Computers and Electrical Engineering 118 (2024) 109370

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

A review of Explainable Artificial Intelligence in healthcare

Zahra Sadeghi [a], Roohallah Alizadehsani [b,*], Mehmet Akif CIFCI [c,d,e], Samina Kausar [f], Rizwan Rehman [g], Priyakshi Mahanta [g], Pranjal Kumar Bora [g], Ammar Almasri [h], Rami S. Alkhawaldeh [i], Sadiq Hussain [j], Bilal Alatas [k], Afshin Shoeibi [l], Hossein Moosaei [m,n], Milan Hladík [o], Saeid Nahavandi [p], Panos M. Pardalos [q,r]

[a] *Institute for Big Data Analytics, Faculty of Computer Science, Dalhousie University, Canada*
[b] *Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Geelong, Australia*
[c] *The Institute of Computer Technology, Tu Wien University, 1040 Vienna, Austria*
[d] *Department of Computer Engineering, Bandirma Onyedi Eylul University, 10200 Balikesir, Türkiye*
[e] *Engineering and Informatics Department, Klaipėdos Valstybinė Kolegija/Higher Education Institution, 92294 Klaipeda, Lithuania*
[f] *University of Kotli Azad Jammu and Kashmir, Kotli Azad Kashmir, Pakistan*
[g] *Centre for Computer Science and Applications, Dibrugarh University, Assam, India*
[h] *Department of Management Information Sys, Al-Balqa Applied University, Salt 19117, Jordan*
[i] *Department of Computer Information Systems, The University of Jordan, Aqaba 77110, Jordan*
[j] *Examination Branch, Dibrugarh University, Dibrugarh, Assam, India*
[k] *Department of Software Eng., Firat University, 23100 Elazig, Turkey*
[l] *Data Science and Computational Intelligence Institute, University of Granada, Spain*
[m] *Department of Informatics, Faculty of Science, Jan Evangelista Purkyně University in Ústí nad Labem, Czech Republic*
[n] *Department of Econometrics, Faculty Informatics and Statistics, Prague University of Economics and Business, Prague, Czech Republic*
[o] *Department of Applied Math, School of CS., Faculty of Math. and Physics, Charles University, Prague, Czech Republic*
[p] *Distinguished Professor, Associate Deputy Vice-Chancellor Research, Swinburne University of Technology, Australia*
[q] *Center for Applied Optimization, Dept. of Industrial and Systems Eng., University of Florida, Gainesville, 32611, USA*
[r] *HSE University, Lab LATNA, Nizhny Novgorod, Rodionova street 136, Russia*

ARTICLE INFO

Keywords:
Explainable AI
Transparent AI
Interpretability
Healthcare

ABSTRACT

Explainable Artificial Intelligence (XAI) encompasses the strategies and methodologies used in constructing AI systems that enable end-users to comprehend and interpret the outputs and predictions made by AI models. The increasing deployment of opaque AI applications in high-stakes fields, particularly healthcare, has amplified the need for clarity and explainability. This stems from the potential high-impact consequences of erroneous AI predictions in such critical sectors. The effective integration of AI models in healthcare hinges on the capacity of these models to be both explainable and interpretable. Gaining the trust of healthcare professionals necessitates AI applications to be transparent about their decision-making processes and underlying logic. Our paper conducts a systematic review of the various facets and challenges of XAI within the healthcare realm. It aims to dissect a range of XAI methodologies and their applications in healthcare, categorizing them into six distinct groups: feature-oriented methods, global methods, concept models, surrogate models, local pixel-based methods, and human-centric approaches. Specifically, this study focuses on the significance of XAI in addressing healthcare-related challenges, underscoring its vital role in safety-critical scenarios. Our objective is to

* Corresponding author.
E-mail address: r.alizadehsani@deakin.edu.au (R. Alizadehsani).

https://doi.org/10.1016/j.compeleceng.2024.109370
Received 27 September 2023; Received in revised form 29 May 2024; Accepted 30 May 2024
Available online 7 June 2024

0045-7906/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

provide an exhaustive exploration of XAI's applications in healthcare, alongside an analysis of relevant experimental outcomes, thereby fostering a holistic understanding of XAI's role and potential in this critical domain.

1. Introduction

In the ever-evolving landscape of artificial intelligence (AI), Explainable Artificial Intelligence (XAI) has emerged as a beacon of trust, compliance, effectiveness, and robustness. XAI embodies methods and techniques that craft AI applications comprehensible not just to domain experts or data scientists, but also to laypersons unfamiliar with AI intricacies. The remarkable strides in deep learning (DL), coupled with its ubiquitous presence in real-world applications, have sparked an insatiable curiosity to unravel the reasoning behind its decisions. Users typically gravitate towards AI models that boast transparency — models whose rationale can be interpreted or elucidated with ease. Before delving deeper, let's demystify two terms often conflated yet distinct in meaning. Interpretability is about providing human-understandable rules that govern a system's decision-making process. In contrast, explainability pertains to crafting an interface that makes the inner workings of AI decision-making accessible and comprehensible to humans. The merit of AI explainability can be viewed through various lenses.

Interpretability involves providing human-understandable rules governing a system's decision-making process, whereas explainability focuses on crafting an interface that makes AI decision-making accessible and comprehensible to humans. The merit of AI explainability can be viewed from various perspectives. It involves creating a human-comprehensible interface to disentangle the internal AI decision-making function [1]. The importance of AI explainability can be discussed from diverse viewpoints [1–4]. For AI-driven healthcare solutions to promote trust, accountability, and transparency, interpretability of AI in healthcare is essential. Interpretability enables clinical adoption, confirms better accuracy, and reduces risks related to errors or biases by empowering physicians to comprehend and mitigate the process by which AI algorithms arrive at their decisions. In addition, interpretability guarantees adherence to ethical and legal requirements, promotes informed consent and patient engagement, and encourages continuous feedback. First, explaining machine learning (ML) models is vital for verifying sensitive models such as those related to the human healthcare system. Medical experts need to ensure the models are trained correctly and the parameters on which they are dependent are consistent with their knowledge. For instance, if the post-hoc analysis results of an ML model conclude that sneezing is a sign of cancer, the medical doctor can immediately imply that the ML model is not trustworthy. Secondly, complex ML models such as deep neural networks are usually trained on very high-dimensional data and encapsulate salient features [5]. Explaining these trained models will provide insightful information for experts in various fields of study such as Physics, Mathematics, and Chemistry.

AI explainability solutions based on post-hoc modelling, and analysis for ML models deciphering can be divided into model agnostic and model-specific methods. Model-agnostic approaches are general purpose and can be applied to almost all ML models regardless of their structure and training mechanism. One of the robust agnostic approaches is sensitivity analysis (SA) which attempts to reveal the contribution and impact of input factors on output prediction by changing input values and observing the amount of variation caused in the output [6]. These methods indicate the sensitivity level of the output on each of the input variables based on different statistical features such as variance derivative and density [7,8]. Sensitivity analysis can be applied globally or locally [9]. Local Sensitivity Analysis (SA) methods focus on the specific impact of minor modifications in input values, assessing how these small changes influence the output of the model. Essentially, they measure the output's sensitivity to these localized input perturbations. On the other hand, Global SA techniques adopt a broader approach. They evaluate the model's response by altering the entire spectrum of potential input values, providing a comprehensive view of how different input scenarios affect the output. This distinction between local and global methods is critical for understanding the varying degrees of influence that input parameters can have on the model's decision-making process [10].

In contrast to model-agnostic approaches, model-specific methods can only be utilized for specific ML models. For example, many explainability mechanisms are developed to analyze trained deep neural networks. These approaches are known as deep network understanding and visualization [11]. Activation Maximization, DeConvNet, inversion, deepDream, feature visualization analysis, and DeepLift are some of the popular methods from this category which attempt to find the contribution of neurons of convolutional neural networks on their final decision through optimization and backpropagation [12–16]. Moreover, specific approaches have been proposed for explaining Graph Neural Networks and Recurrent Neural Networks [17–19].

The causality approach focuses on uncovering cause-effect relationships between variables. Some researchers assert that a comprehensive interpretation of a model necessitates revealing the rationale behind its decisions. In this regard, counterfactual is a promising strategy to find the features contributing to a specific outcome [20]. Another factor used in some AI explainability techniques is the model's ability to engage with end-users. Recent studies have focused on developing hybrid approaches which result in transparent models with representation power of existing black box architectures such as DL. Contextual Explanation Network (CEN), Self-Explaining Neural Network (SENN), BagNet, and TabNet are typical examples of transparent models [21–24]. The end-users' interest in understanding the reasons behind the decisions made by ML models demands further research. Model transparency is especially important in safety-critical applications such as medical domain. Therefore, our focus is on XAI in the healthcare domain and its challenges. The primary goals of our study are listed below:

- XAI methods identification and categorization
- XAI literature review with special focus on healthcare domain

- Ascertainment of XAI challenges and problems in healthcare.

The papers reviewed in this survey are selected using the keywords “Explainable AI” and “Interpretable Machine Learning”, with a focus on “healthcare.” The queries returned multitude of journal and conference articles. For the purpose of this survey, only peer-reviewed articles are considered for a systematic review. The Prisma model in Fig. 1 demonstrates the overall papers reviewed in the study while Table 1 describes the explored databases. To ensure thorough coverage and depth in our research, we adopted a dynamic and iterative approach in our literature review process. We continuously expanded our search, beginning with initial keywords like ’Explainable AI’ and ’Interpretable Machine Learning.’ Each relevant article we found led us to its references, which we scrutinized for additional pertinent studies. This cycle of discovery and exploration was repeated meticulously. We pursued this methodical expansion until it became evident that further searches were yielding no new articles, thereby affirming the comprehensiveness of our literature collection.

As seen in Fig. 1, the process of identifying relevant research papers involved a structured approach encompassing identification, screening, eligibility, and inclusion of selected papers. This methodology is depicted in a step-by-step flow chart in Fig. 1. The search was conducted across various academic databases such as Google Scholar, Elsevier, Springer, and others, as detailed in Table 1. In the initial phase, a total of 324 articles were identified for potential review. However, 45 of these articles, not pertaining to the healthcare sector, were immediately excluded. Further scrutiny of titles and abstracts led to the exclusion of 16 articles, mainly due to their lack of relevance or focus on the core topic. Additionally, preprint versions and duplicates were identified and removed. A thorough evaluation of the quality of the remaining published works resulted in the selection of 200 articles. Out of these, 67 were further excluded as they were not research articles or lacked substantive content, leading to the final selection of 113 comprehensive, relevant articles for an in-depth review.

Table 1 presents the outcome of searches for pertinent articles across various scholarly database search engines. The column labeled "Number of Search Results" reflects the total articles retrieved from each database using specified keywords. The "Number of Relevant Articles" column shows the count of articles that successfully passed the preliminary screening phase, qualifying them for potential

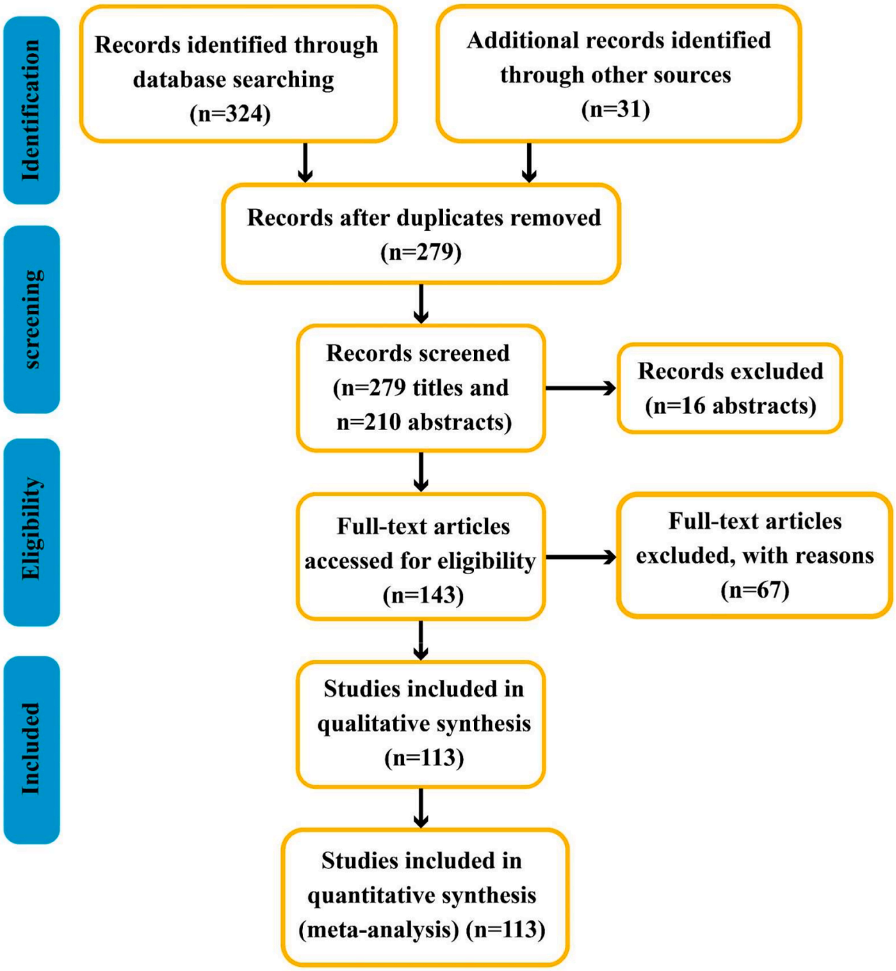
Identification
screening
Eligibility
Included
Records identified through database searching (n=324)
Additional records identified through other sources (n=31)
Records after duplicates removed (n=279)
Records screened (n=279 titles and n=210 abstracts)
Records excluded (n=16 abstracts)
Full-text articles accessed for eligibility (n=143)
Full-text articles excluded, with reasons (n=67)
Studies included in qualitative synthesis (n=113)
Studies included in quantitative synthesis (meta-analysis) (n=113)

Fig. 1. PRISMA Model for the depiction of inclusion and exclusion of records.

