A Systematic Map of Interpretability in Medicine

Hajar Hakkoum[1][a], Ibtissam Abnane[1][b] and Ali Idri[1,2][c]
[1]*Software Project Management Research Team, ENSIAS, Mohammed V University, Rabat, Morocco*
[2]*MSDA, Mohammed VI Polytechnic University, Ben Guerir, Morocco*

Keywords: Explainability, XAI, Medicine, Artificial Intelligence, Machine Learning, Systematic Review.

Abstract: Machine learning (ML) has been rapidly growing, mainly owing to the availability of historical datasets and advanced computational power. This growth is still facing a set of challenges, such as the interpretability of ML models. In particular, in the medical field, interpretability is a real bottleneck to the use of ML by physicians. This review was carried out according to the well-known systematic map process to analyse the literature on interpretability techniques when applied in the medical field with regard to different aspects. A total of 179 articles (1994-2020) were selected from six digital libraries. The results showed that the number of studies dealing with interpretability increased over the years with a dominance of solution proposals and experiment-based empirical type. Additionally, artificial neural networks were the most widely used ML black-box techniques investigated for interpretability.

1 INTRODUCTION

The medical field is a constantly growing domain, and since it is also a very critical one, an error or a misplaced decision might cost a patient's life. Breakthroughs in machine learning (ML) are accelerating the pace of decision-making algorithm development to assist physicians with a second opinion, and therefore reduce potential human errors that may cost the patient life (London, 2019). ML techniques have the potential to increase the survival rate by automating the decision process, providing higher accuracy, responding immediately in emergency cases, and helping minimize the efforts provided by physicians, especially when there is a shortage of medical staff (Hosni et al., 2019).

Two types of ML techniques can be differentiated (Hulstaert., 2020): interpretable (i.e., white-box: the knowledge discovery process is easily explained, such as decision trees (DTs) or linear classifiers), and uninterpretable (i.e., black-box: the knowledge discovery process is not easily explained, such as artificial neural networks (ANNs) and support vector machines (SVM)).

Although black-boxes excel at providing better performance owing to their high and complex computational power and their ability to discover nonlinear relationships in the data, their lack of interpretability is a major problem that explains the present trade-off between accuracy and interpretability of a model (Luo et al., 2019).

ML has long served different medical tasks, yet adoption faces resistance since the medical field still distrust black-box models for no evidence is provided to support their decisions (Pereira et al., 2018). Without any explanation of their outputs, black-box models are dreaded to incorporate harmful biases (London, 2019). Therefore, interpretability can help doctors diagnose issues and check the reliability of ML models by providing insight into the model's reasoning (Barredo Arrieta et al., 2020). Consequently, the reason for misleading the model could be detected.

Studies conducted to curve the accuracy-interpretability trade-off have undergone rapid growth to gain domain-expert trust in black-box models. In particular, in the medical domain, different approaches have been suggested and evaluated (Chuan Chen et al., 2006). To the best of our

[a] https://orcid.org/0000-0002-2881-2196
[b] https://orcid.org/0000-0001-5248-5757
[c] https://orcid.org/0000-0002-4586-4158

Hakkoum, H., Abnane, I. and Idri, A.
A Systematic Map of Interpretability in Medicine.
DOI: 10.5220/0010968700003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 719-726
ISBN: 978-989-758-552-4; ISSN: 2184-4305

knowledge, there is no existing systematic mapping that analyses and summarizes primary studies dealing with the interpretability of ML in the medical field. This motivates the present SMS study due to the relevance and importance of interpretability techniques in the medical field. Consequently, we identified 179 studies published between 1994 and 2020 and reviewed them according to different motivations:

- Identifying papers investigating ML interpretability in medicine.
- Analysing the demographics of the selected papers.
- Enumerating the different black box models whose interpretability was the most interested in.

The paper is organized as follows: Section 2 presents an overview of ML interpretability techniques. Section 3 describes the research methodology used in this study. Section 4 reports the mapping results obtained. Finally, implications and conclusions are presented in Section 6.

2 INTERPRETABILITY

Interpretability has no mathematical measure, but it can be defined as the degree to which a human can predict the outcome of a model or understand the reasons behind its decisions (Kim et al., 2016; Miller, 2019). Many terms can be associated with interpretability, such as comprehensibility and understandability, which can underlie different aspects of interpretability.

