Deriving and Visualizing Predictive Rules for Disease Risk: A Transparent Approach to Medical AI

Enrico Sciacca[1], Giorgia Grasso[1], Damiano Verda[1] and Enrico Ferrari[1,*]

[1]Rulex Innovation Labs, Via Felice Romani 9/2, Genova, Italy

Abstract

The comprehensibility of solutions is critical in AI-supported decision-making processes, particularly in the healthcare sector, where the outcomes of AI techniques must be explained to clinicians to enable informed and reliable decisions. Various approaches have been proposed to address this challenge, primarily by identifying the most influential features in complex models, such as deep neural networks. However, such approaches may oversimplify models, potentially overlooking intricate relationships between clinical, genetic, or other variables. In this paper, we present a rule-based method for deriving diagnostic models from historical patient data. Alongside the rule-generation engine, we introduce a method to assess the impact of each feature on a diagnosis. Additionally, an interactive, intelligent interface generates dashboards that enable users to explore their Explainable AI models and extract meaningful insights. To demonstrate the effectiveness of our approach, we discuss different applications, including a use case involving Acute Myeloid Leukemia data.

Keywords

Explainable AI, Medical AI, Computer-aided diagnosis, Rule-based systems

1. Introduction

The use of Explainable AI (XAI) in medical data analysis pipelines represents a transformative approach to enhancing both the accuracy and trustworthiness of healthcare diagnostics and treatment planning [1, 2]. In fact, the increasing complexity of AI algorithms often renders them opaque, making it difficult for healthcare professionals to understand the reasoning behind their predictions. XAI addresses this challenge by providing transparency and interpretability, which may foster greater confidence in AI-supported medical decisions.

XAI methods in healthcare typically focus on generating human-interpretable explanations for AI predictions. These methods include feature importance analysis, rule-based models, and counterfactual explanations. By making AI more interpretable, XAI improves clinical decision-making, increasing trust among healthcare professionals, and enhancing patient safety. Furthermore, regulatory frameworks, such as the EU's General Data Protection Regulation (GDPR) and the AI Act, underline the need for explainability, making XAI an essential factor in the adoption of AI-supported systems within the healthcare sector.

A crucial aspect of XAI in healthcare is the design of user interfaces that effectively communicate AI-generated insights to medical practitioners by incorporating visualizations, interactive elements, and natural language explanations to enhance interpretability [3]. One common approach involves the visualization of feature relevance [4, 5], where graphical representations highlight which factors (e.g., biomarkers, imaging features, or patient history) contributed most to a model's decision. For instance, heatmaps in medical imaging can indicate the regions that influenced a classification, aiding radiologists in validating AI-generated diagnoses. Also interactive dashboards allow users to explore AI outputs dynamically, by setting adjustable parameters and performing scenario-based and what-if analysis.

Late-breaking work, Demos and Doctoral Consortium, colocated with the 3rd World Conference on eXplainable Artificial Intelligence: July 09–11, 2025, Istanbul, Turkey

*Corresponding author.

enrico.sciacca@rulex.ai (E. Sciacca); giorgia.grasso@rulex.ai (G. Grasso); damiano.verda@rulex.ai (D. Verda); enrico.ferrari@rulex.ai (E. Ferrari)

0009-0005-9032-6366 (E. Sciacca); 0000-0003-2312-0219 (G. Grasso); 0000-0001-9912-3454 (D. Verda); 0000-0002-0666-6597 (E. Ferrari)

Notable examples include platforms such as NEAR [6], which present personalized disease risk scores and their contributing factors through intuitive and clinically oriented visualizations. Additionally, natural language explanations have been introduced to make AI recommendations more accessible to non-experts [7, 8, 9]. Instead of displaying raw probabilities, XAI-enhanced systems generate human-readable justifications, explaining why a particular diagnosis or treatment plan is suggested. This approach is especially beneficial in patient-facing applications, where understanding AI decisions can empower individuals to take an active role in their healthcare.

