Pflügers Archiv - European Journal of Physiology (2025) 477:513–529
https://doi.org/10.1007/s00424-025-03067-7

INVITED REVIEW

Current methods in explainable artificial intelligence and future prospects for integrative physiology

Bettina Finzel[1]

Received: 1 July 2024 / Revised: 14 January 2025 / Accepted: 15 January 2025 / Published online: 25 February 2025

Abstract

Explainable artificial intelligence (XAI) is gaining importance in physiological research, where artificial intelligence is now used as an analytical and predictive tool for many medical research questions. The primary goal of XAI is to make AI models understandable for human decision-makers. This can be achieved in particular through providing inherently interpretable AI methods or by making opaque models and their outputs transparent using post hoc explanations. This review introduces XAI core topics and provides a selective overview of current XAI methods in physiology. It further illustrates solved and discusses open challenges in XAI research using existing practical examples from the medical field. The article gives an outlook on two possible future prospects: (1) using XAI methods to provide trustworthy AI for integrative physiological research and (2) integrating physiological expertise about human explanation into XAI method development for useful and beneficial human-AI partnerships.

Keywords Explainable Artificial Intelligence (XAI) · Physiology · Explainability · Interpretability · Survey

Introduction

Physiology is a diverse field of research within human medicine and beyond. According to Lemoine and Pradeu [57], physiology is an integrative, explanatory science dedicated to overarching phenomena in living beings that are seen as normal (healthy) or pathological (indicative of disease). The British physiologist Noble once pointed out that "physiological analysis requires an understanding of the functional interactions between the key components of cells, organs, and systems, as well as how these interactions change in disease states" [86]. This means physiology is taking a holistic view on living organisms.

Although physiology is a holistic science, it seems that physiological research has not yet exploited the full potential of new technological trends, such as explainable artificial intelligence (XAI) [59], to integrate existing knowledge with new data-intensive analytics. This potential could be utilized in the many research fields that apply physiological methods. This includes research on cardiovascular diseases [33, 49, 84], immunology and homeostasis [102], neural physiology, e.g., for research on Alzheimer's disease [22] and depression [104], physiology of aging [140] and reproduction [117], cell biology [34] as well as the physiology of nutrition, e.g., in research on human adipogenesis [85], diabetes [141], and effects of diet on cardiovascular diseases [33]. Last but not least, physiology plays a crucial role in tumor detection and diagnosis, e.g., based on histopathological findings using either traditional methods (e.g., TMN-based classification) or modern methods (e.g., immunoscore) [57].

Physiology therefore offers a wide range of applications for artificial intelligence (AI), a set of methods and algorithms including techniques that learn models from data (machine learning), perform automated planning, or provide knowledge representation for automated processing [79, 127]. Most of the AI algorithms currently in use belong to machine learning methods. These can be used for example for regression tasks, classification, or clustering [79, 127]. Their approach is predictive rather than descriptive or functional.

AI is already being used in various areas of physiological research, particularly in oncology, e.g., for the detection and classification of breast cancer [27] and colorectal can-

This article is part of the special issue on Artificial Intelligence in Pflügers Archiv-European Journal of Physiology.

✉ Bettina Finzel
bettina.finzel@uni-bamberg.de

[1] Cognitive Systems, University of Bamberg, Weberei 5, 96047 Bamberg, Germany

cer [95]. AI is also applied for the analysis of cardiovascular conditions, e.g., the detection of systolic dysfunction [29] or in neurophysiology, for tasks like the detection of epileptic seizures [13] and stroke prediction [21].

The vast majority of current AI approaches use so-called deep neural networks (DNN), which offer high-performance architectures but remain opaque to humans in their decisions due to their complexity [87, 126, 131].

Explainable artificial intelligence (XAI), as a field of research and at the same time a term for methods that are either inherently interpretable or represent an interpretable extension of opaque methods, aims to counteract this intransparency [131]. XAI is particularly important for the integration of humans in decision-making processes [87]. XAI methods create understanding in recipients of explanations about the behavior and outcomes produced by AI models, e.g., why a certain cancer type was classified in a given microscopy image [87, 126, 130]. Thus, XAI methods serve experts for the validation of AI models, in particular, to examine whether these models have produced results for valid reasons.

AI models depend on input data and may therefore be prone to noise, errors, or incompleteness in data [87]. In some medical areas, such problems can be amplified by small data sets that cannot be representative of large populations, for example, in histopathology, where false tissue annotations in whole slide images, along with very diverse tissue constellations, can lead to limited generalization in AI models [87]. At the same time, XAI methods can help novices to understand the domain of interest with the help of an AI model, given that it was evaluated as representative by experts.

In physiology and its related medical tasks, human-centered XAI can help to integrate human knowledge into AI models as well as to extract human-understandable concepts from AI models [80, 87, 106]. Established rules (such as TMN cancer classification rules [72]) can be integrated in the process of learning AI models to provide human guidance [126, 130]. Analyzing internal representations of AI models and providing human-understandable labels for them can serve as a first approximation of human knowledge [87, 106]. All in all, putting the human into an explanatory dialogue with AI could lead to improved collaboration and better medical decision outcomes [56, 87].

This review article aims for providing an overview on common XAI topics and methods. It therefore contributes a collection of 85 publications that have been retrieved from a systematic review of 200 papers and articles from the fields of medicine and artificial intelligence with a special focus on physiology. It further illustrates solved and discusses open challenges in XAI research using existing practical examples from the medical field. Based on the extensive review, the article provides a short analysis of the current state of XAI usage in physiology and gives an outlook on two possible future prospects: (1) using XAI methods to provide trustworthy AI for integrative physiological research and (2) integrating physiological expertise about human explanation into XAI method development for useful and beneficial human-AI partnerships. Ultimately, this review article's goal is to show how XAI could help to meet the integrative demands of physiological research despite the use of data-intensive (opaque) artificial intelligence methods. To the best of my knowledge, this is the first review article that focuses on XAI for physiology and which discusses the potential of XAI to realize integrative physiology.

The review is structured as follows: first, the research area of XAI is introduced with its specific terminology in Section 2. Then, the methodology behind the review of recent publications that utilize XAI in the field of physiology is briefly explained in Section 3. Subsequently, the results of the research are presented and discussed in Section 4. To illustrate common XAI methods and for the discussion of open challenges, examples of XAI methods are shown afterwards, in particular, for medical applications in physiology (see Section 5). The discussion and conclusion Sections (6 and 7) explore and summarize how physiology and XAI as research areas could mutually benefit from each other in the future.

Explainable artificial intelligence (XAI)

In general, explanatory methods help to create understanding in recipients of explanations (the explainees) about the subject or object that is to be explained (the explanandum) with the help of phenomena or concepts the explanation is based on (the explanans) [57, 126, 131]. In the XAI literature, the component that provides explanations for opaque AI models is often called an XAI method or explainer [131].

According to Carvalho et al. [92] and Schwalbe and Finzel [131], the term explainable artificial intelligence (XAI) has been first used by van Lent et al. in 2004 [59]. The term refers to AI that is capable of explaining itself or that is enhanced by methods that make it understandable to humans. Thus, XAI is the area of research concerned with explaining an AI system's decision [131]. An idea of this kind has already been coined in 1958 when computer scientist McCarthy proposed inherently transparent AI systems using a formal language and problem-solving algorithms [71, 125].

Early transparent methods led to expert systems research in the 1970s and 1980s until the late 1990s [19], which later lost popularity due to integration challenges. Meanwhile, artificial neural networks became prominent, with efforts to make their decisions transparent beginning in the mid-1990s [31].

The introduction of the GDPR in 2018 emphasized the need for transparency in AI, leading to new frameworks that prioritize accountability and the "right of explanation" [14,

91]. XAI gained further attention after Gunning and Aha's publication on explainable AI in 2019, which outlined two main goals: creating interpretable models without sacrificing performance and adopting a user-centric approach to enhance human understanding and trust in AI [32].

Significant contributions to the user-centric perspective include Miller's 2019 paper on social science aspects of AI [78] and Rudin's advocacy for interpretable models [123]. The XAI community now also focuses on evaluating the quality of explanations, recognizing the importance of formalized assessment metrics [75, 122]. Most recent works provide methods that try to map internal representations in AI models to human-understandable concepts [10, 106, 119] and create explanatory dialogues to allow for multi-faceted interaction and explanation processes [56, 87].

A specific terminology has been established in the XAI community for the categorization of XAI methods [75, 131]. It is briefly introduced in the following list of terms. Note that this review does not aim for providing a comprehensive survey of existing XAI methods. Instead, its goal is to point to current XAI usage in physiological research and to open challenges with respect to integrating explanations with human-centered AI in medical fields, such as physiology.

For a more detailed overview, the interested reader may consider, for example, the recent survey paper by Schwalbe and Finzel (2023) that introduces an XAI taxonomy and presents a collection of over 50 XAI methods [131] and the survey article by Ali et al. [75] that complements XAI topics and methods with their formalization. The following list is an excerpt of terminology adapted from Schwalbe and Finzel [131].

