ELEMENT 1: Systematic Literature Review and Report

AUTHOR: ROHITH VADDEPALLY

STUDENT ID: W9529921

EMAIL: W9529921@live.tees.ac.uk

READERSHIP: Dr. Zia Ush Shamszaman

SUBMISSION DATE: 11-05-2022

WORD COUNT: 3,779 excluding references and contents

Table of Contents

INTRODUCTION **3**
INTERPRETABILITY 3
EXPLAINABILITY 3
INTERPRETABILITY IN CANCER DETECTION 4
RESEARCH QUESTION **5**
KEY CONCEPTS **5**
SEARCH RESULTS **5**
IEEE XPLORE 5
SCIENCE DIRECT 8
SCOPUS 10
CRITERIA 11
EVALUATION 11
INCLUSION AND EXCLUSION PROCESS 12
CONCLUSIONS 16
REFERENCES 16

INTRODUCTION

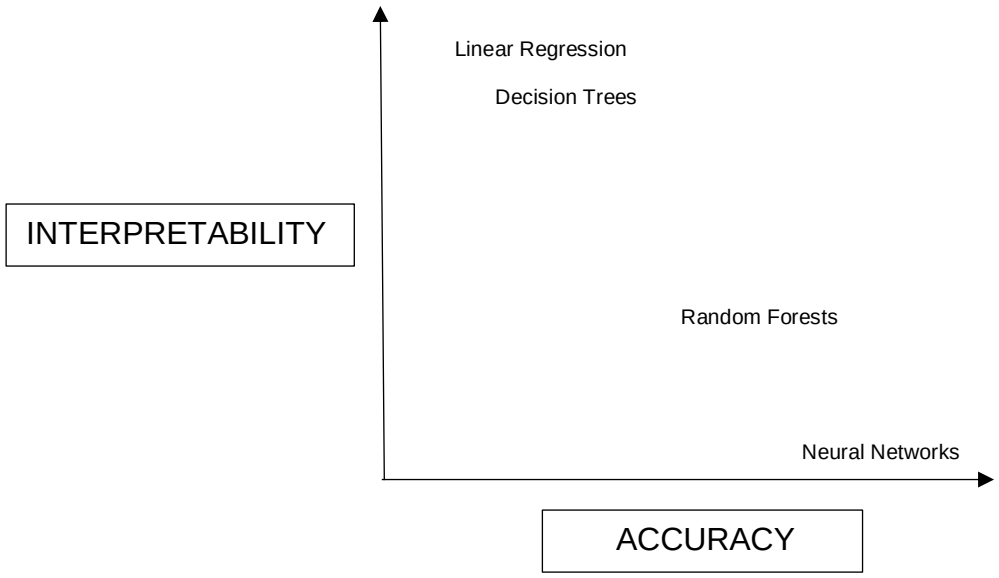
INTERPRETABILITY:

Interpretability in machine learning is when humans can easily understand the decision made by the machine learning algorithm or even if it explains what made the algorithm to make that specific decision. If as humans can tell how model made the decision then it is interpretable model. In this developing world, interpretability in the model is must for the humans to trust it. Machine learning is used in almost every industry today and if there is no interpretability in the decision that it is making it can cause catastrophic failures in making the prediction or show some inaccurate classification reports.

So, what do we gain from the model if it interpretable:

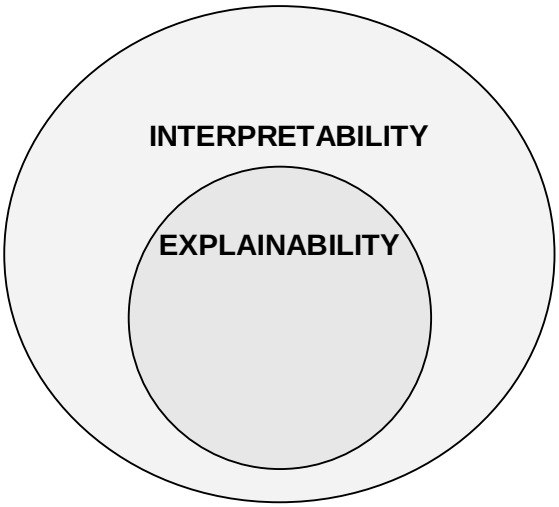
1. Model will be fair.
2. If the model explains it's decision, then it can win the people's trust.
3. Data protection and privacy.