Table 1
Summary of search results and retrieved relevant articles.

Database Engines	Source Address	Number of search results	Number of relevant articles
Elsevier	https://www.elsevier.com	1500	25
Springer	https://www.springer.com	1200	19
Taylor & Francis	https://taylorandfrancis.com	800	13
Semantic Scholar	https://www.sematicscholar.org	500	16
ACM Digital Library	https://www.acm.org	1000	17
IEEE Xplore	https://ieeexplore.ieee.org	2000	23

inclusion in an extensive review. Among the databases, IEEE Xplore yielded the greatest volume of search results (2000), while Taylor & Francis produced the least (800). It is important to note that the quantity of relevant articles does not necessarily correlate directly with the total number of search results. This is because some articles may be filtered out during the screening phase. For instance, despite not having the highest search result count, Elsevier surfaced the most relevant articles for further review.

XAI sheds light on black box ML models to aid with understanding the logic behind their decision making. Black box models may not even be explainable by their designers [25]. Explainability is an influential tool for justifying AI based decisions. It can assist in validating predictions, enhancing models, and gaining new insights into the problem at hand which leads to more trustworthy AI systems.

Research motivations for implementing XAI systems are graphically depicted in Fig. 2. As can be seen, there are several key drivers for XAI development such as increasing model transparency, improving accountability, and enhancing trust in AI systems. Other important drivers include the need for regulatory compliance, the desire for more effective decision-making, and the importance of ethical considerations in AI development [26].

This paper introduces a novel framework that extends beyond traditional approaches in Explainable Artificial Intelligence (XAI). While existing literature predominantly focuses on the technical aspects of explainability and interpretability, our work delves into the integration of these concepts with user-centric design principles. We propose a unique interdisciplinary approach that combines insights from cognitive psychology, user experience (UX) design, and machine learning. This fusion aims to not only make AI systems transparent but also intuitively understandable and usable for a broad spectrum of users, ranging from AI experts to laypersons. Our framework emphasizes the development of explainable AI tools that are not only technically sound but also empathetic to the cognitive and emotional needs of users. By bridging this gap, we contribute to a more inclusive and accessible understanding of AI systems, fostering trust and facilitating wider adoption in various domains, especially in critical sectors like healthcare and finance. This perspective is particularly crucial as AI continues to permeate diverse aspects of our daily lives, making the need for comprehensible and user-friendly AI explanations more pressing than ever.

The remainder of the document is structured as follows: Section I provides an overview of XAI methods. In Section II, we delve into related works on XAI methods. Section III highlights the XAI related tools that could be used in healthcare applications. Section IV focuses on the use of XAI as a perfect decision-making choice in healthcare domain. Section V narrates the implication of XAI in healthcare applications. Section VI focuses on challenges of interpretability using XAI in healthcare. Finally, Section VII concludes the article with the key summary.

2. Explainable artificial intelligence methods

Research in XAI can be categorized into six main groups: Feature-oriented methods, global methods, concept models, surrogate models, local pixel-based methods, and human-centric methods.

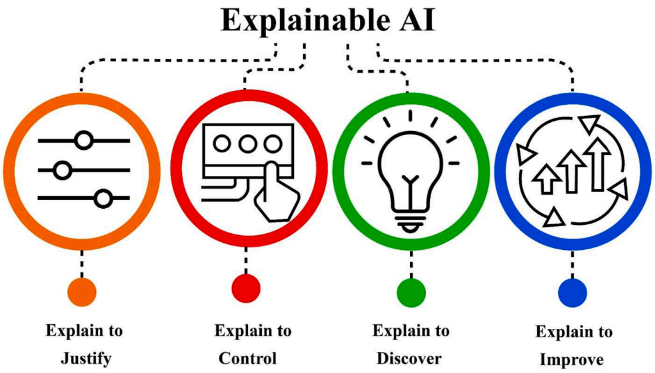

Fig. 2. Motivations for XAI.

2.1. Feature-oriented methods

Shapley Additive exPlanation (SHAP) employs game theory to explain the outcomes of ML techniques. For each sample $x = [x_1,\ldots, x_n]$,contribution of each feature$x_j$tothe prediction f(x) of an ML model is computed using Shapley values by assuming $\{x_1,\ldots, x_n\}$as players in a coalition game [27] expressed as (v, $N=\{x_1,\ldots, x_n\}$). The payoff function v: $2^N \rightarrow$ R, v(∅) = 0maps subset of features (cooperative players) to the real numbers [28]. For a subset of features S, v(S) is equal to the expected sum of payoffs obtained via cooperation of features in S. Once the payoff function is defined, the Shapley value of j-th feature (ϕ_j(v)) can be computedas the average marginal contribution of the j-th feature to the payout as shown in Eq. (1):

$$\phi_j(v) = \sum_{S\subseteq N\setminus\{j\}} w_{S,N}\left(v\left(S\bigcup\{j\}\right) - v(S)\right) \tag{1}$$

where the summation is computed over all possible coalitions S such that j-th player is excluded. Moreover, $w_{S,N}$ is the weight factor computed as shown in Eq. (2):

$$w_{S,N} = \frac{|S|!(|N|-|S|-1)!}{|N|!} = \binom{|\mathrm{N}|}{1, |\mathrm{S}|, |\mathrm{N}|-|\mathrm{S}|-1} \tag{2}$$

where |S| is the cardinality of subset S and $w_{S,N}$is equal to the inverse of multinomial coefficient representing the number of different ways of forming coalition using subset S of N excluding j-th feature (i.e.S⊆N\{j}). Using Shapley values {ϕ_j(v),$j = 1, \ldots, n$}, the SHAP explanation can be computed as shown in Eq. (3):

$$g(z') = \phi_0 + \sum_{j=1}^{n} \phi_j z_j^{'}, \tag{3}$$

where $z' = [z_1', \ldots, z_n'] \in \{0,1\}^n$ is a vectors of zeroes and ones.For $z_j' = 0$, j-th feature is not part of the coalition whereas $z_j' = 1$ means j-th feature is present in the coalition. Moreover, $\phi_0 = E[f(x)]$ is the average predicted value computed over all features.

It is not easy to optimize the SHAP method's implementation for each model type, even if it may be used for several models. The high-level ontology of explainable methods to artificial intelligence is given in Fig. 3. CNNs often produce class activation maps (CAMs) [29]. Per-class weighted linear summation of distinct spatial patterns occurring in an image is indicated by CAMs. Before the output layer, the final convolutional feature map is sent to global average pooling. The input features to a fully connected layer and the output features of a loss function are then produced from this pooled feature map data. By re-projecting the output weights back to the previous convolutional layer, a heatmap depiction of the input image highlights the regions that have a stronger effect on the CNNs' choice. For fully convolutional neural networks, CAMs cannot be used with trained models or those that do not follow the defined architectural guidelines. CAM and Grad-CAM [30,31] are based on the assumption that for each specific class c, the final score f^c of the network can be expressed as in Eq. (4):

$$Y^c = \sum_k w_k^c \sum_i \sum_j A_{ij}^k, \tag{4}$$

Where final score Y^c, w_k^c is the weight corresponding to A^k which is the k-th feature map of the last convolutional layer and A_{ij}^k is the value at row i and column j of A^k. Given Eq. (1), for class c, the saliency map value at each location (i,j) is computed as shown in Eq. (5):

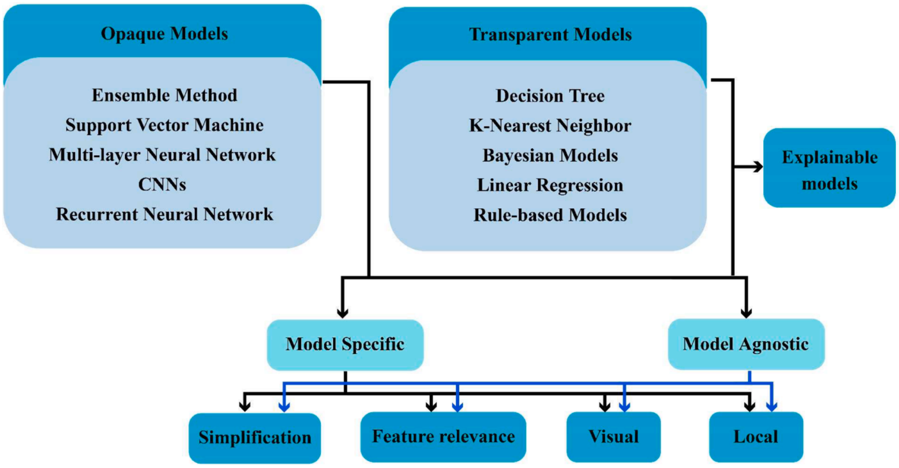

Fig. 3. High-level ontology of XAI approaches.

$$L_{ij}^{c} = \sum_{k} w_{k}^{c} A_{ij}^{k} \tag{5}$$

whereL_{ij}^{c} reflects the importance of location (i,j) for class c. Therefore, L^{c} acts as a visual explanation corresponding to class c that has been predicted by the network. In CAM method, weight values w_{k}^{c} are estimated by training multiple linear classifiers (one per each class). Moreover, CAM assumes that the penultimate layer of the CNN in question is global average pooling which is not always the case.Grad-CAM was proposed to address the shortcomings of CAM. It has been shown that for each feature A^{k}, weight values w_{k}^{c} can be computed as shown in Eq. (6):

$$w_{k}^{c} = \frac{1}{Z} \sum_{i} \sum_{j} \frac{\partial Y^{c}}{\partial A_{ij}^{k}} \tag{6}$$

where Z is number of pixels in A^{k}. Despite addressing issues of CAM method, Grad-CAM yet suffers from drawbacks such as failing to properly localize objects in case input image contains multiple instances with identical class labels. Moreover, the averaging in Eq. (2) is unweighted which leads to localizing only parts of objects instead of their entirety. To overcome these issues, Grad-CAM++ was proposed in which the global averaging in Eq. (2) is reformulated as shown in Eq. (7):

$$w_{k}^{c} = \sum_{i} \sum_{j} \alpha_{ij}^{kc} . relu\left(\frac{\partial Y^{c}}{\partial A_{ij}^{k}} \right), \tag{7}$$

whereα_{ij}^{kc} is the weight corresponding to pixel (i,j) of k-th feature map A^{k}. Using Eq. (6), w_{k}^{c} specifically captures the importance of feature mapA^{k} by using weighted average of partial derivatives. This is in contrast to Eq. (5) in which unweighted averaging is performed. Plugging Eq. (5) into Eq. (3), taking derivatives twice with respect to A_{ij}^{k} on both sides and performing some manipulation yields Eq. (8):

$$\alpha_{ij}^{kc} = \frac{\left(\frac{\partial S^{c}}{\partial A_{ij}^{k}} \right)^{2}}{2\left(\frac{\partial S^{c}}{\partial A_{ij}^{k}} \right)^{2} + \sum_{a} \sum_{b} A_{ab}^{k} \left(\frac{\partial S^{c}}{\partial A_{ij}^{k}} \right)^{3}} \tag{8}$$

Fig. 3 illustrates a critical distinction in the domain of Explainable AI: while feature-oriented approaches are adept at pinpointing the specific input features that influence a decision, they fall short in offering a human-level, comprehensive explanation of the model's underlying reasoning process. These approaches, effective in mapping decision-making to certain features, do not elucidate the 'how' and 'why' behind a model's conclusions. This gap becomes particularly pronounced in sectors like healthcare and finance, where understanding the rationale behind an AI's decision is as important as the decision itself.

2.2. Global methods

Global Attribution Mappings (GAMs) are instrumental in elucidating neural network predictions across diverse populations, offering significant advantages, particularly in identifying and analyzing distinct subpopulations at varied levels of granularity. These mappings effectively create a pairwise rank distance matrix, capturing the intricate relationships between different features. This process involves grouping similar local features using K-medoids clustering, a method that enhances the mapping's precision. Each cluster identified by GAMs is represented by a 'medoid, ' which essentially acts as a reference point. This medoid helps in constructing a global attribution for the cluster by analyzing and synthesizing the patterns observed within it. Consequently, GAMs enable the detailed examination of characteristics unique to various sample groups, thereby enhancing the comprehensiveness of neural network interpretations. In the context of the class most commonly predicted by the network, GAMs employ a specialized technique. They generate a normalized heatmap based on a gradient-based saliency map, which reveals the absolute values of gradients in relation to the input attributes. This heatmap highlights pixels indicating high activation levels, signifying their dominant impact (or high saliency) on the model's decision-making process. Gradient-based saliency maps, like feature-oriented techniques, cannot articulate judgments beyond model diagnosis [32]. Deep attribute maps are explored to increase the explainability of gradient-based algorithms [33]. The model prediction is displayed as a heatmap utilizing the output gradient's important feature, multiplied by relevant input data to compare alternative saliency-based explanation models.