- **Interpretability:** Involves the recognition of constituents of an explanandum and the assignment of meaning to such elements.
- **Explainability:** Refers to the creation of understanding about an explanandum.
- **Model-agnostic explanations:** An explainer produces model-agnostic explanations if it is applicable to any type of model (any type of explanandum).
- **Model-specific explanations:** An explainer produces model-specific explanations if it is applicable to only a certain type of model (a certain type of explanandum). The reason is that such an explainer uses internal model parameters or internal representations to produce explanations.
- **Global explanations:** Provide insights into the reasons for an overall behavior of a model. The explanandum is the model itself. The explanans may be parts of the model or some approximation of the whole model.
- **Local explanations:** Provide insights into the reasons for an outcome of applying a model to input data. The explanandum is the outcome of a model's application to an individual instance. The explanans may be properties of the input instance (its features) and their influence on the model's output.
- **Model induction:** This describes the approach of generating a surrogate model based on inputs, constituents, or outputs of another model in order to approximate the explanandum. This term was specifically coined by Gunning and Aha [32].
- **Deep explanation:** This describes the approach of generating explanations based on constituents or outputs of a model, the explanandum, that is usually a very complex and opaque model such as DNNs. This term was coined by Gunning and Aha [32].
- **Interpretable models:** Such models are constructed from interpretable parts and thus usually inherently transparent (see, for example, [32, 123]).
- **Ante hoc explanations:** Explanations are produced during model creation. This applies to interpretable AI models. The model itself can be used to generate global or local explanations.
- **Post hoc explanations:** Explanations are produced after model creation. This applies to opaque AI models which are made transparent with the help of an external explainer. The explainer is used to generate global or local explanations either by accessing internal representations of the model (deep explanation) or by approximating the model's behavior (surrogate model).
- **Multimodal explanations:** Such explanations are not expressed via a singular modality, e.g., text. Instead, different representations are used (e.g., text in combination with images).
- **Human-centered explanations:** Such explanations address the varying information needs and individual characteristics of human explainees.

This list of terms is not complete as there exists more nuanced terminology for characterizing specialized explainability techniques as well as synonyms that can not be covered altogether in this article. For example, *data explainability* [75] is mentioned as another aspect in the literature to describe methods that enable reasoning on data (e.g., knowledge graphs) or allow the data's direct characterization (e.g., summarization methods). The term *ante hoc explanations* used in this article is also called *model explainability* (see Ali et al. [75]) in other works. As this article does not discuss explanation evaluation methods, the interested reader may find a comprehensive overview in Schwalbe and Finzel [131] and a concise summary of explanation assessment methods in Ali et al. [75].

In the following sections, the recent works applying XAI to physiological use cases will be collected, partially illus-

trated, and discussed with respect to their integrative nature. First, the methodology behind the search is shortly introduced.

Review methodology

For this review paper, relevant journals (with impact factor 2.0 or higher) and conference proceedings from the fields of artificial intelligence and medicine with at least an h-index of 10 were searched with the keywords "explainable artificial intelligence, XAI, explainability, interpretability, physiology." Common platforms (Google Scholar and PubMed) were considered for paper retrieval. Only papers published during the years 2020 to 2024 were included (first considered day, 1.1.2020; last considered day, 1.11.2024). As a further criterion, publications had to have been cited at least *5 times* in the last three years and publications from the past five years at least *10 times* (considering only the year of publication). Preprints and workshop papers were excluded from search results. Papers that referenced applications of XAI in physiology, but did not provide any research outcomes in this area (a survey, method or experimental results), were excluded as well. From the list of results found by PubMed and Google Scholar, only the first 100 entries (sorted by relevance in Google Scholar) have been reviewed. The keywords have been searched with a query using operators. The query was formulated as "(explainable artificial intelligence OR XAI OR explainability OR interpretability) AND physiology" to allow for synonyms and related XAI terms being equally important and to emphasize on physiology.

Review results for XAI in physiology

All retrieved and included papers and articles have either a strong focus on XAI with mentions of physiology or a strong focus on physiology with applications or substantial discussions of XAI.

Querying the data bases with the aforementioned query generated a list of roughly 24,700 results as denoted by Google Scholar and a list of 51,591 results in PubMed.

From the 100 first entries returned by Google Scholar, 60 met the criteria for the minimum number of citations and for the impact factor of the respective journal or the h-index of the conference. From the 100 first entries returned by PubMed, 53 publications met the citation criterion as well as the minimum impact factor or h-index for journals or conferences. The content was reviewed for its depth in XAI focus or physiology. This left 85 articles, of which 18 were survey papers and 67 presented at least one XAI method.

All selected publications are organized in Table 1. It sorts the alphabetically ordered articles according to (1) their article type being a survey or technical contribution using an XAI method, (2) the XAI method(s) they present[1], (3) the journal or conference they were published in[2], and (4) the content focus in medicine (either on physiology or more general with considerable mentions of physiology as application).

Table 1 contains a considerably large number of methodological publications and some survey articles. The majority of the methodological works utilize existing XAI methods, rather than introducing novel ones.

The collection of publications shows neither a prominent research field in terms of journals and conferences, nor a major topic in terms of application areas. The research fields covered in journals and conferences range from multidisciplinary (e.g., IEEE Access), medical (e.g., The Lancet), and microbiological research (e.g., Frontiers in Microbiology) to general, application-oriented research (e.g., Journal of Healthcare Informatics Research).

The publications with a rather broad and general focus and in which physiology plays a subordinate role (denoted by G) cover the topics of health (5 times), mental health (2 times), biology (1 time), and oncology (1 time). The general focus is specified accordingly in Table 1. One of the broad publications (Mei et al. [76]) has a technical focus and was therefore not specified in more detail.

The publications that focus on physiology (P) primarily deal with intensive care medicine (ICU, 5 times), genetics (Genes, 5 times), various forms and measuring instruments of age diagnostics (Age, 4 times), diseases caused by the coronavirus (COVID, 4 times), pain physiology (Pain, 3 times), various biosignals (Biosign., 2 times), Alzheimer's dementia (AD, 2 times), electromyography (EMG, 2 times), sepsis (2 times), and toxicity (2 times).

Further specialist areas are indicated in the table as corresponding focus topics within physiology. At this point, only those are explicitly mentioned from which the focus—without special prior knowledge—cannot be read directly from the abbreviations used in Table 1. These include functional near-infrared spectroscopy (fNRS), gene expression data (GED), heart rate variability (HRV), the integral transmembrane protein aquaporin-4 associated with amyloid burden in aging brains (AQP4), single nucleotide polymorphism in genetics (SNP), whole slide imaging in (histo-)pathology (WSI), non-communicable diseases prediction (NCD), Crohn's disease (CD), whole exome sequencing for leukemia (WES), electroencephalography (EEG), obstructive sleep apnea/hypopnea syndrome (OSAHS), cardiac arrest (CA), peripheral blood mononuclear cells as breast cancer biomarkers (PBMC), hepatocellular carcinoma

[1] Note that for survey articles, the full list is omitted and instead denoted with *various*.

[2] Note that long journal or conference names are presented in abbreviated form for space reasons.

Table 1 Publications (S, survey; M, method) retrieved from query "(explainable artificial intelligence OR XAI OR explainability OR interpretability) AND physiology" with focus (P, physiology; G, general)

#	Reference	Publ. Type	XAI Methods	Published in	Focus
1	Alabdulhafith et al. [97]	M	SHAP, LIME	IEEE Access	P (ICU)
2	Andreu-Perez et al. [143]	M	xMVPA	Commun Biol	P (fNIRS)
3	Anguita-Ruiz et al. [134]	M	Association Rules	PLoS Comput Biol	P (GED)
4	Banerjee et al. [42]	M	SHAP	SN Comp Sci	P (HRV)
5	Beer et al. [26]	M	SHAP	Neurobiol Aging	P (AQP4)
6	Bernard et al. [110]	M	SHAP	Aging Cell	P (PPAge)
7	Boscolo Gal. et al. [139]	S	Various	IEEE Signal Process Mag	P (BrainAge)
8	Boulesteix et al. [96]	M	Regression	Hum Genet	P (SNP)
9	Chan et al. [99]	M	SHAP, PDA, LIME	BMC Med Inform Decis Mak	P (ICU)
10	Chen et al. [54]	M	Integrated Gradients	Cancer Cell	P (WSI)
11	Chen & Chiu [52]	M	Fuzzy Geom. Mean	Patterns	P (Covid)
12	Chen et al. [6]	S	Various	Digital Health	P (NeurImag)
13	Davagdorj et al. [118]	M	DeepSHAP	IEEE Access	P (NCD)
14	Dindorf et al. [93]	M	LIME, SHAP, DeepLift	Sensors	P (Posture)
15	El-Sappagh et al. [137]	M	SHAP	Sci Rep	P (AD)
16	Fellous et al. [16]	S	Various	Front Neurosci	P (NeuroStim)
17	Gao et al. [23]	M	SHAP	Gut Microbes	P (CD)
18	Gimeno et al. [43]	M	MOM	Front Immunol	P (WES)
19	Goodwin et al. [113]	M	SHAP	Nature Neurosci	P (Behavior)
20	Górriz et al. [112]	S	Various	Inf. Fusion	P (Emotion)
21	Gouverneur et al. [3]	M	Grad-CAM	Sensors	P (Pain)
22	Han et al. [89]	M	SHAP, DCA	IEEE ICME Conf.	P (Pain)
23	Hasan et al. [2]	M	SHAP, PDA	Comput Methods Programs Biomed.	P (Driving)
24	He et al. [4]	M	SHAP, PDA	Ecol Ind	P (Seagrass)
25	Hijazi et al. [5]	M	LIME	Sensors	P (Covid)
26	Hossain et al. [11]	S	Various	ACM Comput Surv.	G (Health)
27	Hussain & Jany [12]	M	SHAP, LIME, ACH	Sensors	P (EMG)
28	Islam et al. [18]	M	LIME, Eli5	Sensors	P (EEG)
29	Jaber et al. [20]	M	SHAP	BMC Med Inform Decis Mak	P (Stress)
30	Jiang et al. [25]	M	MDI	IEEE J Biomed Health Inform	P (EMG)
31	Joyce et al. [28]	S	Various	npj Digit. Med.	P (Mental)
32	Juang et al. [30]	M	EFNN	Sleep Med	P (OSAHS)
33	Kalyakulina et al. [36]	S	Various	Ageing Res Rev	P (Age)
34	Keyl et al. [37]	M	LRP, scGeneRAI	Nucleic Acids Res	P (Genes)

