-dose computed tomography (LDCT) which uses less radiation [16].

However, imaging diagnostics has several drawbacks, as pointed out below:

1) The first one is the high false positive rate that varied from 3-30% in Randomised Cohort Trials and from 5% to 51% in cohort studies [17]. This is due to the fact that LDCT can detect benign nodules that are not related to any cancer. Taking into account the invasive nature of the possible treatments, it implies a large cost for the health of these individuals.
2) The second drawback is the invasiveness due to the radioactive dose of LCDT per examination, namely 1.5 mSv [17].
3) The third disadvantage is the price. The annual screening costs between $126,000 - $169,000 per Quality-Adjusted Life-Year (QUALY) for lung cancer patient in USA [18]. This value is beyond the $100,000 threshold that is considered to be as cost effective.

B. Liquid biopsy, a potential game changer

Liquid biopsy is becoming a strong potential new way of diagnosis, cost-effective and minimally invasive. Isolation of biological components is now possible and under research. It aims to study any biological component related to cancer from a blood sample of the individual.

It is known that primary tumour interacts with the blood stream in different ways [9]. It needs nutrients transported by the blood stream to survive, and therefore we may find Circulating Tumour Cells (CTCs), i.e. cells originated from the tumour and made their way to the blood stream. The term "tumour-Educated Platelets" (TEPs) is used when platelets and tumours interacts and exchange information within the blood stream [19], [20]. TEPs can provide tumour's genetic material, more precisely RNA, which can lead to the study of the tumours genetic makeup/gene expression. In addition to the former, their size makes them a suitable solution to be easily extracted and analyzed.

However, lots of challenges remain in cancer diagnosis. One of them is the total transparency of the ML algorithm used to detect cancer. It is of extreme importance to assist clinicians in their work. Strong evidence must be outputted when a sample is classified as cancer. Like lung cancer nodules can be seen from scanner images, cancer driving mutations, genes expression or other features/sources must be explainable to be accepted. It is more important that liquid biopsy pipeline is nearly fully computer-based after isolation of biological components of interests. Thanks to this, **personalized medicine** is becoming a reality [21].

C. Dataset of the study

As commented previously, genomic alterations of cancer can now be studied for diagnosis by means of liquid biopsy. We must take into account that lung cancer is a highly heterogeneous disease. Two major molecular types can be distinguished: Small Cell Lung Cancer (SCLC) and the most commonly Non-Small Cell Lung cancer (NSCLC). In NSCLC, three majors subtypes have been discovered: adenocarcinoma, squamous cell carcinoma and large cell cancer [22]. The current study is focused on NSCLC without taking the subtypes into account.

Best et al. created in 2015 *ThromboSeq*, an ML pipeline detecting cancer using Particle Swarm Optimisation (PSO) [19]. In 2017, they enhanced their pipeline by integrating computed differentially spliced RNA from tumour Educated Platelets (TEPs) [20]. In our study, we use the gene expression dataset of 2017 freely available on Gene Expression Omnibus (GEO) of the NCBI[1] under accession code GSE89843 [20]. The dataset contains 779 individuals. 402 were diagnosed with lung cancer and the remaining 377 are non-cancerous individuals.

The total blood TEP RNA was isolated from the cancer patients and the total platelet RNA from the non-cancer. Furthermore, the RNA was then subjected to complementary DNA (cDNA) synthesis and amplification according to the standard protocol. This cDNA is of high importance as it is the input material for Next Generation Sequencing (NGS) methods, i.e. to extract the whole information from the genes. The first step is the library preparation which includes fragmentation of the cDNA, barcode labeling (important for the sequencing process), and product PCR amplification. Once the library preparation is completed, the library was sequenced with the use of a Illumina HiSeq 2500 machine. The result of the sequencing, i.e sequence of each cDNA, is written in a file. The data analysis can start by removing adaptors for each cDNA and filtered out low quality ones. Then aligning each cDNA to the reference genome (hg19) to identify which sequence belongs to which location in the genome. Lastly, for each gene, sequence reads are aligned to it. Transcription levels of the gene are directly proportional to the read counts.

