JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A COMPREHENSIVE REVIEW OF EXPLAINABLE MACHINE LEARNING MODELS IN CLINICAL APPLICATIONS.

Ajitha R. Subhamathi

NSS College Rajakumari, Idukki, Kerala, India

Abstract: This article presents a comprehensive review of explainable machine learning models, emphasising healthcare applications. It provides an overview of related works, highlighting the significance of transparency and interpretability in clinical imaging systems. The review summarizes a wide range of explainable AI (XAI) techniques employed in the healthcare domain, with special emphasis on their application in medical imaging tasks. Furthermore, it examines the evaluation metrics commonly used to assess the effectiveness and reliability of XAI methods, providing insights into their strengths, limitations, and practical applicability. Ultimately, this review article aims to be a comprehensive reference for researchers focusing on developing or adopting explainable machine learning models aiming to enhance trust, accountability, and usability in healthcare.

Keywords : Explainable AI Models in Healthcare, Explainable AI Models in Medical Imaging, Evaluation Metric for XAI Models in Healthcare.

1. INTRODUCTION

Artificial Intelligence is receiving considerable attention, especially in healthcare, since it improves the accuracy of disease diagnosis and delegates subjective treatment procedures. Exploiting extensive data sets, AI models generate accurate and initial-stage decision-making, especially in complex systems. The comprehensibility of such systems is crucial for ensuring trust and effective application of these models [40] [111]. A major challenge here is the requirement for domain-specific knowledge. Recent research in this direction concentrated on designing explainable machine learning systems to demonstrate the logic behind complex decisions. Most researchers address the problem of interpretability from the human frame of mind. The studies designate that people usually tend to disregard recommendations when they mistrust the system and blindly accept them when they trust it [38]. Enabling informed decision-making nurtures user trust by providing transparency in the decision-making process. The tutorial paper [101] discusses methods to address the interpretability, technical challenges, and potential applications of deep neural network models and explains its predictions. It also provides a framework, guidelines and regulations to make the most efficient use of the layer-wise relevance propagation technique, on real data. A wider use of explainable algorithms is discussed in the introductory paper [94], whereas [39] reviews explainable systems and their applications in predicting solubility, blood-brain barrier permeability, and the scent of molecules. The study in [81] focuses on how explanations of recommendations influence user trust while causability improves user understandability and emotional confidence. Behaviours in computer vision and arcade game tasks are discussed in [90], from naive and short-sighted to well-informed and strategic, concentrating on their explainability and problem-solving approaches. Most works in this direction focus on explanations from the researcher's perspective. Article [96] examines how the organisations utilise interpretability for stakeholder consumption and reveals that explanation techniques focus mostly on the researchers' perspective rather than directly benefiting end users. Despite the extensive observations concerning user choice in selection, development, and evaluation, researchers consider how individuals use specific cognitive biases and social expectations of explanation. It brings about the significance of explanations from a user perspective with improved trust. A framework [96] and a survey in [91] are based on this view.

The evaluative AI framework proposed in [38] utilizes a continuous feedback loop prototype in which both machines and humans contribute to interpreting evidence provided by the decision support system to arrive at conclusions. Counterfactuals or Interactive what-if questions are utilized in many explainable models [68][93][95]. Review in [68] focuses on scope and challenges, the theoretical foundations and computational frameworks are provided in [69], and a system causability scale is proposed in [78] as a quality metric for explainable systems. Machine learning provides insights into model behaviour using various methods such

as the relative contribution of features, counterfactual explanations, or the impact of specific data points. The significance of these methods is analysed in [30]. The interpretability of various visualization and symbolic representation techniques is discussed in [37], highlighting the challenges in constructing timely explanations that exactly reflect the process. Programming techniques for explainable systems are classified in [49]. It also discusses the related frameworks and tools. It focuses on the integration of multiple datasets for better explainability. The model-agnostic system proposed in [102] utilises specific conditions sufficient for predictions in complex models. Similar methods are discussed in [108], proposing individual predictions and corresponding explanations without redundancy. Moreover, illustrates its flexibility in explaining narrative in addition to image-based classification models. The annotation tool for emotional expression analysis introduced in [75] provides sufficient visualization to create more suitable mental models about the machine learning system. Merging of artificial intelligence and blockchain technology is proposed in [41] to provide more secure healthcare facilities. A feature representation space for images, text, and genomics data is constructed in [66]. As an initial connector, the interface utilizes knowledge bases. Different transfer learning models for classification in [53] ensure reliability by utilizing Local Interpretable Model-agnostic Explanations (LIME) for discussing the decision process in a specific classification. A literature review on counterfactuals and causability is presented in [68].

After analysing around a hundred similar research works in the medical field and medical imaging applications, this article presents a comprehensive review of explainable machine learning models, focusing on healthcare applications. It is organized as follows: Section 2 provides a summary of related work. Section 3 summarizes explainable AI techniques in the healthcare domain, and Section 4, in particular, concentrates on medical imaging applications. Section 5 comprehends the evaluation metrics used. Concluding remarks are put in Section 6. The following subsection comprehends the key aspects of explainability.

1.1. Key Aspects of Explainability in Machine Learning Techniques.

1.1.1. Model-Agnostic Approaches: These approaches boost clarity without adjusting the model itself [1][53], moreover, they significantly improve the quality of patient care systems[15]. Various prediction and machine learning techniques, emphasizing supervised, unsupervised, and semi-supervised learning, along with prototyping, evaluation, and instrumentation, are presented in [14].

1.1.2. Post-Hoc Explanations: Techniques such as Shapley values provide insights into feature significance after model training, though they can be affected by data imputation methods [2]. Focusing on post-hoc explanations and their theoretical foundations [60] provides an overview of challenges and possible future directions in this progressive area.

1.1.3. Intrinsic Explanations: Some models, like decision trees, offer built-in interpretability due to their transparent structures [13]. Methods to understand global model structure using optimal local explanations in tree-based models are presented in [71]. The explainable medical recommender system proposed in [34] exploits graphs with overlapping cluster structure to adjust the weight for better accuracy of the recommendations.