Table 1 continued

#	Reference	Publ. Type	XAI Methods	Published in	Focus
35	Khanna et al. [39]	M	Various	Decis Anal J	P (Covid)
36	Khosravi et al. [40]	M	XAI-ED	Comput Educ Artif Intell	P (Biosign.)
37	Kim et al. [41]	M	LRP, SHAP, CAM	Biosensors	P (Biosign.)
38	Kim et al. [44]	M	SHAP	J Med Internet Res	P (CA)
39	Klauschen et al. [45]	S	Various	Annu Rev Pathol	P (Patho)
40	Kumar & Das [46]	M	SHAP	Comput Biol Chem	P (PBMC)
41	Lacalamita et al. [47]	M	SHAP	Int J Mol Sci	P (HCC)
42	Lai et al. [50]	M	SHAP, LIME	Front Immunol	P (AD)
43	Lauritsen et al. [51]	M	DTD, LRP	Nat Commun	P (EWS)
44	Lemańska-P. et al. [55]	M	SHAP, Feature Importance	Cells	P (Sepsis)
45	Li et al. [60]	M	QLattice	npj Microgravity	P (SERCA)
46	Lin et al. [61]	M	SHAP, LIME	Front Med	P (MV)
47	Lisboa et al. [62]	M	EBM	Sci Rep	P (Survival)
48	Liu & Hu [63]	S	Various	Curr Opin Chem Biol	P (RadioGen)
49	Liu et al. [64]	M	SHAP	Lancet Digit Health	P (ICU)
50	Loh et al. [65]	S	Various	Comput Methods Programs Biomed.	G (Health)
51	Lundberg et al. [67]	M	Various	Nature Biomed Eng	P (Blood)
52	Macas et al. [69]	M	Various	Integr Comput-Aided Eng	P (ECG)
53	Madanu et al. [70]	M	Feature Importance	Technologies	P (Pain)
54	Meena & Hasija [73]	M	SHAP	Comput Biol Med	P (Genes)
55	Mei et al. [76]	M	Various	IEEE Trans Evol Comput	G
56	Moulaei et al. [83]	M	SHAP, LIME	Sci Rep	P (Methanol)
57	Novakovsky et al. [88]	M	ExplaiNN	Sci Rep	P (Genes)
58	Novielli et al. [95]	M	SHAP	Front Microbiol	P (CRC)
59	Papadimitroulas et al. [103]	S	Various	Physica Med	G (Onco)
60	Peng et al. [105]	M	SHAP, LIME	J Med Syst	P (Liver)
61	Przepiorka et al. [107]	M	SHAP, LIME	J Neuro-Oncology	P (VS)
62	Qiu et al. [108]	M	SHAP, ENABL Age	Lancet Healthy Longev	P (Age)
63	Ramírez-Mena et al. [111]	M	SHAP	Comput Methods Programs Biomed.	P (PC)
64	Ray et al. [114]	M	Various	Cell Rep Med.	P (Asthma)
65	Roessner et al. [121]	S	Various	Eur Child Adolesc Psychiatry	G (Mental)
66	Rudrapal et al. [124]	M	SHAP, LIME	Mol. Divers.	P (COX-2)
67	Sahoh & Choksuriwong [128]	S	Various	J Ambient Intell Humaniz Comput	G (Health)
68	Sandamal et al. [129]	M	SHAP, LIME	RINENG	P (Fitness)
69	Sganzerla M. et al. [132]	M	SHAP	Sci Rep	P (Proteins)

Table 1 continued

#	Reference	Publ. Type	XAI Methods	Published in	Focus
70	Song et al. [135]	M	SHAP	Nat Commun	P (AKI)
71	Stenwig et al. [136]	M	SHAP	BMC Med Res Methodol	P (ICU)
72	Streich et al. [138]	S	Various	Curr Opin Biotechnol	P (Plants)
73	Talukder et al. [142]	M	Various	Brief Bioinform	P (Genes)
74	Tang et al. [17]	M	Vec2Image	Brief Bioinform	G (Biology)
75	Thorsten-Meyer et al. [1]	M	SHAP	Lancet Digit Health	P (ICU)
76	Tjoa & Guan [15]	S	Various	IEEE Trans. Neural Netw. Learn. Syst.	G (Health)
77	Togo et al. [82]	M	SHAP	J Chem Inf Model	P (Toxicity)
78	Togo et al. [7]	S	Various	Expert Opin Drug Metab Toxicol	P (Toxicity)
79	Veldhuis et al. [109]	M	SHAP, Counterfact	Forensic Sci Int Genet	P (Genes)
80	Wani et al. [35]	M	DeepXplainer	Comput Methods Programs Biomed.	P (LungCancer)
81	Westerlund et al. [120]	S	SHAP, LRP, DTD	Int J Molecular Sci	P (CVD)
82	Wolfe et al. [94]	M	Fuzzy Logic	BMC Genome Biol	P (Genes)
83	Yagin et al. [100]	M	SHAP, LIME	Comput Biol Med	P (Covid)
84	Yang [48]	S	Various	J Healthc Inform Res	G (Health)
85	Zhang [90]	M	LIME	Europ Rev Med Pharmacol Sci	P (Sepsis)

(HCC), early warning scores for critical illness (EWS), cellular calcium homeostasis in muscles through sarcoendoplasmic reticulum calcium ATPase (SERCA), mechanical ventilation in life support (MV), radiogenomics (RadioGen), electrocardiography (ECG), colorectal cancer (CRC), vestibular schwannoma (VS), prostate cancer (PC), bioactivity prediction of phenolic cyclooxygenase-2 inhibitors (COX-2), acute kidney injury (AKI), and cardiovascular disease (CVD).

On closer inspection of the analytical tools utilized in the collected publications, it stands out that some XAI methods were used much more frequently than others. The following numbers were derived for the publications for which the XAI methods are explicitly listed in Table 1. The occurrences in the papers that present a large number of methods simultaneously (mostly surveys) were not counted here, in order to focus more on the application of XAI in physiology. The most frequently used methods include SHapley Additive exPlanations (SHAP) for model-agnostic explanations (40 times) and the Local Interpretable Model-agnostic Explanations (LIME) method (15 times), which are often used in combination. SHAP was mostly used for global explanations and LIME for local explanations. Both methods are model-agnostic and can therefore be widely applied. This probably explains their popularity to a large extent. Layer-wise Relevance Propagation (LRP) as a model-specific explanation method (4 times), Partial Dependency Analysis (PDA) as a model-agnostic explanation method (3 times) as well as Deep Taylor Decomposition (DTD), Class Activation Mapping (CAM, including mostly Grad-CAM and Relevance-CAM in few cases), and feature importance in general (2 times each of the mentioned methods) follow them.

Note that feature importance refers to a wide range of methods that measure and explicate the contribution of features to the final outcome of a computation. There exist different variants, and some of the already mentioned methods, like SHAP, belong to feature importance methods [58], also called attribution methods [131]. The following section briefly introduces the XAI methods that were most common in the reviewed literature.

Selected XAI methods and examples

Subsequently, the most common XAI methods applied in physiology and collected in Table 1 will be shortly introduced and illustrated. Note that methods that appeared only once in the collected literature, introducing XAI applied to physiological use cases (denoted by P), were omitted in this list. The purpose of the following list is to give a concise introduction to the most important XAI methods *applied* to physiological use cases rather than to give a comprehensive introduction of all kinds of existing XAI techniques. The interested reader may find more information in state-of-the-art survey papers on XAI methods [15, 74, 75, 115, 131].

- **SHapley Additive exPlanation (SHAP)** [66]**:** A model-agnostic game-theoretic method for explaining machine learning model predictions. It assigns importance values, called Shapley values, to input features, quantifying their contribution to a specific prediction relative to a baseline. Based on cooperative game theory, SHAP calculates feature contributions by averaging their marginal contributions across all possible subsets of features. SHAP assumes an additive explanation model where the prediction is expressed as the sum of contributions from each feature and a baseline value. Computing exact Shapley values is computationally expensive; therefore, SHAP employs efficient approximations like KernelSHAP (model-agnostic) and TreeSHAP (optimized for tree-based models). SHAP provides local explanations by attributing contributions to input features for individual predictions, while also offering global insights when aggregated across instances. SHAP can be computationally intensive and sensitive to the choice of baseline. SHAP further requires a definition of features on the input to compute importance values on. It serves as a post hoc explanation method.
- **Local Interpretable Model-agnostic Explanations (LIME)** [116]**:** A method for explaining individual predictions of opaque, machine-learned classification models by approximating their behavior locally on input instances with a more simple, interpretable model. LIME works by selecting a specific instance for which the prediction is to be explained, generating a set of randomly perturbed instances around this instance, and using the original model to predict outputs for these perturbations. To focus the explanation on the instance of interest, LIME assigns higher weights to perturbed instances closer to the original. A simple model, such as linear regression, is then trained using weighted features from instances. The interpretable model's coefficients indicate the importance of each feature in the opaque model's prediction for the chosen instance. While LIME provides interpretable explanations, they may be valid only in the vicinity of the instance, not generalizing globally. Furthermore, explanations can vary significantly with different perturbation strategies or parameter settings for feature definition. Additionally, its reliance on perturbations and retraining for each instance can make it computationally expensive. Nevertheless, LIME is suitable for use cases, where the ease of interpretation is of great importance. LIME works with different data types (tabular, text, or images).

For example, in textual data, the contribution of individual words can be computed, while in images, pixels are grouped into superpixels and randomly switched off (e.g., by replacing them with a neutral background color) to derive the contribution of individual image regions. LIME is in general used for local post hoc explanations.