2.3. Concept models

As pointed out by [34], the complex feature spaces of deep neural networks are not necessarily an obstacle; quite the opposite, deep features can be used to our advantage for model interpretability. The researchers have also proposed Concept Activation Vectors (CAVs) to represent neural network internal state in human-understandable format. To this end, directional derivatives are used to measure the sensitivity level of neural network output to user-defined concepts. As an example, suppose a neural network is trained to distinguish pictures of horses from zebras. Using CAT, it is possible to measure the contribution level of the animal having body stripes to being classified as zebra. The hyperplane separating activation values corresponding to concept and random examples is determined. The normal vector of the aforementioned hyperplane is considered as the CAV. Concept Activation Vectors (CAVs) offer a human-friendly conceptual description of the internal state of a neural network which can be for the practical issue of using retinal fundus images to forecast diabetic retinopathy (DR), a condition that is curable but can cause blindness [35].

2.4. Surrogate models

Local interpretable model-agnostic explanation (LIME) constructs locally optimal explanations of ML models using an interpretable surrogate model. While explaining the working mechanism of complex black box models is challenging, it is possible to explain their behavior for a specific input sample. LIME method starts by modifying parts of the given input sample to generate a dataset of perturbed instances similar to the original input but not exactly the same. The perturbation depends on the nature of the input sample. For example, for input of type image, some parts of it can be replaced with gray color to obtain its perturbed counterparts. Figure depicted this process in which the boundaries between image parts (called super-pixels) are shown with yellow color. Fig. 4 shows the five high-level steps of the LIME method for generating explanations for individual predictions made by a black-box classifier. As can be seen, the LIME method allows users to generate explanations for individual predictions made by the black-box classifier by creating an interpretable model that approximates the black-box classifier in the vicinity of the chosen data sample. Determining Alzheimer's disease (AD) was one usage of the LIME approach [36].

2.5. Local, pixel-based methods

Layer-wise relevance propagation (LRP) employs specified propagation rules to explain a multi-layered neural network's output related to the input. The approach produces a heatmap, offering insight into which pixels contributed to the model's prediction and to what extent. Consequently, LRP emphasizes those variables that positively affect a network's choice. LRP may be applied on an already-trained network to simplify the features' decision-making process if the network employs backpropagation rather than forwards propagation. DeconvNet employs a semantic segmentation approach that constructs a deconvolution network and contributes pixels during the classification process [37]. LRP may be very useful in helping physicians explain neural network diagnoses of Alzheimer's Disease (and possibly other disorders) that are made using structural MRI data. In multilayered neural networks, Layer-wise Relevance Propagation (LRP) is a technique that may be used to visualize the contribution of individual pixels to the predictions of kernel-based classifiers over Bag of Words information. To compute the gradient locally, it can be difficult to identify an appropriate reference point. A difficulty that may be taken into consideration in subsequent steps is defining new methods for locating a reference point for calculating the gradient in a predetermined neighborhood [38].

2.6. Human-centric methods

The previous techniques have failed to offer humans logical explanations despite their benefits. They try to "damage limit" the "black box" by "only touching the surface" by using post-hoc indications about features (attribute allocation) or places within an image. This is in contrast to how people think, develop connections, assess similarities, and draw a comparison. Model structure and

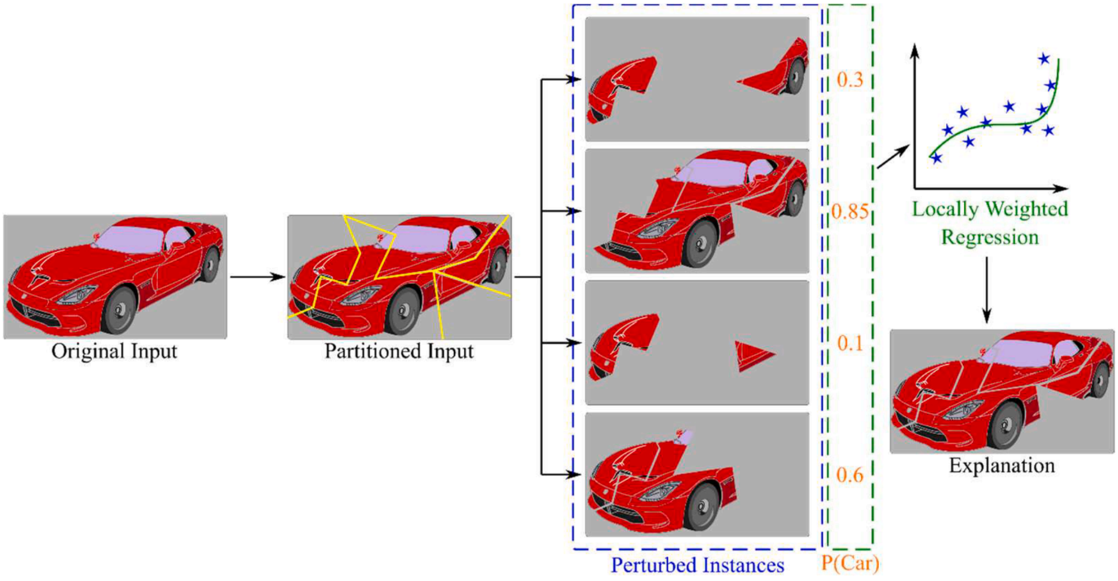

Fig. 4. High level steps of LIME method.

parameters, which are critical to the problem's nature, are utterly ignored by the methods described above so is the logic.

Recently, a fundamentally new approach to explainability was proposed which is based on human-centric (anthropomorphic) understanding rather than pure statistics [39]. Humans assess commodities (e.g., photographs, music, and movies) in their totality, not by feature or pixel. People utilize similarity to relate new data to previously learned and aggregated prototypes [40], whereas statistics employ averages as their foundation.

3. Explainable Artificial Intelligence tools

To facilitate the behavior analysis of AI models, various tools have been developed. Some of these tools are briefly introduced below:

- ELI5 library developed by MIT is a Python library for visualizing and debugging ML models. ELI5 supports various ML frameworks such as Scikit-learn, Keras, LightGBM, etc. The prepared explanations are offered in different data formats such as text, HTML, web dashboard, or JSON. A recent genome sequencing effort has revealed the presence of more than 4300 potential genes responsible for encoding proteins. Using the pool described in [41], create a protein microarray consisting of 673 potential proteins. The protein microarray analysis revealed the identification of 30 antigens that showed differential reactivity. These antigens were then included in the two expression library vaccines. The paper by [42] presents a technique for immunizing alpacas and utilizing molecular biology methods to generate single domain antibodies that target specific antigens. B-cells are utilized to create an immunized library, which is then employed in the process of selecting specific single domain antibodies through panning. The researchers [43] have discovered a new member of the family in the species Hermetia illucens. By employing real-time PCR, the presence of HI-astacin was scarcely observed prior to immunization but became predominantly apparent in the fat body following immunization. [44] conducted simulations using five scRNA-seq library protocols and nine realistic differential expression (DE) setups. They systematically evaluated three mapping, four imputation, seven normalization, and four differential expression testing approaches. This resulted in approximately 3000 pipelines, which allowed them to assess interactions among pipeline steps. The study conducted by [45] reveals that the library preparation method plays a crucial role in detecting symmetric expression differences. The bacterial expression of multiple clones from the library exhibited a significant yield of soluble recombinant proteins. Essential Highlights Initial TBE vaccination elicits a varied degree of CD4+ *T* cell response. [46] aimed to analyze and describe the immune response of human CD4+ *T* cells during the initial stages of TBE immunization. The recommended primary immunization regimen for European TBE vaccines involves administering three vaccine doses over the course of one year. The study conducted by [47] aimed to describe the human CD4 T cell response during primary TBE immunization. This bispecific antibody (bsAb) was obtained through an Omniflic immunization using the antigens CEACAM5 and CEACAM6, which are both significant markers for colorectal cancer. Furthermore, the objective is to specifically introduce the pH-responsive binding for CEACAM5, as stated by [48]. [49] have identified the histone methyltransferase mixed lineage leukaemia 1 (Mll1) as a facilitator of TFH differentiation. Reduced expression of Mll1 hindered the development of TFH cells after acute viral infection or protein immunization.
- AI Fairness 360 from IBM is an open-source library developed to ease the process of detection and alleviation of bias in ML models as well as datasets. This library is available in python and R. Multiple studies have investigated the implementation of the H2O platform in the healthcare sector. [50] created an Internet of Things (IoT) system specifically designed for elderly healthcare. The system's main focus is on monitoring activities in real-time and providing immediate emergency assistance. In 2010, [51] put forward a healthcare information integration and shared platform that is built on Service-Oriented Architecture. This platform facilitates the consolidation of data and enables interoperability between different systems. [52] presented the Healthcare Integration Platform, which employs IHE profiles and EHR standards to guarantee interoperability. In his 2020 publication, [53] conducted a thorough examination of the Human Healthcare Internet of Things, exploring its capacity to enhance the quality of healthcare and its cost-effectiveness. These studies collectively demonstrate the potential of the H2O platform to improve healthcare services. Automated ML, also known as AutoML, seeks to create Machine Learning (ML) models with minimal input from data scientists. AutoML platforms streamline the pre-processing of data by automating tasks such as handling missing values, scaling, and dropping duplicates. Some AutoML platforms, like H2O and Data Robot, also provide feature engineering capabilities [54]. Nevertheless, none of these AutoML platforms possess the capability to incorporate domain knowledge into the machine learning models, a skill that is unique to humans. [55] compares the output of AutoML models with that of a manually created model. Interpretability packages from Microsoft [56] offer several model-agnostic/specific explainers (e.g. SHAP tree explainer and SHAP deep explainer) for ML models based on SHAP framework. These packages are available as part of Responsible AI dashboard of Microsoft Azure Machine Learning. These explainers facilitate the understanding of complex machine learning models by providing insights into feature importance and model predictions, aiding in model transparency and trustworthiness.
- "What If Tool" (WIT) offered by Google [57] provides visual probing for trained ML models behaviors while reducing required coding as much as possible. WIT can be integrated with several platforms such as Jupyter/Colaboratory/Could AI notebooks, TensorBoard, etc. Moreover, analysis tools for various ML problem categories such as binary/multi-classification and regression are available for wide range of data types such as tabular, image, and text. This versatility makes WIT a comprehensive solution for exploring and understanding model behavior across different data modalities and problem types, enhancing interpretability and decision-making in machine learning workflows.
- H2O platform offered by H2O.ai [58] aims to accelerate development and deployment of AI models in various business problems. H2O hides the technical details of AI models from users to enable development of AI applications without writing any codes. This

abstraction allows domain experts and business stakeholders to leverage the power of AI without needing extensive programming knowledge, streamlining the development process and fostering collaboration between data scientists and subject matter experts.

- Distill [59] was research conducted to combine different interpretability techniques to enhance analysis of neural networks decision making process. To this end, distill exploits the complementary roles of interpretability methods such as feature visualization, attribution, and dimensionality reduction by treating them as building blocks of a unified interface. Distill offers a holistic view of neural network behavior, facilitating deeper insights into model decisions and enhancing trust in AI systems for diverse applications.
- Skater developed by Oracle [60] is a python library for analyzing behavior of trained models both globally and locally. In global scenario, the analysis is done based on inference over a complete dataset. In local scenario, the analysis is performed based on prediction for a single sample. Skater's versatility in offering both global and local interpretability allows users to gain insights into model behavior at various levels of granularity, empowering them to understand model decisions comprehensively across different contexts and datasets.

4. Explainable AI for decision makers

XAI examines if ML models can be more understandable to humans and aims to enhance their effectiveness as well as making them feasible for non-experts to apply in diverse settings. Despite several attempts made by researchers, yet the exact definitions of concepts such as interpretability and explainability are not readily available. Transparency and XAI popularity have increasing trend in response to resolving the AI "black box" dilemma. An ML model's output may be better understood using XAI methods. Additionally, explainability is related to the explanation as a means of communication between humans and decision-makers in a reliable and comprehensible manner. Moreover, the increasing demand for high-quality medical services can be addressed by AI-based healthcare systems. Yet, the majority of these AI-based models remain prototypes and never make it to market. The models were unsuccessful to live up to performance expectations, crucial risky judgments, and depended on differentiating traits to anticipate outcomes. Numerous programs such as Google Health have all failed in some way when placed into test or production [61]. This has led to outpouring criticism about AI-based outcomes due to the potentially dire ramifications for humans making high-stakes medical choices. High complexity of the underlying models, large quantity of the datasets, and the huge processing power required to boost the performance of ML models are typical features behind AI models [62]. However, as these models get more complex, it becomes more challenging to grasp their operation, data processing, and decision making which is why they are referred to as opaque or black-box models [63].

There has been an attempt to investigate various explainability methods [53]. First, approaches were grouped into four categories with the following objectives in mind: 1. describing black-box models, 2. assessing black box models, 3.explaining their outcomes, and 4. building transparent black-box models. Moreover, a taxonomy was proposed to express the underlying explanator, input data type, the issue discovered by the approach, and the "opened" black box model. It has been shown that most explanation methods are unable to decipher models [64]. This involves making judgments based on unknown or latent traits. Finally, an explanation is provided for the lack of interpretability techniques in the suggested systems. In addition, an approach was proposed for learning models directly from explanations after recognizing a lack of formality and performance evaluation of interpretability methods. Interpretability framework has also been proposed for developing predictive accuracy, descriptive accuracy, and relevancy [65].