Model-agnostic methods and model-agnostic post-hoc explainability algorithms are acquiring considerable attention as they make machine learning models more interpretable [79]. The balance between post-hoc and ante-hoc explanations, as well as between model-specific and model-agnostic techniques, is discussed in [64].

2. SUMMARY OF RELATED WORK

Comprehension of the basic framework of explainable machine learning models, see [57] [62] [69] [91] [99] [100] and [98]. A systematic review of the futuristic explainable machine learning models are presented in [73] with recommendations for future research directions. Survey article [104] characterises explainable machine learning systems as enigmatic, interpretable, and apprehensible. It also introduces completely transparent systems that use automated reasoning for explanations, and no human interaction in the generative process. Methodologies are outlined in [91] and diverse approaches of deep learning models are categorized in [79] based on their scope, methodology used in the algorithm, and explanation level. A categorization of interpretability in machine learning models is provided in [89], suggesting alternative approaches for interpretability standards and complexities and challenges involved especially in the medical field. Machine learning models in recent scientific works are reviewed in [84] prioritizing applications in the natural sciences areas. A comprehensive review of the explainable recommendation systems focussing research timeline and applications is provided in [72], and a two-dimensional taxonomy is also discussed.

Perturbation-based explainable methods are reviewed in [61][67] focusing on the applications in different data types, from images, video, natural language, software code, and reinforcement learning entities. Article [26] reviews object classification, detection, and tracking methods. The study in [41] categorises the literature on explainable machine learning into data explainability, model explainability, and post-hoc explainability. Also discussed evaluation metrics, open-source packages, and datasets. Using topic modeling, co-occurrence, and network analysis, [103] mapped the research space from diverse domains, such as algorithmic accountability, interpretable machine learning, context-awareness, cognitive psychology, and software learnability. In [80] and [82] a literature review and taxonomy of these methods are presented. See, [80] for links to their programming implementations. The level of understanding of explanations is discussed in [86] under simulation and human-subject experiments to identify optimizing parameters. Relevant work from philosophy, cognitive psychology/science, and social psychology are reviewed in [87]

and discussed its significance on explainable artificial intelligence. Study in this direction from a historical perspective is presented in [74], it also proposes significant criteria for human-understandable explainable systems. Similar study in [43] provides an extensive survey of the heterogeneity of methods for explainability leading to individual explanatory frameworks. A model is discussed in [58], highlighting the main concepts and relations for developing and evaluating explainable approaches relevant to the user. It also demonstrates the successful usage of explainable systems in application scenarios. A similar work is presented in [60]. An overview of explainable machine learning methods most suited for tabular and time series data in the healthcare domain is presented in [54], highlighting the research challenges in this field. Clinical validation, consistency assessment, objective and standardized quality evaluation, and quality assessment from a human perspective are highlighted as key features to ensure effective explanations for the beneficiaries. The review in [77] contributes to formalizing explainable machine learning models in healthcare. See, [50][64][65]92] for other similar works in the healthcare domain.

Table 1: Summary of Review Articles

Reference	**Categorization/Focus**
[33][50] [64] [65] [77][92]	*Review on Explainable models in Health Care Domain*
[33][56] [85][90]	*Review on Evaluation Metrics and Methods*
[58][89] [104]	*Categorization of Interpretability- Opaque, Interpretable, Comprehensible, and Truly Explainable Systems*
[41][42]	*Categorization of Interpretability - data explainability, model explainability, and post-hoc explainability*
[43][103]	*Categorization of Research Domain- algorithmic accountability, interpretable machine learning, context-awareness, cognitive psychology, and software learnability*
[79]	*Scope, methodology used in the algorithm, and explanation level - Based on DNN Model*
[90]	*Characterization of the behaviour of nonlinear explainable models - Spectral Relevance Analysis*
[84]	*Natural sciences domain*
[87]	*Philosophy, Cognitive psychology/Science, and Social Psychology*
[61][67]	*Perturbation-based explainable methods*
[54]	*Tabular and time series data*
[72]	*Research timeline and Applications - review of the explainable recommendation research*
[86]	*The level of understanding of explanations - Optimizing parameters using simulation and human-subject experiments*
[57][59] [60] [69] [74][80] [82] [91] [98] [99] [100]	*Basic Concepts, Challenges, and Research Directions*
[26]	*object classification, detection, and tracking methods*

The reason behind the necessity of explainability influences the relative importance of the different aspects of explainability, concrete recommendations to choose between different classes of explainable systems such as model-based, attribution-based, example-based explanations. Survey articles [33][56][85][90] focus on metrics and methods in this field. Challenges and research

directions of different phases in explainable machine learning systems such as design, development, and deployment are discussed in [59]. Highlighting limitations and future research directions [37] discussed several key areas where reliable explanations are crucial. Focus of review articles in this direction are summarised in Table 1.

3. EXPLAINABLE MODELS IN THE HEALTH CARE DOMAIN.

The complexity of machine learning models often makes their decision-making processes opaque, raising concerns about their reliability and trustworthiness, particularly in critical healthcare applications, where decisions can have profound consequences. By visualizing and explaining the process that leads to a particular decision in machine learning models, healthcare professionals can gain better awareness of the factors contributing to a diagnosis or treatment recommendation. This leads to more informed and effective care, as well as valuable insights for developing novel and fair therapeutic approaches. [31][32] [34][66]. A survey of explainable techniques used in the healthcare sector and medical imaging applications is provided in [33], highlighting the algorithms used to increase interpretability in medical imaging and text analysis. The requirements and operational challenges are presented in [52], and a case study is supported through experimental validation. A method with high accuracy for automated labelling is described in [51], here, the desired level of accuracy is specified by a quantitative threshold on user choice. An overview of case studies in the medical field with open-box architecture is presented in [50], focusing on model enhancements, evaluation metrics, and medical open datasets. It also proposes critical ideas focusing on human-machine collaboration for better explainable solutions. In [54] a human-centered quality assessment is used as a key feature to ensure effective explanations for the end users. It provides an overview of methods that are most suitable for tabular and time series data in the healthcare domain.