In this paper, we propose a machine learning and visualization tool that combines the effectiveness and explanatory power of rule-based systems with the simplicity and immediacy of feature-based visualizations. The proposed approach employs data-driven rules to forecast disease onset in patients, highlighting the complex relations behind the predictions. Moreover, the contribution of each feature to the diagnosis is computed based on the rules. For each patient, a risk score is calculated and decomposed into the contribution of each feature. All this information—the rules, feature relevance, diagnosis, and the contribution of each feature to the diagnosis—is presented in an interactive user interface. The manuscript is structured as follows: Sec. 2 describes the algorithm to generate the rules and determine feature relevance from the data; Sec. 3 presents the tools used to display the results to the end user; Sec. 4 introduces some applications where the approach has been tested; Sec. 5 describes the dashboard developed to visualize the results; and Sec. 6 summarizes the contribution of the presented tools for the XAI community.

2. Logic Learning Machine: an algorithm for data-driven rule generation

Rule-based models are very powerful to explain why a decision was suggested by an AI system. The simplest way to define a rule is by combining a premise and a consequence: **IF** <premise> **THEN** <consequence>. The premise contains the conjunction of some conditions about the inputs while the consequence contains the output of the system, i.e. the quantity to be predicted. This could be an example of rule:

$$\textbf{IF } X = \{A, B\} \textbf{ AND } Y < 10 \textbf{ THEN } Output = C$$

As the example shows, conditions could involve both categorical variables (i.e. for which an ordering cannot be imposed) and numerical attributes. Also, the output could be categorical (in this case the problem is referred to as *classification*) or numerical (*regression*), even if in this paper we only consider the first case. Several approaches have been proposed in literature to generate rules from data. The most popular technique is Decision Tree, that iteratively divides the dataset in smaller subsets. This divide-and-conquer approach provides an easy-to-understand classification, but the classification accuracy is usually poor. For this reason, ensemble approaches, such as Random Forests, aim at combining several rule set to achieve a richer and more accurate model of the system. In general, rules could be disjoint (like in Decision Trees) or overlapping (like in Random Forests). Overlapping rules have been proven [10] to allow a more accurate classification and to better highlight the most relevant features for the classification. In this paper, we focus on Logic Learning Machine (LLM) [11], a method capable of generating accurate overlapping rules. The approach of LLM [12] consists in transforming the data into a Boolean domain where a Boolean function for each output value is reconstructed starting from a portion of its truth table with a method described in [13]. The rules are therefore created through four steps: (a) Discretization; (b) Latticization or Binarization; (c) Synthesis of Positive Boolean functions; (d) Rule generation.

Step (c) constitutes the core of the approach and requires proper algorithms to build a Boolean function that maximizes the accuracy of the model. As a matter of fact, the Boolean function should avoid overfitting, i.e. excessive adherence of the model to the experimental data, and underfitting, i.e. poor effectiveness in describing the available data. To reach this goal, the Shadow Clustering algorithm [13] is usually adopted for Boolean function reconstruction: it makes use of several metrics aimed at

maximizing the ability of the rule to describe the data as well as at minimizing its complexity (i.e. the number of conditions it includes).

2.1. Evaluating the quality of rules

Rules can be characterized by different metrics that measure their effectiveness in describing the training set. Each rule is said to *cover* an example $\mathbf{x}$ if its premise matches $\mathbf{x}$. The examples covered by a rule are called positive, while those not covered are called negative. A true positive is a positive example whose output matches the output of the rule, whereas the output of a false positive differs from that of the rule. Similarly, true negative and false negative cases can be defined. The simplest indicators for a rule are *covering* (or *true positive rate*) and *error* (or *false positive rate*), which account for the fraction of cases correctly and incorrectly described by the rule, respectively. The balance between covering and error is crucial to ensure that rules are general and effective in describing the data. LLM allows for the calibration of the amount of error in a rule, and usually, permitting a small amount of error is beneficial for the quality of the rules.

In addition to these rule-related indicators, metrics for each condition can also be introduced. In particular, we can compute how covering and error change when the studied condition is removed. Since removing a condition corresponds to eliminating a constraint, covering and error usually increase (or remain unchanged). The increase in error, in particular, is a strong indicator of how important the condition is within the rule. If the error increases significantly, the condition cannot be eliminated without heavily affecting the quality of the rule and is therefore essential. Conversely, a small increase in error corresponds to a condition that could be removed without causing significant damage. Let $\Delta E_r(c)$ be the increase of error in rule r after removing condition c.