Johansson et al., 2009 explained the benefits of oracles, which are datasets with corresponding predictions of the black-box model as target values. They compared different dataset setups, such as original instances and oracle instances, or both. In other words, they compared a global surrogate of an ensemble to a DT built on training instances directly. Their experiments were carried out over 26 datasets, including breast cancer, diabetes, hepatitis, and heart, using accuracy and fidelity as performance measures. They showed that there is an accuracy gap between global surrogates and DTs, mainly because global surrogates deliver the accuracy that the black-box offers. Many works used the same approach to decrease the gap between accuracy and interpretability (Krishnan et al., 1999) (Zhou et al., 2004).

On the other hand, local surrogates were also used, although not as widely as global surrogates, to interpret a black-box model's decision for a particular instance. Fan et al., 2020 investigated the use of a factorization machine ANN to predict Cushing's disease recurrence on a newly collected dataset from a hospital in Peking, and they used local interpretable model-agnostic explanations (LIME) (Ribeiro et al., 2016) to address the lack of interpretability. By providing relevant features for each instance the doctors were interested in, LIME revealed the reasoning behind the model's decision for that instance, which allowed the doctors to trust that model. (Hakkoum et al., 2021) also investigated the use of LIME in a medical task that consisted of breast cancer diagnosis using two types of ANNs. They compared the LIME explanations with FI and PDP and showed that the explanations are almost always in agreement with the other two methods. Therefore, it is interesting to mix global and local interpretability techniques to gain more insights that will enable final users (e.g., doctors) to make the final decision.

3 MAPPING PROCESS

A mapping study aims to provide an overview of a research area by identifying the quantity and type of research that has been published in that area. The present mapping process follows the guidelines proposed by (Kitchenham et al., 2007) that are detailed in this section.

3.1 Map Questions

In order to provide insights into the efforts made to leverage the ML interpretability challenge in the medical field, the overall objective was divided into four mapping questions (MQs) presented in Table 1.

Table 1: Mapping questions.

MQs	Motivations
MQ1: What are the publication venues and in which year were the selected studies published?	To check if there is a specific publication channel and identify the number of investigations in interpretability over the years.
MQ2: What type of contributions is being made to this field?	To identify the different types of studies that worked on black-box interpretability.
MQ3: What type of empirical studies were conducted?	To identify the type of evidence that was developed in the selected studies.
MQ4: What are the Black box ML techniques that were subjects of interest?	To enumerate the different black box models that were interpreted and identify the most interested in.

3.2 Search Strategy

To answer the MQs, a search string was defined to provide the maximum manageable coverage. It contained the main terms matching the MQs, along with their synonyms. Synonyms were joined with the OR Boolean and main terms by AND Boolean. The complete set of search strings is defined as follows: ("black box" OR "black-box" OR uninterpretable OR "neural networks" OR "support vector machines" OR "deep learning") AND ("machine learning" OR "data mining" OR "data analytics" OR "knowledge discovery" OR "artificial intelligence" OR prediction OR classification OR clustering OR association) AND (interpretability OR explainability OR understandability OR comprehensibility OR justifiability OR trustworthiness OR XAI).

An automatic search on six selected digital libraries (ScienceDirect, IEEE Xplore, ACM Digital Library, SpringerLink, Wiley, Google Scholar) was performed to extract the primary papers. A secondary search was then performed by scanning the references lists of the relevant papers that satisfy a set of inclusion and exclusion criteria. A new search was performed with the same search string to pick up the newly published articles during the time spent on the first three steps and their article data extraction.

3.3 Study Selection Process

In this phase, the relevant studies addressing the research questions based on their titles, abstracts, and keywords were selected. To achieve this, each of the candidate studies identified in the initial search stage was evaluated by two researchers, using the inclusion (ICs) and exclusion (ECs) criteria, to determine whether it should be retained or rejected.