An explainable strategy using LIME, for leukemia classification is proposed in [53], and reliability with different transfer learning methods, including ResNet101V2, VGG19, and InceptionResNetV2 are compared. Explainable AI solutions in health care are introduced in [65] using multi-modal and multi-centre data fusion, [64] discussed the provision of local explanations. A formalization of explainable AI provided in [77] discusses the clarity of quantitative evaluation metrics and the types of explanations focused on the healthcare domain. A review on [92] summarises the leading psychological theories of explanation. Concentrating in the domain of histopathology [95] presents essential definitions to distinguish between "explainability" and "causability," as well as a use case of deep learning interpretation and human explanation. Blockchain technology is used in the proposed metaverse environment in [41], which provides more secure healthcare facilities as well as explainability and interpretability. See, Table 2 for the summary of related works.

Table 2: Articles in Healthcare Domain –Summary.

Reference	Method	Focus
[16] [17]	*Deep Learning (DL) in natural language processing (NLP)*	*Comparison of explainable and interpretable models, Datasets and Metrics*
[18]	*Machine Learning WorkFlow*	*dimensionality reduction, feature importance, attention mechanisms, knowledge distillation, and surrogate representations*
[20] [19] [33] [52] [24] [25]	*Review on Deep Learning Methods*	*predictions in stroke, heart attack, and cancer detection*
[21] [22] [53]	*model-agnostic approaches*	*Disease Diagnostics, Predictive Analytics, and Personalized Treatment Recommendations.*
[23]	*Rule Based Approaches*	*Ethical Concerns*
[31] [34] [32] [66]	*Therapeutic Approaches*	
[50] [54] [77]	*Human Centered Quality Assessment,*	*model enhancements, evaluation metrics, and medical open datasets*
[51]	*Methods for Automated Labeling*	*Visual Prediction*
[65]	*Deep Learning*	*multi-modal and multi-centre data fusion*
[64]	*Local Explanations*	*post-hoc and ante-hoc explanations, model-specific and model-agnostic techniques*
[92]	*Review*	*psychological theories of explanation*
[95]	*Causability in Deep Learning Models*	*Human Explanation in Histopathology*
[41]	*GradCAM and LIME approaches, Blockchain Technology*	*building block technologies of the metaverse in healthcare*

4. EXPLAINABILITY IN MEDICAL IMAGING APPLICATIONS.

Machine-controlled image classification is crucial in the healthcare domain, where modality-dependent features are essential for model interpretation. Since counterfactual explanation methods result in a high degree of interpretability, complex decision models are significantly influenced by them. Though many existing vision-language models effectively describe image content, they fail to incorporate discriminative image attributes, which are essential, especially in visual predictions [51]. The article [32] explains machine learning in medical imaging, de-emphasising prominence, and provides a classification into Case-based, textual, and auxiliary explanations. See [33] for a summary of explainable AI techniques in medical imaging and text analysis applications. It also provides guidelines to develop better interpretations of deep learning models. Focusing on healthcare and network security, article [37] outlines the challenges of deep learning models in providing trustworthy diagnostic evidence and explanations. Research work in [52] concentrates on an ensemble classification and segmentation architecture for computerized tomography images. In [31] the quality of AI-driven image labelling is compared to that of human radiologists. Modality-specific feature localization explanation model is proposed in [44], which encodes clinical images. A survey in [67] focuses on perturbation-based methods to explore Deep Neural Network models. It compares the applications of perturbation-based methods to different data types, including images, video, natural language, software code, and reinforcement learning entities. An interpretable model for image classification known as Deep Taylor decomposition using generic multilayer neural networks, proposed in [107], explains the decision process based on the contributions of its input elements. Model-agnostic explanation based on the significance of a group of segments in the decision process is proposed in [83] for visual counterfactual explanations. An explainable model for image labelling proposed in [109] is based on reinforcement learning. Observations in [11] signify the need to distinguish causal connections from mere correlations. It introduces the Explainable and Causal Feature Analysis (ECFA) method to enhance model interpretability and reliability in medical diagnostics. [63] presents explainable machine learning approaches for breast cancer diagnosis.

A general solution for the explainability of kernel-based classifiers is proposed in [110] by visualizing the single-pixel contributions as heat maps, analyzing the validity of the decision as well as the areas of potential significance by a human expert. Related works are summarized in Table 3.

Table 3: Research Summary on Medical Imaging Applications

Reference	*XAI model*	*Image*	*Description*
[3]	*CNN*	*Chest radiographs, Synthetic medical image.*	*SEE-GAAN- explainability framework for CNNs*
[4]	*Deep Learning*	*MRI images*	*Combining deep learning, visual attribution algorithms, and natural language explanations.*
[5]	*Deep Learning*	*Brain tumor MRI and COVID-19 chest X-ray datasets*	*Combines qualitative and quantitative assessments*
[55]	*Heat map Visualization*	*COVID-19 chest X-ray*	*Gradient-weighted Class Activation Mapping (Grad-CAM) algorithm for visualization, classification models,- VGG16, VGG19, Xception, InceptionV3, Densenet201, NASNetMobile, Resnet50, and MobileNet,*
[7]	*Deep Learning*		*Survey on interpretability, visualization and significance*
[6] [8] [9] [10] [12] [96] [110][52]	*Visual Explainable AI techniques*	*X-ray, CT Scan*	*User perspective explanation*
[10][a]	*DCNN*	*CT images*	*Gradient-weighted Class Activation Mapping (Grad-CAM), with Deep Convolutional Neural Networks*
[11]	*Causal Feature Analysis (ECFA)*	*Medical Image*	*distinguish causal connections from mere correlations*
[31]	*Experimental Analysis*	*X-ray images*	*Quality Comparison of AI-driven and Expert explanation*
[107]	*DNN*	*MNIST and ILSVRC data sets*	*Deep Taylor decomposition using generic multilayer neural networks*
[32]	*Medical imaging not relying on saliency,*		*Classification - Case-based, textual, and auxiliary explanations.*

[36]	*Deep Learning*	*MRI*	*marker-controlled watershed transformation algorithm.*

5. EVALUATION METRICS.

The efficiency of a prediction model can be highlighted using metrics such as accuracy and sensitivity, especially when it concerns applications involving large and complex datasets, as it is crucial in the selection of a prediction model for a particular application. However, the intelligibility of the prediction models is significant, especially when we focus on systems that interact with laypeople. Metrics to evaluate the interpretability of predictions require considerable research attention in this context, as it remains an unexplored area of research. The following are some of the works in this direction.

