User Experience and Trust Improvement Evaluation on Healthcare oriented Explainable Artificial Intelligence Interventions: Systematic Review

Samantha Cramer, Elina Gastreich-de-Llanes, Jonas Bäcker, Robin Manz, Philip Meyer, Dominik Müller, Ludwig Christian Hinske, Johannes Raffler, Iñaki Soto-Rey

Submitted to: Journal of Medical Internet Research
on: February 25, 2025

Table of Contents

Original Manuscript 5

User Experience and Trust Improvement Evaluation on Healthcare oriented Explainable Artificial Intelligence Interventions: Systematic Review

Samantha Cramer[1] BSc; Elina Gastreich-de-Llanes[1]; Jonas Bäcker[1] BSc; Robin Manz[1] BSc; Philip Meyer[1] BSc; Dominik Müller[1, 2] Dr rer nat; Ludwig Christian Hinske[1] Prof Dr; Johannes Raffler[1] Dr rer nat; Iñaki Soto-Rey[1] Dr rer med

[1]Digital Medicine University Hospital Augsburg Augsburg DE
[2]IT?Infrastructure for Translational Medical Research University of Augsburg Augsburg DE

Corresponding Author:
Iñaki Soto-Rey Dr rer med
Digital Medicine
University Hospital Augsburg
Stenglinstr. 2
Augsburg
DE

Abstract

Background: The field of artificial intelligence (AI) has expanded rapidly in recent years. Generally, AI is viewed as a "black box" since understanding how it came to its presented solution is nearly impossible, which causes mistrust among end-users. This presents a problem, especially when AI is supposed to be implemented in high-stakes decision work environments. An example of such a work environment is the health care system. Additionally, to the general mistrust there are also legal regulations in place in the case of the implementation of AI systems within the health care system. The mistrust and legal regulations create a strong barrier for the widespread implementation of AI methods across the health care sector. To improve trust in the artificial intelligence systems and to fulfill the legal requirements, there has been a need for transparent, interpretable, explainable artificial intelligence systems. Though rather than developing new AI models, many researchers are working on post-hoc explainable artificial intelligence (XAI) systems which could at least provide the legally needed amount of transparency. Nevertheless, to ensure their usability, the created systems must be explainable to the end-user.

Objective: The goal of this systematic review was to identify the number of evaluations done on the usability, user satisfaction, experience and trust of XAI systems in the health care system. We also aimed to find the most used methods for usability/user experience evaluations.

Methods: Following the PRISMA 2020 guidelines, we extracted 6.008 references from four databases. After our concluding our screening steps 134 results remained eligible for the systematic review. The publications were categorized into 26 medical, 102 XAI method and 15 evaluation categories.

Results: 12 of the 15 evaluation categories were user-based. Only 35 of the 134 papers were sorted into user-based evaluation categories. A large portion of the 35 publications used self-designed questionnaires. Only 3 of the 35 presented a User Centered Design-Process. Our hypothesis that XAI is rarely evaluated, let alone developed, in relation to the needs of the end-user was confirmed.

Conclusions: We conclude that there is still a strong need for more involvement of the end-user during the development or at least during the evaluation of the created XAI models. Additionally, we recommend the development of a standardized framework to improve the generalizability of XAI methods. If XAI isn't developed closer to the needs of the end-user, evaluated from the end-user, or at best developed with the users, we expect that the implementation of explainable artificial intelligence in the health care environment will get increasingly hard.

(JMIR Preprints 25/02/2025:73124)
DOI: https://doi.org/10.2196/preprints.73124

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Review

Samantha Cramer[1], BSc; Elina Gastreich-de-Llanes[1]; Jonas Bäcker[1], BSc; Robin Manz[1], BSc; Philip Meyer[1], BSc; Dominik Müller[1,2], Dr. rer. nat.; Ludwig Christian Hinske[1],Prof. Dr. med.; Johannes Raffler[1],Dr. rer. nat.; Iñaki Soto-Rey[1*], Dr. rer. med.

[1] Digital Medicine, University Hospital Augsburg, Augsburg, Germany

[2] IT-Infrastructure for Translational Medical Research, University of Augsburg, Augsburg, Germany

[*]Corresponding author

Registration Number (DOI - Protocols.io):
https://dx.doi.org/10.17504/protocols.io.bp2l69omzlqe/v1

User Experience and Trust Improvement Evaluation on Healthcare oriented Explainable Artificial Intelligence Interventions: Systematic Review

Abstract:

Background: The field of artificial intelligence (AI) has expanded rapidly in recent years. Generally, AI is viewed as a “black box” since understanding how it came to its presented solution is nearly impossible, which causes mistrust among end-users. This presents a problem, especially when AI is supposed to be implemented in high-stakes decision work environments. An example of such a work environment is the health care system. Additionally, to the general mistrust there are also legal regulations in place in the case of the implementation of AI systems within the health care system. The mistrust and legal regulations create a strong barrier for the widespread implementation of AI methods across the health care sector. To improve trust in the artificial intelligence systems and to fulfill the legal requirements, there has been a need for transparent, interpretable, explainable artificial intelligence systems. Though rather than developing new AI models, many researchers are working on post-hoc explainable artificial intelligence (XAI) systems which could at least provide the legally needed amount of transparency. Nevertheless, to ensure their usability, the created systems must be explainable to the end-user.

Objective: The goal of this systematic review was to identify the number of evaluations done on the usability, user satisfaction, experience and trust of XAI systems in the health care system. We also aimed to find the most used methods for usability/user experience evaluations.

Methods: Following the PRISMA 2020 guidelines, we extracted 6.008 references from four databases. After our concluding our screening steps 134 results remained eligible for the systematic review. The publications were categorized into 26 medical, 102 XAI method and 15 evaluation categories.

Results: 12 of the 15 evaluation categories were user-based. Only 35 of the 134 papers were sorted into user-based evaluation categories. A large portion of the 35 publications used self-designed questionnaires. Only 3 of the 35 presented a User Centered Design-

Process. Our hypothesis that XAI is rarely evaluated, let alone developed, in relation to the needs of the end-user was confirmed.

Conclusion: We conclude that there is still a strong need for more involvement of the end-user during the development or at least during the evaluation of the created XAI models. Additionally, we recommend the development of a standardized framework to improve the generalizability of XAI methods. If XAI isn't developed closer to the needs of the end-user, evaluated from the end-user, or at best developed with the users, we expect that the implementation of explainable artificial intelligence in the health care environment will get increasingly hard.

Keywords: Systematic Review, XAI, Usability, Evaluation, PRISMA

Introduction

In recent years the field of artificial intelligence (AI) has grown exponentially. AI aims to imitate cognitive functions to solve real-world problems by learning from previous data, thus consistently gaining knowledge and gradually improving its behavior [1]. One of the most popular fields of AI is the field of neural networks. These networks typically consist of several layers which are connected by a multitude of nonlinear relations. Even if one were to inspect each layer and its contributions to the outcome, it is impossible to fully understand how a neural network came to its decision, which is why these networks are considered "black boxes" [2].

In the healthcare environment, AI-based systems are not expected to act autonomously in patient care but to serve as decision support for medical staff [3] with the purpose of reducing the workload for clinicians and thus improving the patient outcome. AI-based systems have demonstrated utility in various medical departments, including pathology, radiology and dermatology [4,5]. However, it has been demonstrated that AI-based systems are not robust and may encounter failures or develop bias in the face of artificial and natural noise [3]. These kinds of failures or malfunctions could have far-reaching consequences in the healthcare environment, especially if they go unnoticed [2]. Studies show that stakeholders who interact with AI tools to reach a decision may be persuaded to follow the presented recommendations even if these recommendations are not correct, rather than verifying the recommendation [6].

The risks of AI malfunctions and the lack of transparency presented by the systems are a grave cause of mistrust for clinicians [2,4,5]. Additionally, legal regulations such as the European Union's General Data Protection Regulation (GDPR, Article 15) and their AI Act preserve the patients right to receive meaningful information about how a decision was rendered during their care and require AI systems to be developed and used in a traceable and explainable way [2,3,7]. The mistrust and legal regulations create a strong barrier for the widespread implementation of AI methods across the health care sector.

In order to improve trust in the artificial intelligence systems and to fulfill the legal requirements, there has been a need for transparent, interpretable, explainable artificial intelligence systems [1–5,8]. Creating new transparent AI systems is different from creating non-transparent systems, as the desire for transparency contributes an additional layer of complexity that is not necessarily computational. Instead, the creation involves a human factor, namely the end-user to whom the algorithm should be transparent. Consequently, transparency is not a property of the algorithm but a relationship between the AI and the user processing the information [6].

Rather than developing new AI models, many researchers are working on post-hoc explainable artificial intelligence (XAI) systems, which could at least provide a minimal amount of transparency [9]. These post-hoc XAI methods focus on explaining the results of the AI-based systems and providing estimates on how these came to their decisions [8] by reviewing the factors that affect a neural network's output the most [4,10].

However, even when creating post-hoc XAI models, the interpretability and transparency problem is not one that can be solved by a technical approach alone [4]. Similar to the creation of transparent AI systems, the end-users must be involved in the development of post-hoc XAI models as well. It must be ensured that these systems are created to be explainable to the end-user, ensuring their usability. Thus securing that the end-users can evaluate the AI models decision to determine whether or not they should accept the proposed recommendation [3].

The usability of a system is a comprehensive concept which refers to the ease with which users can achieve their goals efficiently, effectively and with satisfaction. While usability itself is not an attribute of a system, there are certain system components that contribute to the usability of a system [11]. Some of these components are learnability, how easy is it to use; efficiency, how quickly can a user perform the desired task; memorability, how easy is it to reestablish the needed skills when returning to the system; effectiveness, how many errors do users make; utility, are all necessary features provided; as well as satisfaction, how pleasant is the system to use [12]. In order to evaluate the usability there are a set of methods to conduct a so-called usability evaluation [13].

To our knowledge, there have been too few usability evaluations conducted to prove that the currently available XAI systems are usable and explainable, thus trustable.