- **Layer-wise Relevance Propagation (LRP)** [101]**:** A method used to interpret the decisions of DNNs. It works by propagating a model's prediction backward through the model's layers to assign relevance scores (importance) to input features (such as individual pixels in an image). These scores represent the contribution of each feature to the network's output. They can be visualized in the form of *heatmaps* (for images or text). The propagation follows specific rules tailored to the structure and parameters of each layer. The common approach for explanation generation on images is to provide heatmaps based on normalizing and plotting the relevance of every individual pixel. LRP allows to express positive relevance on features (contributing to an output) and negative relevance (speaking against a certain output). A key principle of LRP is the conservation of relevance, which ensures that the sum of relevance scores remains consistent as they propagate from one layer to the next. This principle ensures that the input relevance scores explain the full prediction value. The selection of the aforementioned rules can be non-trivial and may significantly affect the quality and interpretability of the produced explanation. LRP is model-specific, relying on the model's structure and learned parameters, which makes it less flexible than some model-agnostic methods. Nevertheless, LRP may be more faithful with respect to representing what a model has learned as it has access to model internals. LRP is a local post hoc explanation method. It can be combined with further methods, such as the t-Distributed Stochastic Neighbor Embedding (t-SNE) method introduced by van der Maaten and Hinton [68], to derive more global explanations.
- **Partial Dependency Analysis (PDA)** [38]**:** Computes the effect of a specific input feature on the final output of a model, usually accompanied by a partial dependency plot for visualization [74, 115]. In principle, PDA isolates the relationship between selected features and the model's output by averaging out the effects of all other features. To perform PDA, one or more features of interest are chosen, and their values are varied across a range. For each combination of values, the model's predictions are averaged over all possible values of the other features in the dataset. This marginalization removes the influence of irrelevant features. The resulting values are then visualized in a partial dependence plot (PDP), which shows how changes in the selected features affect the model's predictions. PDA provides global explanations for a model's behavior [131]. PDA assumes feature independence, which may not hold in cases of highly correlated data, leading to potentially misleading results. Additionally, PDA can be computationally expensive for large datasets or complex models. PDA is a post hoc explanation method.
- **Gradient-weighted Class Activation Mapping (Grad-CAM)** [133]**:** Visualizes for image data which regions of an input image are most important for image classifying convolutional neural network's (CNN) predictions.[3] It produces so-called class-specific activation maps, highlighting areas relevant to the model's decision for a particular class in the form of *heatmaps* (similar to LRP). Grad-CAM first computes a CNN model's prediction on an input image with respect to the target class. Then, gradients of the predicted target class score with respect to feature maps learned by the CNN are calculated. These gradients represent the contribution of each pixel in the learned feature maps to the class score. To create a class activation map for an input image, weights are computed by averaging the gradients over the spatial dimensions of the feature maps. These weights are further used to produce the class activation map from feature maps. Grad-CAM retains only positive contributions in activation maps (as opposed to LRP). Consequently, the final result is a heatmap that highlights the regions most relevant to the target class. Grad-CAM is class-specific and model-specific. It provides local visual explanations and is a post hoc explanation method for CNNs.
- **Deep Taylor Decomposition (DTD)** [81]**:** Is used to interpret the decisions of DNNs. It is an attribution method that decomposes a network's output into contributions from the input features. It builds upon *Taylor expansions*, where the neural network's decision is expressed as a sum of relevance scores assigned to the input features. The idea is to propagate the relevance scores from the output layer back to the input layer, adhering to certain conservation principles. This means that the sum of relevance scores remains consistent across layers, preserving the overall decision value at each stage of the decomposition. Layer-wise Relevance Propagation is built upon the same conservation principle.

Figure 1 illustrates the explanations produced by the selected XAI methods for some of the works collected in Table 1. All produced explanations are visual, either in the form of a plot or a heatmap. The first displayed visual explanation

[3] The principle of convolutional neural networks was introduced by LeCun et al. [53].

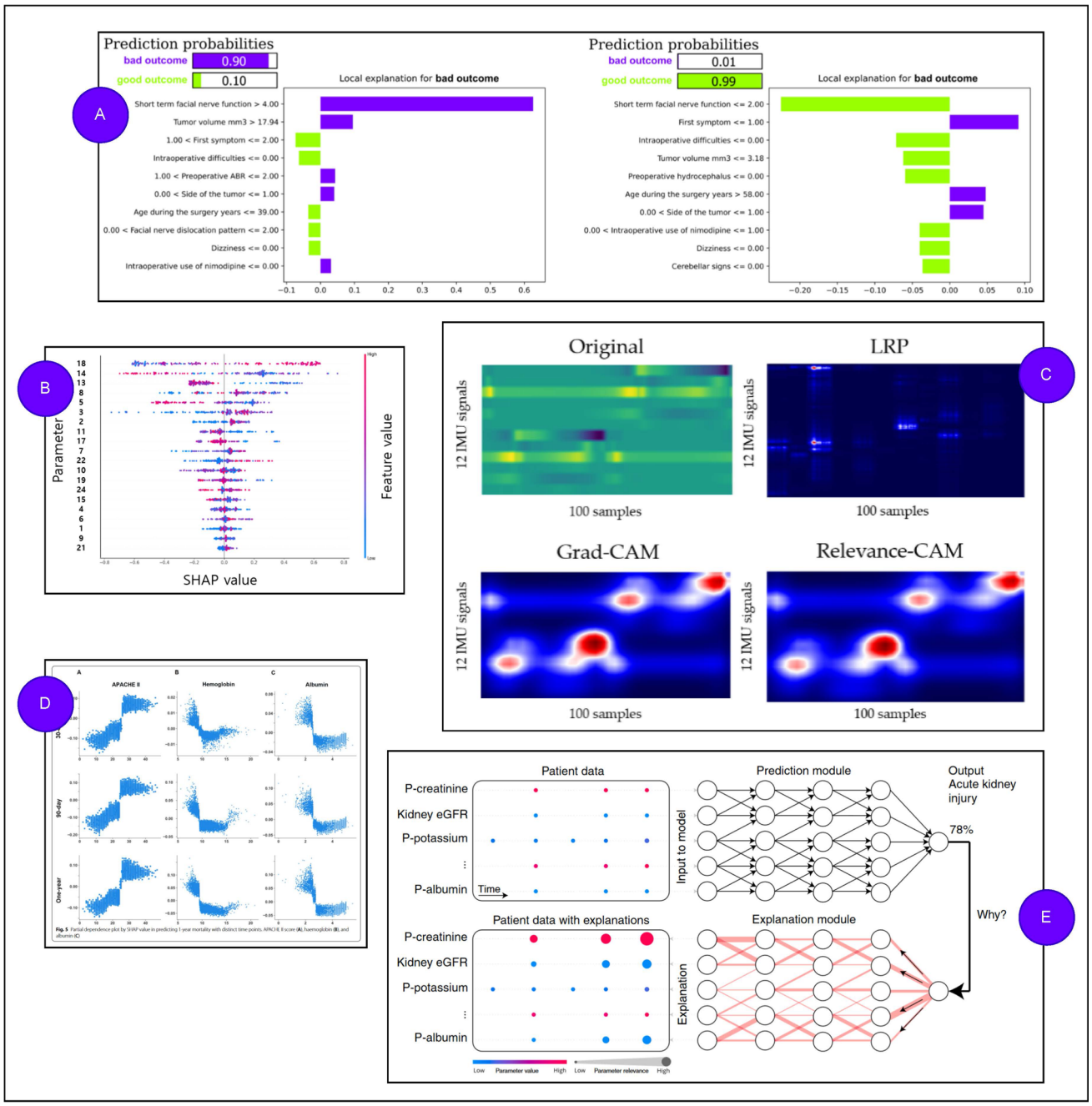

Fig. 1 An illustration of visual explanations for AI-based analysis in physiology presenting **A** Local Interpretable Model-agnostic Explanations from Przepiorka et al. [107], **B** SHapley Additive exPlanations from Kim et al. [41], **C** Layer-wise Relevance Propagation, Gradient-weighted Class Activation Mapping and a relevance-based variant from Kim et al. [41], **D** Partial Dependency Analysis from Chan et al. [99], and **E** Deep Taylor Decomposition from Lauritsen et al. [51]

(A) was produced with LIME for the use case of explaining a classifier for facial nerve functioning after schwannoma surgery as presented in Przepiorka et al. [107]. Another visual explanation (B) was produced with SHAP for sensor-based gait analysis as introduced by Kim et al. [41]. In the same work, Grad-CAM (and Relevance-CAM, a variant of it) and LRP (C) have been used to produce heatmap-based visual explanations. A partial dependency plot (D) was used by Chan et al. [99] to visually explain the mortality prediction of a machine-learned model with the help of PDA. A visual

explanation produced by Deep Taylor Decomposition (E) was presented by Lauritsen et al. [51] for the prediction of acute critical illness from different early warning scores.

After having presented the most important works on XAI in physiology and having illustrated selected methods applied to physiological use cases, the next section discusses which XAI-specific challenges in physiology still remain open from an XAI point of view and what implications the respective findings have for the existing methods and their further development.

Discussion

In the introduction, it was motivated that in addition to existing XAI methods for physiology, open challenges in the further development of XAI for physiology and the improvement of XAI through physiological knowledge should be considered.

One finding is that the majority of the methodological works that have been reviewed utilize existing XAI methods, rather than introducing novel ones. This may suggest that the focus of medical research is currently more on solving a domain-specific problem and justifying the solution with existing XAI methods rather than presenting XAI as a novel solution to physiological problems.

According to Lemoine and Pradeu (2018), clinical phenomena (considered the explanandum) may be first explained by physiological phenomena (which are the explanans here). These physiological phenomena (now considered the explanandum) can then be explained by molecular phenomena (explanans) or by phenomena described by other sciences [57]. Physiology and related fields could therefore benefit from combining XAI methods in an integrative manner to derive new conclusions or achieve the representation of more complex knowledge with the help of explainability.