Techniques based on psychometric evaluation are presented in [27], which also discusses the strengths and weaknesses in the user perspective analysis of satisfaction level, mental model, and trust of an AI system. Quantitative evaluation methods are reviewed in [28], with a focus on compactness and correctness. [37] presents a survey on visualization techniques with improved interpretability, particularly in medical and cybersecurity applications. It also discusses approaches to enhance the clarity and significance of explainable machine learning models, like activation vectors and correlation. Interpretation ability is quantified in [35] by applying techniques used in Bioimaging, focusing on Latent space interpretation and attribution maps. The applicability of explainable models is discussed in [29], concentrating on the features necessary for a specific scenario. An open-source Python toolkit for various evaluation metrics is presented in [45] with supporting tutorials. Guidelines for criteria that are necessarily optimized by explainable models, especially in the healthcare sector, and evaluation metrics are proposed in [46]. Contextualized criteria for assessing explainable models are proposed in [47] based on their relative significance in the typical scenario.

A framework based on cognitive engineering is proposed in [48]. It also discusses methods and metrics for evaluating human workload and trust in explainable systems. More objective metrics are discussed in [50] [54][56] [77] [79] [97] [106] and a unified framework is provided in [105]. Categorization of interpretable machine learning design goals and evaluation methods is presented in [70] and a survey on evaluation metrics is provided in [76]. A similar survey is provided in [88] focusing on its societal impact and in [85] with design guidelines and evaluation methods. A spectral relevance analysis is proposed in [90] to characterize the behaviour of nonlinear explainable models, and in [78] the System Causability Scale is proposed as a quality metric. Summary of related work is presented in Table 4.

Table 4: Evaluation Metrics used

Reference	**Metric**	**Focus**
[27]	Psychometric Evaluation	Strengths and weaknesses in the user perspective analysis
[28]	Quantitative evaluation	Compactness and Correctness
[37]	Visualization Techniques	Approaches to enhance the clarity and significance.
[45]	Quantitative evaluation	Python toolkit
[50] [be] [56] [77] [79] [97] [106] [54]	Quantitative evaluation	*Objective Metrics*
[105]	Quantitative evaluation	*Unified Frameworks*
[46]	*Evaluation metric for health care domain*	*Criteria to optimize evaluation metric*
[70] [76] [88][85]	*Evaluation metric*	*Survey on Evaluation Metric*
[47]	Contextualized criteria for evaluation	Relative significance in specific scenario
[48]	metrics for evaluating human workload and trust	human informational considerations
[78]	*Qualitative Metric*	System Causability Scale

6. CONCLUSION.

The efficiency and acceptability of explainable AI models are assessed by their ability to provide specific, understandable explanations of their decisions. The evaluation methods of XAI can be broadly categorized into human-centered and computer-centered. In the first category, the explanations generated are evaluated by domain experts, and their feedback is collected and analyzed. It is expensive since it requires the service of domain experts, such as clinicians, to evaluate the explanation performance. Whereas in computer-centered explainable models, algorithms are used to assess the quality of the explanation. As human trust is a crucial factor in determining the quality of such systems, especially in the healthcare domain, most researchers concentrated on bridging the trust between humans and AI using more transparent and understandable AI systems. Insights on the decision process enhance the understandability of the reason behind each prediction, which is significant in clinical applications. Another challenge in this domain is biases in medical images. This issue is addressed in [8] by investigating saliency methods, it try to uncover how biases can affect model performance and interpretation. Most of the work points out the significance of human-centered evaluation, promoting user choices, greater awareness, and design free of constraints to appreciate the trust and usability of AI technologies in healthcare [11]. Another challenge is to make the systems suitable for diverse requirements in the healthcare domain in a user-friendly way.

Most of the works emphasize the need for multi-modal explainability to enhance the understanding of the decision-making processes [11][9] and the need for visual explanation methods to ensure trust and effective interpretation, see, [4] [6] [8] [9] [10] [12] [96] [110] [52] [37] [83] [31] [32] [66]. Research in multi-modal embeddings focuses on enhancing trust and understandability of explanations for AI decisions. Another challenge is the limited awareness among participants concerning the value and practical aspects of explainable machine learning models. This gap must be addressed through educational initiatives [10]. Another research focus is the development of unified regulatory frameworks and policies to address the specific legal and ethical issues raised by the explainable models in healthcare [9]. In conclusion, understandable explanations, especially those using visualization techniques, significantly enhance trust in explainable systems, particularly in medical imaging. Since they provide insights into the reason behind decision-making, they encourage clinicians to rely more confidently on AI-powered medical decision-support systems.

REFERENCES

1. Vinayak, Pillai. (2024). 3. Enhancing the transparency of data and ML models using explainable AI (XAI). World Journal of Advanced Engineering Technology and Sciences, doi: 10.30574/wjaets.2024.13.1.04282. Huang, G., Li,

2. Thanh-Truc, Vo., Thu, Nguyen., Hugo, Lewi, Hammer., Michael, Riegler., Pål, Halvorsen. (2024). 4. Explainability of Machine Learning Models under Missing Data. doi: 10.48550/arxiv.2407.00411

3. Kyle, Hasenstab., Lewis, D., Hahn., Nicholas, S.Y., Chao., Albert, Hsiao. (2024). 1. Simulating clinical features on chest radiographs for medical image exploration and CNN explainability using a style-based generative adversarial autoencoder. Dental science reports, doi: 10.1038/s41598-024-75886-0

4. M., Vinoth., V., Jayapradha., K., Anitha., Gowrisankar, Kalakoti., Ezhil, E., Nithila. (2024). 2. Explainable AI for Transparent MRI Segmentation: Deep Learning and Visual Attribution in Clinical Decision Support. International journal of computational and experimental science and engineering, doi: 10.22399/ijcesen.479

5. Yusuf, Brima., Marcellin, Atemkeng. (2024). 3. Saliency-driven explainable deep learning in medical imaging: bridging visual explainability and statistical quantitative analysis. Biodata Mining, doi: 10.1186/s13040-024-00370-4