Misztal-Radecka and Indurkhya [66] proposed a taxonomy to classify explainability of DL which includes explaining deep network processing, explaining deep network representation, and explaining how systems are built. Last but not least, establishing transparent or explicable models, fostering cross-disciplinary cooperation between healthcare practitioners is crucial for information sharing [67].

Many scientists believe that XAI make AI deployment easier in the medical industry since XAI helps with creating trust and understanding among stakeholders about AI system [68]. According to research on the information required when a difficult model is put into a decision-making environment, the information given by XAI frameworks is of primary importance. For instance, it has been observed that physicians are interested in the local, case-specific logic behind a model choice and the model's global properties [69]. A fundamental challenge in the adoption of AI in critical domains like healthcare is the difficulty in trusting the model output, largely due to a lack of understanding of how the model functions in [70]. This trust could be significantly bolstered by gaining deeper insights into the algorithm's development process, thereby enhancing confidence in the outcomes generated by AI applications. Despite the critical need, there has been scant research focused on identifying the specific types of information necessary for effectively integrating AI into decision-making processes, especially in the medical field.

The authors of [71] proposed a taxonomy for neural network interpretability methods with three separate groups. However, the primary focus was on DL. The first category covers methods that mimic data processing to give insights into the links between the model's inputs and outputs. Using a variety of specialized scientific journals, the authors of [72] conducted a comprehensive review of the literature. They underlined the necessity to integrate more formalism into the field of XAI and the interaction between humans and robots. After recognizing the community's predisposition toward examining explainability exclusively through the lens of modelling, they proposed adding explainability into other elements of ML. Finally, they provided a viable research direction which is synthesizing existing explainability methods. The authors of [73] have begun to address the call for additional knowledge about the structure of explanations through conceptual papers that have discussed the origins of explanations, how we are biased in our interpretations of explanations, and how explanations are phenomena that occur in the context of human interaction.

Transparent AI methodologies implemented in medicine are crucial, ensuring the capacity to explain and interpret AI-generated decisions alongside accurate diagnoses and predictions. The development of transparent AI within healthcare marks a significant stride towards accuracy, interpretability, and ethical AI utilization in the medical domain. In radiology and pathology, Transparent AI is being utilized to enhance diagnostic accuracy [74]. Transparent AI is pivotal in CDSS, offering clinicians clear justifications for

AI-generated recommendations. Models embedded with explainable features instill trust among healthcare providers, empowering them to validate suggestions and make more informed, precise decisions [75]. Transparent AI methodologies predict patient outcomes and disease progression, providing understandable insights into prediction factors. This aids healthcare professionals in comprehending AI-based prognoses, enhancing patient care strategies [76]. These techniques also empower researchers to validate predictions, aiding in drug and therapy discoveries. Machine learning models are utilized for prognostication of patient outcomes and disease progression. Transparent AI plays a crucial role in upholding ethical standards within healthcare AI applications. Initiatives focus on crafting AI systems transparent in their handling of patient data, ensuring privacy and ethical compliance.

5. Applications of explainable AI in healthcare

Nowadays, Artificial Intelligence (AI) plays an essential role in pursuing critical systems such as education, healthcare, renewable energy, transportation, and traffic that directly impact our daily lives. In the healthcare domain, the applications of AI techniques are in constant progress [77]. However, AI applications and models in healthcare practice require transparency and explainability since inaccurate predictions may have severe consequences [78]. Clinicians demand to understand AI systems reasoning as a prerequisite for building trust in the predictions and adoption of AI applications [79]. Reliability, accuracy, and transparency are critical requirements, especially for healthcare decision-makers. Therefore, AI researchers and practitioners have focused on explaining the decisions made by AI applications such as ML or Deep Learning (DL). AI algorithms should provide clinicians with understandable explanations about their outputs [80]. For example, in disease diagnosis, XAI can reveal the features that contribute to AI model output on patient's condition.

To understand the relationship between microbial communities and phenotypes, SHapley Additive explanation (SHAP) algorithm was used. The motivation is that SHAP technique can explain the prediction of specific prototype values depending on the outputs of

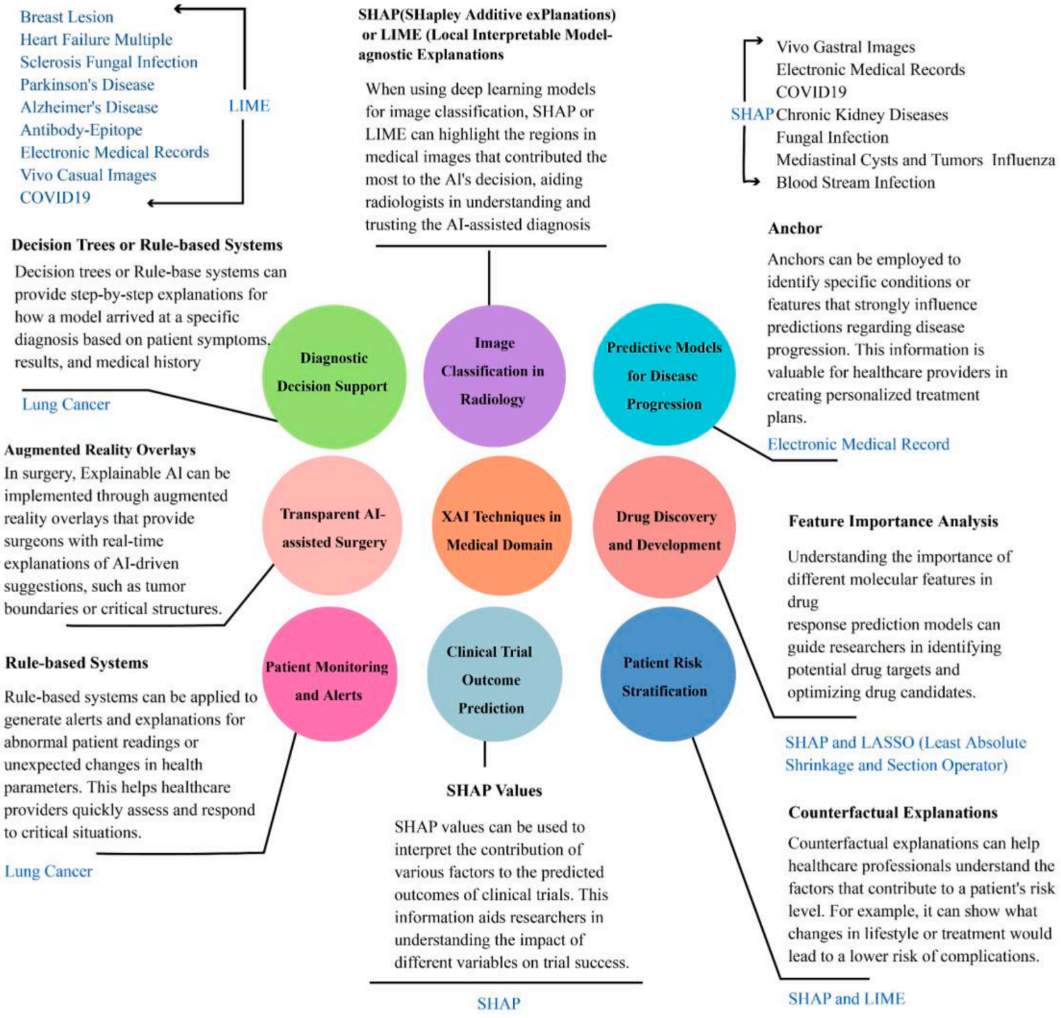
Breast Lesion
Heart Failure Multiple
Sclerosis Fungal Infection
Parkinson's Disease
Alzheimer's Disease
Antibody-Epitope
Electronic Medical Records
Vivo Casual Images
COVID19
LIME
SHAP(SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations)
When using deep learning models for image classification, SHAP or LIME can highlight the regions in medical images that contributed the most to the AI's decision, aiding radiologists in understanding and trusting the AI-assisted diagnosis
SHAP
Vivo Gastral Images
Electronic Medical Records
COVID19
Chronic Kidney Diseases
Fungal Infection
Mediastinal Cysts and Tumors Influenza
Blood Stream Infection
Decision Trees or Rule-based Systems
Decision trees or Rule-base systems can provide step-by-step explanations for how a model arrived at a specific diagnosis based on patient symptoms, results, and medical history
Lung Cancer
Anchor
Anchors can be employed to identify specific conditions or features that strongly influence predictions regarding disease progression. This information is valuable for healthcare providers in creating personalized treatment plans.
Electronic Medical Record
Diagnostic Decision Support
Image Classification in Radiology
Predictive Models for Disease Progression
Augmented Reality Overlays
In surgery, Explainable AI can be implemented through augmented reality overlays that provide surgeons with real-time explanations of AI-driven suggestions, such as tumor boundaries or critical structures.
Transparent AI-assisted Surgery
XAI Techniques in Medical Domain
Drug Discovery and Development
Feature Importance Analysis
Understanding the importance of different molecular features in drug response prediction models can guide researchers in identifying potential drug targets and optimizing drug candidates.
Rule-based Systems
Rule-based systems can be applied to generate alerts and explanations for abnormal patient readings or unexpected changes in health parameters. This helps healthcare providers quickly assess and respond to critical situations.
Lung Cancer
Patient Monitoring and Alerts
Clinical Trial Outcome Prediction
Patient Risk Stratification
SHAP and LASSO (Least Absolute Shrinkage and Section Operator)
SHAP Values
SHAP values can be used to interpret the contribution of various factors to the predicted outcomes of clinical trials. This information aids researchers in understanding the impact of different variables on trial success.
SHAP
Counterfactual Explanations
Counterfactual explanations can help healthcare professionals understand the factors that contribute to a patient's risk level. For example, it can show what changes in lifestyle or treatment would lead to a lower risk of complications.
SHAP and LIME

Fig. 5. Popular XAI Techniques largely applied to Medical Use cases with probable Explanations.

each impactful parameter. The features positively impact predicting the target value if the SHAP values are positive. Dopaminergic imagery techniques like SPECT DaTscan were analyzed for early diagnosis of Parkinson's Disease [81]. The LIME technique was used to accurately classify the Parkinson's disease from the given DaTscan with the appropriate reasoning of the same. Acute critical illness detection is another important use case of XAI approach in medical research. For example, an early warning score system has been proposed [82]. The system was able to explain its prediction with the Electronic Health Record data information using SHAP technique. XAI for the diagnosis of Glioblastoma based on topological and textual features has been investigated as well [83]. The AI model on the fluid-attenuation inversion recovery for the Glioblastoma multiform classification was validated. The local feature relevance to the sample in the test set was computed using LIME method. The Fig. 5 outlines the various XAI techniques commonly employed in medical contexts and hints at their potential explanatory capabilities, particularly in using SHAP and LIME in the field of medical applications.

A computer-aided system with the capability of explainable sentences has been proposed for lung cancer diagnosis [84]. The LIME method was used to develop the local post hoc model, which transforms the critical, relevant feature into natural language. An ensemble clustering based XAI model was proposed for diagnostic analysis of traumatic brain injury [85]. In this model, the expert medical knowledge was combined with the automated data analysis to develop the explainable framework. A model for the COVID-19 detection using the chest X-ray images named COVID-NET has been proposed [86]. This model showed 93.3 % accuracy and a 91.1 % sensitivity on the Covid dataset. The XAI technique Inquire was used to examine the COVID-NET model's output. To predict the post-stroke hospital discharge disposition, an interpretable ML method has been proposed [87]. Linear regression was selected as the baseline model and was compared with the black-box model. The LIME method was used to identify the essential features of the model. For selecting the laser surgery option at an expert level, a multicast XGBoost model has been proposed [88]. The model was validated on the subject who has undergone refractive surgery. An accuracy of 78.9 % was achieved on the external validation dataset. The SHAP method was used to provide a clinical understanding of the ML method.

Ye et al. proposed [89] classification models for COVID-19 equipped with XAI strategies to deliver reliable classification associated with credible explanations for COVID-19. To this end, 380 positive (having COVID-19) and 424 negative CT volumes were obtained. Their model can assist radiologists in determining exact location of lesions in CT scans by providing more diagnostic information. They compared their proposed XAI modules with other models like CAM, SHAP, etc.

Another work [90] has presented a B5G architecture for detecting COVID-19 utilizing CT scan images. The inspection system was developed based on different functionalities of 5 G networks. XAI model was utilized to monitor mask-wearing, social distancing, and body temperature. The suggested healthcare framework utilizes three layers which are edge layer, stakeholder layer, and cloud layer. In the middle edge layer, the Local Interpretable Model-Agnostic (LIMA) model is used by the XAI module Knowledge mapping. Images from X-rays, CT scans, and ultrasounds, along with learnable parameters from different layers of the DL model form the input. Their methodology aids with the reduction of hospital overcrowding, the verification of non-COVID-19 patients, and the processing of sensitive personal data at the edge to protect anonymity.

The researchers of [91] proposed a model for identification and progress prediction of Alzheimer's disease (AD). The approach consists of two layers. The first layer utilizes random forest (RF) to classify patients who have AD in its early stage. The second layer of binary classification is carried out to predict likelihood of mild cognitive impairment (MCI) progression toward AD within three years from the initial diagnosis. In each layer, SHAP is used for providing explanation on the model reasoning. The evaluation of the proposed approach was done on ADNI dataset achieving high reliability of 93.95 % and 87.08 % per layer.