Considering the examples presented in Section 5, a major finding of this work is that many visual explanation methods are applied or developed for physiological use cases. There seems to be a lack of multimodal approaches that combine visualizations and, for example, verbal statements for more expressive explainability. Multimodal explanations [24] as well as interactive explanations, such as bi-directional explanatory dialogues [56, 126, 130], are considered important prerequisites to human-centered explanations and could be implemented with interpretable models that provide ante hoc explanations [87].

The current XAI methods applied to physiology seem to utilize only static explanations rather than interactive ones. Human-centered approaches that integrate the human as an active participant into AI-supported decision-making processes and that allow for introducing human expertise into learned models could be addressed in future development attempts [87]. Another advantage could be that *human-in-the-loop* systems may drive faster implementation of AI-supported analysis [8, 98].

Furthermore, Lemoine and Pradeu (2018) argue that data-driven prediction methods do not seek explanations in the way traditional physiology does. They state that prediction-based approaches in general may "inspire, and be inspired by, physiology, but are not themselves physiological, in that they do not focus on explanation" [57].

XAI could build a bridge by revealing concepts in features underlying the outcome (prediction) of a data-intensive model. These concepts could be the building blocks that form the basis of constructing more complex explanations [10].

As stated by Lemoine and Pradeu in this context, "Clarity can only be achieved by placing knowledge gathered in non-physiological approaches (remark: e.g., statistical, data-driven models) into a framework, by integrating it into a physiological picture" to derive new explanations and knowledge bases [57].

Integrating results and findings produced with the help of XAI is thus not only the way to follow for evaluating data-intensive models against complex knowledge domains like physiology, but XAI could also provide the means for future knowledge discovery in physiology with the help of AI.

It is furthermore worth mentioning that generative artificial intelligence (genAI) has the potential to break new diagnostic ground and make outcomes transparent with the help of XAI [77]. The work considered in this paper is based on more traditional AI and XAI methods, well established and justified for the respective domains. Under consideration of the advancements in regulatory developments, it is expected that there will be an increase in the use of genAI in medicine over the next few years [77].

Finally, the explanation itself is a physiological process with neural and biochemical events occurring in memory, attention, and perception of human cognition. XAI research could benefit from considering neurological constraints and potentials in reasoning about and production of explanations in humans [9].

Conclusion

This article reviewed recent publications using explainable artificial intelligence (XAI) in physiology. It provided a collection of 85 works, introduced and illustrated the most applied XAI methods, and discussed open challenges regarding the interfaces between XAI and physiology. Specifically, physiology could benefit if XAI methods were used to conduct knowledge discovery and to integrate human knowledge with models and data. XAI could benefit from considering

physiological knowledge for improving generated explanations. Ultimately, providing human-centered explanations for AI decisions could pave the way to a more integrative usage of AI in physiology.

Acknowledgements I would like to thank the anonymous reviewers for their helpful comments on the manuscript.

Author contribution The author declares that the conceptualization, draft, and finalization of the manuscript was prepared and written by herself.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability No datasets were generated or analyzed during the current study.

Materials availability Not applicable

Code availability For some of the referenced works, code is available.

Declarations

Ethics approval and consent to participate Not applicable

Consent for publication The author of this paper declares that the images presented in this work have already been published and are referenced accordingly.

Conflict of interest The author declares no competing interests.

References

1. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
2. Alabdulhafith M, Saleh H, Elmannai H, Ali ZH, El-Sappagh S, Hu J-W, El-Rashidy N (2023) A clinical decision support system for edge/cloud ICU readmission model based on particle swarm optimization, ensemble machine learning, and explainable artificial intelligence. IEEE Access 11:100604–100621. https://doi.org/10.1109/ACCESS.2023.3312343
3. Ali S, Abuhmed T, El-Sappagh SHA, Muhammad K, Alonso-Moral JM, Confalonieri R, Guidotti R, Ser JD, Rodríguez ND, Herrera F (2023) Explainable artificial intelligence (XAI): what we know and what is left to attain trustworthy artificial intelligence. Inf Fusion 99:101805. https://doi.org/10.1016/J.INFFUS.2023.101805
4. Andreu-Perez J, Emberson LL, Kiani M, Filippetti ML, Hagras H, Rigato S (2021) Explainable artificial intelligence based analysis for interpreting infant fNIRS data in developmental cognitive neuroscience. Communications biology 4(1):1077
5. Anguita-Ruiz A, Segura-Delgado A, Alcalá R, Aguilera CM, Alcalá-Fdez J (2020) explainable artificial intelligence (XAI) for the identification of biologically relevant gene expression patterns in longitudinal human studies, insights from obesity research. PLoS Comput Biol 16(4):1007792
6. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, Carter RE, Shah ND, Friedman PA, Noseworthy PA (2019) Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. J Cardiovasc Electrophysiol 30(5):668–674
7. Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W (2015) On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS ONE 10(7):0130140. https://doi.org/10.1371/journal.pone.0130140. Accessed 2018-12-06
8. Banerjee JS, Mahmud M, Brown D (2023) Heart rate variability-based mental stress detection: an explainable machine learning approach. SN Comp Sci 4(2):176
9. Beer S, Elmenhorst D, Bischof GN, Ramirez A, Bauer A, Drzezga A (2024) Explainable artificial intelligence identifies an AQP4 polymorphism-based risk score associated with brain amyloid burden. Neurobiol Aging 143:19–29. https://doi.org/10.1016/j.neurobiolaging.2024.08.002
10. Belle V (2017) Logic meets probability: towards explainable ai systems for uncertain worlds. In: 26th International Joint Conference on Artificial Intelligence, pp 5116–5120
11. Bernard D, Doumard E, Ader I, Kemoun P, Pagès J-C, Galinier A, Cussat-Blanc S, Furger F, Ferrucci L, Aligon J, Delpierre C, Pénicaud L, Monsarrat P, Casteilla L (2023) Explainable machine learning framework to predict personalized physiological aging. Aging Cell 22(8):13872
12. Boscolo Galazzo I, Cruciani F, Brusini L, Salih A, Radeva P, Storti SF, Menegaz G (2022) Explainable artificial intelligence for magnetic resonance imaging aging brainprints: grounds and challenges. IEEE Signal Process Mag 39(2):99–116. https://doi.org/10.1109/MSP.2021.3126573
13. Boulesteix A-L, Wright MN, Hoffmann S, König IR (2020) Statistical learning approaches in the genetic epidemiology of complex diseases. Hum Genet 139:73–84
14. Bruckert S, Finzel B, Schmid U (2020) The next generation of medical decision support: a roadmap toward transparent expert companions. Front Artif Intell 3. https://doi.org/10.3389/frai.2020.507973
15. Carvalho DV, Pereira EM, Cardoso JS (2019) Machine learning interpretability: a survey on methods and metrics. Electronics 8(8):832. https://doi.org/10.3390/electronics8080832
16. Carvalho DV, Pereira EM, Cardoso JS (2019) Machine learning interpretability: a survey on methods and metrics. Electronics 8(8):832. https://doi.org/10.3390/electronics8080832
17. Casalicchio G, Molnar C, Bischl B (2019) Visualizing the feature importance for black box models. In: Berlingerio M, Bonchi F, Gärtner T, Hurley N, Ifrim G (eds) Machine Learning and Knowledge Discovery in Databases. Springer, Cham, pp 655–670
18. Chan M-C, Pai K-C, Su S-A, Wang M-S, Wu C-L, Chao W-C (2022) Explainable machine learning to predict long-term mortality in critically ill ventilated patients: a retrospective study in central Taiwan. BMC Med Inform Decis Mak 22(1):75