- IC1: Addressing interpretability or an overview of existing interpretability techniques applied to a medical task.
- IC2: Presenting new interpretability techniques applied to a medical task
- IC3: Evaluating or comparing existing interpretability techniques applied to a medical task.
- IC4: In the case of duplicate papers, only the most recent and complete papers were included. If a paper figures in two libraries, we make use of the order of the digital libraries to choose.
- EC1: Beyond the medical scope
- EC2: Written in a language other than English.
- EC3: Short & abstract paper.
- EC4: Mentioning interpretability or using an interpretability technique or more without it being the main focus.
- EC5: Preprints.

3.4 Data Extraction and Synthesis

To answer the mapping questions, a data extraction form was created and filled for each of the selected papers. The data extracted from each of these studies is listed in Table 2. After extracting the data, it was synthesized and tabulated in a manner consistent with the research questions addressed to be visualized.

Table 2: Extracted data.

MQ	Data Extracted
-	Authors, title, digital library, abstract
MQ1	Publication year Publication Channel (Petersen et al., 2015): Journal, Conference, Book. Source name
MQ2	Research Type (Wieringa et al., 2005): ▪ Evaluation Research (ER) of an (existing/new) interpretability technique applied in the medical field. ▪ Solution Proposal (SP) (or an important improvement) of an interpretability technique applied in the medical field. ▪ Experience (Ex): Personal experience of evaluating an interpretability technique applied in the medical field. ▪ Review (Re): A sum-up of interpretability techniques applied in the medical field. ▪ Opinion papers (OP): The paper contains the author's opinion about interpretability techniques in the medical field.
MQ3	Empirical Methods (Petersen et al., 2015): ▪ Survey: Asking one or several questions to gather the required information. ▪ HBE: using historical existing data in the evaluation. ▪ Case study: An empirical evaluation based on real-world datasets (hospitals/clinics).
MQ4	Name of the black-box technique (ANN / SVM / RF…)

3.5 Study Quality Assessment

Quality assessment (QA) has the potential to limit bias in conducting mapping studies and guide the interpretation of findings (Higgins et al., 2009). Therefore, a questionnaire (Table 3) was designed to improve the selection criteria and ensure the relevance of the papers.

Table 3: QA questions.

Question	Possible answers
QA1: The study presents empirical evidence (about interpretability) that is analysed quantitatively or qualitatively	"Quantitatively" or "Qualitatively"
QA2: The study presents an experimental design that is justifiable and detailed	"Yes", "Partially" or "No"
QA3: The study reports the black-box performance measures	"Yes" or "No"
QA4: The study presents a comparison between the proposed empirical interpretability method and other methods	"Yes" or "No"
QA5: The study explicitly analyses the benefits and limitations of the study	"Yes", "Partially" or "No"
QA6: The study has been published in a recognized and stable publication source	Conferences: **Core2018**: A: +1.5, B: +1, C: +0.5, otherwise: +0 Journals: **JCR**: Q1: +2, Q2: +1.5, Q3 or Q4: +1, otherwise: +0

4 MAPPING RESULTS AND DISCUSSION

This section presents an overview of the selected studies and the results related to the MQs listed in Table 1 along their discussion.

4.1 Overview of the Selected Studies

Figure 1 shows the number of articles obtained at each stage of the selection process. The search in the six electronic databases resulted in 10332 candidate papers (both searches). ICs and ECs criteria selected 240 articles based on the title, abstract, and keywords. In doubt, the full article was read. Personal references and scanning of the retrieved papers references added 41 additional relevant papers. Finally, the QA criteria were applied to select 179 qualified articles published between Aug1994 and Dec 2020. The papers list is available upon request by email to the authors.

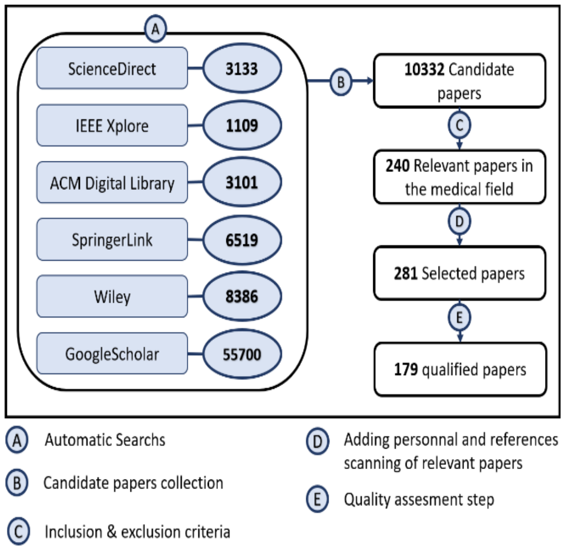

Figure 1: Selection process.