There will be a lot of other factors involved if the model is interpretable. Algorithms like Neural networks, Random forests are less interpretable as they mostly focus on their accuracy and usually be used on large datasets where as algorithms like Linear Regression and Decision Trees are more interpretable as these are used on the smaller datasets. This is also explained in a visual format in the below figure.



EXPLAINABILITY:

We are looking for articles or conferences that are linked to interpretability in machine learning models in cancer detection for this literature study. We're also looking for the term "**Explainability**". If we can understand how a specific node in a complex model technically influences the output, we can say that the model is explainable. The model is interpretable if every component can be explained and we can keep track of each explanation at the same time. Explainability is a concept where it comes within interpretability.



INTERPRETABILITY IN CANCER DETECTION:

Many aspects of oncology have been transformed by machine learning models, including computer-aided diagnosis, medical image segmentation and synthesis, and treatment protocol development. As healthcare industries are dealing with the lives of the patient having a trustful interpretable model is necessary. There are highly complex oncology problems in the industry. Many, machine learning models implementations, particularly deep learning are still a **'Black Box**' in which model outputs are generated from planned inputs in behaviours that aren't justified by oncology-specific theories and knowledge. Interpretable machine learning algorithms, are effective instruments for deciphering the underlying process in a big set of data. When the model is explainable in the cancer detection, the clinics, physicians will be able to understand and know why the model came to a specific decision. That's where we get the term '**Fairness of the model'** which is a very important concept and without these the AI, Machine Learning will be of no use in today's world.

To summarise, interpretability and explainability in the detection of cancer is the area that I am interested to do research. This research is investigating how factors like interpretability, explainability, algorithmic fairness and black box models play their role and their importance in the diagnosis of cancer in healthcare industry.

I, personally think this topic is very rarely discussed at this point of time in the world. As mentioned above as oncology is a healthcare department which is very crucial in the patient's health. It is very necessary that research has to be made on this topic. So, a literature review is the best way to begin a primary or secondary research. I followed some of the research strategies to get as many as journals or conferences papers. All the details of three different databases, inclusion and exclusion process and a conclusion is given below.

RESEARCH QUESTION

Is the machine learning algorithm's decision in the diagnosis of cancer in healthcare industry is interpretable to humans or not?

KEY CONCEPTS

Interpretability in machine learning

Explainability

Algorithmic Fairness

Cancer diagnosis

SEARCH RESULTS

For my research topic below given three databases are selected.

1. IEEE Xplore
2. Science Direct
3. Scopus

The journal articles or conference papers in these databases are peer-reviewed and these three databases have given me more relevant articles then any other databases.

Database 1: IEEE Xplore

S.NO	QUERIES	KEYWORDS	FILTERS	NUMBER OF RESULTS	NOTES
1.	What has been published on Interpretability in machine learning model?	"Interpretable Learning"	None	2,220	Upon the first search on the research topic Interpretable Learning, a lot of Conferences, Journals, Early Access Articles, Magazines and Books were found. Need to apply some filters to narrow down a specific research topic.
2.	What has been published on Interpretability and fairness of it in machine learning model?	"Interpretable Learning" OR "Algorithmic Fairness"	None	2,355	Added a keyword 'Algorithmic Fairness' which gave the results of Conferences (1,632),

					Journals (560), Early Access Articles (133), Magazines (23), Books (7). But need to apply filters to narrow down to a specific topic.
3.	What has been published on Interpretability in machine learning model?	"Interpretable Learning" OR "Algorithmic Fairness" AND "Interpretable Machine Learning Model"	None	2,220	Added a new keyword "Interpretable Machine Learning Model" which gave results of Conferences (1,537), Journals (529), Early Access Articles (130), Magazines (18), Books (6).
4.	What has current peer-reviewed work that has been published on Interpretable machine learning model?	"Interpretable Learning" OR "Algorithmic Fairness" AND "Interpretable Machine Learning Model"	Year: 2017 to 2022	1,791 Conferences (1,187), Journals (457), Early Access Articles (130), Magazines (16), Books (1)	Added a year range filter from 2017-2022 which gave a total number of 1,791 results which are too many.
5.	What has published on Explainability of the model in the articles along with "Interpretability" in it?	"Interpretable Machine Learning Model" OR "Algorithmic Fairness" OR "Explainability of The Model"	Year: 2017 to 2022	1,292 Conferences (878), Journals (309), Early Access Articles (79), Magazines (25), Books (1)	Added a new keyword "Explainability of the Model" as there many results previously.
6.	Which of the papers discussed the topic BLACK BOX and its importance in the interpretability of the model?	"Interpretable Machine Learning Model" OR "Algorithmic Fairness" OR "Explainability of The Model" AND "Black Box Models in Machine Learning"	Year: 2017 to 2022	834 Conferences (554), Journals (216), Early Access Articles (48), Magazines (15), Books (1)	Along with Interpretability of the model, Fairness of the algorithm and explainability of the model, added the keyword "Black Box Models in Machine Learning" which