In this manuscript we explore the scientific research which has been conducted on explainable AI in terms of user experience. The research should include a usability evaluation of XAI focusing on which elements have been implemented to create a trustable XAI system. This goal is to be achieved through a systematic review. Our hypothesis is that XAI is rarely evaluated, let alone developed, in relation to the needs of the end-user but rather based on their technical features and accuracy. Furthermore, we aim to find the most used methods for usability/user experience in XAI evaluations.

Material and Methods

The presented systematic review was conducted based on the PRISMA 2020 guidelines [14]. The goal was to identify the number of evaluations done on the usability, user satisfaction, experience and trust of XAI systems in the health care system. We also aimed to find the most used methods for usability/user experience evaluations.

Search Strategy and Selection of Studies

A comprehensive literature search was carried out in relevant bibliographic web-based databases (PubMed, PubMedCentral, Science Direct, Google Scholar). The used search terms were not restricted to the title but rather were found within the title, abstract or full text. The words included in the search query were "XAI", "X-AI", "Explainable", "AI", "Explainable", "artificial intelligence", "Usability", "satisfaction", "experience", "trust, "Evaluation". The words were combined with the conjunction "AND" and the logical operator "OR". An example of a used search query can be seen in Table 2.1.

Table 1. Example of the search query used during the literature search for the Systematic Review.

Searched words	Conjunction	Disjunction	Example of query
“XAI”, “X-AI”, “Explainable”, “AI”, “Explainable”, “artificial intelligence”, “Usability”, “satisfaction”, ”experience”, “trust, “Evaluation”	and	or	("XAI" OR "X-AI" OR ("Explainable" AND "AI") OR ("Explainable" AND "artificial intelligence")) AND ("Usability" OR "satisfaction" OR "experience" OR "trust") AND "Evaluation"

The results of the initial search were imported into Citavi Version 6 using the “exclude duplicates” option and all results for which a full text could not be retrieved were excluded from further analysis. The titles, abstracts and keywords of the remaining papers were then read to verify the eligibility of each paper. If the read material was deemed irrelevant, the paper was excluded from the process. Papers were generally discarded if they met any of the exclusion criteria (2.3) or were not peer-reviewed publications.

At least two independent reviewers reviewed all papers after which they compared their results.

Inclusion Criteria

The inclusion criteria for this systematic review were fixed. The results had to be a scientific publication written in English and published between 01.01.2017 and 31.07.2023. All included papers were required to involve a medical area, describe and evaluate at least one Explainable Artificial Intelligence (XAI) method or they had to be a literature review about XAI.

Exclusion Criteria

Publications that were not written in English, which were not published in a peer-reviewed scientific journal or were not full scientific publications (for example Abstracts, Posters, Books, or editorials) were excluded. If the depiction of a medical area, a XAI-method or an evaluation of this XAI-method were absent from the publication, the publication was also excluded, as these were irrelevant to the review.

Outcome Assessment

The goal of this systematic review was to identify publications pertinent to user trust/ experience/ satisfaction evaluations of XAI in the health care system. We also aimed to find the most often used methods for the evaluation of usability/user experience of XAI-methods.

Bias Assessment

The results were at risk of bias due to the subjective opinion of the reviewers, which we tried to avoid by having each paper be reviewed by two separate reviewers, individually. Furthermore, risk of bias existed due to the fact we might not have found all fitting publications with our search query, even though we developed it to the best of our knowledge.

Results

Study Selection

During the literature search within PubMed, PubMedCentral, ScienceDirect and Google Scholar 6.008 publications were identified using the search terms and the publication date restrictions (Figure 1).

Combining Results, filtration Full Text

After extracting the results from each database, the export files containing the 6,008 references were imported into Citavi Version 6 using the “exclude duplicates” option. During the import Citavi identified 54 duplicates, which were excluded. In sum, 5,954 papers remained in the Citavi-project.

The first step after creating the Review project was the extraction of the papers for which a full text version could be found. Using the Citavi “Find Full Text”- option and manual internet research we found the full text versions of 5,214 papers, thus excluding the 740 papers for which the full text could not be found or could not be accessed.

Screening based on Title, Keywords and Abstract

During the literature search in the different databases, 5,214 full-text papers qualified for the first screening step.

During the first screening step the papers were included or excluded based on their title, keywords and abstract and if these matched the following inclusion criteria. The papers were included if the reviewed material presented proof of the paper being a scientific publication written in English, including material about a medical area, at least one XAI method and some sort of user-based scientific evaluation. The other option for the inclusion was if the paper was a literature review about XAI.

During this step we identified 762 papers for the temporary inclusion, excluded 4,444 and identified an additional 8 duplicates that were missed by Citavi during the import process.

Screening based on Full Text

The full texts of the remaining 762 publications were then screened during the second screening step. Each paper was categorized into at least one specific medical category, a category about the XAI which was presented and the evaluation type, if this was not possible the paper was excluded from further examination.

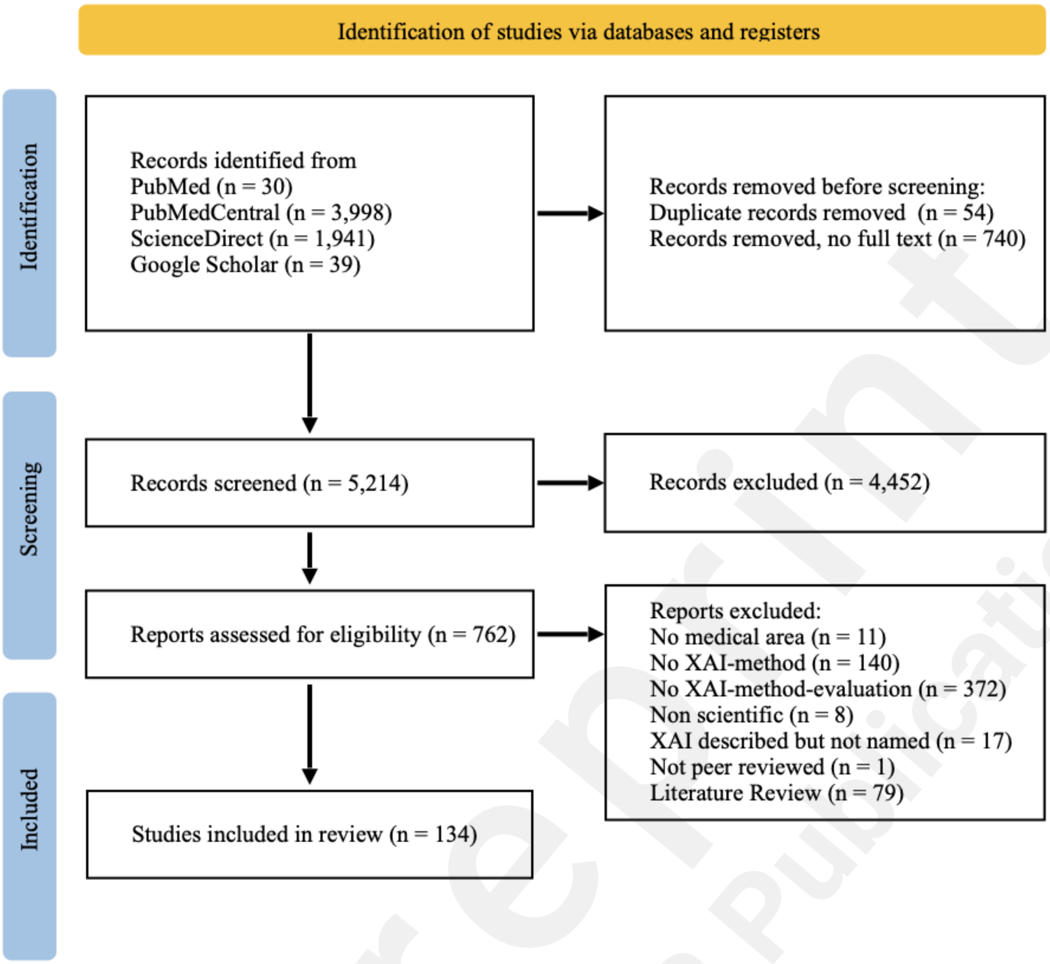
Thus, there were three exclusion categories: no medical area, no XAI-method, no XAI-method-evaluation. During the screening process additional exclusion categories emerged: non scientific, XAI described but not named and not peer reviewed. We also excluded all literature reviews at this point.

Each literature review was extracted and sorted into a category specifically for them, as they would provide possible cross-references later in the review process.

The literature reviews were scanned for cross-references, but no eligible cross-references were identified.

During the second screening step, 628 papers were excluded, leaving 134 results eligible for this systematic review.

Figure 1. Flow Diagram of our PRISMA literature research in the four previously listed databases based on the template of [14]. Out of 6,008 references, 134 were included in this review.



Results of Bias in Studies

The main risk within this review is that the results are based upon the subjective opinion of the reviewers. We aim to avoid this bias by ensuring that each paper is reviewed individually by at least two separate reviewers.

Furthermore, risk of bias exists due to the fact we might not have found all fitting publications with our search query, even though we developed it to the best of our knowledge.

Results of Included Studies

The 134 included studies were categorized into 26 different medical categories visible in Table 3.1. 67 standardized and 35 new XAI methods were identified within the included studies and are visualized in Table 3.2. Lastly, the included studies were classified into 15 evaluation categories listed in Table 3.3. Of the 15 evaluation categories identified, 12 were user-centered. These user-centered categories were based on components which increase the usability of a system, some of which were introduced in the introduction of this paper. The exact categories, as well as the list of papers in each category, are visualized within Table 3.3 starting at the 4th row. A large percentage of the remaining 99 papers were assigned the category "comparison" in combination with "effectiveness" and/or "efficiency" (Table 3.3). 40 papers only described an "effectiveness"- evaluation.

Most of the reviewed papers were categorized into multiple categories of the medical, XAI and evaluation category.

Table 2. Results Table depicting the categorization into medical areas of the included references.