4.2 MQ1

As shown in Figure 2, the 179 selected studies were published in different sources, mainly journals or conferences. 58% (104 papers) of studies were published in journals and 40% (72 papers) in conferences, while the remaining 3 studies were published in books.

Table 4 and 5 present publications venues with more than five papers. It shows that the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining was the publication venue with the most conference qualified papers (9 papers). As for journals "Neurocomputing" was the most frequent one with 9 papers.

Table 4: Journals venues with more than five papers.

Journal venues	Number of papers
Neurocomputing	9
Artificial Intelligence in Medicine	8
Expert Systems with Applications	8

Table 5: Conferences venues with more than five papers.

Conference venues	Number of papers
International conference on knowledge discovery and data mining (SIGKDD)	9
IEEE International Joint Conference on Neural Networks (IJCNN)	7

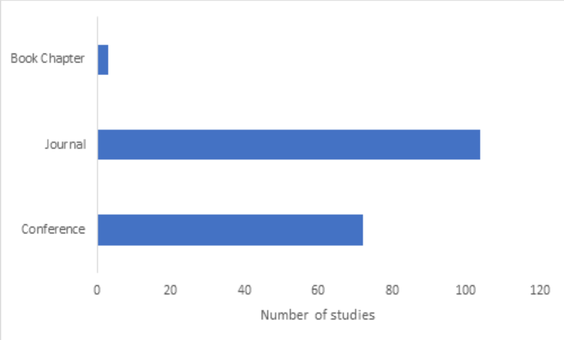

Figure 2: Distribution of publication channels.

Figure 3 displays the number of articles published per year for the period 1994-2020. The number of published studies was low (under 8 papers per year). By 2017 it increased with 12 papers, this increase was also noticed by Barredo Arrieta et al. when analysing Scopus databases to check articles with words such as “XAI” and “interpretability” (Barredo Arrieta et al., 2020). From that year on (2017-2020), 104 articles were published. Moreover, 2019 and 2020 recorded the highest number of published papers with 40 and 38 papers, respectively.

Even though we noticed a variety of sources, it seems that journals did more investigation of black-box model interpretability than conferences, especially in the last two years (2019-2020) with a sum of 50 journal articles and 28 conference articles. Moreover, the highly visible increase in published studies in 2019 is probably due to the awareness of the importance of interpretability, especially in the medical field, backed by sufficient computational power and big datasets that caused models such as ANNs to become achievable with one bottleneck: their lack of interpretability.

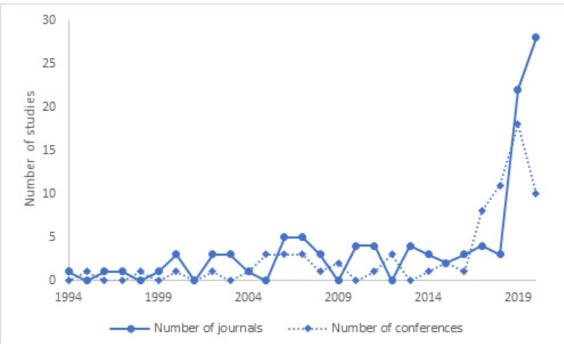

Figure 3: Distribution of the qualified studies per year.

4.3 MQ2

We identified four research types: Solution Proposal (SP), Evaluation Research (ER), Review (Re), Experience papers (Ex), and Opinion papers (OP). As presented in Figure 4, 41% (125 papers) of the selected papers were SP, and they were also ER for evaluating or comparing the proposed interpretability techniques. Moreover, only one study (Augasta et al., 2012) identified both Re and ER. All of the studies (179 papers) were classified as ER since they evaluated or compared proposed or existing interpretability techniques in the medical field. This implies that 53 papers focused solely on evaluating or comparing existing techniques (ER).