					gave us lesser number of results.
7.	What has been published on Health-care sector?	"Interpretable Machine Learning Model" OR "Algorithmic Fairness" OR "Explainability of The Model" AND "Black Box Models in Machine Learning" AND "Abstract: Healthcare"	Year: 2017 to 2022	796 Conferences (523), Journals (210), Early Access Articles (47), Magazines (15), Books (1)	A total of 796 found when searched for health-care sector, but count is too much.
8.	What has been published on Cancer Detection where the ML model is interpreting to the humans?	"Interpretable Machine Learning Model" OR "Algorithmic Fairness" OR "Explainability of The Model" AND "Black Box Models in Machine Learning" AND "Abstract: Healthcare" AND "Cancer Diagnosis"	Year: 2017 to 2022	792 Conferences (519), Journals (210), Early Access Articles (47), Magazines (15), Books (1)	Still, too many results.
9.	How the clinics or Oncologists perform medical diagnostics computing?	"Interpretable Machine Learning Model" OR "Algorithmic Fairness" OR "Explainability of The Model" AND "Black Box Models in Machine Learning" AND "Abstract: Healthcare" AND "Cancer Diagnosis"	Year: 2017 to 2022 Publication Topics: Medical Diagnostic Computing	18 Conferences (12), Journals (6)	A list of Conferences and Journals found.

DATABASE 2: SCIENCE DIRECT

S.NO	QUERIES	KEYWORDS	FILTERS	NUMBER OF RESULTS	NOTES
1.	What has been published on Interpretability in machine learning?	“Interpretable Machine Learning”	None	481	Upon the first search in the database: Science Direct with the Key Word ‘Interpretable Machine Learning’, got the results which include various articles and books.
2.	What has the current most peer-reviewed work that has been published on the Interpretable machine learning?	“Interpretable Machine Learning”	Years: 2017 to 2022	473	Applied the year filter from 2017 to 2022 to get the most recent published articles. But, still have too many results.
3.	What has been published on the Interpretability of the model or Fairness of the decision making?	"Interpretable Machine Learning" OR "Algorithmic Fairness"	Years: 2017 to 2022	593	Added a new keyword “Algorithmic Fairness” which is a very important aspect of a ML model. Results have been increased.
4.	What are the articles that have discussed the importance of the different Interpretable models used in Machine Learning?	"Interpretable Machine Learning" OR "Algorithmic Fairness" OR "Interpretable Models"	Years: 2017 to 2022	2,867	Added a new keyword “Interpretable Models”. But the results have been increased to 2,867 as we searched with “OR” along with Interpretable machine learning and Algorithmic Fairness.
5.	What has published on the topic Explainability of the model in the articles along with “Interpretability” in it?	"Interpretable Machine Learning" OR "Algorithmic Fairness" OR "Interpretable Models" AND “Explainability”	Years: 2017 to 2022	2,235	Added a new keyword “Expalinability” as we want the articles or conference papers which discuss about