Medical Area	Number of References	References
Radiology	33	[15–47]
Neurology	20	[31,41,45,48–63]
COVID*	19	[15–19,21,22,24,28,34,38,42,64–70]
Cardiology	17	[71–87]
Intensive care	11	[37,65,82,83,88–94]
Pulmonology	10	[22,29,30,35,37,93,95–98]
Oncology	10	[46,47,50,99–105]
Gynaecology	9	[39,46,50,80,102,103,105–107]
Pathology	9	[5,99,100,108–113]
Medical Genetics	8	[62,69,103,107,114–117]
Ophthalmology	7	[44,101,118–122]
General Medicine/Diagnosis	7	[123–129]
Endocrinology	6	[119,120,130–133]
Pharmacology	5	[71,116,134–136]
Pediatrics	4	[80,95,101,137]
Gastroenterology	3	[43,104,138]
Orthopedic surgery	3	[26,27,139]
Dentistry	2	[140,141]
Dermatology	2	[106,142]
Nutritional Medicine	2	[143,144]
Psychiatry	2	[41,145]
Anesthesiology	1	[146]
Internal medicine	1	[130]
Anesthesiology	1	[146]
Physiology	1	[139]
General Surgery	1	[147]

* Not a medical category but a subcategory of Pulmonology which was partitioned off due increased publications during the pandemic

Table 3. Results Table depicting the categorization into different XAI-methods groups of the included references.

Explainable AI (XAI)	Number of References	References
SHAP	56	[26,31,39,41,47–

		49,51,52,54,58,60,62,65–67,69,70,73,75,76,78–80,82–87,92–94,98,101–103,105,107,114,116,117,121–123,126,128–130,132,134,135,137,138,144,146]
GradCAM	35	[5,15,16,18–24,27–29,31,33,34,36,38,42,44,45,56,57,61,72,77,81,95,97,99,108,109,119,141,142]
LIME	34	[18,26,28,31,32,47,50,51,54,55,58,62,64,66,67,69,76–78,82,85,93,94,97,101,105,111,123,128,129,131,133,136,140]
Saliency Maps	11	[5,15,17,20,35,36,53,56,101,108,118]
Counterfactuals	9	[5,32,40,58,68,125,130,139,145]
Guided Backpropagation	9	[5,25,31,34,36,41,45,121,142]
Guided Grad-CAM	8	[31,34,41,44,45,113,119,142]
Integrated Gradients	8	[20,31,41,45,52,108,127,128]
Occlusion	8	[28,31,59,77,118–120,127]
Partial Dependence Plot (PDP)	8	[58,65,70,75,82,133,136,138]
Anchors	4	[58,66,67,124]
Layer-Wise Relevance Propagation (LRP)	4	[15,25,32,41]
Permutation Feature Importance (PIMP)	4	[51,75,138,145]
Eli5	3	[58,66,67]
GradCAM++	3	[15,33,38]
QLattice	3	[58,66,67]
CAM	2	[97,110]
deepLift	2	[31,41]
Explainable Boosting Machine (EBM)	2	[84,88]
Feature Importance	2	[48,67]
ScoreCAM	2	[33,38]
SmoothGrad	2	[31,41]
Taylor Decomposition	2	[25,87]
Vanilla Gradients	2	[40,108]
XGBoost as XAI	2	[63,84]
XRAI (based on Integrated Gradients)	2	[108,142]
Automated Concept-Based Explanation (ACE)	1	[113]
Average-CAM	1	[97]
Average-Grad-CAM	1	[97]
Blur Integrated Gradients	1	[108]

BreakDown & iBreakDown	1	[94]
Class-Selective Relevance Maps (CRM)	1	[35]
Concept Attribution	1	[5]
Contrastive Explanations Method (CEM)	1	[145]
DALEX	1	[93]
Deconvolution	1	[31]
DeepLabv3+	1	[106]
deepLift SHAP	1	[41]
DenseASPP	1	[106]
Exact Shapley values	1	[129]
Explaination-by-Example (ExMatchina)	1	[58]
ExpliClas	1	[135]
FasterScore-CAM	1	[38]
Feature Ablation	1	[31]
Feature Permutation	1	[31]
G-Deep-SHAP	1	[50]
Generalized Additive Model (GAM)	1	[74]
Generalized Linear Model (GLM)	1	[74]
Gradient Analysis	1	[41]
GUAJE	1	[135]
Guided Integrated Gradients	1	[108]
Hires-CAM	1	[33]
InputXGradient	1	[41]
Layers-CAM	1	[33]
Logistic Regression Classifiers (LR)	1	[78]
LORE	1	[124]
Network Dissection	1	[57]
Permutation Sample Importance (PFD)	1	[76]
Permutation Testing	1	[78]
Random Forest as XAI	1	[68]
Random Forest Classifier	1	[78]
RefineNet	1	[106]
RELAX	1	[43]
RISE	1	[28]
SHAPHT	1	[63]
Uncertain-CAM	1	[19]
Xgrad-CAM	1	[33]
New Methods	3	[15,36,125]
TWIX	1	[147]
Ada-WHIPS	1	[124]
AggMapNet	1	[116]
Al-CAD	1	[30]
Attention Branch Network (ABN)	1	[110]
Automatic Explanation Method	1	[96]
BI-RADS-NET-V2	1	[46]

BoCSoR XAI	1	[48]
CACHE-Grader	1	[112]
CA-Net	1	[106]
Case-Based Reasoning	1	[102]
COmic	1	[115]
Compensated Integrated Gradients	1	[52]
COVID-GMIC, COVID-GMIC-DRC, COVID-GBM t	1	[37]
COVID-view	1	[16]
Cumulative Fuzzy Class Membership Criterion (CFCMC)	1	[100]
DeepAISE	1	[91]
DeepConsensus	1	[90]
ENDOANGEL-LA	1	[104]
ExplAIn	1	[120]
FSXRF - Feature Selection with Explainable Random Forest	1	[68]
Graph Chart Diagram	1	[80]
MIME	1	[127]
Multi-Input Conv-Att	1	[61]
Natural Language Generation	1	[143]
NDG-CAM	1	[99]
NeuroXAI	1	[45]
SkiNet	1	[142]
Spectro-Temporal Attention (STA)	1	[72]
Text Based Explanations	1	[71]
TraCE	1	[40]
XAIRE	1	[89]
XGBoost Classifier	1	[102]
xVITCOS	1	[17]

Table 4. Results Table depicting the categorization into evaluation types of the included references.

Evaluation Types	**Number of References**	**References**
Effectiveness	100	[5,15,17–24,26–29,31,33,35–41,43,44,46,48–52,55–58,60,63,65–68,70,72–76,78–80,84,85,88,89,91,95–104,106–115,117,119,120,122–132,134,135,138–143,145–147]
Comparison*	67	[5,15,18–20,25,26,28,32–34,36,38,40–42,45,47,48,50,52,54,58,59,62,63,69,70,72,74–78,82–85,87,92–

		94,97,101,102,105,106,108,110,113,116,118,119,121,123–125,127–130,133,136,138,142,144,147]
User Satisfaction	22	[5,16,23,28,30,32,43,44,49,53,61,64,71,81,88,100,102,104,121,125,126,137]
Explainability	14	[5,16,23,28,30–32,49,54,90,102,114,126,143]
Efficiency	14	[31,38,42,44,48,50,52,68,76,106,115,117,124,145]
Trust	11	[5,28,31,32,41,49,64,76,104,125,126]
Utility	8	[28,31,49,53,88,108,126,137]
Learnability	4	[32,53,73,102]
Transparency	4	[30,31,57,126]
Fidelity	2	[90,138]
Generalizability	1	[90]
Interpretability	1	[138]
Persuasiveness	1	[90]
Repeatability	1	[20]
Reproducibility	1	[20]

* not an attribute but rather a comparison of different XAI methods within the paper

Results of syntheses

Of the 5,214 full text papers identified during our initial search, 762 publications were eligible for the second round of screening, describing a medical area, XAI method and some sort of evaluation within their titles, abstracts or keywords (Figure 1). Of these remaining papers, only 134 papers ended up being included for this systematic review, with only 99 describing a technical evaluation of the used XAI methods. Only 34 publications included some sort of user-based evaluation (Table 3.3).

There is not one specific user-based evaluation method that is used to evaluate the XAI methods, though more than a third of the papers, 15 to be exact, use a type of questionnaire to gather feedback about their used XAI methods (Table 3.4). While 15 of these publications used self-designed questionnaires, two used standardized questionnaires additionally and one solely (Table 3.4). A large number of papers also described an evaluation of the improvement of the accuracy of the clinicians when using AI and XAI without focusing on the opinions of the clinicians.

Table 5. listing all used methods for the evaluation of usability/user experience of the different XAI-methods.

Method for the evaluation of usability/user experience	Number of References	Papers using this Method

Self-constructed Survey/Questionnaires	15	[5,28,30–32,49,61,76,100,102,104,121,125,126,130,143]
Interviews / Focus Group	4	[5,31,71,81,102,137]
User Centered Design	3	[16,88,137]
SUS Questionnaire	2	[53,143]
Several Standardized questionnaires (Explanation Satisfaction Scale, Trust in Automation (TiA), Discrete Emotions Questionnaire (DEQ))	1	[32]
Z-Inspection®	1	[64]
Computational explainability checks		
Method testing, working with AI and XAI	5	[23,44,73,76,108]
Other	3	[41,54,57]
Feature based scores	2	[90,138]
Pearson correlation and mean SHAP	1	[114]
Relevance rank accuracy score (ground truth)	1	[43]
Utility	1	[20]

Reporting Bias and Certainty of Evidence

During the screening process, some papers described their methods as being explainable because of the usage of different XAI methods without ever going into detail about how the methods were used or if they improved the usability or explainability of the system for the end-user, thus never proceeding to do any sort of evaluation.

Discussion and Conclusion

Principal Results

We started by doing a literature search on four databases, using a search query to achieve a good consistency of all results throughout the databases.