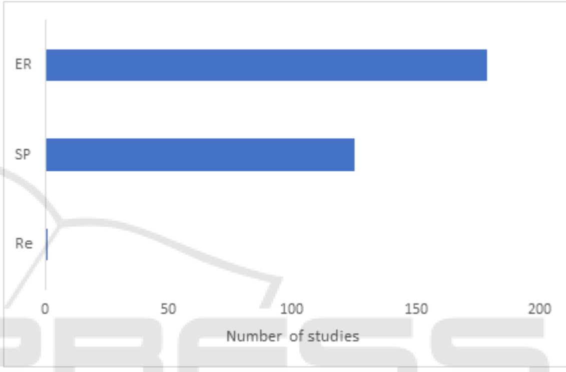

Figure 4: Contribution type of the qualified studies.

The interpretability field is as old as ML itself, yet it has only manifested its importance recently with the emergence of deep networks and datasets availability. For that reason, the number of SP papers is the highest, followed by the ER papers, showing that the interest in interpretability resulted in noticeable efforts in proposing interpretability techniques to unveil the opacity of black-box models. All the qualified papers in this study carried out an empirical evaluation that showed a very high level of maturity within the community. This is also due to our restriction of the field as well as the QA step; therefore, most selected papers needed to perform empirical evaluations in medicine in order to be qualified.

4.4 MQ3

Three types of empirical studies were identified: survey, HBE, and a case study. Figure 5 shows how 76% (142 papers) and 23% (43 papers) were identified as HBE and case studies, respectively, while one study (Liu et al., 2017) presenting 1% of the qualified studies was identified as a survey. It is also important to mention that seven papers used both HBE and case study empirical evaluations. This means that 135 and 36 papers were solely either HBE or case study, respectively.

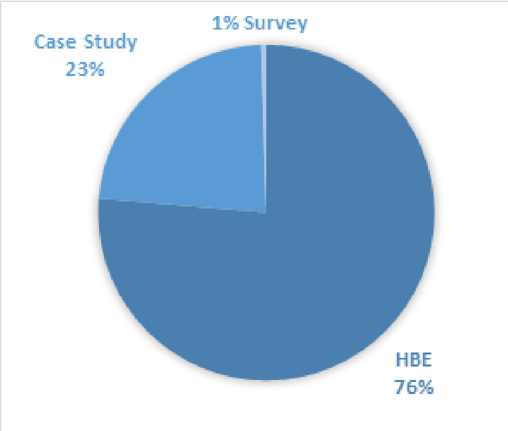

Figure 5: Distribution of empirical study type of the selected papers.

We observe that researchers prefer using public datasets to evaluate their solutions, which can be explained by the fact that the results can be compared with other techniques evaluated on the same datasets and the availability of public datasets in the medical domain. Additionally, it is preferred to evaluate and interpret ML models on historical data before using real-time evaluation. 23% of the qualified studies were empirically evaluated using case study methods that allow a real-life context evaluation. The reason behind this small percentage may be due to the undeveloped interactions and collaborations between academic researchers and physicists, which may require data security to take place. Moreover, since survey evaluations are the least mature empirical evaluation, only one study relied on empirical evaluations based on surveys. Moreover, it might be difficult to collect and validate medical data.

4.5 MQ4

For the black-box models that the selected studies attempted to interpret, we identified three main techniques: ANN, SVM, support vector regression (SVR), and tree ensembles such as RF.

Figure 6 shows the distributions of these techniques, and we can observe that ANN is the most interpreted ML technique (131 articles, 70%), followed by SVM/SVR (in 37 articles, 20%). Finally, even though18 articles (10%) targeted interpreting tree ensembles, 15 of them worked on RF. It is also important to mention that many papers worked on two or more black models at a time.

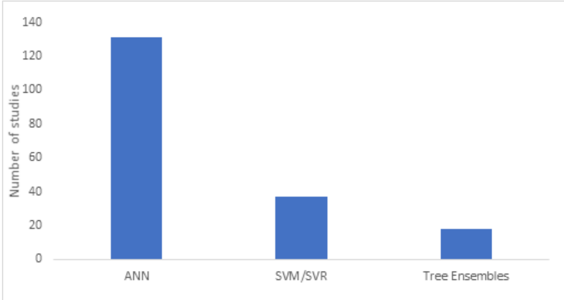

Figure 6: Number of studies per black box techniques.