					the concept of explainability in it, but still need to narrow down to a specific topic.
6.	Which of the papers discussed the topic BLACK BOX and its importance in the interpretability of the model?	"Interpretable Machine Learning" OR "Algorithmic Fairness" OR "Interpretable Models" AND “Explainability” AND "Black Box Models"	Years: 2017 to 2022	869	Added a new keyword “Black Box Models” which reduced the results.
7.	What has been published on Health-care sector?	"Interpretable Machine Learning" OR "Algorithmic Fairness" OR "Interpretable Models" AND “Explainability” AND "Black Box Models" AND “Healthcare”	Years: 2017 to 2022	661	With the added keyword “Healthcare” which gave the articles related to healthcare.
8.	What has been published on Cancer Detection where the ML model is interpreting to the humans?	"Interpretable Machine Learning" OR "Algorithmic Fairness" OR "Interpretable Models" AND “Explainability” AND "Black Box Models" AND “Healthcare”	Years: 2017 to 2022 Title, abstract or author-specified Keywords: Cancer Diagnosis	9	Added a filter Cancer Diagnosis which gave us the results only related to cancer detection in healthcare.
9.	What articles (Review and Research) are published on interpretability of the machine learning model in cancer diagnosis?	"Interpretable Machine Learning" OR "Algorithmic Fairness" OR "Interpretable Models" AND “Explainability” AND "Black Box Models" AND “Healthcare”	Years: 2017 to 2022 Title, abstract or author-specified Keywords: Cancer Diagnosis Article Type: Review articles and Research articles	7	Added filter of article type which includes only Review and Research articles where we got the final set of results.

DATABASE 3: SCOPUS

S.NO	QUERIES	KEYWORDS	FILTERS	NUMBER OF RESULTS	NOTES
1.	What has been published on Interpretability in machine learning?	“Interpretable machine learning”	None	4,920	Upon the first search in the database: Scopus with the Key Word ‘Interpretable machine learning’, got the results which include various articles, conferences, book chapter, review, Editorial.
2.	What has the current most peer-reviewed work that has been published on the Interpretable machine learning?	“Interpretable Machine Learning”	Year: 2017 to 2022	4,877	Applied the year filter from 2017 to 2022 to get the most recent published articles. But, still have too many results.
3.	What are the articles that have discussed the importance of the different Interpretable models used in Machine Learning?	“Interpretable Machine Learning” OR “Interpretable models”	Year: 2017 to 2022	7,087	Added a new keyword “Interpretable models” which increased the number of results.
4.	What has been published on the Interpretability of the model or Fairness of the decision making?	“Interpretable Machine Learning” OR “Interpretable models” OR “Algorithmic fairness”	Year: 2017 to 2022	7,872	Added a new keyword “Algorithmic fairness” which gave the results of 7,872 which are too many. Need to apply filters.
5.	What has published on the topic Explainability of the model in the articles along with “Interpretability” in it?	“Interpretable Machine Learning” OR “Interpretable models” OR “Algorithmic fairness” AND “Explainability”	Year: 2017 to 2022	1,021	After added a keyword “Explainability” got a lesser number of results.
6.	What has been published on Health-care sector where the machine leaning algorithms discuss about interpretability and explainability in it?	“Interpretable Machine Learning” OR “Interpretable models” OR “Algorithmic fairness” AND “Explainability”	Year: 2017 to 2022	292	Searched the concepts of interpretability, fairness and explainability in healthcare sector which

		AND "Healthcare"			gave us the results of 292.
7.	What articles are published on interpretability of the machine learning model in cancer diagnosis?	"Interpretable Machine Learning" OR "Interpretable models" OR "Algorithmic fairness" AND "Explainability" AND "Healthcare" AND "Cancer diagnosis"	Year: 2019 to 2022	22	22 results appeared after searching for detection of cancer where there is interpretability in machine learning algorithms.
8.	What articles are published on interpretability of the machine learning model in cancer diagnosis?	"Interpretable Machine Learning" OR "Interpretable models" OR "Algorithmic fairness" AND "Explainability" AND "Healthcare" AND "Cancer diagnosis"	Year: 2019 to 2022 Keyword**:** Interpretability, Decision Making, Explainability, Health Care, Oncology, Black-box Models, Cancer Diagnosis, Cancer Detection, Clinical Decision Making.	14	Selected some keywords as filters which are related to literature review topic which gave us the results of 14 and these results include only review, article and conference papers.

CRITERIA

EVALUATION

At the beginning of this literature review, I just had a few questions that

1. How a model or algorithm in healthcare industry making a decision in the diagnosis of cancer?
2. Is it interpreting, it's decision to the human or how far the humans are able to understand it's decision?
3. Is the model explainable?
4. Do clinical data scientists, physicians and clinics trust these models?

All these questions, driven me to start a literature review on this specific topic. As this literature review is the starting stage for my research with the time passed, I have gone through the topics like Explainability, Black Box Models, SHAP Models, Medical Diagnostic Computing and Algorithmic Fairness. In each database almost tried to include these concepts or topics in it as they play a vital role in our research topic.