D. Bioinformatic methodology before classification

The selected dataset for the study [20] contains 779 samples with related 4,635 values representing the abundance of the gene transcripts (features). This type of dataset is called gene expression count matrix due to the expression of genes are represented by the number of RNA sequences belonging to each gene (discrete values). Gene expression is a regulatory process by which information goes through the following stages: DNA-RNA-proteins. The expression of a given gene is frequently estimated by the abundance of the RNA that it is transcribed of that gene by the cells. Thus, the more RNA of a gene A is detected, the more gene A is expressed. The

[1]https://www.ncbi.nlm.nih.gov/gds

human genome contains about 21,000 genes, but only 4,635 have been monitored in the TEPs.

The pipeline continues with the following steps. Firstly low count genes have been filtered with the method of Chen & Smyth [23], leading to a subset of 1,585. Secondly, the count matrix was normalised by trimmed mean of Mvalues (TMM). Before classification, a feature selection preprocessing step was used, according to the gini score computation [24]. This score was obtained for every gene, then taking as threshold the 10% of the maximum value found over all features.

In summary, the characteristics of the dataset used in the experimental study are the following ones. There is a total number of 779 instances divided into two classes, namely 377 samples that represent non-cancer patients, and 402 samples for lung cancer cases. The total amount of input attributes used for training comprises about 200 genes, depending on how feature selection is computed on the different partitions; whereas test partitions contain the whole 1,585 genes. Finally, the range of each gene/attribute ranges approximately between 0.0 and 2.0 after the TMM normalization step.

IV. Experimental Study

In this section, we will show the good properties of EFS to address the lung cancer prediction problem. To this end, we will first present the methods and parameters that will be employed to extract the knowledge from the data (Subsection IV-A). Then, we will show the experimental results in terms of accuracy and interpretability, and we will carry out an empirical evaluation of the rule-based methods in accordance with these metrics (Subsection IV-B). Finally, we present a brief discussion on the issues of interpretability achieved by the EFS model in the selected biomedical application (Subsection IV-C).

A. Methods and parameters

As introduced at the beginning of this work, the EFS method that we propose to apply for reaching the a high degree of interpretability and explainability is FARC-HD. The reasons behind this choice are the good capabilities towards this objective that are provided by this classification algorithm, such as a low number of rules and antecedents, among others.

Regarding state-of-the-art methods in ML that are applied to lung cancer classification, they may vary from one study to another [25]. In a recent review on the topic, authors stressed two points [26]: on the one hand "simple is often better" and, on the other hand, "ensemble methods produce robust results". Taking this into account, we will make use of both decision trees and random forest for the experimental analysis.

Below, we provide a brief description of these learning algorithms, whose implementations have been taken from two well-known ML software packages, namely KEEL [27], and scikit-learn [28].

1) **EFS**: FARC-HD [10] extracts fuzzy association rules by limiting the order of the associations. The former constraint is used as a "pre-screening" for high quality candidate rules during learning, which allows the achievement of more interpretable rules, i.e. a low number of rules with few antecedents. Finally, an evolutionary rule selection and lateral tuning procedure is applied for improving the classification accuracy of the final rule set.
   A standard configuration has been selected. This includes five fuzzy sets per attribute, with product t-norm and winning rule inference to determine the output class. Minimum support and confidence were set to 0.05 and 0.8 respectively. The maximum depth of the tree when discovering frequent items was set to 3, and the pre-screening parameter (k) was set to 2.
2) **Decision Tree (DT)**: CART [11] creates a binary tree, finding for each node (i.e. in a greedy manner) the feature and threshold that yield the largest information gain at each node for categorical targets.
   We have selected the gini-index criterion for splitting nodes, with a maximum depth of 5 levels for the tree (to ensure a certain level of interpretability). The minimum number of samples to make a split was set to 3, and the minimum number of samples for a leave is 1.
3) **Ensemble System**: Random Forest [12] is composed of a set of different DTs, each of which is built by the CART algorithm. In order to consider diversity among these different trees or estimators, a bagging approach is considered, in which examples are randomly taken with replacement into each "bag" or new training set. To add even more diversity, a random feature selection is applied within each bag.
   The parameters in this case are exactly the same as in the case of the DT. By being an ensemble system 10 different estimators have been considered to build the forest, and a number of `log2` features are selected within each bag.