19. Chen ZS, Galatzer-Levy IR, Bigio B, Nasca C, Zhang Y et al (2022) Modern views of machine learning for precision psychiatry. Patterns 3(11)
20. Chen T-CT, Chiu M-C (2022) Evaluating the sustainability of smart technology applications in healthcare after the COVID-19 pandemic: a hybridising subjective and objective fuzzy group decision-making approach with explainable artificial intelligence. Digital Health 8:20552076221136380
21. Chen RJ, Lu MY, Williamson DF, Chen TY, Lipkova J, Noor Z, Shaban M, Shady M, Williams M, Joo B et al (2022) Pan-cancer integrative histology-genomic analysis via multimodal deep learning. Cancer Cell 40(8):865–878
22. Chormai P, Herrmann J, Müller K-R, Montavon G (2024) Disentangled explanations of neural network predictions by finding relevant subspaces. IEEE Transactions on Pattern Analysis and Machine Intelligence (2024)
23. Council AUPP (2017) Statement on algorithmic transparency and accountability. ACM, Commun
24. Craven MW, Shavlik JW (1992) Visualizing learning and computation in artificial neural networks. Int J Artif Intell Tools 1(03):399–425
25. Davagdorj K, Bae J-W, Pham V-H, Theera-Umpon N, Ryu KH (2021) Explainable artificial intelligence based framework for non-communicable diseases prediction. IEEE Access 9:123672–123688. https://doi.org/10.1109/ACCESS.2021.3110336
26. Deane KE, Brunk MG, Curran AW, Zempeltzi MM, Ma J, Lin X, Abela F, Aksit S, Deliano M, Ohl FW et al (2020) Ketamine anaesthesia induces gain enhancement via recurrent excitation in granular input layers of the auditory cortex. J Physiol 598(13):2741–2755
27. Dindorf C, Konradi J, Wolf C, Taetz B, Bleser G, Huthwelker J, Werthmann F, Bartaguiz E, Kniepert J, Drees P, Betz U, Fröhlich M (2021) Classification and automated interpretation of spinal posture data using a pathology-independent classifier and explainable artificial intelligence (XAI). Sensors 21(18). https://doi.org/10.3390/s21186323
28. El-Sappagh S, Alonso JM, Islam SR, Sultan AM, Kwak KS (2021) A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. Sci Rep 11(1):2660
29. Estruch R, Ros E, Salas-Salvadó J, Covas M-I, Corella D, Arós F, Gómez-Gracia E, Ruiz-Gutiérrez V, Fiol M, Lapetra J, Lamuela-Raventos RM, Serra-Majem L, Pintó X, Basora J, Muñoz MA, Sorlí JV, Martínez JA, Martínez-González MA (2013) Primary prevention of cardiovascular disease with a Mediterranean diet. N Engl J Med 368(14):1279–1290. https://doi.org/10.1056/NEJMoa1200303
30. Fellous J-M, Sapiro G, Rossi A, Mayberg H, Ferrante M (2019) Explainable artificial intelligence for neuroscience: behavioral neurostimulation. Front Neurosci 13:1346
31. Finzel B (2024) Human-centered explanations: lessons learned from image classification for medical and clinical decision making. KI-Künstliche Intelligenz, 1–11
32. Finzel B, Hilme P, Rabold J, Schmid U (2024) Telling more with concepts and relations: exploring and evaluating classifier decisions with CoReX. CoRR arxiv:2405.01661. https://doi.org/10.48550/ARXIV.2405.01661
33. Finzel B, Tafler DE, Thaler AM, Schmid U (2021) Multimodal explanations for user-centric medical decision support systems. In: Doyle TE, Kelliher A, Samavi R, Barry B, Yule SJ, Parker S, Noseworthy MD, Yang Q (eds) Proceedings of the AAAI 2021 Fall Symposium on Human Partnership with Medical AI: Design, Operationalization, and Ethics (AAAI-HUMAN 2021), Virtual Event, November 4-6, 2021. CEUR Workshop Proceedings, vol. 3068. CEUR-WS.org. https://ceur-ws.org/Vol-3068/short2.pdf
34. Finzel B, Tafler DE, Thaler AM, Schmid U (2021) Multimodal explanations for user-centric medical decision support systems. In: HUMAN@AAAI Fall Symposium. https://api.semanticscholar.org/CorpusID:246751352
35. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451
36. Gao S, Gao X, Zhu R, Wu D, Feng Z, Jiao N, Sun R, Gao W, He Q, Liu Z et al (2023) Microbial genes outperform species and SNVS as diagnostic markers for Crohn's disease on multicohort fecal metagenomes empowered by artificial intelligence. Gut Microbes 15(1):2221428
37. Gimeno M, San José-Enériz E, Villar S, Agirre X, Prosper F, Rubio A, Carazo F (2022) Explainable artificial intelligence for precision medicine in acute myeloid leukemia. Front Immunol 13:977358
38. Goodman B, Flaxman S (2017) European union regulations on algorithmic decision-making and a "right to explanation". AIMag 38(3):50–57. https://doi.org/10.1609/aimag.v38i3.2741
39. Goodwin NL, Choong JJ, Hwang S, Pitts K, Bloom L, Islam A, Zhang YY, Szelenyi ER, Tong X, Newman EL et al (2024) Simple behavioral analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience. Nature Neurosci 1–14
40. Górriz JM, Álvarez-Illán I, Álvarez-Marquina A, Arco JE, Atzmueller M, Ballarini F, Barakova E, Bologna G, Bonomini P, Castellanos-Dominguez G, Castillo-Barnes D, Cho SB, Contreras R, Cuadra JM, Domínguez E, Domínguez-Mateos F, Duro RJ, Elizondo D, Fernández-Caballero A, Fernandez-Jover E, Formoso MA, Gallego-Molina NJ, Gamazo J, González JG, Garcia-Rodriguez J, Garre C, Garrigós J, Gómez-Rodellar A, Gómez-Vilda P, Graña M, Guerrero-Rodriguez B, Hendrikse SCF, Jimenez-Mesa C, Jodra-Chuan M, Julian V, Kotz G, Kutt K, Leming M, de Lope J, Macas B, Marrero-Aguiar V, Martinez JJ, Martinez-Murcia FJ, Martínez-Tomás R, Mekyska J, Nalepa GJ, Novais P, Orellana D, Ortiz A, Palacios-Alonso D, Palma J, Pereira A, Pinacho-Davidson P, Pinninghoff MA, Ponticorvo M, Psarrou A, Ramírez J, Rincón M, Rodellar-Biarge V, Rodríguez-Rodríguez I, Roelofsma PHMP, Santos J, Salas-Gonzalez D, Salcedo-Lagos P, Segovia F, Shoeibi A, Silva M, Simic D, Suckling J, Treur J, Tsanas A, Varela R, Wang SH, Wang W, Zhang YD, Zhu H, Zhu Z, Ferrández-Vicente JM (2023) Computational approaches to explainable artificial intelligence: advances in theory, applications and trends. Information Fusion 100. https://doi.org/10.1016/j.inffus.2023.101945
41. Gouverneur P, Li F, Shirahama K, Luebke L, Adamczyk WM, Szikszay TM, Luedtke K, Grzegorzek M (2023) Explainable artificial intelligence (XAI) in pain research: understanding the role of electrodermal activity for automated pain recognition. Sensors 23(4). https://doi.org/10.3390/s23041959
42. Grochowska KM, Gomes GM, Raman R, Kaushik R, Sosulina L, Kaneko H, Oelschlegel AM, Yuanxiang P, Reyes-Resina I, Bayraktar G et al (2023) Jacob-induced transcriptional inactivation of CREB promotes Aβ-induced synapse loss in alzheimer's disease. EMBO J 42(4)
43. Gunning D, Aha D (2019) Darpa's explainable artificial intelligence (XAI) program. AI Mag 40(2):44–58
44. Han F, Cheng J, Liao S, Deng Y (2022) Building trust for post-operative pain estimation: towards explainable machine-learning prediction based on multimodal indicators. In: 2022 IEEE International Conference on Multimedia and Expo (ICME), pp 01–06. https://doi.org/10.1109/ICME52920.2022.9859635
45. Hasan MM, Watling CN, Larue GS (2024) Validation and interpretation of a multimodal drowsiness detection system using explainable machine learning. Comput Methods Pro-