Moreover, for considering a journal or conference paper, it must be answering one of the above questions including the key concepts in like Interpretability, Explainability of the algorithms in the cancer diagnosis and should be relevant to these topics. Found total of 39 journals or conference papers in the three databases IEEE Xplore, Science Direct and Scopus. My list of journals or conference papers include Empirical Research as from a few articles it is studied that the research approach is evidence based. Next inclusion and exclusion process will be presented below. The below inclusion and exclusion process will be done based systematic review of each paper and furtherly select more relevant papers based on the criteria.

INCLUSION AND EXCLUSION PROCESS

S.NO	Research Paper Title	Included/Excluded	Reasons
1.	Interpretable Machine Learning with Boosting by Boolean Algorithm	Included	As the conference paper is discussing interpretability of the Boolean machine learning algorithm and applying it on Breast cancer dataset which is very much relevant to the research topic.
2.	Intuitive and interpretable visual communication of a complex statistical model of disease progression and risk	Excluded	Although the conference paper is discussing about interpretability in visual communication, the paper does not focus on the cancer detection.
3.	Diabetes prognosis using white-box machine learning framework for interpretability of results	Excluded	This conference paper is excluded from the final list because it focuses on diabetes prognosis rather than cancer detection.
4.	Explainable Machine Learning for Breast Cancer Diagnosis	Included	This conference paper is included as it is very relevant to the research topic as it is discussing the concept explainability for the breast cancer diagnosis.
5.	Exploring Machine Learning Algorithms to Identify Heart Failure Patients: the Tuscany Region Case Study	Excluded	This conference paper is not relevant and will not be useful to the research.
6.	Improving the Correctness of Medical Diagnostics Based on Machine Learning With Coloured Petri Nets	Included	This journal article is included because it discusses the improvement of decision-making process and demonstrated the effectiveness of the

			algorithm using breast cancer datasets.
7.	A new machine learning technique for predicting traumatic injuries outcomes based on the vital signs	Excluded	This conference paper is not relevant and will not be useful to the research.
8.	CIDMP: Completely Interpretable Detection of Malaria Parasite in Red Blood Cells using Lower-dimensional Feature Space	Excluded	This conference paper is not relevant as it focused on Pathology rather than Oncology.
9.	A Precision Environment-Wide Association Study of Hypertension via Supervised Cadre Models	Excluded	Not a relevant journal article.
10.	Interpretable Prediction of Diabetes from Tabular Health Screening Records Using an Attentional Neural Network	Excluded	This conference paper is not relevant although it discusses interpretability but on diabetes which is not used for the research.
11.	An Interpretable Model for ECG Data Based on Bayesian Neural Networks	Included	This Journal article would be helpful for the research.
12.	Interpretability methods of machine learning algorithms with applications in breast cancer diagnosis	Included	This conference paper is very relevant to the research as it discusses the key concepts like interpretability, Black box in breast cancer diagnosis.
13.	Non-Invasive Meningitis Diagnosis Using Decision Trees	Excluded	The Journal article is mainly focused on Meningococcal Disease which is not useful for the research.
14.	Interpretable Quantitative Description of the Digital Clock Drawing Test for Parkinson's Disease Modelling	Excluded	The Conference paper is mainly focused on Parkinson's Disease which is not relevant for the research.
15.	SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection	Included	This Journal article is suitable and used for the research as it told that an interpretable SetSVM model can be used for thyroid cancer, liver cancer and melanoma. It also had total citations of 11.
16.	Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification	Included	This Journal article is included as it discusses about the image segmentation technique on chest