Only 20% of the selected papers used SVM/SVR. This may be explained by the fact that since ANNs are based on the human brain analogy, they are more appealing to medical researchers than SVM/SVR, which has a purely mathematical basis and therefore can be more complicated to implement and understand than ANNs that have a mathematical representation close to that of the brain process (Wu et al., 2008).

Figure 7 goes more in-depth, presenting the distribution of the ANN types. We noticed that 30% of the articles (39 papers) focused on interpreting convolutional neural networks (CNN) and 20% (27 papers) worked on a multilayer perceptron (MLP) network with only three layers, which is why we excluded it from deep ANNs (DNNs). Moreover, 12% (16 papers) used recurrent neural networks (RNNs), 12% (16 papers) were articles interpreting ANNs without specifying their types, and 11% interpreted DNNs (15 papers). Note that CNNs are a type of DNNs, but we referred by DNNs to studies that used a deep network with no convolutions.

Additionally, 8% of the selected articles investigated the interpretability of ANN ensembles, 1% of radial basis function network (RBFN) models, and 2% of probabilistic neural network (PNN) models. The remaining 4% included other ANNs such as neural logic nets (Chia et al., 2006) and artificial hydrocarbon networks (Ponce et al., 2017).

Interest in interpretability was intrigued by the emergence and use of deep networks in general and CNNs, in particular, mainly for recognition and detection tasks. Therefore, most of the selected studies focused on interpreting ANNs, and more specifically, CNNs (as well as DNNs and RNNs). Moreover, 20% of studies worked on interpreting MLPs, which can be explained by the fact that MLPs are the simplest form of complex DNNs because they consist of only three layers, and the interest was probably starting with the simplest form.

5 IMPLICATIONS AND CONCLUSIONS

This study was undertaken as an SMS to investigate the interpretability of black-box models. In this study, an automatic search was performed in six digital libraries. A total of 179 papers published between 1994 and 2020 were qualified for the investigation of interpretability in the medical field.

Different sources were identified for future publications, which could be useful for researchers. Most of the qualified papers proposed a solution along with its evaluation (usually HBE), which shows the huge interest in debunking interpretability as well as the high maturity of the community. Nevertheless, researchers are encouraged to attempt to validate their proposals or evaluations in real-world scenarios (e.g., clinics, hospitals) by implementing their proposed ensembles in a decision support system. As to ML techniques, ANNs were the most appealing black-box technique for investigating interpretability. More efforts should be put into interpreting SVM/SVR models and tree ensembles because they are widely used as ANNs.

To use ML efficiently in domains such as medicine, the entire community should break down the barrier of interpretability, which will solve the bottleneck of lack of ML transparency.

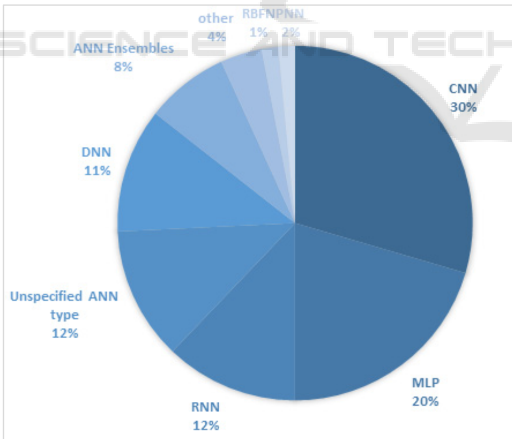

Figure 7: Distribution of ANNs types.

ACKNOWLEDGEMENTS

The authors would like to thank the Moroccan Ministry of Higher Education and Scientific Research, and CNRST.

REFERENCES

Augasta, M. G., & Kathirvalavakumar, T. (2012). Rule extraction from neural networks—A comparative study. *International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012). Salem, India*, 404–408. Retrieved from https://ieeexplore.ieee.org/abstract/document/6208380/

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. doi: 10.1016/j.inffus.2019.12.012

Chia, H. W. K., Tan, C. L., & Sung, S. Y. (2006). Enhancing knowledge discovery via association-based evolution of neural logic networks. *IEEE Transactions on Knowledge and Data Engineering*, *18*(7), 889–901. doi: 10.1109/TKDE.2006.111