Sepsis must be diagnosed as soon as possible since delayed treatment leads to patient's irreversible organ damage increasing the mortality rate. The authors of [92] tackled early diagnosis of Sepsis based on health records obtained from Cardiology Challenge 2019. Using 168 features collected on hourly basis, an explainable AI model was developed for sepsis diagnosis. Gradient-boosting-trees model called XGBoots was used in K-fold cross-validation setup to forecast sepsis and provide interpretable sepsis risk in the ICU.

The authors of [93] have used XAI to investigate the relationship between tumor immune cell composition and breast cancer survival rates. Using EPIC, TIMER, CIBERSORT and xCell computational approaches, from TIMER2.0 and TCGA breast invasive cancer data, first, they extracted immune cell from a RNA bulk sequencing data. According to the proposed XAI techniques, the most significant cells responsible for breast cancer are the M0 macrophages, B cells, CD8+, and NK T cells. Their model demonstrated that by increasing the fraction of B cell with CD8+ *T* and NK T cell, their points of inflection might increase survival rate of breast cancer patients up to 18 %. The authors of [94] have studied and exploited different XAI methods in the healthcare sector. LIME and SHAP feature-based techniques have been applied to a heart disease dataset with 70+ features taken from UCI ML Repository.

The authors of [95] have adopted Magnetic Resonance Imaging (MRI) scan of brain with 1901 different subjects obtained from IXI, ADNI and AIBL repositories. To categorize patients suffering from AD, they trained an analytical model constructed on chronological and brain age data. They have argued that this model offers superior performance compared to other ML methods for females and males with 88 % and 92 % accuracy. The authors developed a methodology for performing regression and classification tasks while retaining the input space's of morphological semantics and giving a feature score to quantify each morphological region's detailed contribution to the final outcome.

Apart from importance of interpretability, the authors of [96] have recognized the necessity of ethics of AI by presenting a survey on ethical solutions for deployment of AI in different application domains. The survey points out some concerns regarding AI taking over our lives. For example, while automation using AI is beneficial in terms of lowering the production cost, it may lead to unemployment in human workforce. Moreover, companies that utilize AI will receive much higher profits faster compared to companies running on human workforce. This is unfair for companies that cannot afford AI automation. The authors also discuss the details of preparing high quality data which is necessary for training robust and reliable AI models.

The authors present a system in [97] for content-based image retrieval (CBIR) of Video frames related to minimally invasive surgery

Table 2
Summary of various XAI methods in digital healthcare and medicine, including their ML and XAI Methods.

Disease/ Type of Images or documents	ML Methods	XAI Methods	Refs	Year	Accuracy	Challenges/ Observations
Epilepsy	Bagged tree-based classifier (BTBC), DT, RF, XGB, LR, NB	SHAP	[99]	2024	Mean Accuracy 99.50 %	Patient-independent multimodel data using the proposed framework
Heart arrhythmia	CNN-LSTM	SHAP	[100]	2024	Mean diagnostic accuracy of 98.24 %	Accurate detection of cardiac arrhythmias
Parkinson's disease	GBT, NN, SVM-RBF, Linear SVM, Logistic regression, and RF	SHAP	[101]	2024	Best Accuracy 85 %	Highlights the challenge posed by missing data in EHR system
Lung Cancer	CNN and XGBoost	SHAP	[102]	2024	Accuracy of 97.43 %	Minimize the risk of error or bias in the results
Glaucoma	VGG19, ResNet50, InceptionV3, Xception	Saliency maps	[103]	2024	Accuracy of 96.62 %	No specific collection of spatially well-defined and consistently located features
Retinal fundus images	ResNet50	SIDU, GRAD-CAM	[104]	2021	AUC 0.6605	Eye-tracker in the medical domain particularly for the screening of retinal diseases, DR and quality assessment for retinal images
Autism Spectrum Disorder (ASD)	SVM	feature importance score, Visual	[105]	2021	Accuracy of toddler ASD dataset has an 98.27 ± 1.12 (for Log FT method)	The feature importance score is critical for explainability
Hepatitis	LR, DT, kNN, SVM, RF	SHAP, LIME, PDP	[106]	2021	Highest accuracy (91.9 %)	The proposed framework combining the global and local interpretable methods improve the transparency of complex models
Traumatic brain injury (TBI) identification	k-means, spectral clustering, Gaussian mixture	Quality assessment of clustering features.	[107]	2020	f-measure 64.11 %	A new consensus function in clustering ensemble has been introduce
Colorectal cancer diagnosis	CNN	Visual explanation	[108]	2020	Accuracy 92.74 %	Although many ML methods showed exceptional performances, the majority of them are not able to rationalize their decisions
Automatic recognition of instruments in laparoscopy videos	CNN	Activation Maps	[109]	2019	Maximum precision of 99 % on the validation dataset	Automatic identification of instruments in laparoscopy videos carriages many challenges that need to be addressed, like detecting multiple instruments appearing in various representations and in different lighting conditions.
Decision Support System for Prostate Cancer	Gradient-boosting algorithm	SHAP	[110]	2020	AUC 0.869	To validate a risk calculator for prostate cancer (PCa) and clinically significant PCa (csPCa) using XAI
ECG Based Hypoglycaemia	PCA	Grad-Cam	[111]	2020	Average accuracy 89.3 ± 4.5	Due to strong inter-subject heterogeneity, previous studies based on a cohort of subjects failed to deploy electrocardiogram (ECG)-based hypoglycemic detection systems reliably
Inflammatory bowel disease diagnostic	Linear SVM, RF, Nearest Shrunken Centroids (NSC), and Logistic Regression with L2 regularization (LR).	Feature Marginalization	[112]	2017	Accuracy of RF - 99 % and 81 %, for SVG and IBD datasets, respectively	microbiota data is challenging as It is high-dimensional, sparse and comprises of high inter-personal variation
Macromolecular Complexes	CNN	ML-CAM	[113]	2018	Mean accuracy 88 %	A computer-aided diagnostic approach for CLE imaging of human glioma with feature localization function
Allergy diagnosis	Decision Tree, SVM and Random Forest	Post-hoc XAI and CDSS	[114]	2021	Accuracy of RF 83.07 %	Computer-aided framework for allergy diagnosis which is capable of handling comorbidities
Glaucoma Diagnosis	CNN	ML-CAM	[115]	2022	Accuracy: 93.5 %	Both histogram equalization (HE) and contrast-limited adaptive HE (CLAHE) were used to enhance colored fundus image data
Pneumonia identification	VGG16	Grad-CAM	[116]	2022	Highest accuracy of VGG16 reaches 95.6 %	Traditional deep learning methods for pneumonia identification take less account of the influence of the lung X-

(*continued on next page*)

Table 2 (*continued*)

Disease/ Type of Images or documents	ML Methods	XAI Methods	Refs	Year	Accuracy	Challenges/ Observations
						ray image background on the model's testing effect
Alzheimer's disease	VGG-16 and CNN	Feature importance score, Visual	[117]	2022	10-fold cross-validation accuracy of 98.81 %	Identify the stages of Alzheimer using Layer wise relevance propagation method in XAI

(MIS) videos. In the proposed method, descriptors were extracted that were semantic in nature from mentioned video frames. In Case-Based Reasoning (CBR), a pool of labelled samples is available and labelling a new query sample is done based on the labels of its similar counterparts fetched from the pool. Contrary to DL, CBR reasoning is clear since it is primarily based on some type of similarity measure between labelled samples and the query sample. Therefore, the authors of [98] proposed a CBR method for breast cancer diagnosis equipped with a user interface for providing visual explanations. Prostate cancer is very common among men worldwide and its early diagnosis is vital to patient's survival chance. To make the results interpretable, LIME approach was utilized. It is beneficial to close this section by summarizing XAI research papers related to medicine and healthcare in Table 2. The applied ML and XAI approaches as well as studied diseases are listed for each paper. For instance, in epilepsy detection, Bagged Tree-Based Classifier (BTBC), Decision Trees (DT), Random Forests (RF), and others are employed, with SHAP providing interpretability. The mean accuracy achieved is 99.50 %, focusing on patient-independent multimodal data [99]. Heart arrhythmia diagnosis utilizes a CNN-LSTM model with SHAP, achieving a mean diagnostic accuracy of 98.24 %, aiming for accurate arrhythmia detection [100]. Parkinson's disease detection involves Gradient Boosted Trees (GBT), Neural Networks, among others, with SHAP, achieving an 85 % accuracy, addressing challenges posed by missing data in EHR systems [101]. The rest of table continues with similar disease-specific ML methods and XAI techniques, each targeting high accuracy while addressing specific challenges unique to the disease or dataset, emphasizing the importance of explainability and robustness in medical AI systems.

6. The challenge of interpretability in healthcare

In this section, we investigate the challenge of XAI in healthcare. We discuss the barriers that prevents the widespread application of XAI in healthcare and medicine.

6.1. User-Centric explanations

Understanding internal process of ML methods and outcomes in healthcare systems is essential to make such critical systems more trustworthy for end-users. The challenges that emerge in making ML models explainable are listed below.

- Analysis of complex ML models requires background in advanced mathematics and statistics.
- So far, healthcare systems have failed to fulfill the design and functional requirements for successful deployment of ML models in medical domain.
- The end-users' desire towards interpretability widens the gap between development of complex black box models and human-readable explanations.
- Making ML models more transparent is likely to make them less efficient in achieving their objectives. This is because high-performance models consist of many layers with complex interconnections. Therefore, tracing the training process on millions of samples is almost impossible.
- XAI methods only highlight the regions relevant to the ML model outputs without determining the features that have caused the relevancy of those regions.
- While considerable effort has been put into making models transparent, the appropriate evaluation of provided explanations is still an open issue. Moreover, some researchers have doubts about reliability of XAI methods considering them to be misleading.

As a result, end-users must be involved in developing ML models to bridge the gap between user needs and expectations and design support of the products. In applications such as detecting objects (e.g. vehicles, animals), recognizing actions, controlling robots, etc. lack of user participation may not be much of an issue. This is because experts can analyze XAI methods outputs to debug models and determine training data gaps on their own. However, in medical domain, the situation is different. Even if XAI methods provide plausible explanations, only clinicians can analyze XAI outputs and understand the cause of fail cases for ML models. Therefore, ML experts always have to rely on clinicians for debugging and improving their models. Considering that clinicians are usually busy with their own tasks, collaborating with them would be challenging.

Another challenge encountered in applying XAI in the medical domain is the fact that ML experts are usually comfortable with mathematical explanation outputs. On the contrary, clinicians prefer to receive explanations in visual form [118]. Such requirement puts limitations on the output format of XAI methods. However, it's imperative to note that advancements in explainable AI have the potential to contribute significantly to achieving Sustainable Development Goals (SDGs) in healthcare. By aligning XAI methodologies with SDG 3, which aims to ensure healthy lives and promote well-being for all at all ages, we can leverage these technologies to enhance diagnostic accuracy, treatment effectiveness, and overall healthcare delivery. The goals set forth in SDG 3, particularly pertaining to the attainment of universal health coverage and the facilitation of access to vital medications and immunizations, are intricately linked to the utilization of XAI within healthcare. Through the implementation of XAI methodologies, healthcare systems stand to enrich comprehension and transparency in medical decision-making, thereby ensuring fair access to high-quality care and bolstering patient outcomes. Moreover, XAI holds the capacity to offer intelligible insights into vaccine safety and effectiveness, assuaging concerns and fostering community trust. By shedding light on the decision-making mechanisms of these models, XAI augments their transparency and dependability, ultimately playing a pivotal role in advancing SDG 3′s objectives for immunization and curbing preventable fatalities. Thus, while the preference for visual explanations poses a challenge, integrating XAI solutions within the framework of SDGs underscores the importance of overcoming such barriers for the greater benefit of global health initiatives.

6.2. *Performance vs. transparency tradeoff*

XAI in healthcare systems impacts how end-users comprehend ML model decisions. Therefore, it is essential to balance the trade-off between the model's complexity and accuracy. Explainability is inversely related to performance of AI systems. Increasing the model transparency improves the ability to analyze its decisions [119]. Consequently, the XAI models divide into black-box AI, grey box AI, and white box AI. DL and ensembles approaches are included in the black box, statistical models are included in the grey box, and graphical models, linear models, rule-based models, and decision trees are included in the white box [120]. The black box of AI in healthcare systems is not transparent, making it difficult to provide acceptable reasoning for fair decisions and end-user trustworthiness [121]. The gray-box AI maintains average balance between transparency and explainability, while the white-box AI has high explainability with low-performance models. In an ideal healthcare system, models that combine high explainability with acceptable performance are preferred. However, this introduces a trade-off between the model's ability to identify understandable patterns and its flexibility in accurately fitting data. This trade-off is crucial and should be discussed with end-users to fully understand the clinical and human implications of any potential misclassifications. Balancing these aspects is key to developing AI systems that are both effective and trusted in healthcare contexts [122].