Of the 6,008 resulting papers, we first extracted the 54 duplicates automatically using the Citavi import system. After the remaining 5,954 papers were imported, we searched for the full text versions of the papers. We found the full text versions of 5,214 results. These results were then filtered based on set requirements based on their title, keywords and the abstract. During the first screening step, we excluded 4,452 papers, leaving 762 publications to be screened during the second step. The full texts were analyzed and categorized based on the presented medical area, XAI method and evaluation type. If the publication could not be sorted into all three categories at least once, the paper was excluded. After the second screening step, only 134 papers were eligible to be included in our systematic review. During this last step, only 34 of the 134 papers were sorted into user-based evaluation categories.

This confirmed our hypothesis that XAI is rarely evaluated, let alone developed, in relation to the needs of the end-user. Especially when reviewing the 34 papers sorted into user-based evaluation categories, it became clear that XAI is very rarely developed in relation to the end-user. Only 3 of the 6,008 papers that were selected by our initial search presented a User Centered Design-Process, thus a user-centered development of the XAI.

There was not one specific usability evaluation method which was prominent within the publications presenting a user-based evaluation and there was no detailed information on which specific elements of the XAI provided the users with additional trust in the systems.

Limitations

The results are at risk of bias due to the subjective opinion of the reviewers, which we aimed to avoid by having each paper reviewed by at least two reviewers separately before they discussed their individual categorization.

An additional limitation is the restricted number of databases that we searched. There might have been more papers published within other databases that would have fit our search query as well.

Lastly, there were some papers which were difficult to classify into specific medical/XAI/evaluation categories during the final screening step. The reason for the categorization issue varied. Most often than not, the issue arose around the categorization into a specific evaluation category. While some papers mentioned explainability and evaluation within their titles, keywords or abstracts, they only had very vague effectiveness evaluations which didn't refer to the ground truth and did not compare their results with any other methods within their full texts. Other papers included user-based evaluation keywords but did not explicitly describe any user involvement. [58] is an example of such a paper. While they use the keywords fidelity and understandability, they do not describe any user involvement and base their XAI grading solely on technical features and their own opinion.

Comparison with Prior Work

The scientific community has been aware of the need for transparency of AI systems long before this review [1–5,8]. Nevertheless, only a handful of studies have attempted to bridge the gap between technical implementation and user experience of XAI in the health care domain [5]. Many publications that strive to learn about the explainability of XAI methods chose to conduct interviews with the end users (Table 3.4) [148] or interview design experts to review the development side of usability [149]. However, there are some publications, including [150] which try to summarize the discovered knowledge about usability as we have done here. [150] for instance focuses on extracting the explanation type and the effect of it on the user while we focused on the specific user-based evaluation types.

Conclusion and Future Work

This paper presented the results of a systematic review done based on the PRISMA 2020 guidelines. During the review, the hypothesis was confirmed that XAI is rarely evaluated, let alone developed, in relation to the needs of the end-user and that there is no standardized user-based evaluation method.

Based on this review we conclude that there is still a desperate need for more involvement of the end-user during the development or at least during the evaluation of the created XAI models. This would ensure a safer usage of the XAI after the development phase, as users would have been in contact with the system early on and could be more aware of its limitations. This awareness could improve the relationship and thus the trust of the end-users in the system and make for a smoother implementation of these developed systems in the work environment.

Additionally, to improve the development process, we advise the creation of a standardized framework or evaluation method to support the generalizability of XAI methods.

If XAI isn't developed closer to the needs of the end-user, evaluated from the end-user or at best developed with the users, we expect that the implementation of explainable artificial intelligence in the health care environment will get increasingly hard.

Acknowledgements

SC coordinated the review team, reviewed all the publications and drafted the manuscript. EGdL, JB, RM, PM and DM took part in the review process and helped to draft the manuscript. LCH provided helpful feedback for the conception and conduction of the study and helped to draft the manuscript. JR took part in the review process, co-led the EKIPRO Projekt and helped to draft the manuscript. ISR conceived the idea of the manuscript, took part in the review process, co-led the EKIPRO Projekt and helped to draft the manuscript.

Funding

This work was conducted in the context of the EKIPRO project funded by the Intramural Research and Young Scientist Grant from the Faculty of Medicine at the University of Augsburg.

Conflicts of Interest

None to declare.

References

1. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery 2019;9(4). doi: 10.1002/widm.1312

2. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Medical Image Analysis 2022;79:102470. doi: 10.1016/j.media.2022.102470

3. Di Jin, Sergeeva E, Weng W-H, Chauhan G, Szolovits P. Explainable Deep Learning in Healthcare: A Methodological Survey from an Attribution View. arXiv; 2021. doi: 10.48550/ARXIV.2112.02625

4. Gille F, Jobin A, Ienca M. What we talk about when we talk about trust: Theory of trust for AI in healthcare. Intelligence-Based Medicine 2020;1–2:100001. doi: 10.1016/j.ibmed.2020.100001

5. Evans T, Retzlaff CO, Geißler C, Kargl M, Plass M, Müller H, Kiehl T-R, Zerbe N, Holzinger A. The explainability paradox: Challenges for xAI in digital pathology. Future Generation

Computer Systems 2022;133:281–296. doi: 10.1016/j.future.2022.03.009

6. Chen H, Gomez C, Huang C-M, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. npj Digital Medicine 2022;5(1). doi: 10.1038/s41746-022-00699-2

7. Publications Office of the European Union L-2985 Luxembourg LUXEMBOURG. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)Text with EEA relevance.

8. Schoonderwoerd TAJ, Jorritsma W, Neerincx MA, van den Bosch K. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. International Journal of Human-Computer Studies 2021;154:102684. doi: 10.1016/j.ijhcs.2021.102684

9. Mourby M, Ó Cathaoir K, Collin CB. Transparency of machine-learning in healthcare: The GDPR & European health law. Computer Law & Security Review 2021;43:105611. doi: 10.1016/j.clsr.2021.105611

10. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. Journal of the American Medical Informatics Association 2020;27(4):592–600. doi: 10.1093/jamia/ocz229

11. ISO. Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts. 2018; Available from: https://www.iso.org/obp/ui/en/#iso:std:iso:9241:-11:ed-2:v1:en

12. Nielsen J. Usability 101: Introduction to Usability. Available from: https://www.nngroup.com/articles/usability-101-introduction-to-usability/

13. Rosenzweig E. Usability Inspection Methods. Successful User Experience: Strategies and Roadmaps Elsevier; 2015. p. 115–130. doi: 10.1016/b978-0-12-800985-7.00006-5ISBN:978-0-12-800985-7

14. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;n71. doi: 10.1136/bmj.n71

15. Shi W, Tong L, Zhu Y, Wang MD. COVID-19 Automatic Diagnosis With Radiographic Imaging: Explainable Attention Transfer Deep Neural Networks. IEEE Journal of Biomedical and Health Informatics 2021;25(7):2376–2387. doi: 10.1109/jbhi.2021.3074893

16. Jadhav S, Deng G, Zawin M, Kaufman AE. COVID-view: Diagnosis of COVID-19 using Chest CT. IEEE Transactions on Visualization and Computer Graphics 2022;28(1):227–237. doi: 10.1109/tvcg.2021.3114851

17. Mondal AK, Bhattacharjee A, Singla P, Prathosh AP. xViTCOS: Explainable Vision Transformer Based COVID-19 Screening Using Radiography. IEEE Journal of Translational Engineering in Health and Medicine 2022;10:1–10. doi: 10.1109/jtehm.2021.3134096

18. Agrawal S, Honnakasturi V, Nara M, Patil N. Utilizing Deep Learning Models and Transfer Learning for COVID-19 Detection from X-Ray Images. SN Computer Science 2023;4(4). doi: 10.1007/s42979-022-01655-3

19. Aldhahi W, Sull S. Uncertain-CAM: Uncertainty-Based Ensemble Machine Voting for Improved COVID-19 CXR Classification and Explainability. Diagnostics 2023;13(3):441. doi: 10.3390/diagnostics13030441

20. Arun N, Gaw N, Singh P, Chang K, Aggarwal M, Chen B, Hoebel K, Gupta S, Patel J, Gidwani M, Adebayo J, Li MD, Kalpathy-Cramer J. Assessing the Trustworthiness of Saliency Maps for Localizing Abnormalities in Medical Imaging. Radiology: Artificial Intelligence 2021;3(6). doi: 10.1148/ryai.2021200267

21. Aviles-Rivero AI, Sellars P, Schönlieb C-B, Papadakis N. GraphXCOVID: Explainable deep graph diffusion pseudo-Labelling for identifying COVID-19 on chest X-rays. Pattern Recognition 2022;122:108274. doi: 10.1016/j.patcog.2021.108274

22. Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays. Computer Methods and Programs in Biomedicine 2020;196:105608. doi: 10.1016/j.cmpb.2020.105608

23. Cabitza F, Campagner A, Ronzio L, Cameli M, Mandoli GE, Pastore MC, Sconfienza LM, Folgado D, Barandas M, Gamboa H. Rams, hounds and white boxes: Investigating human–AI collaboration protocols in medical diagnosis. Artificial Intelligence in Medicine 2023;138:102506. doi: 10.1016/j.artmed.2023.102506

24. Chung J, Kim D, Choi J, Yune S, Song KD, Kim S, Chua M, Succi MD, Conklin J, Longo, Maria G. Figueiro, Ackman JB, Petranovic M, Lev MH, Do S. Prediction of oxygen requirement in patients with COVID-19 using a pre-trained chest radiograph xAI model: efficient development of auditable risk prediction models via a fine-tuning approach. Scientific Reports 2022;12(1). doi: 10.1038/s41598-022-24721-5

25. Coupet M, Urruty T, Leelanupab T, Naudin M, Bourdon P, Maloigne CF, Guillevin R. A multi-sequences MRI deep framework study applied to glioma classfication. Multimedia Tools and Applications 2022;81(10):13563–13591. doi: 10.1007/s11042-022-12316-1