Chuan Chen, Youqing Chen, & Junbing He. (2006). Neural Network Ensemble Based Ant Colony Classification Rule Mining. *First International Conference on Innovative Computing, Information and Control - Volume I (ICICIC'06)*, *3*, 427–430. doi: 10.1109/ICICIC.2006.477

Fan, Y., Li, D., Liu, Y., Feng, M., Chen, Q., & Wang, R. (2020). Toward better prediction of recurrence for Cushing's disease: a factorization-machine based neural approach. *International Journal of Machine Learning and Cybernetics 2020 12:3*, *12*(3), 625–633. doi: 10.1007/S13042-020-01192-6

Hakkoum, H., Idri, A., & Abnane, I. (2021). Assessing and Comparing Interpretability Techniques for Artificial Neural Networks Breast Cancer Classification. *Https://Doi.Org/10.1080/21681163.2021.1901784*. doi: 10.1080/21681163.2021.1901784

Higgins, J., & Green, S. (2009). *Cochrane Handbook for Systematic Reviews of Interventions*. Retrieved from http://handbook-5-1.cochrane.org/v5.0.2/

Hosni, M., Abnane, I., Idri, A., Carrillo de Gea, J. M., & Fernández Alemán, J. L. (2019). Reviewing ensemble classification methods in breast cancer. In Computer Methods and Programs in Biomedicine (Vol. 177, pp. 89–112). Elsevier Ireland Ltd. doi: 10.1016/j.cmpb.2019.05.019

Hulstaert., L. (2020). *Black-box vs. white-box models*. Retrieved from https://towardsdatascience.com/machine-learning-interpretability-techniques-662c723454f3

Johansson, U., & Niklasson, L. (2009). Evolving decision trees using oracle guides. *2009 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2009 - Proceedings*, 238–244. doi: 10.1109/CIDM.2009.4938655

Kim, B., Khanna, R., & Koyejo, O. (2016). Examples are *not Enough, Learn to Criticize! Criticism for Interpretability*.

Kitchenham, B., & Charters., S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering*. Retrieved from https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf

Krishnan, R., Sivakumar, G., & Bhattacharya, P. (1999). Extracting decision trees from trained neural networks. *Pattern Recognition*, *32*, 1999–2009.

Liu, N., Kumara, S., & Reich, E. (2017). Explainable data-driven modeling of patient satisfaction survey data. *2017 IEEE International Conference on Big Data (Big Data)*, 3869–3876. doi: 10.1109/BigData.2017.8258391

London, A. J. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Center Report*, *49*(1), 15–21. doi: 10.1002/hast.973

Luo, Y., Tseng, H.-H., Cui, S., Wei, L., Haken, R. K. T., & Naqa, I. E. (2019). Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR|Open*, *1*(1). doi: https://doi.org/10.1259/bjro.20190021

Miller, T. (2019). Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence*, *267*, 1–38.

Pereira, S., Meier, R., McKinley, R., Wiest, R., Alves, V., Silva, C. A., & Reyes, M. (2018). Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation. *Medical Image Analysis*, *44*, 228–244. doi: 10.1016/j.media.2017.12.009

Petersen, K., Vakkalanka, S., & Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, *64*. doi: 10.1016/j.infsof.2015.03.007

Ponce, H., & de Lourdes Martinez-Villaseñor, M. (2017). Interpretability of artificial hydrocarbon networks for breast cancer classification. *2017 International Joint Conference on Neural Networks (IJCNN)*, 3535–3542. doi: 10.1109/IJCNN.2017.7966301

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should i trust you?” Explaining the predictions of any classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 1135–1144. doi: 10.1145/2939672.2939778

Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2005). Requirements engineering paper classification and evaluation criteria : a proposal and a discussion. *Requirements Eng*, *11*, 102–107. doi: 10.1007/s00766-005-0021-6

Wu, T. K., Huang, S. C., & Meng, Y. R. (2008). Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities. *Expert Systems with Applications*, *34*(3), 1846–1856. doi: 10.1016/j.eswa.2007.02.026

Zhou, Z.-H., & Jiang, Y. (2004). Nec4.5: neural ensemble based c4.5. *IEEE Transactions on Knowledge and Data Engineering*, *16*(6), 770–773. doi: 10.1109/TKDE.2004.11