Explainable Artificial Intelligence in Healthcare: Opportunities, Gaps and Challenges and a Novel Way to Look at the Problem Space

Explainable Artificial Intelligence in Healthcare: Opportunities, Gaps and Challenges and a Novel Way to Look at the Problem Space

Petra Korica[0000-0002-9066-2926], Neamat El Gayar[0000-0003-1467-1115] and Wei Pang[0000-0002-1761-6659]

School of Mathematical and Computer Sciences
Heriot-Watt University
{pk2005, n.elgayar, w.pang}@hw.ac.uk

Abstract. Explainable Artificial Intelligence (XAI) is rapidly becoming an emerging and fast-growing research field; however, its adoption in healthcare is still at the early stage despite the potential that XAI can bring to the application of AI in this industry. Many challenges remain to be solved, including setting standards for explanations, the degree of interaction between different stakeholders and the models, the implementation of quality and performance metrics, the agreement on standards for safety and accountability, its integration into clinical workflows, and IT infrastructure. This paper has two objectives. The first one is to present summarized outcomes of a literature survey and highlight the state-of-the-art for explainability including gaps, challenges, and opportunities for XAI in healthcare industry. For easier comprehension and onboarding to this research field we suggest a synthesized taxonomy for categorizing explainability methods. The second objective is to ask the question if applying a novel way of looking at explainability problem space, through a specific problem/domain lens, and automating that approach in an AutoML similar fashion, would help mitigate the challenges mentioned above. In the literature there is a tendency to look at the explainability of AI from model-first lens, which puts concrete problems and domains aside. For example, the explainability of a patient's survival model is treated the same as explaining a hospital cost procedure calculation. With a well-identified problem/domain that XAI should be applied to, the scope is clear and well-defined, enabling us to (semi-) automatically find suitable models, optimize their parameters and their explanations, metrics, stakeholders, safety/accountability level, and suggest means of their integration into clinical workflow.

Keywords: Artificial Intelligence, Machine Learning, Interpretability, Explainability, Explainable AI, XAI, AI in Healthcare.

1 Introduction

The lack of explainability and transparency of state-of-the-art artificial intelligence (AI) systems is one of the main reasons why AI is not yet (fully) trusted and hence

widely deployed in many real-world problems, including healthcare problems. World Health Organization (WHO) recently published a set of guidelines for the future of AI in healthcare, which include "ensuring explainability" [1]. Healthcare can uniquely benefit from the applications of AI, e.g., saving a significant number of lives through early diagnosis or drug development; however, it also poses unique challenges and risks if AI is not carefully implemented and used. Healthcare benefits from AI models working together with humans, assisting humans in decision making, for example, in diagnosis and prognosis. Explainable AI models can be of high value in holding AI models and assisted decision-making systems accountable [2], which is vital for their applications in healthcare. Accountable Machine Learning is another emerging topic that is related to the explainability of AI, and the machine learning (ML) systems. ML systems encompass ML models, supporting infrastructure, hardware, and processes etc. Explainability in this context is used to enable the system to justify its outputs, meaning its decisions, predictions and even the reasoning process in some cases, which is an important part of accountability of a system [3].

The first objective and main contribution of this paper is to present the outcomes of the literature survey that we have done and to create a taxonomy that summarizes as many aspects of XAI as possible into a ”simple” structure. The guiding idea is to categorize and make the vast knowledge quicker addressable, foster joint understanding and easier comprehensibility of the field as well as to highlight the state-of-the-art for explainability as well as the gaps, challenges, and opportunities explainable artificial intelligence faces in the healthcare industry.

In the literature there is a tendency to look at the explainability from model-first lens approach, putting the concrete problems and domains aside. Examples of such approach would be: 1) applying black-box XAI methods after training without considering how to reformulate the problem to be explainable, 2) without considering the users of the model, 3) without considering the recipients of the explanation and 4) applying the same XAI methods to different problems in different domains. Considering this, our second objective is to pose a question whether applying a novel way of looking at the explainability problem space by applying a so-called ”problem/domain lens” approach would help the adoption of explainable AI in healthcare. The proposed approach is looking at the domain and problem that the ML models are trying to solve and based on that offers guidance on what XAI methods with what parameters, what form of explanations, what stakeholders, what regulations should be looked at. Furthermore, automating that approach in an AutoML similar way, would help mitigate the challenges mentioned above.