6. Izegbua, E., Ihongbe., Shereen, Fouad., Taha, F., Mahmoud., Arvind, Rajasekaran., B.S., Bhatia. (2024). 4. Evaluating Explainable Artificial Intelligence (XAI) techniques in chest radiology imaging through a human-centered Lens. PLOS ONE, doi: 10.1371/journal.pone.0308758

7. Deepshikha Bhati, Fnu Neha, Md. Amiruzzaman, A Survey on Explainable Artificial Intelligence (XAI) Techniques for Visualizing Deep Learning Models in Medical Imaging, J. Imaging 2024, 10(10), 239; https://doi.org/10.3390/jimaging10100239

8. Salamata Konate, Interpretability of machine learning models for medical image analysis, 2024, 10.5204/thesis.eprints.248812

9. Marey, A., Arjmand, P., Alerab, A.D.S. et al. Explainability, transparency and black box challenges of AI in radiology: impact on patient care in cardiovascular radiology. Egypt J Radiol Nucl Med 55, 183 (2024). https://doi.org/10.1186/s43055-024-01356-2.

10. Teuku Rizky Noviandy, Aga Maulana, Teuku Zulfikar, Asep Rusyana, Seyi Samson Enitan, Rinaldi Idroes,Explainable Artificial Intelligence in Medical Imaging: A Case Study on Enhancing Lung Cancer Detection through CT Images, Indonesian Journal of Case Reports, Vol. 2 No. 1 (2024). https://doi.org/10.60084/ijcr.v2i1.150

11. Daniel Flores-Araiza, Armando Villegas-Jimenez, Francisco Lopez-Tiro, Miguel González-Mendoza, Rosa-Maria Rodriguez-Guéant, Jacques Hubert, Gilberto Ochoa-Ruiz, Christian Dau, On the Link Between Model Performance and Causal Scoring of Medical Image Explanations,IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS),2024, DOI: 10.1109/CBMS61543.2024.00009

12. Alec, Parise., Brian, Mac, Namee. Explainable Interactive Machine Learning Using Prototypical Part Networks for Medical Image Analysis. Frontiers in artificial intelligence and applications,(2024). 10. , doi: 10.3233/faia240214

13. Sebastian, Ordyniak., Giacomo, Paesani., Mateusz, Rychlicki., Stefan, Szeider. (2024). 5. Explaining Decisions in ML Models: a Parameterized Complexity Analysis. arXiv.org, doi: 10.48550/arxiv.2407.15780

14. Daniel, Ståhl. (2024). 1. New horizons in prediction modelling using machine learning in older people's healthcare research. Age and Ageing, doi: 10.1093/ageing/afae201

15. Pokkuluri, Kiran, Sree. "Machine Learning for Quality in Health Care: A Comprehensive Review". (2023). 4. doi: 10.26717/bjstr.2023.51.008138.

16. Guangming, Huang., Yingya, Li., Shoaib, Jameel., Yunfei, Long., Giorgos, Papanastasiou. (2024). 1. From Explainable to Interpretable Deep Learning for Natural Language Processing in Healthcare: How Far from Reality?. Computational and structural biotechnology journal, doi: 10.1016/j.csbj.2024.05.004

17. Adla, Padma., Vasavi, Chithanuru., Posham, Uppamma., R., Vishnukumar. (2024). 2. Exploring Explainable AI in Healthcare. Advances in healthcare information systems and administration book series, doi: 10.4018/979-8-3693-5468-1.ch011

18. Shantha, Visalakshi, Upendran. (2024). 3. Explainable AI in Healthcare. Advances in healthcare information systems and administration book series, doi: 10.4018/979-8-3693-5468-1.ch004

19. Guangming, Huang., Yunfei, Long., Yingya, Li., Giorgos, Papanastasiou. (2024). 4. From explainable to interpretable deep learning for natural language processing in healthcare: how far from reality?. arXiv.org, doi: 10.48550/arxiv.2403.11894

20. S, Ishwarya., Anitha, S., Pillai. (2024). 5. Explainable Artificial Intelligence in Healthcare -A Review. doi: 10.1109/icict60155.2024.10544745

21. Adewale, Abayomi, Adeniran., Amaka, Peace, Onebunne., Paul, William. (2024). 6. Explainable AI (XAI) in healthcare: Enhancing trust and transparency in critical decision-making. World Journal Of Advanced Research and Reviews, doi: 10.30574/wjarr.2024.23.3.2936

22. R., S., Thakur. (2024). 7. Explainable AI: Developing Interpretable Deep Learning Models for Medical Diagnosis. International Journal For Multidisciplinary Research, doi: 10.36948/ijfmr.2024.v06i04.25281

23. Amit, Kumar., Eshan, Jaiswal, Kanishk, Gupta, Kartik, Chaudhary., Pratyush, Rai., Er., Radha, -. (2024). 8. Explainable Artificial Intelligence in Healthcare. International Journal For Multidisciplinary Research, doi: 10.36948/ijfmr.2024.v06i02.18735

24. Veena, Grover., Mahima, Dogra. (2024). 9. Challenges and Limitations of Explainable AI in Healthcare. Advances in healthcare information systems and administration book series, doi: 10.4018/979-8-3693-5468-1.ch005

25. Siva, Raja, Sindiramutty., Wee, Jing, Tee., Sumathi, Balakrishnan., Simran, Kaur., Rajan, Thangaveloo., Husin, Jazri., Navid, Ali, Khan., Abdalla, Hassan, Gharib., Amaranadha, Reddy, Manchuri. (2024). 10. Explainable AI in Healthcare Application. Advances in computational intelligence and robotics book series, doi: 10.4018/978-1-6684-6361-1.ch005

26. Jian-Xun Mi,Xiaoyi Jiang,Lin Luo,Yun Gao, Toward explainable artificial intelligence: A survey and overview on their intrinsic properties, Neurocomputing 563(2):126919 DOI:10.1016/j.neucom.2023.126919

27. Robert R. Hoffman,Shane T. Mueller,Gary Klein,Jordan A. Litman Measures for explainable AI: Explanation goodness, user satisfaction, mental models, curiosity, trust, and human-AI performance, Frontiers of Computer Science, 2023, https://doi.org/10.3389/fcomp.2023.1096257 ·

28. Meike Nauta,Jan Trienes et.al. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI ACM Computing Surveys 55(13s),DOI:10.1145/3583558

29. Gesina Schwalbe,Bettina Finzel, A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts Data Mining and Knowledge Discovery(2021), 1-59,https://api.semanticscholar.org/CorpusID:245124075

30. Mohammad Nagahisarchoghaei, et.al., An Empirical Survey on Explainable AI Technologies: Recent Trends, Use-Cases, and Categories from Technical and Application Perspectives Electronics, 2023, 12(5), 1092;https://doi.org/10.3390/electronics12051092

31. Susanne Gaube,et.al., Non-task expert physicians benefit from correct explainable AI advice when reviewing X-raysScientific Reports,Sci Rep. 2023 Jan 25;13(1):1383. doi: 10.1038/s41598-023-28633-w. PMID: 36697450; PMCID: PMC9876883.