			radiographs for heart and lung cancer diagnosis.
17.	Revealing Common and Rare Patterns for Peritoneal Dialysis Eligibility Decisions with Association Discovery and Disentanglement	Excluded	Not a relevant conference paper.
18.	A Hierarchical Fuzzy System for Risk Assessment of Cardiovascular Disease	Excluded	Not a relevant conference paper as it is related Cardiovascular disease and does not discuss any key concepts.
19.	An improved random forest-based rule extraction method for breast cancer diagnosis	Included	It is relevant as discusses about interpretability in breast cancer diagnosis.
20.	Artificial intelligence: A promising frontier in bladder cancer diagnosis and outcome prediction	Included	This Review article would be useful to the research topic.
21.	Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images	Included	This Research article is useful to the research as discusses the explainability in decision making.
22.	Artificial intelligence for COVID-19: battling the pandemic with computational intelligence	Excluded	Not a relevant Review article.
23.	NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images	Excluded	Not a relevant Research article
24.	An ISHAP-based interpretation-model-guided classification method for malignant pulmonary nodule	Included	This Research article is relevant to the topic as it discusses one of the main key concepts.
25.	Segmentation of cellular patterns in confocal images of melanocytic lesions in vivo via a multiscale encoder-decoder network (MED-Net)	Excluded	Segmentation is involved but it is too irrelevant to the research topic.
26.	The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending	Excluded	This Journal would be helpful for the research, but it's not anywhere related to cancer diagnosis.
27.	Interpretability in the medical field: A systematic mapping and review study	Included	This Review article would be useful for our research.
28.	Using Explainable Machine Learning to Explore the Impact of Synoptic Reporting on Prostate Cancer	Included	This Research article is very much relevant to the research topic.
29.	Deep learning in cancer diagnosis, prognosis and treatment selection	Included	This Review article would be useful for the research as it discussed explainability in deep learning model.

30.	Artificial intelligence for clinical oncology	Included	This Review article would be useful for the research.
31.	A review of explainable deep learning cancer detection models in medical imaging	Included	This article is included as it discussed the concept explainability in cancer detection.
32.	Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications	Excluded	The Research article had discussed one of the key concepts, but not in cancer detection.
33.	A survey of explainable artificial intelligence decision	Excluded	This research article is relevant but not useful.
34.	A Systematic Review of Human-Computer Interaction and Explainable Artificial Intelligence in Healthcare with Artificial Intelligence Techniques	Included	This paper's objective is relevant to the topic.
35.	Explainable artificial intelligence: a comprehensive review	Included	This Research article would be useful to the research.
36.	Development of an Explainable Prediction Model of Heart Failure Survival by Using Ensemble Trees	Excluded	This conference paper is excluded, although one of the key concepts is discussed but on Heart failure.
37.	Interpretability of machine learning-based prediction models in healthcare	Included	This Review article is useful for the research and it had total citations of 46.
38.	Learning optimized risk scores	Excluded	This article is too irrelevant.
39.	Toward Efficient Automation of Interpretable Machine Learning	Included	This conference paper is included as key concepts are covered in this and it has 8 citations.

TOTAL INCLUDED: 20

TOTAL EXCLUDED: 19

CONCLUSIONS

By doing this systematic literature review, it allowed me to become acquainted with current knowledge in my chosen topic, as well as the field's bounds and limitations. I also identified knowledge gaps and unsolved challenges that my study can solve. Most importantly also got to know to select the correct keywords, Document type, Article type and also to apply right filters. It improved my research methodology and usage of strategies. Finally, it has helped me conceptualize my research problem clearly and precisely.

My literature review search was good some of the times, but sometimes in some databases found too many irrelevant articles and it because I couldn't continue with the database ACM Digital Library. The three databases I used were IEEE Xplore, Science Direct and Scopus where I learned the most. At the beginning it didn't go as expected because of the ACM Digital Library database as there were many articles and those are irrelevant ones. The limitations that I have encountered were in some of the databases I thought I was going on a different route than expected in the literature search as I was getting wide range of results. I had to use different keywords which are related to the research topic and also different filters to narrow down to lesser number of results. By the end of this literature review I and sure that I used all the research strategies, keywords, filters and also the criteria to select that specific journal or conference paper correctly and wisely.

REFERENCES

A. Khan *et al.* (2020; 2020) 'CIDMP: Completely Interpretable Detection of Malaria Parasite in Red Blood Cells using Lower-dimensional Feature Space; A Precision Environment-Wide Association Study of Hypertension via Supervised Cadre Models', - *2020 International Joint Conference on Neural Networks (IJCNN); - IEEE Journal of Biomedical and Health Informatics.* doi: 10.1109/IJCNN48605.2020.9206885; 10.1109/JBHI.2019.2918070.

F. Almaghrabi, D. -L. Xu and J. -B. Yang (2019) 'A new machine learning technique for predicting traumatic injuries outcomes based on the vital signs', - *2019 25th International Conference on Automation and Computing (ICAC)*. doi: 10.23919/IConAC.2019.8895012.

G. Casalino *et al.* (2020) 'A Hierarchical Fuzzy System for Risk Assessment of Cardiovascular Disease', - *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*. doi: 10.1109/EAIS48028.2020.9122750.