The rest of the paper is organised as follows: after looking at current research in the Section 2, we will identify the challenges in Section 3. This is followed by Section 4, in which we will show an example of the proposed approach for the healthcare industry.

2 Current Outlook on XAI

Explainable AI (XAI) and its application to healthcare is an emerging research field influenced by regulatory frameworks and research programs such as DARPA XAI [4], GDPR [5] and impending Artificial Intelligence Act [6]. Although this field is not completely new, it has been exponentially growing since the launch of DARPA XAI program [4] in 2017. For example, between February and July of 2021, there were around 6,500 new papers published mentioning "Explainable AI" and around 8,500 mentioning "Interpretable AI" [7, 8]. This is an incredible rate of increase of publications which shows the difficulties to keep track of for the researchers in the field.

XAI is a multi-disciplinary research field influenced by psychology, philosophy, sociology, human-computer interface, and education, where the recipient of the explanation is important. However, this field is still immature in its applications to real-life situations, and there is work to be done by the research community to establish mutual understanding and standards for definitions, safety, metrics, etc. Let's look at the proposed definitions for XAI. In the literature, e.g. [9], the "Interpretability" is defined as the quality or feature of a model providing enough expressive data to understand how the model works. In literature the term domain is noted as relevant for the interpretability [9, 10], with some authors such as [10] are arguing that interpretability is inherently domain specific.

On the other hand, "Explainability" can be defined as the ability of models to summarize the reasons for their behavior, gain the trust of users or produce insights into the causes of their decisions [11]. Explainability entails the definition of an explanation. "Explanation" is defined in the literature [12, 13] as a statement, fact or situation that tells one why something happened. There are several desired facts to explanations such as accuracy, fidelity, consistency, stability, comprehensibility, relevance, and certainty [14]. A good explanation needs to be produced responsibly and evaluated rigorously. In [15] the authors bring an important aspect in addition to the definitions above: "The explainability of an AI systems behavior needs to consider different dimensions: 1) who is the receiver of that explanation, 2) why that explanation is needed, and 3) in which context and other situated information the explanation is presented." This again brings us to the domain and problem relevancy.

Model interpretability can be intrinsic or post-hoc (external). Intrinsic means that the interpretability is achieved by restricting the complexity of the ML model at the time of building while post-hoc means that interpretability is achieved by applying methods that analyze the model after training and generate an explanation [14]. In case of intrinsically interpretable models, the explanation is the model itself.

In the literature, we found somewhat different but related categorizations of the methodology used for applying XAI. We consolidated the different points of view and combined them into following categorization of XAI that is adapted from [9, 14, 16-21, 29]:

- **Model Design:** Intrinsic, post-hoc.
- **Scope:** Local (explaining just one instance of prediction), global (explaining entire model behavior).
- **Relevancy to Model:** Model-specific, model-agnostic.
- **Methodology:** Example-based, simplification-based, feature relevance-based, perturbation-based, back propagation-, gradient-based, ontology-based.
- **Timing:** Pre-model, in-model, post-model.
- **Presentation:** Visualization, text explanation, mathematical explanation.
- **Data Type:** Text, tabular, image, graphs.

So far, many applications of XAI are on deep learning models due to their powerful but opaque nature with the goal to provide easy model-agnostic explanation of already designed models. In the literature, most used approaches are post-hoc model-agnostic, which means that they can be applied to "any" black-box models after they have been trained (e.g., in literature we see high usage of XAI methods focused on local instances such as LIME [22], feature relevance-based SHAP [23] or Partial Dependence Plot [24], backpropagation or gradient-based such as Attribution Maps or DeepLIFT [25]). The main benefit of post-hoc specific models is their flexibility to be applied to any existing or new machine learning models without the need to understand the structure and internal functioning of the model. The opposite of post-hoc methods are intrinsically interpretable models where the model itself can be used for explanation, see [10] for details. An example of those are the visual transformers which are gaining traction due to the explanations given by their built-in attention map, and we see their usage increasing in healthcare applications, for example in [38, 39]. We also see increased usage of model-specific methods (such as Grad-CAM [26]) for deep learning networks. Model-specific XAI methods are generally using some of the model-specific architecture to create explainability such as removing the last layer and creating a class activation heatmap, as described in [26].