Finally, for carrying out a proper validation of the experimental results, the data were partitioned using a 5-fold cross validation procedure, that is, using the 80% of the data for training and the remaining 20% for testing, and repeating this process 5 times.

B. Results and analysis

The accuracy values obtained by each classification system and the information regarding interpretability are shown in Table I. Specifically, we have included both the training and test prediction scores, the number of nodes for the solutions based on DT, and the number of rules and antecedents. In the case of DTs, the number of rules is equal to the number of total leaves, whereas the number of antecedents is computed as the average depth of the tree.

Observing these results we may excel the goodness of the FARC-HD EFS approach, as it obtains the best accuracy overall in the test partitions. Furthermore, the robustness of the EFS solution is also emphasized by the low deviation shown in the accuracy values with respect to the DT and RF.

Focusing on the main topic of this work, namely the capabilities of EFS to achieve XAI ML models, we must

TABLE I
EXPERIMENTAL RESULTS FOR ACCURACY AND INTERPRETABILITY ON THE LUNG CANCER DATA. RESULTS OF THE EFS, DT AND ENSEMBLE SYSTEM ARE SHOWN FROM LEFT TO RIGHT.

	EFS	***DT***	***Ensemble***
Accuracy_{Tr}	0.9143 ± 0.0075	0.8280 ± 0.0174	**0.9974 ± 0.0022**
Accuracy_{Tst}	**0.7819 ± 0.0382**	0.7205 ± 0.0751	0.7744 ± 0.0587
#Nodes	-	45.40 ± 7.92	44.96 ± 1.75
#Rules	23.40 ± 3.71	23.20 ± 3.96	22.98 ± 0.87
#Antecedents	2.7283 ± 0.1061	3.6787 ± 0.1447	3.6788 ± 0.0407

also conclude that FARC-HD outputs a very compact and interpretable model, comparable to the one obtained by DT in terms of absolute number of rules. In the case of RF, each tree has a similar number of rules (branches) than for FARC-HD and CART, but we must acknowledge that we are using 10 different estimators, thus losing a high degree of explainability, i.e. the inference mechanism is impossible to follow.