grams Biomed 243:107925. https://doi.org/10.1016/j.cmpb.2023.107925
46. He B, Zhao Y, Mao W (2022) Explainable artificial intelligence reveals environmental constraints in seagrass distribution. Ecol Ind 144:109523. https://doi.org/10.1016/j.ecolind.2022.109523
47. Hijazi H, Abu Talib M, Hasasneh A, Bou Nassif A, Ahmed N, Nasir Q (2021) Wearable devices, smartphones, and interpretable artificial intelligence in combating COVID-19. Sensors 21(24). https://doi.org/10.3390/s21248424
48. Holzinger A (2016) Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics 3(2):119–131. https://doi.org/10.1007/S40708-016-0042-6
49. Horne Z, Muradoglu M, Cimpian A (2019) Explanation as a cognitive process. Trends Cogn Sci 23(3):187–199
50. Hossain MI, Zamzmi G, Mouton PR, Salekin MS, Sun Y, Goldgof D (2023) Explainable ai for medical data: current methods, limitations, and future directions. ACM Comput Surv. https://doi.org/10.1145/3637487
51. Hussain I, Jany R (2024) Interpreting stroke-impaired electromyography patterns through explainable artificial intelligence. Sensors 24(5). https://doi.org/10.3390/s24051392
52. Hussein R, Palangi H, Ward RK, Wang ZJ (2019) Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. Clin Neurophysiol 130(1):25–37. https://doi.org/10.1016/j.clinph.2018.10.010
53. Islam MS, Hussain I, Rahman MM, Park SJ, Hossain MA (2022) Explainable artificial intelligence model for stroke prediction using EEG signal. Sensors 22(24). https://doi.org/10.3390/s22249859
54. Islam MS, Hussain I, Rahman MM, Park SJ, Hossain MA (2022) Explainable artificial intelligence model for stroke prediction using EEG signal. Sensors 22(24):9859
55. Jaber D, Hajj H, Maalouf F, El-Hajj W (2022) Medically-oriented design for explainable AI for stress prediction from physiological measurements. BMC Med Inform Decis Mak 22(1):38
56. Jackson P (1998) Introduction to expert systems, 3rd edn. Addison-Wesley Longman Publishing Co., Inc, USA
57. Jiang X, Nazarpour K, Dai C (2023) Explainable and robust deep forests for EMG-force modeling. IEEE J Biomed Health Inform 27(6):2841–2852. https://doi.org/10.1109/JBHI.2023.3262316
58. Jones MA, Islam W, Faiz R, Chen X, Zheng B (2022) Applying artificial intelligence technology to assist with breast cancer diagnosis and prognosis prediction. Frontiers in Oncology 12. https://doi.org/10.3389/fonc.2022.980793
59. Joyce DW, Kormilitzin A, Smith KA, Cipriani A (2023) Explainable artificial intelligence for mental health through transparency and interpretability for understandability. npj Digital Medicine 6(1):6
60. Juang C-F, Wen C-Y, Chang K-M, Chen Y-H, Wu M-F, Huang W-C (2021) Explainable fuzzy neural network with easy-to-obtain physiological features for screening obstructive sleep apnea-hypopnea syndrome. Sleep Med 85:280–290. https://doi.org/10.1016/j.sleep.2021.07.012
61. Kalyakulina A, Yusipov I, Moskalev A, Franceschi C, Ivanchenko M (2024) explainable artificial intelligence (XAI) in aging clock models. Ageing Res Rev 93. https://doi.org/10.1016/j.arr.2023.102144
62. Keyl P, Bischoff P, Dernbach G, Bockmayr M, Fritz R, Horst D, Blüthgen N, Montavon G, Müller K-R, Klauschen F (2023) Single-cell gene regulatory network prediction by explainable ai. Nucleic Acids Res 51(4):20–20
63. Khanna VV, Chadaga K, Sampathila N, Prabhu S, Chadaga R (2023) A machine learning and explainable artificial intelligence triage-prediction system for covid-19. Decis Anal J 7:100246. https://doi.org/10.1016/j.dajour.2023.100246
64. Khosravi H, Shum SB, Chen G, Conati C, Tsai Y-S, Kay J, Knight S, Martinez-Maldonado R, Sadiq S, Gašević D (2022) Explainable artificial intelligence in education. Comput Educ Artif Intell 3:100074. https://doi.org/10.1016/j.caeai.2022.100074
65. Kim J-K, Bae M-N, Lee K, Kim J-C, Hong SG (2022) Explainable artificial intelligence and wearable sensor-based gait analysis to identify patients with osteopenia and sarcopenia in daily life. Biosensors 12(3). https://doi.org/10.3390/bios12030167
66. Kim YK, Koo JH, Lee SJ, Song HS, Lee M (2023) Explainable artificial intelligence warning model using an ensemble approach for in-hospital cardiac arrest prediction: Retrospective cohort study. J Med Internet Res 25:48244
67. Klauschen F, Dippel J, Keyl P, Jurmeister P, Bockmayr M, Mock A, Buchstab O, Alber M, Ruff L, Montavon G, Müller K-R (2024) Toward explainable artificial intelligence for precision pathology. Annu Rev Pathol 19:541–570. https://doi.org/10.1146/annurev-pathmechdis-051222-113147
68. Kumar S, Das A (2023) Peripheral blood mononuclear cell derived biomarker detection using explainable artificial intelligence (xai) provides better diagnosis of breast cancer. Comput Biol Chem 104:107867. https://doi.org/10.1016/j.compbiolchem.2023.107867
69. Lacalamita A, Serino G, Pantaleo E, Monaco A, Amoroso N, Bellantuono L, Piccinno E, Scalavino V, Dituri F, Tangaro S, Bellotti R, Giannelli G (2023) Artificial intelligence and complex network approaches reveal potential gene biomarkers for hepatocellular carcinoma. Int J Mol Sci 24(20). https://doi.org/10.3390/ijms242015286
70. Lai Y, Lin P, Lin F, Chen M, Lin C, Lin X, Wu L, Zheng M, Chen J (2022) Identification of immune microenvironment subtypes and signature genes for alzheimer's disease diagnosis and risk prediction based on explainable machine learning. Front Immunol 13:1046410
71. Lauritsen SM, Kristensen M, Olsen MV, Larsen MS, Lauritsen KM, Jørgensen MJ, Lange J, Thiesson B (2020) Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat Commun 11(1):3852
72. LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, Jackel LD (1989) Handwritten digit recognition with a back-propagation network. In: Touretzky DS (ed) Advances in Neural Information Processing Systems 2, [NIPS Conference, Denver, Colorado, USA, November 27-30, 1989], pp 396–404. Morgan Kaufmann. http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network
73. Lemańska-Perek A, Krzyżanowska-Golab D, Kobylińska K, Biecek P, Skalec T, Tyszko M, Gozdzik W, Adamik B (2022) Explainable artificial intelligence helps in understanding the effect of fibronectin on survival of sepsis. Cells 11(15). https://doi.org/10.3390/cells11152433
74. Lemoine M, Pradeu T (2018) Dissecting the meanings of "physiology" to assess the vitality of the discipline. Physiology 33(4):236–245. (PMID: 29873600). https://doi.org/10.1152/physiol.00015.2018
75. Lent M, Fisher W, Mancuso M (2004) An explainable artificial intelligence system for small-unit tactical behavior. In: Proc. 2004 Nat. Conf. Artificial Intelligence, pp 900–907. AAAI Press; MIT Press
76. Li K, Desai R, Scott RT, Steele JR, Machado M, Demharter S, Hoarfrost A, Braun JL, Fajardo VA, Sanders LM et al (2023) Explainable machine learning identifies multi-omics signatures of muscle response to spaceflight in mice. npj Microgravity 9(1):90
77. Lin M-Y, Li C-C, Lin P-H, Wang J-L, Chan M-C, Wu C-L, Chao W-C (2021) Explainable machine learning to predict successful weaning among patients requiring prolonged mechanical venti-

lation: a retrospective cohort study in central taiwan. Front Med 8:663739
78. Lisboa PJ, Jayabalan M, Ortega-Martorell S, Olier I, Medved D, Nilsson J (2022) Enhanced survival prediction using explainable artificial intelligence in heart transplantation. Sci Rep 12(1):19525
79. Liu Q, Hu P (2022) Extendable and explainable deep learning for pan-cancer radiogenomics research. Curr Opin Chem Biol 66:102111
80. Liu X, Hu P, Yeung W, Zhang Z, Ho V, Liu C, Dumontier C, Thoral PJ, Mao Z, Cao D et al (2023) Illness severity assessment of older adults in critical illness using machine learning (elder-icu): an international multicentre study with subgroup bias evaluation. The Lancet Digital Health 5(10):657–667
81. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR (2022) Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput Methods Programs Biomed 226:107161. https://doi.org/10.1016/j.cmpb.2022.107161
82. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30, pp 4765–4774. Curran Associates, Inc.. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
83. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, Liston DE, Low DK-W, Newman S-F, Kim J et al (2018) Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomed Eng 2(10):749–760
84. Maaten L, Hinton G (2008) Visualizing data using t-sne. J Mach Learn Res 9(11)
85. Macas B, Garrigós J, Martínez JJ, Ferrández JM, Bonomini MP (2024) An explainable machine learning system for left bundle branch block detection and classification. Integr Comput-Aided Eng 31(1):43–58
86. Madanu R, Abbod MF, Hsiao F-J, Chen W-T, Shieh J-S (2022) Explainable ai (xai) applied in machine learning for pain modeling: A review. Technologies 10(3). https://doi.org/10.3390/technologies10030074
87. McCarthy J (1958) Programs with Common Sense. In: Proceedings of the Teddington Conference on the Mechanisation of Thought Processes, pp 77–84
88. Meena J, Hasija Y (2022) Application of explainable artificial intelligence in the identification of squamous cell carcinoma biomarkers. Comput Biol Med 146:105505. https://doi.org/10.1016/j.compbiomed.2022.105505
89. Mei Y, Chen Q, Lensen A, Xue B, Zhang M (2023) Explainable artificial intelligence by genetic programming: A survey. IEEE Trans Evol Comput 27(3):621–641. https://doi.org/10.1109/TEVC.2022.3225509
90. Meskó B, Topol EJ (2023) The imperative for regulatory oversight of large language models (or generative ai) in healthcare. Digit Med 6:120. https://doi.org/10.1038/s41746-023-00873-0
91. Miller T (2019) Explanation in artificial intelligence: Insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007
92. Mitchell TM (1997) Machine Learning, 1st edn. McGraw-Hill Inc, USA
93. Mohammed A, Geppert C, Hartmann A, Kuritcyn P, Bruns V, Schmid U, Wittenberg T, Benz M, Finzel B (2022) Explaining and evaluating deep tissue classification by visualizing activations of most relevant intermediate layers. Curr Direct Biomed Eng 8(2):229–232. https://doi.org/10.1515/cdbme-2022-1059
94. Montavon G, Lapuschkin S, Binder A, Samek W, Müller K-R (2017) Explaining nonlinear classification decisions with deep taylor decomposition. Pattern Recogn 65:211–222
95. Moulaei K, Afrash MR, Parvin M, Shadnia S, Rahimi M, Mostafazadeh B, Evini PET, Sabet B, Vahabi SM, Soheili A et al (2024) Explainable artificial intelligence (xai) for predicting the need for intubation in methanol-poisoned patients: a study comparing deep and machine learning models. Sci Rep 14(1):15751
96. Niemann B, Haufs-Brusberg S, Puetz L, Feickert M, Jaeckstein MY, Hoffmann A, Zurkovic J, Heine M, Trautmann E-M, Müller CE, Tönjes A, Schlein C, Jafari A, Eltzschig HK, Gnad T, Blüher M, Krahmer N, Kovacs P, Heeren J, Pfeifer A (2022) Apoptotic brown adipocytes enhance energy expenditure via extracellular inosine. Nature 609:361–368. https://doi.org/10.1038/s41586-022-05041-0
97. Noble D (2002) Modeling the heart-from genes to cells to the whole organ. Science 295(5560):1678–1682. https://doi.org/10.1126/science.1069881. https://www.science.org/doi/pdf/10.1126/science.1069881
98. Novakovsky G, Fornes O, Saraswat M, Mostafavi S, Wasserman WW (2023) Explainn: interpretable and transparent neural networks for genomics. Genome Biol 24(1):154
99. Novielli P, Romano D, Magarelli M, Bitonto PD, Diacono D, Chiatante A, Lopalco G, Sabella D, Venerito V, Filannino P et al (2024) Explainable artificial intelligence for microbiome data analysis in colorectal cancer biomarker identification. Front Microbiol 15:1348974
100. Padmanabhan S, Tran TQB, Dominiczak AF (2021) Artificial intelligence in hypertension. Circ Res 128(7):1100–1118
101. Paludan SR, Pradeu T, Masters SL, Mogensen TH (2021) Constitutive immune mechanisms: mediators of host defence and immune regulation. Nat Rev Immunol 21:137–150. https://doi.org/10.1038/s41577-020-0391-5
102. Papadimitroulas P, Brocki L, Chung NC, Marchadour W, Vermet F, Gaubert L, Eleftheriadis V, Plachouris D, Visvikis D, Kagadis GC et al (2021) Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data harmonization. Physica Med 83:108–121
103. Peng J, Zou K, Zhou M, Teng Y, Zhu X, Zhang F, Xu J (2021) An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. J Med Syst 45(5):61
104. Poeta E, Ciravegna G, Pastor E, Cerquitelli T, Baralis E (2023) Concept-based explainable artificial intelligence: A survey. arXiv preprint arXiv:2312.12936
105. Przepiorka L, Kujawski S, Wójtowicz K, Maj E, Marchel A, Kunert P (2024) Development and application of explainable artificial intelligence using machine learning classification for long-term facial nerve function after vestibular schwannoma surgery. J Neuro-Oncology 1–13
106. Qiu W, Chen H, Kaeberlein M, Lee S-I (2023) Explainable biological age (enabl age): an artificial intelligence framework for interpretable biological age. The Lancet Healthy Longevity 4(12):711–723
107. Ramírez-Mena A, Andrés-León E, Alvarez-Cubero MJ, Anguita-Ruiz A, Martinez-Gonzalez LJ, Alcala-Fdez J (2023) Explainable artificial intelligence to predict and identify prostate cancer tissue by gene expression. Comput Methods Programs Biomed 240:107719. https://doi.org/10.1016/j.cmpb.2023.107719
108. Ray A, Das J, Wenzel SE (2022) Determining asthma endotypes and outcomes: Complementing existing clinical practice with modern machine learning. Cell Reports Medicine 3(12)
109. Ribeiro MT, Singh S, Guestrin C (2016) Why should I trust you?: Explaining the predictions of any classifier. In: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining. KDD'16, pp 1135–1144. ACM. https://doi.org/10.1145/2939672.2939778
110. Rodríguez-Nuevo A, Torres-Sanchez A, Duran JM, De Guirior C, Martínez-Zamora MA, Böke E (2022) Oocytes maintain ros-free mitochondrial metabolism by suppressing complex i. Nature 607:756–761. https://doi.org/10.1038/s41586-022-04979-5