Explainability of a machine learning model needs to be presented in different ways to different stakeholders as they will need a personalized view and level of depth. Following a workflow of a usual ML project, the stakeholders could be summarized into the following categories which are synthetized from [12, 20, 27]: Model Builder (data scientist/ML developer, ML ops or IT/Dev ops engineer), Model Breakers (domain experts, business decision makers, auditors) and Model Consumers (provider of the solution/service using the ML model, end user of the model).

3 Opportunities, Gaps and Challenges

With reference to above mentioned points, below is an attempt to outline the most important areas that the future research in XAI should address. We haven't classified the points in opportunities, gaps, and challenges as we believe that each of the points represents a gap and offers an opportunity or challenge based on how well the point is addressed. All below-mentioned points are relevant for applying XAI to the healthcare domain:

- **Definition of the terms:** Agree upon definitions of interpretability and explainability, agree upon vocabulary and taxonomy of XAI methods, see [9-11, 14, 16-21, 29].
- **Explanations:** Reach an agreement what are good, human-friendly explanations, perform more exploration of the human-computer-interaction aspect of explanations such as visualization, using concepts or ontologies, etc. Create benchmarking models for the quality of explanations, automatic generations of explanations from the ML model, and reducing human subjectivity. Investigate legal implications of explanations provided, etc. See [12-15, 17, 29] for details.
- **Quality and performance metrics:** Create frameworks for the evaluation of performance and definition of agreed upon metrics for measuring and benchmarking XAI methods, see [9, 10, 13, 14]. An interesting interpretability challenge was found in the user study carried out with data scientists by [28] with the results of the study indicating that data scientists tend to over-trust and not correctly use the interpretability tools which can have dangerous implications.
- **ML Ops:** Integration and automation of XAI methods into ML model life cycle and deployment model, see [12, 20, 27, 28].
- **Safety:** We need to further research on security of explainability of AI, including methods to prohibit the fooling of XAI methods through perturbations and randomized input as well as methods to mitigate inferring private and sensitive information through explanations, see [9, 10, 13, 17, 19-21]. In addition, XAI methods could be relevant to expose risks entailed in the large parameter natural language models such as GPT-3 [34] whose usage is growing. We are just becoming aware of additional security risks that could be impactful in healthcare. In [35] the authors claim that such models can memorize parts of the training data within their parameters. This means that it is possible to carry out attacks to retrieve potentially privacy-sensitive information that were present in the training data.
- **Regulations:** Create regulatory and legislative framework for XAI involving fairness, protection of privacy, truthfulness of explanations and accountability, see [10, 14, 18].
- **Human - ML model collaboration:** Are there ways to work on creating explainability jointly and interactively using novel XAI methods and create a dialog and collaboration between human and the ML model to be interpreted? There is a lack

of methods to generate feedback from human to ML model (the so-called machine teaching). Define a clear framework of integration into human workflows. See [9, 10, 14, 18] for details.

In addition, we foresee the need for specific adjustments in healthcare such as specialized interfaces for various medical situations, e.g., for emergency room or surgery preparations, more rigorous definitions of data privacy, more comprehensive processing guidelines to ensure patient safety as well as a more rigorous definition of the level of explanation to be provided. The topics of integrating XAI into healthcare workflows, accountability, and safety of the XAI methods used are very important for healthcare. Furthermore, creating and implementing a special auditing process or framework with safeguards checks for XAI for healthcare could be beneficial to facilitate fairness, safety, stability, and fidelity of XAI.

Another potential restriction is to audit the quality of the data used with the goal to ensure that the data that the machine learning systems have been trained on is not of low quality, not biased or perhaps even wrongly annotated data and to ensure that there is no privacy-sensitive information leak. We believe that data access is a bigger issue in XAI in healthcare, and healthcare research overall as it is very difficult to obtain data in the first place. In fact, an interesting area for the future research could be the explainability of the synthetic data in healthcare. As there are difficulties in obtaining the access and sharing health data for research, we believe that next to federated learning, there will be more realistic looking synthetic health data produced. We believe this could be an interesting area of research due to difficulties with obtaining and cleaning training data as the research community will need to prove the quality of training the model using synthetic data and that models trained in such a fashion generate well on real patients' data without the risk of exposing any privacy-sensitive information.