2.2. Computing the relevance of features

The importance of conditions in a rule can, in turn, be used to compute the relevance of attributes associated with each condition. By summing $\Delta E_r(c)$ over all the rules and all the conditions associated with attribute x_i, the *absolute relevance* $R(x_i)$ of x_i is computed. If the sum is limited to the rules whose output class is y, the *relative relevance* $R_y(x_i)$ is computed.

Additionally, it is possible to compute a relevance $R^{S_i}(x_i)$ for each subset $S_i \subset X_i$ of the domain of x_i by summing $\Delta E_r(c)$ only for the conditions that contains S_i. For categorical inputs this corresponds to evaluate the relevance of each possible value, while for ordered variables, a relevance score can be associated with every interval. Similarly also the relative relevance $R_y^{S_i}(x_i)$ can be associated with each subset of X_i.

3. Rule visualization, evaluation and feature ranking

Understanding and interpreting the rules generated by AI models is crucial, especially in medical applications where explainability is fundamental. The Rulex Platform[1] provides multiple tools to visualize, evaluate, and manage rules, making it easier for users to explore the relationships between variables and assess the impact of different conditions. In this work we show how four key components can support stakeholders in making more informed decisions: the *Feature Ranking*, the *Rule Viewer*, the *Rule Manager*, and *Rulex Studio*. These tools are currently used in several fields, such as logistics and financial services, but here we focus specifically on the applications in healthcare problems.

Feature Ranking. The Feature Ranking tool provides a quantitative assessment of how different input attributes contribute to the predictive model, directly leveraging the metrics introduced in Sec. 2. Specifically, it utilizes the concept of error variation ($\Delta E_r(c)$) discussed in Sec. 2.2 to determine the importance of each condition within a rule. By aggregating these variations across all rules, the

[1] https://www.rulex.ai/

system computes both the absolute relevance of an attribute—reflecting its overall impact on model predictions—and the relative relevance, which quantifies its importance within a specific output class.

The Feature Ranking interface allows users to explore these computed relevance scores interactively. Attributes can be sorted based on different criteria, such as their absolute importance or their specific contribution to a given diagnosis. Additionally, for categorical variables, the tool evaluates the relevance of individual values, while for numerical attributes, it highlights key threshold intervals that significantly affect the model's decision-making process.

Rule Viewer. The Rule Viewer provides an interactive environment for visually exploring rule-based models, offering intuitive representations of extracted rules and their relationships with input attributes. Figure 1 shows an example of the Rule Viewer applied to the Alzheimer's Disease dataset presented in Sec. 4.1.

The central rule chart constitutes the core visualization element. The outer circular ring represents input attributes, sorted by their absolute relevance, with each segment corresponding to a specific attribute. Attributes are colored and annotated to distinguish nominal (N), integer (I) and continuous (C) types. Within this outer ring, circles represent individual rules, grouped according to the predicted output class (e.g., *Control* or *Sick*). The size of each circle is proportional to the rule's coverage, while the hole in the center reflects the associated error rate. Hovering over any rule (as shown in the figure for Rule #12) reveals its logical conditions, displays detailed metrics such as covering and error, and highlights the corresponding attributes on the outer ring to visually link the rule to its defining features.

The interactive settings panel on the right-hand side enables users to dynamically adjust visualization parameters, such as the number of attributes displayed, attribute sorting criteria, and relevance thresholds.