26. Wang H, Wu W, Han C, Zheng J, Cai X, Chang S, Shi J, Xu N, Ai Z. Prediction Model of Osteonecrosis of the Femoral Head After Femoral Neck Fracture: Machine Learning–Based Development and Validation Study. JMIR Medical Informatics 2021;9(11):e30079. doi: 10.2196/30079

27. Gao Y, Soh, Nicholas Yock Teck, Liu N, Lim G, Ting D, Cheng LT-E, Wong KM, Liew C, Oh HC, Tan JR, Venkataraman N, Goh SH, Yan YY. Application of a deep learning algorithm in the detection of hip fractures. iScience 2023;26(8):107350. doi: 10.1016/j.isci.2023.107350

28. Goel K, Sindhgatta R, Kalra S, Goel R, Mutreja P. The effect of machine learning explanations

on user trust for automated diagnosis of COVID-19. Computers in Biology and Medicine 2022;146:105587. doi: 10.1016/j.compbiomed.2022.105587

29. Gozzi N, Giacomello E, Sollini M, Kirienko M, Ammirabile A, Lanzi P, Loiacono D, Chiti A. Image Embeddings Extracted from CNNs Outperform Other Transfer Learning Approaches in Classification of Chest Radiographs. Diagnostics 2022;12(9):2084. doi: 10.3390/diagnostics12092084

30. Hong S, Hwang EJ, Kim S, Song J, Lee T, Jo GD, Choi Y, Park CM, Goo JM. Methods of Visualizing the Results of an Artificial-Intelligence-Based Computer-Aided Detection System for Chest Radiographs: Effect on the Diagnostic Performance of Radiologists. Diagnostics 2023;13(6):1089. doi: 10.3390/diagnostics13061089

31. Jin W, Li X, Fatehi M, Hamarneh G. Guidelines and evaluation of clinical explainable AI in medical image analysis. Medical Image Analysis 2023;84:102684. doi: 10.1016/j.media.2022.102684

32. Mertes S, Huber T, Weitz K, Heimerl A, André E. GANterfactual—Counterfactual Explanations for Medical Non-experts Using Generative Adversarial Learning. Frontiers in Artificial Intelligence 2022;5. doi: 10.3389/frai.2022.825565

33. Mukhtorov D, Rakhmonova M, Muksimova S, Cho Y-I. Endoscopic Image Classification Based on Explainable Deep Learning. Sensors 2023;23(6):3176. doi: 10.3390/s23063176

34. Niranjan K, Shankar Kumar S, Vedanth S, Chitrakala S. An Explainable AI driven Decision Support System for COVID-19 Diagnosis using Fused Classification and Segmentation. Procedia Computer Science 2023;218:1915–1925. doi: 10.1016/j.procs.2023.01.168

35. Rajaraman S, Sornapudi S, Alderson PO, Folio LR, Antani SK, Huo Y. Analyzing inter-reader variability affecting deep ensemble learning for COVID-19 detection in chest radiographs. PLOS ONE 2020;15(11):e0242301. doi: 10.1371/journal.pone.0242301

36. Saleem H, Shahid AR, Raza B. Visual interpretability in 3D brain tumor segmentation network. Computers in Biology and Medicine 2021;133:104410. doi: 10.1016/j.compbiomed.2021.104410

37. Shamout FE, Shen Y, Wu N, Kaku A, Park J, Makino T, Jastrzębski S, Witowski J, Wang D, Zhang B, Dogra S, Cao M, Razavian N, Kudlowitz D, Azour L, Moore W, Lui YW, Aphinyanaphongs Y, Fernandez-Granda C, Geras KJ. An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department. npj Digital Medicine 2021;4(1). doi: 10.1038/s41746-021-00453-0

38. Suri J, Agarwal S, Chabert G, Carriero A, Paschè A, Danna P, Saba L, Mehmedović A, Faa G, Singh I, Turk M, Chadha P, Johri A, Khanna N, Mavrogeni S, Laird J, Pareek G, Miner M, Sobel D, Balestrieri A, Sfikakis P, Tsoulfas G, Protogerou A, Misra D, Agarwal V, Kitas G, Teji J, Al-Maini M, Dhanjil S, Nicolaides A, Sharma A, Rathore V, Fatemi M, Alizad A, Krishnan P, Nagy F, Ruzsa Z, Fouda M, Naidu S, Viskovic K, Kalra M. COVLIAS 2.0-cXAI: Cloud-Based Explainable Deep Learning System for COVID-19 Lesion Localization in Computed Tomography Scans. Diagnostics 2022;12(6):1482. doi: 10.3390/diagnostics12061482

39. Teng LY, Mattar, Citra Nurfarah Zaini, Biswas A, Hoo WL, Saw SN. Interpreting the role of nuchal fold for fetal growth restriction prediction using machine learning. Scientific Reports 2022;12(1). doi: 10.1038/s41598-022-07883-0

40. Thiagarajan JJ, Thopalli K, Rajan D, Turaga P. Training calibration-based counterfactual explainers for deep learning models in medical image analysis. Scientific Reports 2022;12(1). doi: 10.1038/s41598-021-04529-5

41. Thomas AW, Ré C, Poldrack RA. Benchmarking explanation methods for mental state decoding with deep learning models. NeuroImage 2023;273:120109. doi: 10.1016/j.neuroimage.2023.120109

42. Ullah F, Moon J, Naeem H, Jabbar S. Explainable artificial intelligence approach in combating real-time surveillance of COVID19 pandemic from CT scan and X-ray images using ensemble model. The Journal of Supercomputing 2022;78(17):19246–19271. doi: 10.1007/s11227-022-04631-z

43. Wickstrøm KK, Østmo EA, Radiya K, Mikalsen KØ, Kampffmeyer MC, Jenssen R. A clinically motivated self-supervised approach for content-based image retrieval of CT liver images. Computerized Medical Imaging and Graphics 2023;107:102239. doi: 10.1016/j.compmedimag.2023.102239

44. Xu F, Jiang L, He W, Huang G, Hong Y, Tang F, Lv J, Lin Y, Qin Y, Lan R, Pan X, Zeng S, Li M, Chen Q, Tang N. The Clinical Value of Explainable Deep Learning for Diagnosing Fungal Keratitis Using in vivo Confocal Microscopy Images. Frontiers in Medicine 2021;8. doi: 10.3389/fmed.2021.797616

45. Zeineldin RA, Karar ME, Elshaer Z, Coburger ·Jan, Wirtz CR, Burgert O, Mathis-Ullrich F. Explainability of deep neural networks for MRI analysis of brain tumors. International Journal of Computer Assisted Radiology and Surgery 2022;17(9):1673–1683. doi: 10.1007/s11548-022-02619-x

46. Zhang B, Vakanski A, Xian M. BI-RADS-NET-V2: A Composite Multi-Task Neural Network for Computer-Aided Diagnosis of Breast Cancer in Ultrasound Images With Semantic and Quantitative Explanations. IEEE Access 2023;11:79480–79494. doi: 10.1109/access.2023.3298569

47. Zuo Y, Liu Q, Li N, Li P, Zhang J, Song S. Optimal 18F-FDG PET/CT radiomics model development for predicting EGFR mutation status and prognosis in lung adenocarcinoma: a multicentric study. Frontiers in Oncology 2023;13. doi: 10.3389/fonc.2023.1173355

48. Alfeo AL, Zippo AG, Catrambone V, Cimino MGCA, Toschi N, Valenza G. From local counterfactuals to global feature importance: efficient, robust, and model-agnostic explanations for brain connectivity networks. Computer Methods and Programs in Biomedicine 2023;236:107550. doi: 10.1016/j.cmpb.2023.107550

49. Antoniadi AM, Galvin M, Heverin M, Wei L, Hardiman O, Mooney C. A Clinical Decision Support System for the Prediction of Quality of Life in ALS. Journal of Personalized Medicine 2022;12(3):435. doi: 10.3390/jpm12030435

50. Chen H, Lundberg SM, Lee S-I. Explaining a series of models by propagating Shapley values. Nature Communications 2022;13(1). doi: 10.1038/s41467-022-31384-3

51. Choi BK, Choi YJ, Sung M, Ha W, Chu MK, Kim W-J, Heo K, Kim KM, Park YR. Development and validation of an artificial intelligence model for the early classification of the aetiology of meningitis and encephalitis: a retrospective observational study. eClinicalMedicine 2023;61:102051. doi: 10.1016/j.eclinm.2023.102051

52. Kawai Y, Tachikawa K, Park J, Asada M. Compensated Integrated Gradients for Reliable Explanation of Electroencephalogram Signal Classification. Brain Sciences 2022;12(7):849. doi: 10.3390/brainsci12070849

53. Leonardi G, Montani S, Striani M. Explainable process trace classification: An application to stroke. Journal of Biomedical Informatics 2022;126:103981. doi: 10.1016/j.jbi.2021.103981

54. Lombardi A, Diacono D, Amoroso N, Monaco A, Tavares, João Manuel R. S., Bellotti R, Tangaro S. Explainable Deep Learning for Personalized Age Prediction With Brain Morphology. Frontiers in Neuroscience 2021;15. doi: 10.3389/fnins.2021.674055

55. Magesh PR, Myloth RD, Tom RJ. An Explainable Machine Learning Model for Early Detection of Parkinson's Disease using LIME on DaTSCAN Imagery. Computers in Biology and Medicine 2020;126:104041. doi: 10.1016/j.compbiomed.2020.104041

56. Malhotra R, Singh Saini B, Gupta S. An interpretable feature-learned model for overall survival classification of High-Grade Gliomas. Physica Medica 2023;110:102591. doi: 10.1016/j.ejmp.2023.102591

57. Natekar P, Kori A, Krishnamurthi G. Demystifying Brain Tumor Segmentation Networks: Interpretability and Uncertainty Analysis. Frontiers in Computational Neuroscience 2020;14. doi: 10.3389/fncom.2020.00006