J. Li and O. Arandjelović (2017) 'Intuitive and interpretable visual communication of a complex statistical model of disease progression and risk', - *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. doi: 10.1109/EMBC.2017.8037782.

N. Neuhaus and B. Kovalerchuk (2019) 'Interpretable Machine Learning with Boosting by Boolean Algorithm', - *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. doi: 10.1109/ICIEV.2019.8858511.

P. F. Khan and K. Meehan (2021) 'Diabetes prognosis using white-box machine learning framework for interpretability of results', - *2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC)*. doi: 10.1109/CCWC51732.2021.9375927.

P. Karatza *et al.* (2021; 2020) 'Interpretability methods of machine learning algorithms with applications in breast cancer diagnosis; Non-Invasive Meningitis Diagnosis Using Decision Trees', - *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); - IEEE Access.* doi: 10.1109/EMBC46164.2021.9630556; 10.1109/ACCESS.2020.2966397.

P. Zhou *et al.* (2020) 'Revealing Common and Rare Patterns for Peritoneal Dialysis Eligibility Decisions with Association Discovery and Disentanglement', - *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. doi: 10.1109/BIBM49941.2020.9313486.

S. Nömm *et al.* (2018; 2019; 2020) 'Interpretable Quantitative Description of the Digital Clock Drawing Test for Parkinson's Disease Modelling; SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection; Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification', - *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV); - IEEE Journal of Biomedical and Health Informatics; - IEEE Access.* doi: 10.1109/ICARCV.2018.8581074; 10.1109/JBHI.2018.2803793; 10.1109/ACCESS.2020.2995567.

S. Panicacci *et al.* (2019; 2021) 'Exploring Machine Learning Algorithms to Identify Heart Failure Patients: the Tuscany Region Case Study; Improving the Correctness of Medical Diagnostics Based on Machine Learning With Coloured Petri Nets', - *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS); - IEEE Access.* doi: 10.1109/CBMS.2019.00088; 10.1109/ACCESS.2021.3121092.

T. Brito-Sarracino *et al.* (2019) 'Explainable Machine Learning for Breast Cancer Diagnosis', - *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*. doi: 10.1109/BRACIS.2019.00124.

Y. Oba *et al.* (2021; 2021) 'Interpretable Prediction of Diabetes from Tabular Health Screening Records Using an Attentional Neural Network; An Interpretable Model for ECG Data Based on Bayesian Neural Networks', - *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA); - IEEE Access.* doi: 10.1109/DSAA53316.2021.9564151; 10.1109/ACCESS.2021.3071731.

A. New *et al.* (2020) *A Precision Environment-Wide Association Study of Hypertension via Supervised Cadre Models; An Interpretable Model for ECG Data Based on Bayesian Neural Networks; Non-Invasive Meningitis Diagnosis Using Decision Trees; SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection; Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification.*

C. Liu *et al.* (2020) *SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection; Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification.*

E. Sogancioglu *et al.* (2020) *Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification.*

M. Nauman *et al.* (2020) *Improving the Correctness of Medical Diagnostics Based on Machine Learning With Coloured Petri Nets; A Precision Environment-Wide Association Study of Hypertension via Supervised Cadre Models; An Interpretable Model for ECG Data Based on Bayesian Neural Networks; Non-Invasive Meningitis Diagnosis Using Decision Trees; SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection; Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification.*

Q. Hua *et al.* (2020) *An Interpretable Model for ECG Data Based on Bayesian Neural Networks; Non-Invasive Meningitis Diagnosis Using Decision Trees; SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection; Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification.*

V. M. Lelis *et al.* (2020) *Non-Invasive Meningitis Diagnosis Using Decision Trees; SetSVM: An Approach to Set Classification in Nuclei-Based Cancer Detection; Cardiomegaly Detection on Chest Radiographs: Segmentation Versus Classification.*

Borhani, S., Borhani, R. and Kajdacsy-Balla, A. (2022) 'Artificial intelligence: A promising frontier in bladder cancer diagnosis and outcome prediction', *Critical Reviews in Oncology/Hematology*, 171, pp. 103601. doi: https://doi.org/10.1016/j.critrevonc.2022.103601.

He, W. *et al.* (2022) 'An ISHAP-based interpretation-model-guided classification method for malignant pulmonary nodule', *Knowledge-Based Systems*, 237, pp. 107778. doi: https://doi.org/10.1016/j.knosys.2021.107778.