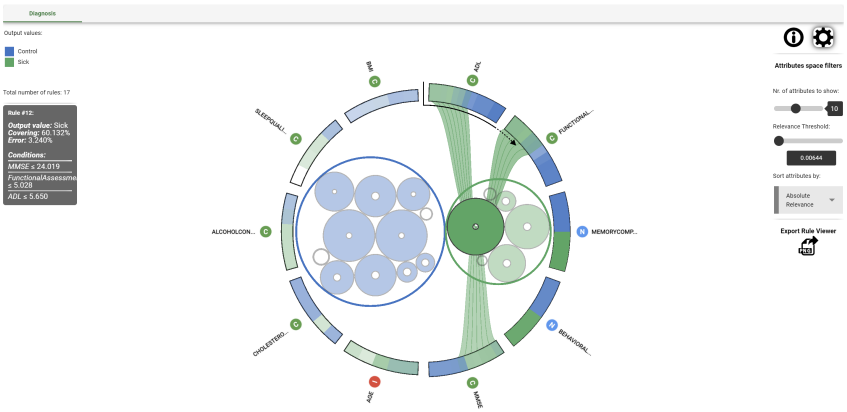

Figure 1: Example of the Rule Viewer applied to Alzheimer's Disease prediction. The left panel shows detailed information for a specific rule (Rule #12), while the right panel provides interactive settings to dynamically adjust visualization parameters.

Rule Manager. The Rule Manager task offers a structured interface for inspecting, refining, and optimizing rule-based models. It is especially valuable when working with large rulesets that require manual adjustments. The tool provides a spreadsheet-like environment where users can filter, sort, and modify rules, adjusting conditions or output values as needed. Additionally, a history tracking feature allows easy review and reversal of changes, ensuring flexibility and control over rule modifications.

Rulex Studio. Rulex Studio is an advanced visualization tool that enables users to create interactive dashboards for analyzing AI-generated models. Unlike previous tasks, which focus primarily on rule inspection and refinement, Rulex Studio provides a broader visualization framework, supporting data exploration and model presentation. The main component of Rulex Studio is the View, a customizable workspace where users can design graphical representations of their data. The platform includes various options for adjusting layout configurations, importing views, and integrating plots. This flexibility makes it an ideal tool for presenting explainable AI models to clinicians and other stakeholders.

4. Selected applications in healthcare problems

Medical AI applications must be designed to meet specific clinical needs, ensuring that predictive models provide transparent, reliable, and actionable insights for healthcare professionals. In this study, we focus on two case studies: Alzheimer's disease and Acute Myeloid Leukemia (AML), both severe conditions where early detection and prevention are crucial for improving patient outcomes. By applying our rule-based approach to these diseases, we aim to demonstrate how interpretable predictive models can support clinicians in identifying at-risk individuals and making informed decisions to enhance early intervention strategies.

4.1. Alzheimer's Disease

For Alzheimer's disease, we used a dataset [14] consisting of health records from 2,149 patients aged 60 to 90, with 35 attributes covering demographics, lifestyle, medical, and cognitive factors. Key variables included cognitive assessments like the Mini-Mental State Examination (MMSE) and Activities of Daily Living (ADL) scores, as well as functional and behavioral indicators of cognitive decline. The dataset included a binary Alzheimer's diagnosis, enabling a classification task to distinguish between healthy (*Control*) and at-risk (*Sick*) individuals.

To develop an interpretable predictive model, we applied LLM, conducting an extensive fine-tuning of its parameters. We evaluated multiple configurations using a 70-30% training-test split, selecting the optimal setup to balance accuracy and generalization. The final model produced 11 rules for the Control class and 7 for the Sick class, with each rule containing between 1 and 15 conditions. The model's final performance, summarized in Table 1, demonstrated strong predictive capability, confirming the effectiveness of rule-based learning for Alzheimer's risk assessment.

	Control	**Sick**
Training	92.6%	93.9%
Test	90.6%	89.2%

Table 1
Classification performance of the Logic Learning Machine model on the Alzheimer's dataset. Accuracy values are reported for both the training and test sets, distinguishing between Control and Sick classes.

4.2. Acute Myeloid Leukemia

For Acute Myeloid Leukemia (AML), we utilized a dataset derived from the study by the Weizmann Institute of Science [15], containing 497 patient records, of which only 83 corresponded to individuals who later developed AML. The dataset included age, sex, and the Variant Allele Frequencies (VAF) of 22 genes, known to be associated with clonal hematopoiesis and leukemia onset.