4 Recommendations for Applying a Problem/Domain Approach to Interpretability and Explainability

In section 3, we highlighted the main gaps/issues for XAI which are slowing down the adoption of AI in healthcare. As stated previously, the state-of-the-art approach is looking at explainability through "model lens": model-agnostic interpretability, model-specific interpretability, and intrinsically interpretable models. The same "model lens" is then applied to any problem and any domain. See [16, 18, 36-39] for examples of such previous work. We believe that this way of applying XAI often creates unnecessary complex situations as different problems require different levels of explainability, and they have different stakeholders and regulations that often might

require different explanations. For example, in healthcare domain itself, the scenarios can be very different: explainability of a model in an emergency room scenario, dealing with life-or-death situations, will have different requirements compared to explainability of a model in drug discovery. The importance of the domain for XAI has also been noted by authors in e.g. [10, 18]. To address this, we suggest applying a "problem/domain lens" approach to the explainability from the start and using this approach to (semi-) automatically find the most suitable XAI model for a problem in a domain and if none is found, develop a new model by combining or redesigning existing ones or developing completely new problem/domain explainable models considering relevant explanations, performance, and safety metrics, etc.

While this approach will not mitigate all issues stated, we expect improvements by capitalizing on the knowledge of the problem/domain in our approach. This makes it easier to design quality, performance, safety level, needed explanations level and facilitate human and model collaboration. Let us look at an example in medical imaging:

Problem definition: Using AI to detect and classify type and potential abnormality of white blood cells in patients' blood smear images with suspicion of haematological malignancy.

Potential machine learning algorithms: We know the common machine learning algorithms applied for such problems – e.g., CNN, Capsule Nets, Visual Transformers, etc. Note that we can also design intrinsically interpretable deep learning models for image recognition such as adapted versions of CNN like [30] or different architectures like [31, 38-39].

Potential XAI methods: If we are not using an intrinsically interpretable model, we can use some of the common XAI methods like SHAP [23], LIME [22], gradient-based Grad-CAM [26] or attribution-propagation such as layer-wise relevance propagation (LRP) [32], attention map [38-39] and others to verify that the model is looking at the right pixels/patches in the image for its decision. This is something that our approach could (semi-) automatically suggest given the problem/domain. Figure 1 below shows one of the experiments we conducted in exploring suitable XAI methods for the problem we defined above. The figure shows the results of using CNN and Grad-CAM for Acute Lymphoblastic Leukaemia (ALL) classifying lymphoblasts on a blood-smear image of ALL-IDB database [40].

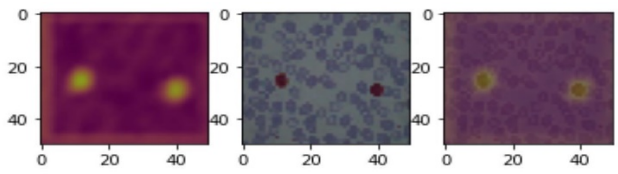

Fig. 1. An example of Grad-CAM implementation on ALL blood smears

Metrics: We can identify the KPIs that are relevant for this problem and domain, such as accuracy, number of patients that pathologist can perform this analysis per day, turnaround time from patient reception to diagnosis, response time for blood smear analysis. See [33] for more examples.
Stakeholders: For this problem the types of stakeholders such as doctors (pathologist, haematologist), nurses, patients, and insurance companies. Also, we need to consider that we might need to discuss with stakeholders what XAI methods would be acceptable/preferable to them.
Safety and regulatory level: We need to be aware of the regulations, laws, and standards such as hospital regulations, insurance regulations, GDPR or future Artificial Intelligence Act, etc.

Similar to automated machine learning (AutoML) [41], we are working on expanding this idea to conceptualize a framework/algorithm to automatize XAI for practitioners. The example above is just a short example, however having the knowledge of the problem and the domain lets us tailor a unique approach of the explainability and explanations that are relevant, using industry language and standards which makes them easy to use for the stakeholders. Using the AutoML or AutoXAI idea we could add a degree of automatism in selecting the right XAI method, parameter optimization, metrics, stakeholders, and regulatory suggestions, and selecting the right level and type of explanations.