32. Katarzyna Borys,et.al. Explainable AI in medical imaging: An overview for clinical practitioners â€“ Beyond saliency-based XAI approaches European Journal of Radiology, (V) 162110786 May 2023

33. Ahmad Chaddad,Jihao Peng,Jian Xu,Ahmed Bouridane, Survey of Explainable AI Techniques in Healthcare Sensors (Basel), 2023 Jan 5;23(2):634. doi: 10.3390/s23020634.

34. Arun Kumar Sangaiah,Samira Rezaei,Amir Javadpour,Weizhe Zhang, Explainable AI in big data intelligence of community detection for digitalization e-healthcare services, Applied Soft Computing, Volume 136, March 2023, 110119

35. Lisa Anita De Santi,Filippo Bargagna,Maria Filomena Santarelli,Vincenzo Positano Evaluating Explanations of an Alzheimer's Disease 18F-FDG Brain PET Black-Box Classifier Explainable Artificial Intelligence,2023

36. Tahamina Yesmin,Pinaki Pratim Acharjya, Identification and Segmentation of Medical Images by Using Marker-Controlled Watershed Transformation Algorithm, XAI, and ML IRMA international, 2023, DOI:10.4018/978-1-6684-7524-9.ch003

37. Wenli Yang, et.al. Survey on Explainable AI: From Approaches, Limitations and Applications Aspects Human-Centric Intelligent Systems, 2023, V. 3, pages 161–188, (2023)

38. Tim Miller, Explainable AI is Dead, Long Live Explainable AI!, "Conference on Fairness, Accountability and Transparency, Pages 333-342 https://doi.org/10.1145/3593013.3594001"

39. Geemi P. Wellawatte,Heta A. Gandhi,Aditi Seshadri,Andrew D. White A Perspective on Explanations of Molecular Prediction Models Journal of Chemical Theory and Computation, v.19(8), March 2023

40. Ričards Marcinkevičs, Julia E. Vogt, Interpretable and explainable machine learning: A methods-centric overview with concrete examples, WIERS Data mining and Knowledge Discovery., 2023 https://doi.org/10.1002/widm.1493

41. Sikandar Ali,et.al. Metaverse in Healthcare Integrated with Explainable AI and Blockchain: Enabling Immersiveness, Ensuring Trust, and Providing Patient Data Security Italian National Conference on Sensors, 2023, 23(2), 565; https://doi.org/ 10.3390/ s23020565

42. Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, Francisco Herrera, Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence, Information Fusion, V. 99, 2023, 101805, ISSN 1566-2535, https://doi.org/10.1016/j.inffus.2023.101805.

43. Simon D Duque Anton, et.al., On Explainability in AI-Solutions: A Cross-Domain Survey, Lecture Notes in Computer Science, SAFECOMP 2022 Workshops (2022) 235-246

44. Weina Jin,Xiaoxiao Li,Ghassan Hamarneh Evaluating Explainable AI on a Multi-Modal Medical Imaging Task: Can Existing Algorithms Fulfill Clinical Requirements? arXiv:2210.05173 2022

45. Anna HedstrÃm,Leander Weber,et.al. Quantus: An Explainable AI Toolkit for Responsible Evaluation of Neural Network Explanations Journal of Machine Learning Research 24 (2023) 1-11

46. Weina Jin,et.al, Guidelines and evaluation of clinical explainable AI in medical image analysis "Medical Image Analysis, 84, ISSN: 1361-8415, pages 102684 2023"

47. Yunfeng Zhang, et.al. Connecting Algorithmic Research and Usage Contexts: A Perspective of Contextualized Evaluation for Explainable AI arXiv:2206.10847, 2022

48. Lindsay Sanneman,Julie A. Shah, The Situation Awareness Framework for Explainable AI (SAFE-AI) and Human Factors Considerations for XAI Systems International Journal of Human-computer Interaction, Volume 38, 2022 - Issue 18-20: 2022

49. Rudresh Dwivedi, et.al., Explainable AI (XAI): Core Ideas, Techniques and Solutions ACM Computing Surveys, Vol. 55, No. 9 ,2023

50. Ruey-Kai SheuMayuresh Sunil Pardeshi,A Survey on Medical Explainable AI (XAI): Recent Progress, Explainability Approach, Human Interaction and Scoring System Sensors 2022, 22(20), 8068; https://doi.org/10.3390/s22208068

51. Doyun Kim, et.al., Accurate auto-labeling of chest X-ray images based on quantitative similarity to an explainable AI model. Nature Communications volume 13, Article number: 1867 (2022)

52. Deepti Saraswat, et.al. Explainable AI for Healthcare 5.0: Opportunities and Challenges IEEE Access, Digital Object Identifier 10.1109/ACCESS.2022.3197671

53. Wahidul Hasan Abir, et.al., Explainable AI in Diagnosing and Anticipating Leukemia Using Transfer Learning Method Computational Intelligence and Neuroscience, arXiv:2312.00487 2023

54. Flavio Di Martino, et.al., Explainable AI for clinical and remote health applications: a survey on tabular and time series data "Artificial Intelligence Review, v. 56, pages 5261–5315, (2023)"