This study is part of the broader SInISA project[2], which aims to develop predictive screening methods for hematological diseases through genetic and AI-driven analysis. Given the small dataset size and the class imbalance, we conducted fine-tuning of Logic Learning Machine parameters using a 70-30% training-test split, but classification performance was lower compared to the Alzheimer's model. The final model generated 40 rules, with 24 rules for the AML class and 16 for the Control class, each containing up to 10 conditions. Despite the classification challenges, the extracted rules aligned with findings from the Weizmann study, reinforcing the known biological patterns: mutations in certain genes (e.g., DNMT3A, TET2, SRSF2, ASXL1, TP53) are commonly found in healthy aging individuals but exceed a critical threshold in those likely to develop AML. This result supports the SInISA project's goal of identifying early genetic markers for leukemia risk, highlighting the potential of rule-based models in predictive medicine.

[2]https://progettosinisa.it/

5. A Dashboard for prediction and model evaluation

To make the results of explainable models more accessible and clinically meaningful, we developed an interactive dashboard using Rulex Studio, a flexible environment for building custom data visualizations. The dashboard supports both model evaluation and in-depth exploration of individual predictions, with the primary goal of providing transparent, patient-specific risk assessments.

A key innovation of our approach is the transition from binary classification to the computation of a personalized disease risk. Based on the analysis of the Logic Learning Machine rule conditions and feature relevance, each patient is assigned a percentage risk score that quantifies the likelihood of disease onset. The following subsections illustrate the dashboard's main functionalities, using examples based on the Alzheimer's use case.

5.1. Personalized risk assessment

The central feature of the dashboard is the ability to compute and visualize a personalized disease risk for each patient. The patient's risk is calculated starting from a baseline risk, derived from the training dataset. Each attribute of the patient contributes positively or negatively to this baseline, based on the relevance of its specific value. These contributions are derived from the relevance analysis of Logic Learning Machine at the condition level.

As illustrated in Fig. 2, the top-left panel shows the final risk score, with a visual indicator of how far it deviates from the base value. Below, a radial contribution chart displays the impact of each attribute, distinguishing between those increasing (red) and decreasing (green) the predicted risk. On the right, a dual plot shows the cumulative effect of each variable, sorted by relevance: the upper section tracks the progressive build-up of the final score, while the lower plot quantifies the individual contributions of all features. This representation provides a clear explanation of both the prediction and its underlying rationale. A similar strategy has been adopted in platforms such as NEAR [6], reinforcing the importance of combining quantitative risk scores with intuitive, patient-specific visualizations.

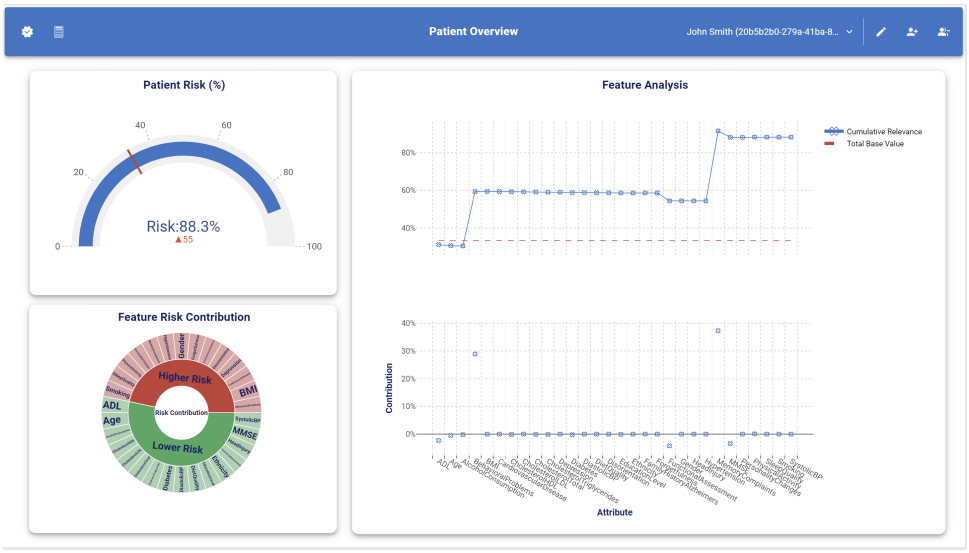

Figure 2: Patient-specific risk analysis panel. The overview shows the final risk estimate (top-left), while the radial chart (bottom-left) decomposes contributions by attribute. The right-hand plots show the cumulative relevance and feature-wise impact relative to the base risk.