5 Conclusion and Future Work

In conclusion, it is important to address the opportunities and the challenges of explainable AI thus enabling wider use and deployment of explainable machine learning assisted decision-making support systems in healthcare. As stated in the introductory chapter, the main objective and contribution of this paper is two-fold: we communicated the summary of our field survey and for easier comprehension and onboarding to this research field we suggest a synthesized taxonomy for categorizing explainability methods and a summary of opportunities, gaps, and challenges for applying explainability of AI to healthcare.

Secondly, we are investigating with our research question how looking at the problem, the context, and the domain where the ML model will be applied can simplify, streamline, and personalize the applications of XAI methods. Furthermore, we are exploring whether this approach be done in an AutoML similar fashion, as AutoXAI, to support the ML practitioners in applying the right “configuration settings” to a XAI application.

Our initial experience of the work in progress using this approach shows promise. In our future work, we will be assessing the results of applying this problem/domain

lens on explainability and pointing out gaps as well as suggesting future areas of research.

References

1. World Health Organization (WHO). https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use. Last accessed on 2021/07/14.
2. Aurangzeb, A. M., Eckert, C., Teredesai, A.: Interpretable Machine Learning in Healthcare. In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, pp. 559-560 (2018).
3. Pang, W., Markovic, M., Naja, I., Fung, C.P. and Edwards, P.: On Evidence Capture for Accountable AI Systems. In: SICSA Workshop on eXplainable Artificial Intelligence (XAI) (2021).
4. Gunning, D., Aha, D.: Explainable artificial intelligence (XAI) program. AI Magazine, 40(2), 44-58 (2019)
5. European Law General Data Protection Regulation. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434. Last accessed 2021/07/27.
6. European Commission Artificial Intelligence Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206. lLst accessed on 2021/07/18.
7. Dimensions query “Explainable AND Artificial Intelligence”. https://app.dimensions.ai/analytics/publication/overview/timeline?search_mode= content&search_text=explainable%20AND%20%22artificial%20intelligence%22&search_type= kws&search_field=full_search. Last accessed 2021/07/14.
8. Dimensions query “Interpretable AND Artificial Intelligence”. https://app.dimensions.ai/ analyics/publication/overview/timeline?search_mode=content&search_text=interpretable % 20AND % 20 % 22artificial% 20intelligence% 22 & search _type = kws & search _field= full_search. Last accessed 2021/07/14.
9. Das, A., Rad, P.: Opportunities and challenges in explainable artificial intelligence (XAI): A survey. ArXiv preprint arXiv:2006.11371 (2020).
10. Rudin, C.: Stop explaining black box machine learning models for high stakes decisions anduse interpretable models instead. Nature Machine Intelligence 1(5), 206–215 (2019).
11. Gilpin, L. H, Bau, D., Yuan, B. Z. Bajwa, A., Specter M., Kagal L.: Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA), pp. 80–89, IEEE (2018).
12. Derek, D., Schulz, S. Besold, T. R.: What does explainable AI really mean? A new conceptualization of perspectives. ArXiv preprint arXiv:1710.00794 (2017).
13. Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. ArXiv preprint arXiv:1702.08608 (2017).
14. Molnar, C.: Interpretable Machine Learning, A Guide for Making Black Box Models Explainable. Leanpub, Monee, IL, USA (2020)