111. Roessner V, Rothe J, Kohls G, Schomerus G, Ehrlich S, Beste C (2021) Taming the chaos?! Using eXplainable Artificial Intelligence (XAI) to tackle the complexity in mental health research. Springer. Eur Child Adolesc Psychiatry 30:1143–1146. https://doi.org/10.1007/s00787-021-01836-0
112. Rosenfeld A (2021) Better metrics for evaluating explainable artificial intelligence. In: Dignum F, Lomuscio A, Endriss U, Nowé A (eds) AAMAS '21: 20th International Conference on Autonomous Agents and Multiagent Systems, Virtual Event, United Kingdom, May 3-7, 2021, pp 45–50. ACM. https://doi.org/10.5555/3463952.3463962. https://www.ifaamas.org/Proceedings/aamas2021/pdfs/p45.pdf
113. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1(5):206–215. https://doi.org/10.1038/s42256-019-0048-x
114. Rudrapal M, Kirboga KK, Abdalla M, Maji S (2024) Explainable artificial intelligence-assisted virtual screening and bioinformatics approaches for effective bioactivity prediction of phenolic cyclooxygenase-2 (cox-2) inhibitors using pubchem molecular fingerprints. Molecular Diversity, 1–20
115. Russell S, Norvig P (2020) Artificial Intelligence: A Modern Approach (4th Edition). Pearson. http://aima.cs.berkeley.edu/
116. Sahoh B, Choksuriwong A (2023) The role of explainable artificial intelligence in high-stakes decision-making systems: a systematic review. J Ambient Intell Humaniz Comput 14(6):7827–7843
117. Sandamal K, Arachchi S, Erkudov VO, Rozumbetov KU, Rathnayake U (2024) Explainable artificial intelligence for fitness prediction of young athletes living in unfavorable environmental conditions. Results in Engineering 23:102592
118. Schmid U, Finzel B (2020) Mutual explanations for cooperative decision making in medicine. KI-Künstliche Intelligenz 34(2):227–233
119. Schwalbe G, Finzel B (2023) A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. Data Min Knowl Disc. https://doi.org/10.1007/s10618-022-00867-8
120. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proc. 2017 IEEE Int. Conf. Computer Vision, pp 618–626. IEEE. https://doi.org/10.1109/ICCV.2017.74 . arxiv:1610.02391
121. Sganzerla Martinez G, Perez-Rueda E, Kumar A, Sarkar S, de Avila E Silva S (2023) Explainable artificial intelligence as a reliable annotator of archaeal promoter regions. Scientific Reports 13(1):1763
122. Song X, Yu AS, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, Hu Y, Liu M (2020) Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. Nat Commun 11(1):5668
123. Stenwig E, Salvi G, Rossi PS, Skjærvold NK (2022) Comparative analysis of explainable machine learning prediction models for hospital mortality. BMC Med Res Methodol 22(1):53
124. Streich J, Romero J, Gazolla JGFM, Kainer D, Cliff A, Prates ET, Brown JB, Khoury S, Tuskan GA, Garvin M et al (2020) Can exascale computing and explainable artificial intelligence applied to plant biology deliver on the united nations sustainable development goals? Curr Opin Biotechnol 61:217–225
125. Sun-Wang JL, Ivanova S, Zorzano A (2020) The dialogue between the ubiquitin-proteasome system and autophagy: Implications in ageing. Ageing Res Rev 64:101203. https://doi.org/10.1016/j.arr.2020.101203
126. Sun-Wang JL, Yarritu-Gallego A, Ivanova S, Zorzano A (2021) The ubiquitin-proteasome system and autophagy: self-digestion for metabolic health. Trends in Endocrinology & Metabolism 32(8):594–608
127. Talukder A, Barham C, Li X, Hu H (2020) Interpretation of deep learning in genomics and epigenomics. Brief Bioinform 22(3):177
128. Tang H, Yu X, Liu R, Zeng T (2022) Vec2image: an explainable artificial intelligence model for the feature representation and classification of high-dimensional biological data by vector-to-image conversion. Brief Bioinform 23(2):584
129. Thorsen-Meyer H-C, Nielsen AB, Nielsen AP, Kaas-Hansen BS, Toft P, Schierbeck J, Strøm T, Chmura PJ, Heimann M, Dybdahl L et al (2020) Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. The Lancet Digital Health 2(4):179–191
130. Tjoa E, Guan C (2021) A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE Transactions on Neural Networks and Learning Systems 32(11):4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314
131. Togo MV, Mastrolorito F, Ciriaco F, Trisciuzzi D, Tondo AR, Gambacorta N, Bellantuono L, Monaco A, Leonetti F, Bellotti R et al (2022) Tiresia: an explainable artificial intelligence platform for predicting developmental toxicity. J Chem Inf Model 63(1):56–66
132. Togo MV, Mastrolorito F, Orfino A, Graps EA, Tondo AR, Altomare CD, Ciriaco F, Trisciuzzi D, Nicolotti O, Amoroso N (2024) Where developmental toxicity meets explainable artificial intelligence: state-of-the-art and perspectives. Expert Opinion on Drug Metabolism & Toxicology 20(7):561–577
133. Veldhuis MS, Ariëns S, Ypma RJ, Abeel T, Benschop CC (2022) Explainable artificial intelligence in forensics: Realistic explanations for number of contributor predictions of dna profiles. Forensic Sci Int Genet 56:102632
134. Vogler M, Zieseniss A, Hesse AR, Levent E, Tiburcy M, Heinze E, Burzlaff N, Schley G, Eckardt KU, Willam C, Katschinski DM (2015) Pre- and post-conditional inhibition of prolyl-4-hydroxylase domain enzymes protects the heart from an ischemic insult. Pflügers Arch Eur J Physiol 467:2141–2149. https://doi.org/10.1007/s00424-014-1667-z5
135. Wani NA, Kumar R, Bedi J (2024) Deepxplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. Comput Methods Programs Biomed 243:107879. https://doi.org/10.1016/j.cmpb.2023.107879
136. Westerlund AM, Hawe JS, Heinig M, Schunkert H (2021) Risk prediction of cardiovascular events by exploration of molecular data with explainable artificial intelligence. Int J Molecular Sci 22(19). https://doi.org/10.3390/ijms221910291
137. Wittekind C, Bootz F, Meyer H-J (2004) Tumoren des Verdauungstraktes. In: Wittekind C, Bootz F, Meyer H-J (eds) TNM Klassifikation Maligner Tumoren. International Union Against Cancer, pp 53–88. Springer
138. Wolfe JC, Mikheeva LA, Hagras H, Zabet NR (2021) An explainable artificial intelligence approach for decoding the enhancer histone modifications code and identification of novel enhancers in drosophila. Genome Biol 22:1–23
139. Wout O, Boiero Sanders M, Funk J, Prumbaum D, Raunser S, Bieling P (2024) Molecular mechanism of actin filament elongation by formins. Science 384(6692):9560. https://doi.org/10.1126/science.adn9560
140. Yagin FH, Cicek B, Alkhateeb A, Yagin B, Colak C, Azzeh M, Akbulut S (2023) Explainable artificial intelligence model for identifying covid-19 gene biomarkers. Comput Biol Med 154:106619. https://doi.org/10.1016/j.compbiomed.2023.106619
141. Yang CC (2022) Explainable artificial intelligence for predictive modeling in healthcare. Journal of healthcare informatics research 6(2):228–239

142. Zhang T-Y, Zhong M, Cheng Y-Z, Zhang M-W (2023) An interpretable machine learning model for real-time sepsis prediction based on basic physiological indicators. Europ Rev Med Pharmacol Sci 27(10)
143. Zieseniss A, Hesse AR, Jatho A, Krull S, Hölscher M, Vogel S, Katschinski DM (2015) Cardiomyocyte-Specific Transgenic Expression of Prolyl-4-Hydroxylase Domain 3 Impairs the Myocardial Response to Ischemia. Cell Physiol Biochem 36(3):843–851. https://doi.org/10.1159/000430260