58. Nayebi A, Tipirneni S, Foreman B, Reddy CK, Subbian V. An Empirical Comparison of Explainable Artificial Intelligence Methods for Clinical Data: A Case Study on Traumatic Brain Injury. 2022; Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10148324/

59. Pathak S, Lu C, Nagaraj SB, van Putten M, Seifert C. STQS: Interpretable multi-modal Spatial-Temporal-seQuential model for automatic Sleep scoring. Artificial Intelligence in Medicine, publisher=Elsevier BV 2021 Apr;114, ISSN=0933-3657, url=http://dx.doi.org/10.1016/j.artmed.2021.102038. doi: 10.1016/j.artmed.2021.102038

60. Qiu S, Miller MI, Joshi PS, Lee JC, Xue C, Ni Y, Wang Y, de Anda-Duran I, Hwang PH, Cramer JA, Dwyer BC, Hao H, Kaku MC, Kedar S, Lee PH, Mian AZ, Murman DL, O'Shea S, Paul AB, Saint-Hilaire M-H, Alton Sartor E, Saxena AR, Shih LC, Small JE, Smith MJ, Swaminathan A, Takahashi CE, Taraschenko O, You H, Yuan J, Zhou Y, Zhu S, Alosco ML, Mez J, Stein TD, Poston KL, Au R, Kolachalama VB. Multimodal deep learning for Alzheimer's disease dementia assessment, volume=13. Nature Communications 2022 Jun;(1, journal=Nature Communications). doi: 10.1038/s41467-022-31037-5

61. Ruengchaijatuporn N, Chatnuntawech I, Teerapittayanon S, Sriswasdi S, Itthipuripat S, Hemrungrojn S, Bunyabukkana P, Petchlorlian A, Chunamchai S, Chotibut T, Chunharas C. An explainable self-attention deep neural network for detecting mild cognitive impairment using multi-input digital drawing tasks. 2022 Aug;14, ISSN=1758-9193(1, journal=Alzheimer's Research&Therapy). doi: 10.1186/s13195-022-01043-2

62. Sekaran K, Alsamman AM, George Priya Doss C, Zayed H. Bioinformatics investigation on blood-based gene expressions of Alzheimer's disease revealed ORAI2 gene biomarker susceptibility: An explainable artificial intelligence-based approach. Metabolic Brain Disease 2023 Feb;38, ISSN=1573-7365(4). doi: 10.1007/s11011-023-01171-0

63. Wang Y, Yang Y, Li S, Su Z, Guo J, Wei P, Huang J, Kang G, Zhao G. Automatic Localization of Seizure Onset Zone Based on Multi-Epileptogenic Biomarkers Analysis of Single-Contact from Interictal SEEG. Bioengineering 2022 Dec;9(12, journal=Bioengineering). doi: 10.3390/bioengineering9120769

64. Allahabadi H, Amann J, Balot I, Beretta A, Binkley C, Bozenhard J, Bruneault F, Brusseau J, Candemir S, Cappellini LA, Chakraborty S, Cherciu N, Cociancig C, Coffee M, Ek I, Espinosa-Leal L, Farina D, Fieux-Castagnet G, Frauenfelder T, Gallucci A, Giuliani G, Golda A, van Halem I, Hildt E, Holm S, Kararigas G, Krier SA, Kuhne U, Lizzi F, Madai VI, Markus AF, Masis S, Mathez EW, Mureddu F, Neri E, Osika W, Ozols M, Panigutti C, Parent B, Pratesi F, Moreno-Sanchez PA, Sartor G, Savardi M, Signoroni A, Sormunen H-M, Spezzatti A, Srivastava A, Stephansen AF, Theng LB, Tithi JJ, Tuominen J, Umbrello S, Vaccher F, Vetter D, Westerlund M, Wurth R, Zicari RV. Assessing Trustworthy AI in Times of COVID-19: Deep Learning for Predicting a Multiregional Score Conveying the Degree of Lung Compromise in COVID-19 Patients. IEEE Transactions on Technology and Society 2022;3(4):272–289. doi: 10.1109/tts.2022.3195114

65. Cavallaro M, Moiz H, Keeling MJ, McCarthy ND, Althouse BM. Contrasting factors associated with COVID-19-related ICU admission and death outcomes in hospitalised patients by means of Shapley values. PLOS Computational Biology 2021;17(6):e1009121. doi: 10.1371/journal.pcbi.1009121

66. Chadaga K, Prabhu S, Bhat V, Sampathila N, Umakanth S, Chadaga R. Artificial intelligence for diagnosis of mild–moderate COVID-19 using haematological markers. Annals of Medicine 2023;55(1). doi: 10.1080/07853890.2023.2233541

67. Khanna VV, Chadaga K, Sampathila N, Prabhu S, P. RC. A machine learning and explainable artificial intelligence triage-prediction system for COVID-19. Decision Analytics Journal 2023;7:100246. doi: 10.1016/j.dajour.2023.100246

68. Rostami M, Oussalah M. A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest. Informatics in Medicine Unlocked 2022;30:100941. doi: 10.1016/j.imu.2022.100941

69. Yagin FH, Cicek \.Ipek Balikci, Alkhateeb A, Yagin B, Colak C, Azzeh M, Akbulut S. Explainable artificial intelligence model for identifying COVID-19 gene biomarkers. Computers in Biology and Medicine 2023;154:106619. doi: 10.1016/j.compbiomed.2023.106619

70. Zucco AG, Agius R, Svanberg R, Moestrup KS, Marandi RZ, MacPherson CR, Lundgren J, Ostrowski SR, Niemann CU. Personalized survival probabilities for SARS-CoV-2 positive patients by explainable machine learning. Scientific Reports 2022;12(1). doi: 10.1038/s41598-022-17953-y

71. Alahmadi A, Davies A, Royle J, Goodwin L, Cresswell K, Arain Z, Vigo M, Jay C. An explainable algorithm for detecting drug-induced QT-prolongation at risk of torsades de pointes (TdP) regardless of heart rate and T-wave morphology. Computers in Biology and Medicine 2021;131:104281. doi: 10.1016/j.compbiomed.2021.104281

72. Elul Y, Rosenberg AA, Schuster A, Bronstein AM, Yaniv Y. Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning–based ECG analysis. Proceedings of the National Academy of Sciences 2021;118(24). doi: 10.1073/pnas.2020620118

73. Katsushika S, Kodera S, Sawano S, Shinohara H, Setoguchi N, Tanabe K, Higashikuni Y, Takeda N, Fujiu K, Daimon M, Akazawa H, Morita H, Komuro I. An explainable artificial intelligence-enabled electrocardiogram analysis model for the classification of reduced left ventricular function. European Heart Journal - Digital Health 2023;4(3):254–264. doi: 10.1093/ehjdh/ztad027

74. Kawakami M, Karashima S, Morita K, Tada H, Okada H, Aono D, Kometani M, Nomura A, Demura M, Furukawa K, Yoneda T, Nambo H, Kawashiri M. Explainable Machine Learning for Atrial Fibrillation in the General Population Using a Generalized Additive Model — A Cross-Sectional Study —. Circulation Reports 2022;4(2):73–82. doi: 10.1253/circrep.cr-21-0151

75. Moreno-Sánchez, Pedro A., year=2023, month=aug. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Frontiers in Cardiovascular Medicine 10, ISSN=2297-055X. doi: 10.3389/fcvm.2023.1219586

76. Interpretable heartbeat classification using local model-agnostic explanations on ECGs. Computers in Biology and Medicine 2021 Jun;133, ISSN=0010-4825. doi: 10.1016/j.compbiomed.2021.104393

77. Ogbomo-Harmitt S, Muffoletto M, Zeidan A, Qureshi A, King AP, Aslanidi O. Exploring interpretability in deep learning prediction of successful ablation therapy for atrial fibrillation. Frontiers in Physiology 2023 Mar;14, ISSN=1664-042X. doi: 10.3389/fphys.2023.1054401

78. Rouhi R, Clausel M, Oster J, Lauer F. An Interpretable Hand-Crafted Feature-Based Model for Atrial Fibrillation Detection. 2021 May;12, ISSN=1664-042X. doi: 10.3389/fphys.2021.657304,

79. Rousset A, Dellamonica D, Menuet R, Lira Pineda A, Sabatine MS, Giugliano RP, Trichelair P, Zaslavskiy M, Ricci L. Can machine learning bring cardiovascular risk assessment to the next level? A methodological study using FOURIER trial data. European Heart Journal - Digital Health 2021 Nov;3, ISSN=2634-3916. doi: 10.1093/ehjdh/ztab093,

80. Sakai A, Komatsu M, Komatsu R, Matsuoka R, Yasutomi S, Dozen A, Shozu K, Arakaki T,

Machino H, Asada K, Kaneko S, Sekizawa A, Hamamoto R. Medical Professional Enhancement Using Explainable Artificial Intelligence in Fetal Cardiac Ultrasound Screening. Biomedicines 2022;10(3):551. doi: 10.3390/biomedicines10030551

81. Storås AM, Andersen OE, Lockhart S, Thielemann R, Gnesin F, Thambawita V, Hicks SA, Kanters JK, Strümke I, Halvorsen P, Riegler MA. Usefulness of Heat Map Explanations for Deep-Learning-Based Electrocardiogram Analysis. Diagnostics 2023;13(14):2345. doi: 10.3390/diagnostics13142345

82. Sun R, Wang X, Jiang H, Yan Y, Dong Y, Yan W, Luo X, Miu H, Qi L, Huang Z. Prediction of 30-day mortality in heart failure patients with hypoxic hepatitis: Development and external validation of an interpretable machine learning model. Frontiers in Cardiovascular Medicine 2022;9. doi: 10.3389/fcvm.2022.1035675

83. Yang B, Zhu Y, Lu X, Shen C. A Novel Composite Indicator of Predicting Mortality Risk for Heart Failure Patients With Diabetes Admitted to Intensive Care Unit Based on Machine Learning. Frontiers in Endocrinology 2022;13. doi: 10.3389/fendo.2022.917838