Additionally, the number of antecedents per rule in the case of FARC-HD is below 3 on average. That means that every rule can be easily managed by the end-user or expert, comprising a few features (genes in this case) and providing very useful knowledge. As an example, Fig. 1 illustrates a complete RB obtained by FARC-HD, where we may observe that 13 and 11 rules are considered for predicting "Non Cancer" and "Cancer" conditions, respectively. Another advantage is related to the use of linguistic variables so that no crisp threshold values are used on the antecedent conditions, allowing a closer semantic comprehension of the model.

```
R1: IF NREP IS med AND YBX1 IS low AND LAPTM4B IS med: NONCANCER CF: 0.875
R2: IF NSA2 IS verylow AND HNRNPR IS verylow AND PRKCB IS verylow: NONCANCER CF: 0.84
R3: IF NSA2 IS verylow AND PRKCB IS verylow AND LINC00892 IS med: NONCANCER CF: 0.875
R4: IF NSA2 IS verylow AND COX7B IS high AND ZNF542 IS verylow: NONCANCER CF: 0.92
R5: IF CLINT1 IS verylow AND PPCS IS med AND RBBP4 IS verylow: NONCANCER CF: 0.80
R6: IF LTBP3 IS med AND SLA IS verylow AND EVI5 IS low: NONCANCER CF: 0.75
R7: IF LTBP3 IS med AND SLA IS verylow AND DCTN1 IS verylow: NONCANCER CF: 0.80
R8: IF HNRNPR IS verylow AND PRKCB IS verylow: NONCANCER CF: 0.85
R9: IF HNRNPR IS verylow AND ATPAF1 IS low AND PARVB IS med: NONCANCER CF: 0.825
R10: IF ENKUR IS low AND LINC01151 IS med: NONCANCER CF: 0.80
R11: IF PARVB IS verylow AND FNBP1 IS low AND PTPRF IS low: NONCANCER CF: 0.80
R12: IF COX7B IS high AND ZNF542 IS verylow: NONCANCER CF: 0.925
R13: IF RBBP4 IS verylow AND LINC00892 IS high AND RP11-556E13.1 IS verylow: NONCANCER CF:
0.825

R14: IF DHRS7 IS verylow AND RPL19 IS med AND UXS1 IS verylow: CANCER CF: 0.975
R15: IF RPN2 IS verylow AND CCDC53 IS med AND RASA2 IS med: CANCER CF: 0.925
R16: IF RTN2 IS verylow AND WHAMMP2 IS verylow: CANCER CF: 0.80
R17: IF MAST4 IS verylow AND PPP1R12C IS verylow AND LINC01088 IS med: CANCER CF: 0.875
R18: IF LGALS3BP IS med AND WASF1 IS med AND PARD3 IS low: CANCER CF: 0.825
R19: IF KCTD20 IS verylow AND BAD IS verylow AND LINC01151 IS low: CANCER CF: 0.725
R20: IF NCOA2 IS verylow AND PARVB IS verylow AND PTPRF IS verylow: CANCER CF: 0.90
R21: IF DCK IS verylow AND YBX1 IS med AND AC079807.4 IS verylow: CANCER CF: 0.875
R22: IF DCK IS low AND PFKM IS verylow AND TCP1 IS verylow: CANCER CF: 0.675
R23: IF PARK7 IS verylow AND MEF2C-AS1 IS verylow AND MTRNR2L2 IS verylow: CANCER CF: 0.80
R24: IF CREB3 IS verylow AND RBM3 IS med: CANCER CF: 0.925
```

Fig. 1. Example of RB obtained by FARC-HD algorithm.

Finally, regarding the number of genes used within the model, about 25 genes are considered for "Non Cancer" condition and 30 genes for the "Cancer" prediction. This implies a very low number if we contrast it versus the original number of genes selected at the beginning of the data extraction procedure, i.e. more than 4,000. Furthermore, only 3 genes are overlapped in the "Cancer" vs. "Non Cancer" rules (PARVB, PTPRF and LINC01151), but the linguistic label associated in each case is very different, i.e. "Medium" vs. "Very low".

C. Discussion on interpretability issues

Throughout this work contribution we have stressed the relevance of EFS in the context of the emerging field of XAI [6], [29]. It is straightforward to acknowledge the need of an interpretable system for any diagnosis application. We must take into account that the medical doctor must have the central role within this task, using the AI model only as a decision support system.

With this purpose, we have shown the benefits of a predictive model obtained by FARC-HD algorithm in the context of lung cancer prediction. Below, we enumerate the good capabilities and advantages of this type of FRBCS with respect to other type of ML solutions:

1) First, the benefits of using natural language for explaining any choice taken by the AI system must be stressed [14]. In this particular case, this is related with the use of linguistic terms in the antecedents of the rules, which should be meaningful for physicians [13].
2) Second, the working procedure of FARC-HD is designed to obtain a compact fuzzy rule-based system in both terms of number of rules and number of antecedents per rule. The first point, i.e. low number of rules, is achieved in two different stages of the algorithm: first due to a filtering procedure based on the actual rule coverage, and second by means of a evolutionary rule selection mechanism. The second point, namely a reduced antecedent length, is due to the associative rule learning design, that is, selecting only those combination of input attributes that imply a high support and confidence values.
3) Finally, we must focus on the inference mechanism, which determines the reason why for the prediction. In this sense, using a winning rule scheme implies that the decision only depends on a single antecedent description. In addition to the former, the use of rule weights is also positive as it adds a confidence degree that may help the human user to have a certain trust on the system.