55. Nillmani; Sharma, N.; Saba, L.; Khanna, N.N.; Kalra, M.K.; Fouda, M.M.; Suri, J.S. Segmentation-Based Classification Deep Learning Model Embedded with Explainable AI for COVID-19 Detection in Chest X-ray Scans. Diagnostics 2022, 12, 2132. https://doi.org/10.3390/diagnostics12092132

56. Jianlong Zhou,et.al., Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics Electronics 2021, 10(5), 593; https://doi.org/10.3390/electronics10050593

57. Gesina Schwalbe,Bettina Finzel., XAI Method Properties: A (Meta-)study. Corpus ID: 234742123, arXiv.org 2021

58. Markus Langer,et.al. What do we want from Explainable Artificial Intelligence (XAI)? â€“ A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI researchArtificial Intelligence arXiv:2102.07817 [cs.AI] 2021

59. Waddah Saeed,Christian W. Omlin, Explainable AI (XAI): A Systematic Meta-Survey of Current Challenges and Future Opportunities Elsevier Knowledge-Based Systems,V. 263, 5 March 2023, 110273

60. Wojciech Samek,et.al. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications Proceedings of the IEEE , V: 109 (3) 2021

60. Wojciech Samek,et.al. Explaining Deep Neural Networks and Beyond: A Review of Methods and Applications Proceedings of the IEEE V109 : 3, pages 247 - 278, 2021

61. Gargi Joshi,Gargi Joshi,Rahee Walambe,Rahee Walambe,Ketan Kotecha,Ketan Kotecha A Review on Explainability in Multimodal Deep Neural Nets IEEE Access, v.9, pages. 59800 - 59821, 2021

62. Giulia Vilone, Luca Longo, Notions of explainability and evaluation approaches for explainable artificial intelligence Information Fusion,v. 76, December 2021, Pages 89-106

63. Pascal Bourdon,Olfa Ben Ahmed,Thierry Urruty,Khalifa Djemal,Christine Fernandez-Maloigne Explainable AI for Medical Imaging: Knowledge Matters hal-03612280 , version 1 (17-03-2022) 202`

64. Anna Markella Antoniadi,et.al. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review Applied Sciences, 2021, 11(11), 5088; https://doi.org/10.3390/app11115088

65. Guang Yang,Qinghao Ye,Jun Xia Unbox the Black-box for the Medical Explainable AI via Multi-modal and Multi-centre Data Fusion: A Mini-Review, Two Showcases and BeyondInformation Fusion, v. 77, January 2022, Pages 29-52

66. Andreas Holzinger, et.al., Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI Information Fusion, V. 71, July 2021, Pages 28-37

67. Maksims Ivanovs,Roberts Kadikis,Kaspars Ozols., Perturbation-based methods for explaining deep neural networks: A survey Pattern Recognition Letters, V. 150, October 2021, Pages 228-234

68. Yu-Liang Chou,Catarina Moreira,Peter Bruza,Chun Ouyang,Joaquim Jorge,. Counterfactuals and Causability in Explainable Artificial Intelligence: Theory, Algorithms, and Applications. Information Fusion, V. 81, May 2022, Pages 59-83

69. Ilia Stepin, et.al., A Survey of Contrastive and Counterfactual Explanation Generation Methods for Explainable Artificial Intelligence IEEE Access, v.9, pages. 59800 - 59821, 2021

70. MohseniSina,ZareiNiloofar,Eric D. Ragan., A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems arXiv:1811.11839 [cs.HC], 2020

71. Scott Lundberg,et.al., From Local Explanations to Global Understanding with Explainable AI for Trees. Nature Machine Intelligence v. 2, pages56–67 (2020)

72. Yongfeng Zhang,Xu Chen., Explainable Recommendation: A Survey and New Perspectives Foundations and Trends in Information Retrieval, v. 14: 1, Pages 1 - 101, 2020

73. Giulia Vilone,Luca Longo., Explainable Artificial Intelligence: a Systematic Review arXiv:2006.00093 [cs.AI], 2020

74. Roberto Confalonieri,et.al., A historical perspective of explainable Artificial Intelligence Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery, https://doi.org/10.1002/widm.1391, 2020

75. Alexander Heimerl,et.al., Unraveling ML Models of Emotion with NOVA: Multi-Level Explainable AI for Non-Experts IEEE Transactions on Affective Computing, V.. 13, NO. 3, JULY-SEPTEMBER 2022

76. Xiao-Hui Li,et.al,. A Survey of Data-driven and Knowledge-aware eXplainable AI IEEE Transactions on Knowledge and Data Engineering, Volume 34, Issue 1, Pages 29 - 49, 2022

77. Aniek F. Markus,Jan A. Kors,Peter R. Rijnbeek The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. Journal of Biomedical Informatics, Volume 113, January 2021, 103655

78. Andreas Holzinger, et.al., Measuring the Quality of Explanations: The System Causability Scale (SCS): Comparing Human and Machine Explanations. Springer Nature Link, Volume 34, pages 193–198, (2020)

79. Arun Das,Paul Rad., Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey Semantic Scholar, Corpus ID: 219965893, June2020.

80. Pantelis Linardatos,Vasilis Papastefanopoulos,Sotiris Kotsiantis., Explainable AI: A Review of Machine Learning Interpretability Methods Entropy 2021, 23(1), 18; https://doi.org/10.3390/e23010018

81. Donghee Shin,Dong-Hee Shin, The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI, Int. J. Hum. Comput. Stud. 1 February 2021, DOI:10.1016/j.ijhcs.2020.102551Corpus ID: 225106328

82. Alejandro Barredo Arrieta,et.al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI Information Fusion, Volume 58, June 2020, Pages 82-115

83. Tom Vermeire,et.al Explainable Image Classification with Evidence Counterfactual Springer Nature Link, Volume 25, pages 315–335, (2022)

84. Ribana Roscher,Bastian Bohn,Marco F. Duarte,Jochen Garcke Explainable Machine Learning for Scientific Insights and Discoveries, IEEE Access, vol. 8, pp. 42200-42216, 2020, doi: 10.1109/ACCESS.2020.2976199

85. Sina Mohseni,Niloofar Zarei,Eric D. Ragan A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems CM Transactions on Interactive Intelligent Systems (TiiS), Volume 11, Issue 3-4, Article No.: 24, Pages 1 - 45