84. Lolak S, Attia J, McKay GJ, Thakkinstian A. Comparing Explainable Machine Learning Approaches With Traditional Statistical Methods for Evaluating Stroke Risk Models: Retrospective Cohort Study. JMIR Cardio 2023;7:e47736. doi: 10.2196/47736

85. Zheng Y, Guo Z, Zhang Y, Shang J, Yu L, Fu P, Liu Y, Li X, Wang H, Ren L, Zhang W, Hou H, Tan X, Wang W. Rapid triage for ischemic stroke: a machine learning-driven approach in the context of predictive, preventive and personalised medicine. EPMA Journal 2022;13(2):285–298. doi: 10.1007/s13167-022-00283-4

86. Zignoli A. Machine Learning Models for the Automatic Detection of Exercise Thresholds in Cardiopulmonary Exercising Tests: From Regression to Generation to Explanation. Sensors 2023;23(2):826. doi: 10.3390/s23020826

87. Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, Frey D, Stoean R. Opening the black box of artificial intelligence for clinical decision support: A study predicting stroke outcome. PLOS ONE 2020;15(4):e0231166. doi: 10.1371/journal.pone.0231166

88. Hegselmann S, Ertmer C, Volkert T, Gottschalk A, Dugas M, Varghese J. Development and validation of an interpretable 3 day intensive care unit readmission prediction model using explainable boosting machines. Frontiers in Medicine 2022;9. doi: 10.3389/fmed.2022.960296

89. Rivera AJ, Muñoz JC, Pérez-Goody MD, de San Pedro BS, Charte F, Elizondo D, Rodríguez C, Abolafia ML, Perea A, Del Jesus MJ. XAIRE: An ensemble-based methodology for determining the relative importance of variables in regression tasks. Application to a hospital emergency department. Artificial Intelligence in Medicine 2023;137:102494. doi: 10.1016/j.artmed.2023.102494

90. Salman S, Payrovnaziri SN, Liu X, Rengifo-Moreno P, He Z. DeepConsensus: Consensus-based Interpretable Deep Neural Networks with Application to Mortality Prediction. 2020 International Joint Conference on Neural Networks (IJCNN) Institute of Electrical and Electronics Engineers (IEEE); 2020. p. 1–8. doi: 10.1109/ijcnn48605.2020.9206678

91. Shashikumar SP, Josef CS, Sharma A, Nemati S. DeepAISE – An interpretable and recurrent neural survival model for early prediction of sepsis. 2021 Mar;113, ISSN=0933-3657. doi: 10.1016/j.artmed.2021.102036,

92. Stenwig E, Salvi G, Rossi PS, Skjærvold NK. Comparative analysis of explainable machine learning prediction models for hospital mortality. BMC Medical Research Methodology 2022 Feb;22, ISSN=1471-2288, url=http://dx.doi.org/10.1186/s12874-022-01540-w. doi: 10.1186/s12874-022-01540-w,

93. Wang Z, Zhang L, Huang T, Yang R, Cheng H, Wang H, Yin H, Lyu J. Developing an explainable machine learning model to predict the mechanical ventilation duration of patients with ARDS in intensive care units. Heart & Lung 2023;58:74–81. doi: 10.1016/j.hrtlng.2022.11.005

94. Zhang L, Wang Z, Zhou Z, Li S, Huang T, Yin H, Lyu J. Developing an ensemble machine learning model for early prediction of sepsis-associated acute kidney injury. iScience 2022;25(9):104932. doi: 10.1016/j.isci.2022.104932

95. Magrelli S, Valentini P, de Rose C, Morello R, Buonsenso D. Classification of Lung Disease in Children by Using Lung Ultrasound Images and Deep Convolutional Neural Network. Frontiers in Physiology 2021;12. doi: 10.3389/fphys.2021.693448

96. Zeng S, Arjomandi M, Luo G. Automatically Explaining Machine Learning Predictions on Severe Chronic Obstructive Pulmonary Disease Exacerbations: Retrospective Cohort Study. JMIR Medical Informatics 2022;10(2):e33043. doi: 10.2196/33043

97. Rajaraman S, Candemir S, Kim I, Thoma G, Antani S. Visualization and Interpretation of Convolutional Neural Network Predictions in Detecting Pneumonia in Pediatric Chest Radiographs. Applied Sciences 2018;8(10):1715. doi: 10.3390/app8101715

98. Liu Y-C, Cheng H-Y, Chang T-H, Ho T-W, Liu T-C, Yen T-Y, Chou C-C, Chang L-Y, Lai F. Evaluation of the Need for Intensive Care in Children With Pneumonia: Machine Learning Approach. JMIR Medical Informatics 2022;10(1):e28934. doi: 10.2196/28934

99. Altini N, Brunetti A, Puro E, Taccogna MG, Saponaro C, Zito FA, de Summa S, Bevilacqua V. NDG-CAM: Nuclei Detection in Histopathology Images with Semantic Segmentation Networks and Grad-CAM. Bioengineering 2022;9(9):475. doi: 10.3390/bioengineering9090475

100. Sabol P, Sinčák P, Hartono P, Kočan P, Benetinová Z, Blichárová A, Verbóová Ľ, Štammová E, Sabolová-Fabianová A, Jašková A. Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. Journal of Biomedical Informatics 2020;109:103523. doi: 10.1016/j.jbi.2020.103523

101. Aldughayfiq B, Ashfaq F, Jhanjhi NZ, Humayun M. Explainable AI for Retinoblastoma Diagnosis: Interpreting Deep Learning Models with LIME and SHAP. Diagnostics 2023;13(11):1932. doi: 10.3390/diagnostics13111932

102. Gu D, Su K, Zhao H. A case-based ensemble learning system for explainable breast cancer recurrence prediction. Artificial Intelligence in Medicine 2020;107:101858. doi:

10.1016/j.artmed.2020.101858

103. Huang W, Suominen H, Liu T, Rice G, Salomon C, Barnard AS. Explainable discovery of disease biomarkers: The case of ovarian cancer to illustrate the best practice in machine learning and Shapley analysis. Journal of Biomedical Informatics 2023;141:104365. doi: 10.1016/j.jbi.2023.104365

104. Li J, Zhu Y, Dong Z, He X, Xu M, Liu J, Zhang M, Tao X, Du H, Di Chen, Huang L, Shang R, Zhang L, Luo R, Zhou W, Deng Y, Huang X, Li Y, Chen B, Gong R, Zhang C, Li X, Wu L, Yu H. Development and validation of a feature extraction-based logical anthropomorphic diagnostic system for early gastric cancer: A case-control study. eClinicalMedicine 2022;46:101366. doi: 10.1016/j.eclinm.2022.101366

105. Sathyan A, Weinberg AI, Cohen K. Interpretable AI for bio-medical applications. Complex Engineering Systems 2022;2(4):18. doi: 10.20517/ces.2022.41

106. Gu R, Wang G, Song T, Huang R, Aertsen M, Deprest J, Ourselin S, Vercauteren T, Zhang S. CA-Net: Comprehensive Attention Convolutional Neural Networks for Explainable Medical Image Segmentation. IEEE Transactions on Medical Imaging 2021;40(2):699–711. doi: 10.1109/tmi.2020.3035253

107. Liu X, Morelli D, Littlejohns TJ, Clifton DA, Clifton L. Combining machine learning with Cox models to identify predictors for incident post-menopausal breast cancer in the UK Biobank. Scientific Reports 2023 Jun;13(1). doi: 10.1038/s41598-023-36214-0

108. Dolezal JM, Wolk R, Hieromnimon HM, Howard FM, Srisuwananukorn A, Karpeyev D, Ramesh S, Kochanny S, Kwon JW, Agni M, Simon RC, Desai C, Kherallah R, Nguyen TD, Schulte JJ, Cole K, Khramtsova G, Garassino MC, Husain AN, Li H, Grossman R, Cipriani NA, Pearson AT. Deep learning generates synthetic cancer histology for explainability and education. npj Precision Oncology 2023;7(1). doi: 10.1038/s41698-023-00399-4

109. Fu H, Mi W, Pan B, Guo Y, Li J, Xu R, Zheng J, Zou C, Zhang T, Liang Z, Zou J, Zou H. Automatic Pancreatic Ductal Adenocarcinoma Detection in Whole Slide Images Using Deep Convolutional Neural Networks. Frontiers in Oncology 2021;11. doi: 10.3389/fonc.2021.665929

110. Krishna S, Suganthi SS, Bhavsar A, Yesodharan J, Krishnamoorthy S. An interpretable decision-support model for breast cancer diagnosis using histopathology images. Journal of Pathology Informatics 2023;14:100319. doi: 10.1016/j.jpi.2023.100319

111. Palatnik de Sousa, Iam, Maria Bernardes Rebuzzi Vellasco, Marley, Costa da Silva, Eduardo. Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases. Sensors 2019;19(13):2969. doi: 10.3390/s19132969

112. Peyster EG, Arabyarmohammadi S, Janowczyk A, Azarianpour-Esfahani S, Sekulic M, Cassol C, Blower L, Parwani A, Lal P, Feldman MD, Margulies KB, Madabhushi A. An automated computational image analysis pipeline for histological grading of cardiac allograft rejection. European Heart Journal 2021;42(24):2356–2369. doi: 10.1093/eurheartj/ehab241

113. Sauter D, Lodde G, Nensa F, Schadendorf D, Livingstone E, Kukuk M. Validating Automatic Concept-Based Explanations for AI-Based Digital Histopathology. Sensors 2022;22(14):5346. doi: 10.3390/s22145346

114. Bifarin OO, Ashraf I. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. PLOS ONE 2023 May;18(5). doi: 10.1371/journal.pone.0284315

115. Ditz JC, Reuter B, Pfeifer N. COmic: convolutional kernel networks for interpretable end-to-end learning on (multi-)omics data. Bioinformatics 2023 Jun;39, ISSN=1367-4811(Supplement_1). doi: 10.1093/bioinformatics/btad204