Thanks to all these characteristics, we must consider the fuzzy classifier from FARC-HD to be a very suitable solution comprising both a good predictive performance and interpretability. The computational cost involved in the optimization step of the EFS has also shown to be necessary to achieve an appropriate balance between both former capabilities. In this sense, developing ad-hoc multi-objective approaches and/or using specific feature selection techniques for reducing the number of involved genes to the minimum, can be regarded as some interesting topics to be taken into account as future research lines.

V. CONCLUDING REMARKS

New applications addressed by Data Science are no longer focused on achieving the highest accuracy, but also to make it explainable for researchers and practitioners. Along with the different paradigms in ML, those based in EFS have the advantage of preserving the comprehensibility of fuzzy

systems, together with a boost in prediction via evolutionary optimization. This way, it allows handing XAI learning models including transparency, understanding and comprehensibility.

In this work, we have included a case study on a biomedical scenario. Specifically, we have selected a lung cancer prediction problem based on liquid biopsy, and we have applied different rule learning methods to check whether EFSs can provide a good trade-off between accuracy and interpretability. Experimental results have shown that EFS are a highly recommended solution, as they achieved the highest test accuracy, together with the best interpretability features. Among others, we have stressed a low number of rules, also with few linguistic fuzzy antecedents, and an easy-to-follow inference mechanism.

For the first time, a cancer diagnostic tool such as liquid biopsy can use the powerful and relevant knowledge contained in a genome of an individual. Abnormalities detected can be better studied than standard X-ray scan outputting a lot of genomic data. As pointed out above, we have stressed how the good capabilities of EFS allow using the former genomic information to extract useful and comprehensible knowledge. This synergy between liquid biopsy biotechnology and XAI will surely lead to personalised interpretable medicine, ensuring adequate and better diagnostic tools and treatments.

REFERENCES

[1] S. Ramírez-Gallego, A. Fernández, S. García, M. Chen, and F. Herrera, "Big data: Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce." *Information Fusion*, vol. 42, pp. 51–61, 2018.

[2] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, 2015.

[3] D. Castelvecchi, "Can we open the black box of AI?" *Nature*, vol. 538, no. 7623, pp. 20–23, 2016.

[4] A. Tan, D. Naiman, L. Xu, R. Winslow, and D. Geman, "Simple decision rules for classifying human cancers from gene expression profiles," *Bioinformatics*, vol. 21, no. 20, pp. 3896–3904, 2005.

[5] S. F. Bryce Goodman, "European union regulations on algorithmic decision-making and a "right to explanation"," *AI Magazine*, vol. 38, no. 3, 2017. [Online]. Available: https://arxiv.org/abs/1606.08813v3

[6] A. Fernandez, M. J. del Jesus, O. Cordon, F. Marcelloni, and F. Herrera, "Evolutionary fuzzy systems for explainable artificial intelligence: Why, when, what for, and where to?" *IEEE Computational Intelligence Magazine*, vol. 14, no. 1, pp. 69–81, 2019.

[7] M. J. Gacto, R. Alcala, and F. Herrera, "Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures," *Inform. Sciences*, vol. 181, no. 20, pp. 4340–4360, 2011.

[8] A. Fernandez, V. Lopez, M. J. del Jesus, and F. Herrera, "Revisiting evolutionary fuzzy systems: Taxonomy, applications, new trends and challenges," *Knowlegde Based Systems*, vol. 80, pp. 109–121, 2015.

[9] I. Domínguez-Vigil, A. Moreno-Martínez, W. J.Y., M. Roehrl, and H. Barrera-Saldaña, "The dawn of the liquid biopsy in the fight against cancer." *Oncotarget*, vol. 9, no. 2, pp. 2912–2922, jan 2017.

[10] J. Alcala-Fdez, R. Alcala, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 857–872, 2011.

[11] L. Breiman, J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*. Chapman and Hall (Wadsworth and Inc.), 1984.

[12] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[13] J. M. Alonso, C. Castiello, and C. Mencar, "Interpretability of fuzzy systems: Current research trends and prospects," in *Handbook of Computational Intelligence*, J. Kacprzyk and W. Pedrycz, Eds. Springer, 2015, pp. 219–237.