5.2. What-if scenarios and patient editing

Beyond evaluating existing patients, the dashboard allows users to manually edit or create new patient profiles, enabling powerful "what-if" analysis. This feature is especially useful for exploring hypothetical situations and assessing how changes in specific attributes influence the final risk of the disease.

Each time a variable is modified, the system instantly recalculates the overall risk and updates the visual explanations, allowing users to quickly assess the impact of individual changes. This makes it possible to simulate the effects of lifestyle or clinical interventions and to identify which variables drive risk the most, supporting more informed and personalized prevention strategies.

5.3. Automated report generation

The dashboard also includes a report generation feature that produces a two-page summary for each patient. This report consolidates all key insights: patient's input data, predicted risk score, the contribution of individual features, and supporting visualizations such as risk indicator and feature contribution chart. Each section of the report is accompanied by clear captions explaining how to interpret the data, making it suitable for sharing with clinicians or for integration into broader diagnostic workflows. Reports can be generated for patients in the dataset or for manually created profiles, supporting both retrospective analysis and scenario testing.

5.4. Model summary and rule overview

While the main focus of the dashboard is on personalized patient analysis, it also includes a global view summarizing the behavior and internal logic of the model. This view provides access to essential evaluation metrics, such as feature ranking and rule visualization, which offer insight into how the model generalizes across the dataset. The feature ranking panel highlights the attributes that most influence the model's predictions, computed using the relevance scores assigned by the Logic Learning Machine. These scores reflect the importance of each feature value in shaping the classification rules and, consequently, the calculated risk. The ranking can be customized and sorted according to different criteria. The rule manager displays all the rules extracted by the model, color-coded by class and organized by complexity and coverage. Users can filter rules based on the attributes they include, inspect their logic, and visualize associated metrics such as covering and error.

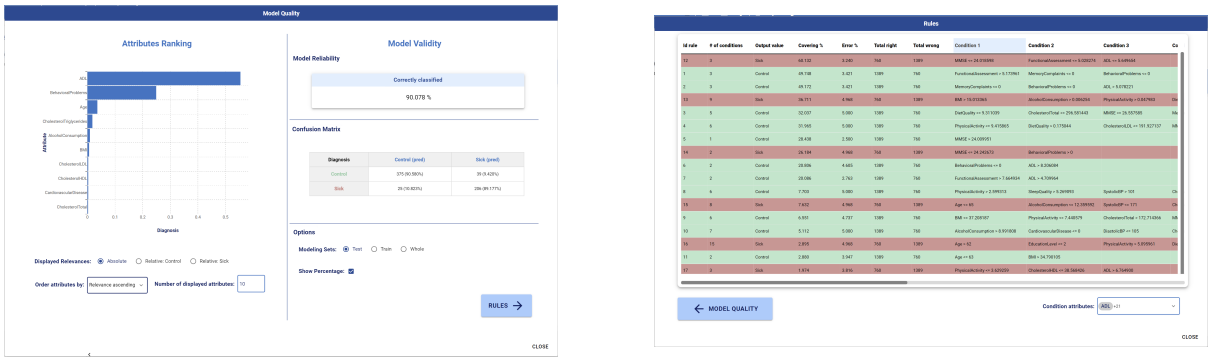

Figure 3: Visualization of model performance and rule-based decision logic in the dashboard. The left panel displays the Feature Ranking and confusion matrix, while the right panel shows the Rule Manager, where extracted rules are color-coded by class and can be filtered based on attributes.

6. Contribution to XAI Community

This work presents tools to enhance transparency in AI-supported medical diagnosis. The rule-based approach is essential to convey complex relations in a human-readable format. Nonetheless, effective AI algorithms may not be sufficient to enhance trust among decision-makers. The approach proposed in this work aims to overcome this limitation by providing both data-driven rules to support the diagnosis and a user interface to enable stakeholders to utilize the derived insights. The combination of powerful XAI algorithms and an interactive interface may ultimately foster users' trust, as shown in the two case studies presented.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4 in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] T. Hulsen, Explainable artificial intelligence (XAI): concepts and challenges in healthcare, AI 4 (2023) 652–666.