Jiménez-Sánchez, D. *et al.* (2022) 'NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images', *Medical Image Analysis*, 78, pp. 102384. doi: https://doi.org/10.1016/j.media.2022.102384.

Kose, K. *et al.* (2021) 'Segmentation of cellular patterns in confocal images of melanocytic lesions in vivo via a multiscale encoder-decoder network (MED-Net)', *Medical Image Analysis*, 67, pp. 101841. doi: https://doi.org/10.1016/j.media.2020.101841.

Sabol, P. *et al.* (2020) 'Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images', *Journal of Biomedical Informatics*, 109, pp. 103523. doi: https://doi.org/10.1016/j.jbi.2020.103523.

Wang, S. *et al.* (2020) 'An improved random forest-based rule extraction method for breast cancer diagnosis', *Applied Soft Computing*, 86, pp. 105941. doi: https://doi.org/10.1016/j.asoc.2019.105941.

Xu, Z. *et al.* (2022) 'Artificial intelligence for COVID-19: battling the pandemic with computational intelligence', *Intelligent Medicine*, 2(1), pp. 13-29. doi: https://doi.org/10.1016/j.imed.2021.09.001.

Dikmen, M. and Burns, C. (2022) 'The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending', *International Journal of Human Computer Studies*, 162 doi: 10.1016/j.ijhcs.2022.102792.

Gulum, M.A., Trombley, C.M. and Kantardzic, M. (2021) 'A review of explainable deep learning cancer detection models in medical imaging', *Applied Sciences (Switzerland)*, 11(10) doi: 10.3390/app11104573.

Hakkoum, H., Abnane, I. and Idri, A. (2022) 'Interpretability in the medical field: A systematic mapping and review study', *Applied Soft Computing*, 117 doi: 10.1016/j.asoc.2021.108391.

Janssen, F.M. *et al.* (2022) 'Using Explainable Machine Learning to Explore the Impact of Synoptic Reporting on Prostate Cancer', *Algorithms*, 15(2) doi: 10.3390/a15020049.

Kann, B.H., Hosny, A. and Aerts, H. J. W. L. (2021) 'Artificial intelligence for clinical oncology', *Cancer Cell*, 39(7), pp. 916-927. doi: 10.1016/j.ccell.2021.04.002.

Kong, X., Tang, X. and Wang, Z. (2021) 'A survey of explainable artificial intelligence decision', *Xitong Gongcheng Lilun Yu Shijian/System Engineering Theory and Practice*, 41(2), pp. 524-536. doi: 10.12011/SETP2020-1536.

Kovalerchuk, B. and Neuhaus, N. (2019) 'Toward Efficient Automation of Interpretable Machine Learning', . Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85062613534&doi=10.1109%2fBigData.2018.8622433&partnerID=40&md5=973e1b11949ecabb5ed8c82b783d7524 doi: 10.1109/BigData.2018.8622433.

Minh, D. *et al.* (2021) 'Explainable artificial intelligence: a comprehensive review', *Artificial Intelligence Review*, doi: 10.1007/s10462-021-10088-y.

Moreno-Sanchez, P.A. (2020) 'Development of an Explainable Prediction Model of Heart Failure Survival by Using Ensemble Trees', . Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85103815191&doi=10.1109%2fBigData50022.2020.9378460&partnerID=40&md5=82dbef59d9fea48bda3bf4ca82a0b752 doi: 10.1109/BigData50022.2020.9378460.

Nazar, M. *et al.* (2021) 'A Systematic Review of Human-Computer Interaction and Explainable Artificial Intelligence in Healthcare with Artificial Intelligence Techniques', *IEEE Access*, 9, pp. 153316-153348. doi: 10.1109/ACCESS.2021.3127881.

Samek, W. *et al.* (2021) 'Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications', *Proceedings of the IEEE*, 109(3), pp. 247-278. doi: 10.1109/JPROC.2021.3060483.

Stiglic, G. *et al.* (2020) 'Interpretability of machine learning-based prediction models in healthcare', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5) doi: 10.1002/widm.1379.

Tran, K.A. *et al.* (2021) 'Deep learning in cancer diagnosis, prognosis and treatment selection', *Genome Medicine*, 13(1) doi: 10.1186/s13073-021-00968-x.

Ustun, B. and Rudin, C. (2019) 'Learning optimized risk scores', *Journal of Machine Learning Research*, 20.