[14] J. M. Alonso, A. Ramos-Soto, E. Reiter, and K. van Deemter, "An exploratory study on the benefits of using natural language for explaining fuzzy rule-based systems," in *FUZZ-IEEE*. IEEE, 2017, pp. 1–6.

[15] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, nov 2018.

[16] M. Oudkerk et al., "European position statement on lung cancer screening." *The Lancet. Oncology*, vol. 18, no. 12, pp. e754–e766, 2017.

[17] P. B. Bach, J. N. Mirkin, T. K. Oliver, C. G. Azzoli, D. A. Berry, O. W. Brawley, T. Byers, G. A. Colditz, M. K. Gould, J. R. Jett, A. L. Sabichi, R. Smith-Bindman, D. E. Wood, A. Qaseem, and F. C. Detterbeck, "Benefits and harms of CT screening for lung cancer: a systematic review." *JAMA*, vol. 307, no. 22, pp. 2418–29, jun 2012.

[18] P. M. McMahon, C. Y. Kong, C. Bouzan, M. C. Weinstein, L. E. Cipriano, A. C. Tramontano, B. E. Johnson, J. C. Weeks, and G. S. Gazelle, "Cost-effectiveness of computed tomography screening for lung cancer in the United States." *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*, vol. 6, no. 11, pp. 1841–8, nov 2011.

[19] M. Best et al., "RNA-Seq of Tumor-Educated Platelets Enables Blood-Based Pan-Cancer, Multiclass, and Molecular Pathway Cancer Diagnostics," *Cancer Cell*, vol. 28, no. 5, pp. 666–676, nov 2015.

[20] M. G. Best et al., "Swarm Intelligence-Enhanced Detection of Non-Small-Cell Lung Cancer Using Tumor-Educated Platelets." *Cancer cell*, vol. 32, no. 2, pp. 238–252.e9, aug 2017.

[21] M. Hamburg and F. Collins, "The path to personalized medicine," *New England Journal of Medicine*, vol. 363, no. 4, pp. 301–304, 2010.

[22] D. Sharma, T. G. Newman, and W. S. Aronow, "Lung cancer screening: history, current perspectives, and future directions." *Archives of medical science : AMS*, vol. 11, no. 5, pp. 1033–43, oct 2015.

[23] Y. Chen, A. T. L. Lun, and G. K. Smyth, "From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline." *F1000Research*, vol. 5, p. 1438, 2016.

[24] Y. Saeys, T. Abeel, and Y. Van De Peer, "Robust feature selection using ensemble feature selection techniques," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5212 LNAI, no. PART 2, pp. 313–325, 2008.

[25] J. Ko, S. Baldassano, P.-L. Loh, K. Kording, B. Litt, and D. Issadore, "Machine learning to detect signatures of disease in liquid biopsies-a user's guide," *Lab on a Chip*, vol. 18, no. 3, pp. 395–405, 2018.

[26] D. Camacho, K. Collins, R. Powers, J. Costello, and J. Collins, "Next-generation machine learning for biological networks," *Cell*, vol. 173, no. 7, pp. 1581–1592, 2018.

[27] I. Triguero, S. González, J. M. Moyano, S. Garcia, J. Alcala-Fdez, J. Luengo, A. Fernandez, M. J. del Jesus, L. Sánchez, and F. Herrera, "KEEL 3.0: An open source software for multi-stage analysis in data mining," *International Journal of Computational Intelligence Systems*, vol. 10, pp. 1238–1249, 2017.

[28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[29] J. M. Alonso, C. Castiello, and C. Mencar, "A bibliometric analysis of the explainable artificial intelligence research field." in *IPMU (1)*, ser. Communications in Computer and Information Science, J. Medina, M. Ojeda-Aciego, J. L. V. Galdeano, D. A. Pelta, I. P. Cabrera, B. Bouchon-Meunier, and R. R. Yager, Eds., vol. 853. Springer, 2018, pp. 3–15.