15. Ferreira, J. J., Monteiro, M. S.: What Are People Doing About XAI User Experience? A Survey on AI Explainability Research and Practice. In: Design, User Experience, and Usability. Design for Contemporary Interactive Environments, pp. 56–73. Springer (2020).
16. Tjoa, E., Guan C.: A survey on explainable artificial intelligence (XAI): Toward medical XAI". IEEE Transactions on Neural Networks and Learning Systems, 1–21 (2020).
17. Longo, L., Goebel, R., Lecue, F., Kieseberg, P, Holzinger., A.: Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions. In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction, pp. 1–16. Springer (2020).
18. Adadi, A, Berrada. M.: Peeking inside the black box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access (6), 52138–52160 (2018).
19. Carvalho, D. V., Pereira, E. M., Cardoso: Machine Learning Interpretability: A Survey on Methods and Metrics". Electronics 8(8), 832 (2019).
20. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R. and Chatila, R.: Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion (58), 82-115 (2020).
21. Linardatos, P., Papastefanopoulos, V. and Kotsiantis, S.: Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy 23(1), 18 (2021).
22. Ribeiro, M. T., Singh, S., Guestrin, C.: Why should i trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 1135–1144 (2016).
23. Lundberg, S., Lee, S. I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS), pp. 4765-4774 (2017).
24. Friedman, J. H.: Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189–1232 (2001).
25. Avanti, S., Greenside, P., Kundaje A.: Learning important features through propagating activation differences. In: International Conference on Machine Learning. PMLR, pp. 3145–3153 (2017).
26. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision, pp. 618-626. IEEE (2017).
27. Hong, S, R., Hullman, J., Bertini E.: Human factors in model interpretability: Industry practices, challenges, and needs. In: Proceedings of the ACM on Human-Computer Interaction 4 CSCW1, pp. 1–26 (2020).
28. Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., Wortman Vaughan J.: Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–14 (2020).
29. Carrilo, A., Cantu, L. F., Noriega, A.: Individual Explanations in Machine Learning models: A Survey for Pratictioners. ArXiv preprint arXiv:2104.04144 (2021).
30. Chen, C., Li, O., Tao, C., Barnett, A.J., Su, J. and Rudin, C.: This looks like that: deep learning for interpretable image recognition. arXiv preprint arXiv:1806.10574 (2018).

31. Singh, G., Yow, K. C.: These do not Look Like Those: An Interpretable Deep Learning Model for Image Recognition. IEEE Access (9), 41482-41493 (2021).
32. Eitel, F., Soehler, E., Bellmann-Strobl, J., Brandt, A.U., Ruprecht, K., Giess, R.M., Kuchling, J., Asseyer, S., Weygandt, M., Haynes, J.D. and Scheel, M.: Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation. NeuroImage Clinical (24), 102003 (2019).
33. Royal College of Pathologists, Key Performance Indicators in Pathology. https://www.rcpath.org/uploads/assets/e7b7b680-a957-4f48-aa78e601e42816de/Key-Performance-Indicators-in-Pathology-Recommendations-from-the-Royal-College-of-Pathologists.pdf. Last accessed on 2021/07/25.
34. Floridi, L. and Chiriatti, M.: GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30(4), pp.681-694 (2020).
35. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, Roberts A, Brown T, Song D, Erlingsson U, Oprea A.: Extracting training data from large language models. ArXiv preprint arXiv:2012.07805 (2020).
36. Shaban-Nejad A., Michalowski M., Buckeridge D.L.: Explainability and Interpretability: Keys to Deep Medicine. In: Shaban-Nejad A., Michalowski M., Buckeridge D.L. (eds) Explainable AI in Healthcare and Medicine. Studies in Computational Intelligence, vol 914. Springer, Cham (2021).
37. Harsha, N., Jenkins, S., Koch, P., Caruana R: Interpretml: A unified framework for machine learning interpretability. ArXiv preprint arXiv:1909.09223 (2019).
38. Matsoukas, Christos, M., Haslum, J. F., Söderberg, M., Smith, K: Is it Time to Replace CNNs with Transformers for Medical Images?. ArXiv preprint arXiv:2108.09038 (2021), Accepted at ICCV-2021: Workshop on Computer Vision for Automated Medical Diagnosis (CVAMD).
39. Shi, Wenqi, S., Tong, L, Zhu Y, Wang, M. D.: COVID-19 Automatic Diagnosis with Radiographic Imaging: Explainable Attention Transfer Deep Neural Networks. IEEE Journal of Biomedical and Health Informatics (25), 2376-2386 (2021).
40. Labati, R. D., Piuri, V., Scotti F.: All-IDB: The acute lymphoblastic leukemia image database for image processing". In: 2011 18th IEEE International Conference on Image Processing, pp. 2045 -2048. IEEE (2011).
41. Hutter F, Kotthoff L, Vanschoren J.: Automated machine learning: methods, systems, challenges. Springer Nature (2019).