116. Shen WX, Liu Y, Chen Y, Zeng X, Tan Y, Jiang YY, Chen YZ. AggMapNet: enhanced and explainable low-sample omics deep learning with feature-aggregated multi-channel networks. Nucleic Acids Research 2022;50(8):e45–e45. doi: 10.1093/nar/gkac010

117. Tan, Cedric Chih Shen, Acman M, van Dorp L, Balloux F. Metagenomic evidence for a polymicrobial signature of sepsis. Microbial Genomics 2021;7(9). doi: 10.1099/mgen.0.000642

118. Lucieri A, Dengel A, Ahmed S. Translating theory into practice: assessing the privacy implications of concept-based explanations for biomedical AI. Frontiers in Bioinformatics 2023;3. doi: 10.3389/fbinf.2023.1194993

119. Preston FG, Meng Y, Burgess J, Ferdousi M, Azmi S, Petropoulos IN, Kaye S, Malik RA, Zheng Y, Alam U. Artificial intelligence utilising corneal confocal microscopy for the diagnosis of peripheral neuropathy in diabetes mellitus and prediabetes. Diabetologia 2021;65(3):457–466. doi: 10.1007/s00125-021-05617-x

120. Quellec G, Al Hajj H, Lamard M, Conze P-H, Massin P, Cochener B. ExplAIn: Explanatory artificial intelligence for diabetic retinopathy diagnosis. Medical Image Analysis 2021;72:102118. doi: 10.1016/j.media.2021.102118

121. Singh A, Jothi Balaji J, Rasheed MA, Jayakumar V, Raman R, Lakshminarayanan V. Evaluation of Explainable Deep Learning Methods for Ophthalmic Diagnosis. Clinical Ophthalmology 2021;Volume 15:2573–2581. doi: 10.2147/opth.s312236

122. Yoo TK, Ryu IH, Choi H, Kim JK, Lee IS, Kim JS, Lee G, Rim TH. Explainable Machine Learning Approach as a Tool to Understand Factors Used to Select the Refractive Surgery Technique on the Expert Level. Translational Vision Science & Technology 2020;9(2):8. doi: 10.1167/tvst.9.2.8

123. Erdeniz SP, Schrempf M, Kramer D, Rainer PP, Felfernig A, Tran T, Burgstaller T, Lubos S. Computational Evaluation of Model-Agnostic Explainable AI Using Local Feature Importance in Healthcare. p. 114–119.

124. Hatwell J, Gaber MM, Atif Azad RM. Ada-WHIPS: explaining AdaBoost classification with applications in the health sciences. BMC Medical Informatics and Decision Making 2020 Oct;20, ISSN=1472-6947. doi: 10.1186/s12911-020-01201-2,

125. Woodcock C, Mittelstadt B, Busbridge D, Blank G. The Impact of Explanations on

Layperson Trust in Artificial Intelligence–Driven Symptom Checker Apps: Experimental Study. Journal of Medical Internet Research 2021 Nov;23, ISSN=1438-8871:e29386. doi: 10.2196/29386,

126. Kovalchuk SV, Kopanitsa GD, Derevitskii IV, Matveev GA, Savitskaya DA. Three-stage intelligent support of clinical decision making for higher trust, validity, and explainability. Journal of Biomedical Informatics 2022;127:104013. doi: 10.1016/j.jbi.2022.104013

127. Resta M, Monreale A, Bacciu D. Occlusion-Based Explanations in Deep Recurrent Models for Biomedical Signals. Entropy 2021;23(8):1064. doi: 10.3390/e23081064

128. Rietberg MT, van Nguyen B, Geerdink J, Vijlbrief O, Seifert C. Accurate and Reliable Classification of Unstructured Reports on Their Diagnostic Goal Using BERT Models. Diagnostics 2023;13(7):1251. doi: 10.3390/diagnostics13071251

129. Roder J, Maguire L, Georgantas R, Roder H. Explaining multivariate molecular diagnostic tests via Shapley values. BMC Medical Informatics and Decision Making 2021;21(1). doi: 10.1186/s12911-021-01569-9

130. Lenatti M, Carlevaro A, Guergachi A, Keshavjee K, Mongelli M, Paglialonga A, Ijaz MF. A novel method to derive personalized minimum viable recommendations for type 2 diabetes prevention based on counterfactual explanations. PLOS ONE 2022;17(11):e0272825. doi: 10.1371/journal.pone.0272825

131. Magboo, Ma. Sheila A., Magboo, Vincent Peter C. Feature Importance Measures as Explanation for Classification Applied to Hospital Readmission Prediction. Procedia Computer Science 2022;207:1388–1397. doi: 10.1016/j.procs.2022.09.195

132. Xie P, Yang C, Yang G, Jiang Y, He M, Jiang X, Chen Y, Deng L, Wang M, Armstrong DG, Ma Y, Deng W. Mortality prediction in patients with hyperglycaemic crisis using explainable machine learning: a prospective, multicentre study based on tertiary hospitals. Diabetology & Metabolic Syndrome 2023;15(1). doi: 10.1186/s13098-023-01020-1

133. Zhang Y, Zhang X, Razbek J, Li D, Xia W, Bao L, Mao H, Daken M, Cao M. Opening the black box: interpretable machine learning for predictor finding of metabolic syndrome. BMC Endocrine Disorders 2022;22(1). doi: 10.1186/s12902-022-01121-4

134. Jiang D, Wu Z, Hsieh C-Y, Chen G, Liao B, Wang Z, Shen C, Cao D, Wu J, Hou T. Could graph neural networks learn better molecular representation for drug discovery? A comparison study of descriptor-based and graph-based models. Journal of Cheminformatics 2021;13(1). doi: 10.1186/s13321-020-00479-8

135. Martínez-Agüero S, Soguero-Ruiz C, Alonso-Moral JM, Mora-Jiménez I, Álvarez-Rodríguez J, Marques AG. Interpretable clinical time-series modeling with intelligent feature selection for early prediction of antimicrobial multidrug resistance. Future Generation Computer Systems 2022;133:68–83. doi: 10.1016/j.future.2022.02.021

136. Zhu X, Hu J, Xiao T, Huang S, Wen Y, Shang D. An interpretable stacking ensemble learning framework based on multi-dimensional data for real-time prediction of drug concentration: The example of olanzapine. Frontiers in Pharmacology 2022;13. doi:

10.3389/fphar.2022.975855

137. Barda AJ, Horvat CM, Hochheiser H. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. BMC Medical Informatics and Decision Making 2020;20(1). doi: 10.1186/s12911-020-01276-x

138. Moreno-Sánchez PA. Data-Driven Early Diagnosis of Chronic Kidney Disease: Development and Evaluation of an Explainable AI Model. IEEE Access 2023;11:38359–38369. doi: 10.1109/access.2023.3264270

139. Dindorf C, Ludwig O, Simon S, Becker S, Fröhlich M. Machine Learning and Explainable Artificial Intelligence Using Counterfactual Explanations for Evaluating Posture Parameters. Bioengineering 2023;10(5):511. doi: 10.3390/bioengineering10050511

140. Ferrer-Sánchez A, Bagan J, Vila-Francés J, Magdalena-Benedito R, Bagan-Debon L. Prediction of the risk of cancer and the grade of dysplasia in leukoplakia lesions using deep learning. Oral Oncology 2022;132:105967. doi: 10.1016/j.oraloncology.2022.105967

141. Oztekin F, Katar O, Sadak F, Yildirim M, Cakar H, Aydogan M, Ozpolat Z, Talo Yildirim T, Yildirim O, Faust O, Acharya UR. An Explainable Deep Learning Model to Prediction Dental Caries Using Panoramic Radiograph Images. Diagnostics 2023;13(2):226. doi: 10.3390/diagnostics13020226

142. Singh RK, Gorantla R, Allada, Sai Giridhar Rao, Narra P, Qiu Y. SkiNet: A deep learning framework for skin lesion diagnosis with uncertainty estimation and explainability. PLOS ONE 2022;17(10):e0276836. doi: 10.1371/journal.pone.0276836

143. Dragoni M, Donadello I, Eccher C. Explainable AI meets persuasiveness: Translating reasoning results into behavioral change advice. Artificial Intelligence in Medicine 2020;105:101840. doi: 10.1016/j.artmed.2020.101840

144. Brankovic A, Hendrie GA, Baird DL, Khanna S. Predicting Disengagement to Better Support Outcomes in a Web-Based Weight Loss Program Using Machine Learning Models: Cross-Sectional Study. Journal of Medical Internet Research 2023;25:e43633. doi: 10.2196/43633

145. Mishra S, Tripathy HK, Kumar Thakkar H, Garg D, Kotecha K, Pandya S. An Explainable Intelligence Driven Query Prioritization Using Balanced Decision Tree Approach for Multi-Level Psychological Disorders Assessment. Frontiers in Public Health 2021;9. doi: 10.3389/fpubh.2021.795007

146. Twick I, Zahavi G, Benvenisti H, Rubinstein R, Woods MS, Berkenstadt H, Nissan A, Hosgor E, Assaf D. Towards interpretable, medically grounded, EMR-based risk prediction models. Scientific Reports 2022;12(1). doi: 10.1038/s41598-022-13504-7

147. Kiyasseh D, Laca J, Haque TF, Miles BJ, Wagner C, Donoho DA, Anandkumar A, Hung AJ. A multi-institutional study using artificial intelligence to provide reliable and fair feedback to surgeons. Communications Medicine 2023;3(1). doi: 10.1038/s43856-023-00263-3

148. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. “Hello AI”: Uncovering the Onboarding

Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. Proceedings of the ACM on Human-Computer Interaction 2019;3(CSCW):1–24. doi: 10.1145/3359206

149. Liao QV, Gruen D, Miller S. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. 2020; doi: 10.48550/ARXIV.2001.02478

150. Haque ABahalulKM, Islam AKMN, Mikalef P. Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. Technological Forecasting and Social Change 2023;186:122120. doi: 10.1016/j.techfore.2022.122120