86. Isaac Lage,Emily Chen,Jeffrey He,Menaka Narayanan,Been Kim,Sam Gershman,Finale Doshi-Velez An Evaluation of the Human-Interpretability of Explanation, arXiv:1902.00006 [cs.LG], 2029

87. Tim Miller, Explanation in artificial intelligence: Insights from the social sciences Artificial Intelligence, V. 267, February 2019, Pages 1-38

88. Diogo Vieira Carvalho,Eduardo M. Pereira,Jaime S. Cardoso.,Machine Learning Interpretability: A Survey on Methods and Metrics Electronics 2019, 8(8), 832; https://doi.org/10.3390/electronics8080832

89. Erico Tjoa,Cuntai Guan, A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. IEEE Trans Neural Netw Learn Syst. 2021 Nov;32(11):4793-4813, doi: 10.1109/TNNLS.2020.3027314. Epub 2021 Oct 27

90. Sebastian Lapuschkin,et.al., Unmasking Clever Hans predictors and assessing what machines really learn. Nature Communications arXiv:1902.10178 [cs.AI] 2019

91. Alejandro Barredo Arrieta, et.al., Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. Information Fusion, Volume 58, June 2020, Pages 82-115

92. Shane T. Mueller,Robert R. Hoffman,William J. Clancey,Abagail K Emery,Gary Klein., Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications, and Bibliography for Explainable AI. Semantic Scholar, Corpus ID: 59606335, 2019

93. Ruth M. J. Byrne, Counterfactuals in Explainable Artificial Intelligence (XAI): Evidence from Human Reasoning. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, Survey track. Pages 6276-6282. https://doi.org/10.24963/ijcai.2019/876

94. Samek, W., Müller, KR. (2019). Towards Explainable Artificial Intelligence. In: Samek, W., Montavon, G., Vedaldi, A., Hansen, L., Müller, KR. (eds) Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science(), vol 11700. Springer, Cham. https://doi.org/10.1007/978-3-030-28954-6_1

95. Andreas Holzinger,Georg Langs,Helmut Denk,Kurt Zatloukal,Heimo MÃller, Causability and explainability of artificial intelligence in medicine., Wiley Interdisciplinary Reviews-Data Mining and Knowledge Discovery, Volume9, Issue4, July/August 2019

96. U. Bhatt,A. Xiang,S. Sharma,A. Weller,A. Taly,Y. Jia,J. Ghosh,R. Puri,J.M.F. Moura,P. Eckersley, Explainable machine learning in deployment, Semantic Scholar, DOI:10.1145/3351095.3375624Corpus ID: 202572724

97. Robert R. Hoffman,Shane T. Mueller,Gary Klein,Jordan A. Litman. Metrics for Explainable AI: Challenges and Prospects. arXiv: Artificial Intelligence

98. K.D. Filip ,B.Mario ,H.Nikica., Explainable artificial intelligence: A survey 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, DOI: 10.23919/MIPRO.2018.8400040

99. Leilani H. Gilpin,et.al. Explaining Explanations: An Overview of Interpretability of Machine Learning IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), 2018, DOI: 10.1109/DSAA.2018.00018

100. Amina Adadi,Mohammed Berrada, Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI) IEEE Access 17 September 2018, DOI:10.1109/ACCESS.2018.2870052Corpus ID: 52965836

101. Grégoire Montavon, Wojciech Samek,Klaus-Robert Müller Methods for interpreting and understanding deep neural networks, Digital Signal Processing, Volume 73, February 2018, Pages 1-15

102. Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, Anchors: High-Precision Model-Agnostic Explanations. Thirty-Second AAAI Conference on Artificial Intelligence , Vol. 32 No. 1 (2018), DOI: https://doi.org/10.1609/aaai.v32i1.11491

103. Ashraf Abdul,Jo Vermeulen,Danding Wang,Brian Y. Lim,Mohan Kankanhalli, Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda, "Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Paper No.: 582, Pages 1 – 18, https://doi.org/10.1145/3173574.3174156"

104. Derek Doran, Sarah Schulz,Tarek R. Besold, What Does Explainable AI Really Mean? A New Conceptualization of Perspectives, arXiv: Artificial Intelligence, :1710.00794 [cs.AI], 2017

105. Scott Lundberg,Su-In Lee, A unified approach to interpreting model predictions, Proceedings of the 31st International Conference on Neural Information Processing Systems, Pages 4768 - 477, 2017

106. F. Doshi-Velez, Been Kim, Towards A Rigorous Science of Interpretable Machine Learning Semminar Scholar: Corpus ID: 11319376, 2017

107. "Grégoire Montavon, Sebastian Lapuschkin, Alexander Binder, Wojciech Samek, Klaus-Robert Müller, "Explaining nonlinear classification decisions with deep Taylor decomposition, Pattern Recognition, Volume 65, May 2017, Pages 211-222

108. Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin, Authors Info & Claims, "Why Should I Trust You?": Explaining the Predictions of Any Classifier, KDD '16: Pro. of the 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, Pages 1135 - 1144, https://doi.org/10.1145/2939672.293977

109. Lisa Anne Hendricks, Zeynep Akata,Marcus Rohrbach,Jeff Donahue, Bernt Schiele, Trevor Darrell Generating Visual Explanations, Computer Vision – ECCV 2016, pages 3-19, 2016

110. Sebastian Bach, Alexander Binder,Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller,Wojciech Samek, On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLoS One. 2015 Jul 10;10(7):e0130140. doi: 10.1371/journal.pone.0130140

111. Todd Kulesza, Simone Stumpf,Margaret Burnett,Sherry Yang,Irwin Kwan,Weng-Keen Wong, Too much, too little, or just right? Ways explanations impact end users' mental models 2013 IEEE Symposium on Visual Languages and Human Centric Computing, DOI:10.1109/VLHCC.2013.6645235Corpus ID: 6960803

[A] El-Sayed A. El-Dahshan, Mahmoud M. Bassiouni, Smith K. Khare, Ru-San Tan, U. Rajendra Acharya, ExHyptNet: An explainable diagnosis of hypertension using EfficientNet with PPG signals, Expert Systems with Applications,V. 239, 2024,122388, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.122388.