[2] A. Vellido, The importance of interpretability and visualization in machine learning for applications in medicine and health care, Neural computing and applications 32 (2020) 18069–18083.

[3] S. Khairat, A. Dukkipati, H. Lauria, T. Bice, D. Travers, S. Carson, The impact of visualization dashboards on quality of care and clinician satisfaction: Integrative literature review, JMIR Hum Factors (2018).

[4] B. Remeseiro, V. Bolon-Canedo, A review of feature selection methods in medical applications, Computers in biology and medicine 112 (2019) 103375.

[5] N. Pudjihartono, T. Fadason, A. W. Kempa-Liehr, J. M. O'Sullivan, A review of feature selection methods for machine learning-based disease risk prediction, Frontiers in Bioinformatics 2 (2022) 927312.

[6] K. Kassem, M. Sperti, A. Cavallo, A. M. Vergani, D. Fassino, M. Moz, A. Liscio, R. Banali, M. Dahlweid, L. Benetti, F. Bruno, G. Gallone, O. De Filippo, M. Iannaccone, F. D'Ascenzo, G. M. De Ferrari, U. Morbiducci, E. Della Valle, M. A. Deriu, An innovative artificial intelligence-based method to compress complex models into explainable, model-agnostic and reduced decision support systems with application to healthcare (NEAR), Artificial Intelligence in Medicine 151 (2024) 102841. doi:https://doi.org/10.1016/j.artmed.2024.102841.

[7] S. Wang, Z. Zhao, X. Ouyang, T. Liu, Q. Wang, D. Shen, Interactive computer-aided diagnosis on medical image using large language models, Communications Engineering 3 (2024) 133.

[8] S. Lee, J. Lee, J. Park, J. Park, D. Kim, J. Lee, J. Oh, Deep learning-based natural language processing for detecting medical symptoms and histories in emergency patient triage, The American Journal of Emergency Medicine 77 (2024) 29–38.

[9] C. Li, Y. Zhang, Y. Weng, B. Wang, Z. Li, Natural language processing applications for computer-aided diagnosis in oncology, Diagnostics 13 (2023) 286.

[10] S. Parodi, C. Manneschi, D. Verda, E. Ferrari, M. Muselli, Logic learning machine and standard supervised methods for hodgkin's lymphoma prognosis using gene expression data and clinical variables, Health informatics journal 24 (2018) 54–65.

[11] C. Muselli, D. Verda, E. Ferrari, C. T. Gaggiotti, M. Muselli, Rulex platform: leveraging domain knowledge and data-driven rules to support decisions in the fintech sector through explainable ai models, in: L. Longo, W. Liu, G. Montavon (Eds.), XAI-2024 Late-breaking Work, Demos and Doctoral Consortium Joint Proceedings, CEUR: Workshop Proceedings, 2024.

[12] M. Muselli, Switching neural networks: A new connectionist model for classification, in: B. Apolloni, M. Marinaro, G. Nicosia, R. Tagliaferri (Eds.), Neural Nets, Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 23–30.

[13] M. Muselli, E. Ferrari, Coupling logical analysis of data and shadow clustering for partially defined positive boolean function reconstruction, IEEE Transactions on Knowledge and Data Engineering 23 (2011) 37–50. doi:10.1109/TKDE.2009.206.

[14] R. E. Kharoua, Alzheimer's disease dataset, 2024. URL: https://www.kaggle.com/dsv/8668279. doi:10.34740/KAGGLE/DSV/8668279.

[15] S. Abelson, G. Collord, S. W. Ng, O. Weissbrod, N. Mendelson Cohen, E. Niemeyer, N. Barda, P. C. Zuzarte, L. Heisler, Y. Sundaravadanam, et al., Prediction of acute myeloid leukaemia risk in healthy individuals, Nature 559 (2018) 400–404.