6.3. *Balancing requirements of interpretability*

Considering that interpretability is a complicated and nuanced term with no single definition, several requirements for an ideal interpretable ML system should be stated specially for healthcare applications. In general, the ML explanation is related to model's soundness (or optimality) and ability to be comprehended by the user [123]. In addition to soundness and comprehension, it is crucial to cover the explanation scope of models from a local (or instance-based) or global level. The global level reduces model performance, while the local level increases the time complexity to provide a comprehensive explanation [124–126]. The challenge is to ensure soundness, comprehension, and scope of the model and satisfy trustworthiness about the black and grey box AI model working mechanisms. The soundness and comprehension requirements are balanced based on the sensitivity of the application domain and the amount of which the end-user is anticipated to identify the ML model's interpretability.

6.4. *Assistive intelligence*

The ultimate objective of ML algorithms is removal of humans from decision making in various application domains [127]. However, in safety-critical applications such as healthcare, the decision making cannot be left to ML systems entirely. Supervision of human experts is necessary to avoid catastrophe in case wrong decisions are made by the ML system. While ML methods cannot be fully trusted with patients' lives, they can act as medical assistants for human experts accelerating medical data analysis and useful knowledge extraction [128]. Healthcare systems require precise data to make robust decisions so human-in-the-loop framework as well as XAI mechanisms are needed. For instance, such assistive intelligence is supported for classification of diabetic retinal fundus images [129], medical writing for the scientific medical community [130], and digital tourism [131], which could be tailored for healthcare professionals.

The article in [132] systematically analyzes XAI models in healthcare, focusing on trends and future directions. It explores XAI methodologies and proposes strategies for deriving trustworthy AI in healthcare, addressing concerns such as security, performance, and communication issues. Loh et al. [133] review highlights the importance of XAI in healthcare to enhance trust in AI models, and it identifies areas such as abnormality detection in biosignals and key text identification in clinical notes that warrant further attention from the XAI research community, aiming to foster the development of a comprehensive smart city cloud system. The work in [134] critically reviews previous studies on interpretability methods like LRP, UMAP, LIME, SHAP, and Grad-CAM, discussing their usability and reliability, and explores how AI and ML technologies can revolutionize healthcare services, offering insights for researchers and decision-makers.

Gupta et al. [135] review the utilization of techniques such as LIME, SHAP, PDP, and decision trees, and identify opportunities for further research in less explored medical data areas. Jung et al. [136] summarized the findings of comprehensive framework and standardized approaches to evaluate XAI's explanation effectiveness across diverse AI stakeholders. The work in [137] presents a survey of recent XAI techniques in healthcare and medical imaging applications, aiming to mimic human judgment and interpretation skills while providing insights into the black-box model of deep learning decision-making for clinical tasks. The article [138] highlights the increasing precision in using XAI and describing ML pipelines alongside the growing demand for experts bridging informatics and medical domains.

The work [139] provides a comprehensive overview of XAI frameworks tailored for IoT applications in healthcare, addressing the need for trustworthiness in the growing Internet of Medical Things (IoMT). It systematically explores XAI services for healthcare IoT, including security enhancement and edge XAI structures, while highlighting the potential of sixth generation (6 G) communication services, aiming to enhance transparency and reliability in future healthcare IoT implementations. The review in [140] explores the current state and potential of XAI in revolutionizing drug discovery, addressing the need for transparency and interpretability in increasingly complex AI and ML models. The systematic review [141] explores the trustworthiness and explainability of AI applications in healthcare, employing data fusion techniques to bridge gaps in current research, offering guidelines for policymakers and addressing methodological aspects to enhance trustworthiness. Band et al. [134] discussed the transformative potential of AI and ML technologies in healthcare, and provided insights into the usability and reliability of these methods alongside XAI.

The article in [142] systematically analyzes XAI models in healthcare, focusing on trends and future directions. It explores XAI methodologies and proposes strategies for deriving trustworthy AI in healthcare, addressing concerns such as security, performance, and communication issues. Loh et al. [143] review highlights the importance of XAI in healthcare to enhance trust in AI models, and it identifies areas such as abnormality detection in biosignals and key text identification in clinical notes that warrant further attention from the XAI research community, aiming to foster the development of a comprehensive smart city cloud system. The work in [144] critically reviews previous studies on interpretability methods like LRP, UMAP, LIME, SHAP, and Grad-CAM, discussing their usability and reliability, and explores how AI and ML technologies can revolutionize healthcare services, offering insights for researchers and decision-makers.

Our article on using XAI for personalized healthcare stands apart by addressing the gaps identified in the referenced articles. Specifically, while [142] systematically analyzes XAI models in healthcare and proposes strategies for trustworthy AI, our article focuses on the application of XAI for personalized healthcare, offering practical insights into its implementation. Additionally, while [143] highlights the importance of XAI in healthcare to enhance trust in AI models, our article delves deeper into the specific areas of abnormality detection in biosignals and key text identification in clinical notes, providing detailed guidance on leveraging XAI techniques for these tasks. Furthermore, while [144] reviews interpretability methods and explores the potential of AI and ML technologies in revolutionizing healthcare services, our article goes beyond by offering a comprehensive framework for integrating XAI into personalized healthcare, catering to both researchers and decision-makers in the healthcare domain. In [142], the authors review the potential of explainable XAI methods in diagnostic imaging, particularly for MRI, CT, and PET imaging, highlighting the need for systematic quality assessment and suggesting future directions. The study in [143,144] provides a systematic meta-survey of challenges and future research directions in XAI, organized into two themes, aiming to guide future exploration in the XAI field and advance its adoption in critical medical domains.

The systematic review by Fontes et al. [145] explores the potential of example-based XAI techniques in medical imaging, analyzing recent studies to enhance accuracy, transparency, and usability in clinical practice. The discussion in [146] highlights the contentious issue of XAI in healthcare applications, emphasizing the need for comprehensible and transparent patient-centric approaches. In [147] the need for standardized metrics for evaluating XAI systems, especially in the healthcare domain, to ensure transparency and domain-specific relevance is summarized. The survey in [148] explores XAI methods beyond saliency-based approaches, aiming to provide a diverse understanding of XAI techniques applicable to healthcare professionals and facilitate cross-disciplinary exchange, categorized into case-based, textual, and auxiliary explanations. In addition to this, the authors in [149] introduce modified deep learning models, MobileNetV2 and DenseNet201, augmented with additional convolutional layers, to enhance skin cancer detection efficiency in detecting both benign and malignant cases. The Inception-ResNetV2 model enhanced with local binary patterns [150] is used for precise lung and colon cancer diagnosis, achieving 99.98\% accuracy. Here, employing XAI through SHAP enhances transparency in deep learning model decision-making, promising a transformative impact on cancer diagnosis accuracy.

7. Conclusion

AI makes an impact in every sphere of life by inducing a significant paradigm shift in the healthcare sector and education [151,152] and has revolutionized data access and analytical methods. With the emergence of deep learning, there have been notable advancements in decision-making and prediction algorithms, particularly in their ability to achieve high performance. However, a critical issue with these advancements is the often-opaque nature of the algorithms, posing challenges in understanding their decision-making processes. This opacity has spurred the growth of Explainable AI (XAI), a field dedicated to making AI decisions transparent and comprehensible.

XAI has not only gained prominence for its ability to clarify the mechanisms of AI learning but has also been instrumental in refining the performance of deep networks. Developers are increasingly leveraging XAI techniques to decode the learning strategies of advanced methods like graph neural networks, generative adversarial networks, self-supervised learning, and reinforcement learning. This paper has undertaken an exhaustive literature review focused on XAI, particularly in healthcare applications. We have systematically categorized XAI methodologies, surveying 26 papers related to medical diagnosis and surgery that utilize XAI. The paper also highlights the challenges of implementing XAI in healthcare and emphasizes the importance of model explainability in this sector. It underscores the need for a human-centered approach in designing and developing XAI methods, facilitating the interpretation of AI models by both patients and medical professionals. While XAI has made significant inroads in healthcare, particularly in diagnosis and surgery, the integration of comprehensive explainability tools with ML and DL methods remains a challenge, especially in critical areas like surgery.

The future of AI in healthcare should prioritize the development of self-explanatory techniques that obviate the need for post-hoc explanations. Such approaches could lead to AI solutions in healthcare that are transparent and easily understandable by both professionals and patients. These transparent AI solutions promise further advancements in the health sector, paving the way for new treatments and approaches in medicine. Moreover, the reliability and consistency of AI-assisted medical applications will be enhanced if users can comprehend the rationale behind treatment plans. Thus, a collaborative effort between data scientists and medical experts is essential for creating more effective XAI applications, which will significantly influence medical procedures and treatments, increasing patient satisfaction by providing insights into disease causation and the impact of various medications.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for the research described in the article.

Acknowledgement

The work of H. Moosaei was supported by the Czech Science Foundation Grant 22–19353S.

Milan Hladík was supported by the Czech Science Foundation Grant 22–11117S.

The work of P.M. Pardalos PM was conducted within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE).

References

[1] Arrieta AB, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115.
[2] Angelov PP, Soares EA, Jiang R, Arnold NI, Atkinson PM. Explainable artificial intelligence: an analytical review. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2021;11(5):e1424.
[3] Samek W., Wiegand T., Müller K.-R. Explainable artificial intelligence: understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv: 170808296. 2017.
[4] Shahroudnejad A. A survey on understanding, visualizations, and explanation of deep neural networks. arXiv preprint arXiv:210201792. 2021.
[5] Hendrickson R.G. Survey of sensitivity analysis methodology. 1984.
[6] Chen X, Molina-Cristóbal A, Guenov MD, Riaz A. Efficient method for variance-based sensitivity analysis. Reliab Eng Syst Saf 2019;181:97–115.
[7] Kucherenko S, Song S, editors. Derivative-based global sensitivity measures and their link with sobol'sensitivity indices. monte carlo and quasi-monte carlo methods. Leuven, Belgium: MCQMC; 2016. April 2014Springer.
[8] Plischke E, Borgonovo E, Smith CL. Global sensitivity measures from given data. Eur J Oper Res 2013;226(3):536–50.
[9] Iooss B., Lemaître P. A review on global sensitivity analysis methods. Uncertainty management in simulation-optimization of complex systems: algorithms and applications. 2015:101–22.
[10] Van Stein B, Raponi E, Sadeghi Z, Bouman N, Van Ham RC, Bäck T. A comparison of global sensitivity analysis methods for explainable AI with an application in genomic prediction. IEEE Access 2022;10:103364–81.
[11] Zeiler MD, Fergus R, editors. Visualizing and understanding convolutional networks. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13. Springer; 2014.
[12] Simonyan K., Vedaldi A., Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv preprint arXiv: 13126034. 2013.
[13] Deep residual learning for image recognition. In: He K, Zhang X, Ren S, Sun J, editors. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
[14] Understanding deep image representations by inverting them. In: Mahendran A, Vedaldi A, editors. Proceedings of the IEEE conference on computer vision and pattern recognition; 2015.
[15] Mordvintsev A., Olah C., Tyka M. Inceptionism: going deeper into neural networks. 2015.
[16] Sadeghi Z, editor. editor An Information Analysis Approach into Feature Understanding of Convolutional Deep Neural Networks. Machine Learning, Optimization, and Data Science: 5th International Conference, LOD 2019, Siena, Italy, September 10–13, 2019, Proceedings 5. Springer; 2019.
[17] Shrikumar A, Greenside P, Kundaje A, editors Learning important features through propagating activation differences. International conference on machine learning; 2017: PMLR.
[18] Yuan H, Yu H, Gui S, Ji S. Explainability in graph neural networks: a taxonomic survey. IEEE Trans Pattern Anal Mach Intell 2022.
[19] Arras L., Montavon G., Müller K.-R., Samek W. Explaining recurrent neural network predictions in sentiment analysis. arXiv preprint arXiv:170607206. 2017.
[20] Chou Y-L, Moreira C, Bruza P, Ouyang C, Jorge J. Counterfactuals and causability in explainable artificial intelligence: theory, algorithms, and applications. Information Fusion 2022;81:59–83.
[21] Al-Shedivat M, Dubey A, Xing E. Contextual explanation networks. J Mach Learning Res 2020;21(1):7950–93.
[22] Alvarez Melis D, Jaakkola T. Towards robust interpretability with self-explaining neural networks. Adv Neural Inf Process Syst 2018;31.
[23] Brendel W., Bethge M. Approximating cnns with bag-of-local-features models works surprisingly well on imagenet. arXiv preprint arXiv:190400760. 2019.
[24] Tabnet: attentive interpretable tabular learning. In: Arik SÖ, Pfister T, editors. Proceedings of the AAAI Conference on Artificial Intelligence; 2021.
[25] Adadi A, Berrada M. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access 2018;6:52138–60.
[26] Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.
[27] Bach S, Binder A, Montavon G, Klauschen F, Müller K-R, Samek W. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 2015;10(7):e0130140.
[28] " Why should i trust you?" Explaining the predictions of any classifier. In: Ribeiro MT, Singh S, Guestrin C, editors. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016.
[29] Learning deep features for discriminative localization. In: Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A, editors. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
[30] Grad-cam: visual explanations from deep networks via gradient-based localization. In: Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Proceedings of the IEEE international conference on computer vision; 2017.

[31] Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks. In: Chattopadhay A, Sarkar A, Howlader P, Balasubramanian VN, editors. 2018 IEEE winter conference on applications of computer vision (WACV). IEEE; 2018.

[32] Scenario-Based Requirements Elicitation for User-Centric Explainable AI: a Case in Fraud Detection. In: Cirqueira D, Nedbal D, Helfert M, Bezbradica M, editors. Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 84, WG 89, WG 129 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings 4. Springer; 2020.

[33] Alicioglu G, Sun B. A survey of visual analytics for Explainable Artificial Intelligence methods. Comput Graph 2022;102:502–20.

[34] Interpretability beyond feature attribution: quantitative testing with concept activation vectors (tcav). In: Kim B, Wattenberg M, Gilmer J, Cai C, Wexler J, Viegas F, editors. International conference on machine learning. PMLR; 2018.

[35] Goel T, Sharma R, Tanveer M, Suganthan PN, Maji K, Pilli R. Multimodal neuroimaging based alzheimer's disease diagnosis using evolutionary RVFL Classifier. IEEE J Biomed Health Inform 2023.

[36] Kamal MS, Northcote A, Chowdhury L, Dey N, Crespo RG, Herrera-Viedma E. Alzheimer's patient analysis using image and gene expression data and explainable-AI to present associated genes. IEEE Trans Instrum Meas 2021;70:1–7.

[37] Biffi C, Cerrolaza JJ, Tarroni G, Bai W, De Marvao A, Oktay O, Rueckert D. Explainable anatomical shape analysis through deep hierarchical generative models. IEEE Trans Med Imaging 2020;39(6):2088–99.

[38] Gauch JM. Image segmentation and analysis via multiscale gradient watershed hierarchies. IEEE Trans Image Process 1999;8(1):69–79.

[39] Carson DL, Botha FC. Preliminary analysis of expressed sequence tags for sugarcane. Crop Sci 2000;40(6):1769–79.

[40] Kinimi E, Muyldermans S, Vincke C, Odongo S, Kock R, Parida S, Misinzo G. Development of nanobodies targeting peste des petits ruminants virus: the prospect in disease diagnosis and therapy. Animals 2021;11(8):2206.

[41] Shin Hak Sup, Park; Soon-Ik. Novel Attacin From Hermetia Illucens: CDNA Cloning, Characterization, And Antibacterial Properties. Prep Biochem Biotechnol 2019;(IF: 3).

[42] Vieth B, Parekh S, Ziegenhain C, Enard W, Hellmann; I. A systematic evaluation of single cell RNA-Seq analysis pipelines. Bio.Bioinformatics 2019;(IF: 4).

[43] Ferrari Davide, Garrapa Valentina, Locatelli Massimo, Bolchi; Angelo. A novel nanobody scaffold optimized for bacterial expression and suitable for the construction of ribosome display libraries. Mol Biotechnol 2019;(IF: 3).

[44] Varnaitė Renata, Blom Kim, Lampen Margit H, Vene Sirkka, Thunberg Sarah, Lindquist Lars, Ljunggren Hans-Gustaf, Rombo Lars, Askling Helena H, Gredmark-Russ; Sara. Magnitude and functional profile of the human CD4+ T cell response throughout primary immunization with tick-borne encephalitis virus vaccine. J Immunol 2020.

[45] Varnaitė R, Blom K, Lampen MH, Vene S, Thunberg S, Lindquist L, Gredmark-Russ S. Magnitude and functional profile of the human CD4+ T cell response throughout primary immunization with tick-borne encephalitis virus vaccine. J Immunol 2020;204(4):914–22.

[46] Li Xue-Peng, Zhang; Jian. A live attenuated edwardsiella tarda vaccine induces immunological expression pattern in japanese flounder (Paralichthys Olivaceus) in the early phase of immunization. Comp Biochem Physiol Toxicol & 2020.

[47] Steffen Hinz; "Tailor-made Antibodies by Multidimensional Functional Screening", 2021.

[48] Bélanger Simon, Haupt Sonya, Faliti Caterina E, Getzler Adam, Choi Jinyong, Diao Huitian, Karunadharma Pabalu P, Bild Nicholas A, Pipkin Matthew E, Crotty Shane. The chromatin regulator Mll1 supports T follicular helper cell differentiation by controlling expression of Bcl6, LEF-1, and TCF-1. J Immunol 2022.

[49] Baio Gianluca. SurvHE: survival analysis for health economic evaluation and cost-effectiveness modeling. J Stat Softw 2020;(IF: 3).

[50] Karthik Raja S, Shahul Hameed G. Serum vitamin D3 deficiency among cases with extensive tinea corporis infection. Int J Res Dermatol 2020.

[51] Xie F, Chakraborty B, Ong M, Goldstein B, Liu Nan. AutoScore: a Machine Learning–Based Automatic Clinical Score Generator and Its Application to Mortality Prediction Using Electronic Health Records. JMIR Med Inform 2020;(IF: 3).

[52] Musa Taha Hussein, Ahmad Tauseef, Li Wei, Kawuki Joseph, Wana Mohammed Nasiru, Musa Hassan Hussein, Wei Pingmin. A bibliometric analysis of global scientific research on scrub typhus. Biomed Res Int 2020;(IF: 3).

[53] Dehnoei S. A stochastic optimization approach for staff scheduling decisions at inpatient clinics (Doctoral dissertation. Université d'Ottawa/University of Ottawa; 2020.

[54] Enitan Comfort Bosede, Osadolor Humphrey Benedo, Digban Kester Awharentomah, Osakue Eguagie Osareniro, Alao James, Adejumo Esther Ngozi, Ladipo Oluwakemi Anike, Ufuoma Chukwuani, Ogar Francis Awusha, Okoro Chigozie Victor. Evaluation of serum lipid profile at different trimesters among pregnant women in ovia north-east local government area, edo state, nigeria. Int J Med Sci Health Res 2021.

[55] Hreńczuk Marta, Gruszkiewicz Patrycja, Małkowski Piotr. Knowledge, opinions, and attitudes of students of warsaw universities toward hematopoietic stem cell transplantation. Transplant Proc 2021.

[56] Nahavandi S, Alizadehsani R, Nahavandi D, Lim CP, Kelly K, Bello F. Machine learning meets advanced robotic manipulation. Inf Fusion 2024;105:102221.

[57] Carcagnì P, Leo M, Del Coco M, Distante C, De Salve A. Convolution neural networks and self-attention learners for alzheimer dementia diagnosis from brain MRI. Sensors 2023;23(3):1694.

[58] Kaplan A, Haenlein M. Siri, Siri, in my hand: who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Bus Horiz 2019;62(1):15–25.

[59] Zhang Y, Tiňo P, Leonardis A, Tang K. A survey on neural network interpretability. IEEE Transactions on Emerging Topics in Comput Intell 2021;5(5):726–42.

[60] Zhang JD, Xue C, Kolachalama VB, Donald WA. Interpretable machine learning on metabolomics data reveals biomarkers for parkinson's disease. ACS Cent Sci 2023.

[61] Warman A., Warman P.I., Sharma A., Parikh P., Warman R., Viswanadhan N., et al. Interpretable artificial intelligence for COVID-19 diagnosis from chest CT reveals specificity of ground-glass opacities. medRxiv. 2020:2020.05. 16.20103408.

[62] Zhu S, Fan W, Yang S, Pardalos PM. Scheduling operating rooms of multiple hospitals considering transportation and deterioration in mass-casualty incidents. Ann Oper Res 2023;321(1–2):717–53.

[63] KAMBOMBO M. Optimized patient flow process-A case of outpatient and surgical departments in sub-saharan africa healthcare systems (Doctoral dissertation. University of Rwanda (College of science and Technology; 2022.

[64] Mutanu L. A self-learning framework for validation of runtime adaptation in service-oriented systems. United Kingdom: Lancaster University; 2017.

[65] Horáček J, Koucký V, Hladík M. Novel approach to computerized breath detection in lung function diagnostics. Comput Biol Med 2018;101:1–6.

[66] Oestreich MA, Wyler F, Frauchiger BS, Latzin P, Ramsey KA. Breath detection algorithms affect multiple-breath washout outcomes in pre-school and school age children. PLoS One 2022;17(10):e0275866.

[67] Moreno Escobar, J.S. (2021). Aplicación de registro y reporte automático de parámetros fisiológicos para la trazabilidad psicofisiológica en personas post COVID-19.

[68] Nahavandi, S., Alizadehsani, R., Nahavandi, D., Mohamed, S., Mohajer, N., Rokonuzzaman, M., & Hossain, I. (2022). A comprehensive review on autonomous navigation. arXiv preprint arXiv:2212.12808.

[69] Nasarian E, Alizadehsani R, Acharya UR, Tsui KL. Designing interpretable ML system to enhance trust in healthcare: a systematic review to proposed responsible clinician-AI-collaboration framework. Information Fusion 2024:102412.

[70] Explainable artificial intelligence: concepts, applications, research challenges and visions. In: Longo L, Goebel R, Lecue F, Kieseberg P, Holzinger A, editors. Machine Learning and Knowledge Extraction: 4th IFIP TC 5, TC 12, WG 84, WG 89, WG 129 International Cross-Domain Conference, CD-MAKE 2020, Dublin, Ireland, August 25–28, 2020, Proceedings. Springer; 2020.

[71] Classification of road traffic accident data using machine learning algorithms. In: Kumeda B, Zhang F, Zhou F, Hussain S, Almasri A, Assefa M, editors. 2019 IEEE 11th international conference on communication software and networks (ICCSN). IEEE; 2019.

[72] Behrad F, Abadeh MS. An overview of deep learning methods for multimodal medical data mining. Expert Syst Appl 2022;200:117006.

[73] PAGE-Net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data. In: Hao J, Kosaraju SC, Tsaku NZ, Song DH, Kang M, editors. Pacific Symposium on Biocomputing 2020. World Scientific; 2019.

[74] Xiang A, Wang F, editors. Towards interpretable skin lesion classification with deep learning models. amia annual symposium proceedings. American Medical Informatics Association; 2019.

[75] Karimi M, Wu D, Wang Z, Shen Y. DeepAffinity: interpretable deep learning of compound–protein affinity through unified recurrent and convolutional neural networks. Bioinformatics 2019;35(18):3329–38.

[76] Tjoa E, Guan C. A survey on explainable artificial intelligence (xai): toward medical xai. IEEE Trans Neural Netw Learn Syst 2020;32(11):4793–813.

[77] Carrieri AP, Haiminen N, Maudsley-Barton S, Gardiner L-J, Murphy B, Mayes AE, et al. Explainable AI reveals changes in skin microbiome composition linked to phenotypic differences. Sci Rep 2021;11(1):1–18.

[78] Magesh PR, Myloth RD, Tom RJ. An explainable machine learning model for early detection of Parkinson's disease using LIME on DaTSCAN imagery. Comput Biol Med 2020;126:104041.

[79] Olsen LSKM, MJ LMLKJ, Lange J, Thiesson B. Explainable artificial intelligence model to predict acute critical illness from electronic health records. Nat Commun 2020;11(1):1.

[80] Rucco M, Viticchi G, Falsetti L. Towards personalized diagnosis of glioblastoma in fluid-attenuated inversion recovery (FLAIR) by topological interpretable machine learning. Mathematics 2020;8(5):770.

[81] Meldo A, Utkin L, Kovalev M, Kasimov E. The natural language explanation algorithms for the lung cancer computer-aided diagnosis system. Artif Intell Med 2020;108:101952.

[82] Yeboah D, Steinmeister L, Hier DB, Hadi B, Wunsch DC, Olbricht GR, et al. An explainable and statistically validated ensemble clustering model applied to the identification of traumatic brain injury subgroups. IEEE Access 2020;8:180690–705.

[83] Arun N, Gaw N, Singh P, Chang K, Aggarwal M, Chen B, Kalpathy-Cramer J. Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging. Radiology: Artificial Intelligence 2021;3(6):e200267.

[84] Predicting post-stroke hospital discharge disposition using interpretable machine learning approaches. In: Cho J, Alharin A, Hu Z, Fell N, Sartipi M, editors. 2019 IEEE International Conference on Big Data (Big Data). IEEE; 2019.

[85] Yoo TK, Ryu IH, Choi H, Kim JK, Lee IS, Kim JS, et al. Explainable machine learning approach as a tool to understand factors used to select the refractive surgery technique on the expert level. Translational Vision Sci Technol 2020;9(2):8. -.

[86] Explainable AI for COVID-19 CT classifiers: an initial comparison study. In: Ye Q, Xia J, Yang G, editors. 2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS). IEEE; 2021.

[87] Hossain MS, Muhammad G, Guizani N. Explainable AI and mass surveillance system-based healthcare framework to combat COVID-I9 like pandemics. IEEE Netw 2020;34(4):126–32.

[88] El-Sappagh S, Alonso JM, Islam S, Sultan AM, Kwak KS. A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease. Sci Rep 2021;11(1):1–26.

[89] Yang M, Liu C, Wang X, Li Y, Gao H, Liu X, et al. An explainable artificial intelligence predictor for early detection of sepsis. Crit Care Med 2020;48(11). e1091-e6.

[90] Chakraborty D, Ivan C, Amero P, Khan M, Rodriguez-Aguayo C, Başağaoğlu H, et al. Explainable artificial intelligence reveals novel insight into tumor microenvironment conditions linked with better prognosis in patients with breast cancer. Cancers (Basel) 2021;13(14):3450.

[91] Dave D., Naik H., Singhal S., Patel P. Explainable ai meets healthcare: a study on heart disease dataset. arXiv preprint arXiv:201103195. 2020.

[92] Komatsu M, Sakai A, Komatsu R, Matsuoka R, Yasutomi S, Shozu K, et al. Detection of cardiac structural abnormalities in fetal ultrasound videos using deep learning. Applied Sciences 2021;11(1):371.

[93] Pappalardo M, Starnoni M, Franceschini G, Baccarani A, De Santis G. Breast cancer-related lymphedema: recent updates on diagnosis, severity and available treatments. J Pers Med 2021;11(5):402.

[94] Varzandian A, Razo MAS, Sanders MR, Atmakuru A, Di Fatta G. Classification-biased apparent brain age for the prediction of Alzheimer's disease. Front Neurosci 2021;15:673120.

[95] Ethics of artificial intelligence: research challenges and potential solutions. In: Jameel T, Ali R, Toheed I, editors. 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE; 2020.

[96] XAI-CBIR: explainable AI system for content-based retrieval of video frames from minimally invasive surgery videos. In: Chittajallu DR, Dong B, Tunison P, Collins R, Wells K, Fleshman J, et al., editors. 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019). IEEE; 2019.

[97] Lamy J-B, Sekar B, Guezennec G, Bouaud J, Séroussi B. Explainable artificial intelligence for breast cancer: a visual case-based reasoning approach. Artif Intell Med 2019;94:42–53.

[98] Hassan MR, Islam MF, Uddin MZ, Ghoshal G, Hassan MM, Huda S, et al. Prostate cancer classification from ultrasound and MRI images using deep learning based Explainable Artificial Intelligence. Future Generation Comput Syst 2022;127:462–72.

[99] Ahmad I, Yao C, Li L, Chen Y, Liu Z, Ullah I, Chen S. An efficient feature selection and explainable classification method for EEG-based epileptic seizure detection. J Inf Securit Applications 2024;80:103654.

[100] Alamatsaz N, Tabatabaei L, Yazdchi M, Payan H, Alamatsaz N, Nasimi F. A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection. Biomed Signal Process Control 2024;90:105884.

[101] Amini, M., Bagheri, A., Piri, S., & Delen, D. (2024). A hybrid ai framework to address the issue of frequent missing values with application in ehr systems: the case of parkinson's disease.

[102] Wani NA, Kumar R, Bedi J. DeepXplainer: an interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence. Comput Method Programs Biomed 2024;243:107879.

[103] Sigut J, Fumero F, Estévez J, Alayón S, Díaz-Alemán T. In-depth evaluation of saliency maps for interpreting convolutional neural network decisions in the diagnosis of glaucoma based on fundus imaging. Sensors 2024;24(1):239.

[104] Expert level evaluations for explainable AI (XAI) methods in the medical domain. In: Muddamsetty SM, Jahromi MN, Moeslund TB, editors. Pattern Recognition ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part III. Springer; 2021.

[105] An XAI based autism detection: the context behind the detection. In: Biswas M, Kaiser MS, Mahmud M, Al Mamun S, Hossain MS, Rahman MA, editors. Brain Informatics: 14th International Conference, BI 2021, Virtual Event, September 17–19, 2021, Proceedings 14. Springer; 2021.

[106] Peng J, Zou K, Zhou M, Teng Y, Zhu X, Zhang F, et al. An explainable artificial intelligence framework for the deterioration risk prediction of hepatitis patients. J Med Syst 2021;45:1–9.

[107] Mahmoudi MR, Akbarzadeh H, Parvin H, Nejatian S, Rezaie V. Alinejad-Rokny H. Consensus function based on cluster-wise two level clustering. Artif Intell Rev 2021;54:639–65.

[108] Sabol P, Sinčák P, Hartono P, Kočan P, Benetinová Z, Blichárová A, et al. Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. J Biomed Inform 2020;109:103523.

[109] Kletz S, Schoeffmann K, Husslein H. Learning the representation of instrument images in laparoscopy videos. Healthc Technol Lett 2019;6(6):197–203.

[110] Suh J, Yoo S, Park J, Cho SY, Cho MC, Son H, et al. Development and validation of an explainable artificial intelligence-based decision-supporting tool for prostate biopsy. BJU Int 2020;126(6):694–703.

[111] Porumb M, Stranges S, Pescapè A, Pecchia L. Precision medicine and artificial intelligence: a pilot study on deep learning for hypoglycemic events detection based on ECG. Sci Rep 2020;10(1):170.

[112] Eck A, Zintgraf LM, de Groot E, de Meij TG, Cohen TS, Savelkoul P, et al. Interpretation of microbiota-based diagnostics by explaining individual classifier decisions. BMC Bioinf 2017;18(1):1–13.

[113] Weakly-supervised learning-based feature localization for confocal laser endomicroscopy glioma images. In: Izadyyazdanabadi M, Belykh E, Cavallo C, Zhao X, Gandhi S, Moreira LB, et al., editors. Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. Springer; 2018.

[114] Kavya R, Christopher J, Panda S, Lazarus YB. Machine learning and XAI approaches for allergy diagnosis. Biomed Signal Process Control 2021;69:102681.

[115] Deperlioglu O, Kose U, Gupta D, Khanna A, Giampaolo F, Fortino G. Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: analysis with doctor evaluation. Future Generation Comput Syst 2022;129:152–69.
[116] Sheu RK, Pardeshi MS. A Survey on Medical Explainable AI (XAI): recent Progress, Explainability Approach, Human Interaction and Scoring System. Sensors 2022;22(20):8068.
[117] Alzheimer's Disease Analysis using Explainable Artificial Intelligence (XAI. In: Sudar KM, Nagaraj P, Nithisaa S, Aishwarya R, Aakash M, Lakshmi SI, editors. 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). IEEE; 2022.
[118] Shin Hak Sup, Park Soon-Ik. Novel Attacin From Hermetia Illucens: CDNA Cloning, Characterization, And Antibacterial Properties. Prep Biochem Biotechnol. 2019;(IF 3).
[119] Jerby-Arnon L, Regev A. DIALOGUE maps multicellular programs in tissue from single-cell or spatial transcriptomics data. Nat Biotechnol 2022;40(10): 1467–77.
[120] Li B, Qin X, Mi LZ. Nanobodies: from structure to applications in non-injectable and bispecific biotherapeutic development. Nanoscale 2022;14(19):7110–22.
[121] Zent O, Jilg W, Plentz A, Schwarz TF, Frühwein N, Kuhr HB, Banzhoff A. Kinetics of the immune response after primary and booster immunization against tick-borne encephalitis (TBE) in adults using the rapid immunization schedule. Vaccine 2003;21(32):4655–60.
[122] Kleiter I, Jilg W, Bogdahn U, Steinbrecher A. Delayed humoral immunity in a patient with severe tick-borne encephalitis after complete active vaccination. Infection 2007;35:26–9.
[123] Odusami M, Maskeliūnas R, Damaševičius R. An intelligent system for early recognition of Alzheimer's disease using neuroimaging. Sensors 2022;22(3):740.
[124] Hinz SC, Elter A, Rammo O, Schwämmle A, Ali A, Zielonka S, Kolmar H. A generic procedure for the isolation of pH-and magnesium-responsive chicken scFvs for downstream purification of human antibodies. Front Bioeng Biotechnol 2020;8:688.
[125] Ekuma G, Hier DB, Obafemi-Ajayi T. An Explainable Deep Learning Model for Prediction of Severity of Alzheimer's Disease. In: 2023 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE; 2023. p. 1–8.
[126] Incerti, D., & Jansen, J.P. (2021). hesim: health economic simulation modeling and decision analysis. arXiv preprint arXiv:2102.09437.
[127] Bocheva G, Slominski RM, Slominski AT. The impact of vitamin D on skin aging. Int J Mol Sci 2021;22(16):9097.
[128] Musa IH, Afolabi LO, Zamit I, Musa TH, Musa HH, Tassang A, Li W. Artificial intelligence and machine learning in cancer research: a systematic and thematic analysis of the top 100 cited articles indexed in Scopus database. Cancer Control 2022;29. 10732748221095946.
[129] Mishra RK. Deep Learning Model for Multiclass Classification of Diabetic Retinal Fundus Images Using Gradient Descent Optimization. In: Advances in Signal Processing and Communication Engineering: Select Proceedings of ICASPACE 2021. Springer Nature Singapore; 2022. p. 27–35.
[130] Mishra RK, Urolagin S, Jothi JAA, Gaur P. Deep hybrid learning for facial expression binary classifications and predictions. Image Vis Comput 2022;128: 104573.
[131] Mishra RK, Jothi JAA, Urolagin S, Irani K. Knowledge based topic retrieval for recommendations and tourism promotions. Int J Inf Manage Data Insights 2023; 3(1):100145.
[132] Bharati S, Mondal MRH, Podder P. A review on explainable artificial intelligence for healthcare: why, how, and when? IEEE Transactions on Artificial Intelligence 2023.
[133] Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). Comput Method Programs Biomed 2022;226:107161.
[134] Band SS, Yarahmadi A, Hsu CC, Biyari M, Sookhak M, Ameri R, Liang HW. Application of explainable artificial intelligence in medical health: a systematic review of interpretability methods. Inf Med Unlock 2023:101286.
[135] Gupta J, Seeja KR. A Comparative Study and Systematic Analysis of XAI Models and their Applications in Healthcare. Arch Comput Meth Eng 2024:1–26.
[136] Jung J, Lee H, Jung H, Kim H. Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: a systematic review. Heliyon 2023.
[137] Chaddad A, Peng J, Xu J, Bouridane A. Survey of explainable AI techniques in healthcare. Sensors 2023;23(2):634.
[138] Allgaier J, Mulansky L, Draelos RL, Pryss R. How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare. Artif Intell Med 2023;143:102616.
[139] Jagatheesaperumal SK, Pham QV, Ruby R, Yang Z, Xu C, Zhang Z. Explainable AI over the internet of things (IoT): overview, state-of-the-art and future directions. IEEE Open J Communicat Societ 2022;3:2106–36.
[140] Alizadehsani R, Oyelere SS, Hussain S, Jagatheesaperumal SK, Calixto RR, Rahouti M, De Albuquerque VHC. Explainable artificial intelligence for drug discovery and development-a comprehensive survey. IEEE Access 2024.
[141] Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Deveci M. A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Information Fusion 2023.
[142] de Vries BM, Zwezerijnen GJ, Burchell GL, van Velden FH, Menke-van der Houven van Oordt CW, Boellaard R. Explainable artificial intelligence (XAI) in radiology and nuclear medicine: a literature review. Front Med (Lausanne) 2023;10:1180773.
[143] Saeed W, Omlin C. Explainable AI (XAI): a systematic meta-survey of current challenges and future opportunities. Knowl Based Syst 2023;263:110273.
[144] Ali S, Akhlaq F, Imran AS, Kastrati Z, Daudpota SM, Moosa M. The enlightening role of explainable artificial intelligence in medical & healthcare domains: a systematic literature review. Comput Biol Med 2023:107555.
[145] Fontes M, De Almeida JDS, Cunha A. Application of example-based explainable artificial intelligence (xai) for analysis and interpretation of medical imaging: a systematic review. IEEE Access 2024;12:26419–27.
[146] Bharati S, Mondal MRH, Podder P, Kose U. Explainable artificial intelligence (XAI) with IoHT for smart healthcare: a review. Interpretable Cognitive Internet of Things for Healthcare 2023:1–24.
[147] Pietilä E, Moreno-Sánchez PA. When an Explanation is not Enough: an Overview of Evaluation Metrics of Explainable AI Systems in the Healthcare Domain. In: Mediterranean Conference on Medical and Biological Engineering and Computing. Springer Nature Switzerland; 2023. p. 573–84.
[148] Borys K, Schmitt YA, Nauta M, Seifert C, Krämer N, Friedrich CM, Nensa F. Explainable AI in medical imaging: an overview for clinical practitioners–Beyond saliency-based XAI approaches. Eur J Radiol 2023:110786.
[149] Hosny KM, Said W, Elmezain M, Kassem MA. Explainable deep inherent learning for multi-classes skin lesion classification. Appl Soft Comput 2024:111624.
[150] Alsubai S. Transfer learning based approach for lung and colon cancer detection using local binary pattern features and explainable artificial intelligence (AI) techniques. PeerJ Comput Sci 2024;10:e1996.
[151] Agarwal N, Tayal DK. FFT based ensembled model to predict ranks of higher educational institutions. Multimed Tools Appl 2022;81(23):34129–62.
[152] Agarwal N, Tayal DK. A new model based on the extended COPRAS method for improving performance during the accreditation process of Indian Higher Educational Institutions. Comput Appl Eng Educ 2023;31(3):728–54.

Zahra Sadeghi holds a Ph.D. degree in computer engineering, specializing in artificial intelligence and robotics. She has extensive experience in prestigious universities and industry collaborations, tackling diverse Machine Learning challenges. Her research interest includes Machine Learning, Computer Vision, Data mining and cognitive psychology.

Roohallah Alizadehsani obtained a Bachelor of Science and then a Master of Science degree in Computer Engineering-Software from Sharif University of Technology. Roohallah's research interests include mainly in the areas of Data Mining, Machine Learning, Bioinformatics, Heart Disease, Skin Disease, Diabetes Disease, Hepatitis Disease, and Cancer Disease. He is currently a Research Fellow at Deakin University of Australia.

Mehmet Akif Cifci is an accomplished Associate Professor in the field of computer science, currently contributing his expertise at TU Wien in Austria. Akif began his career with Topkapı University and Bandırma Onyedi Eylül University in Turkey, further enriching his experience at Klaipeda University in Lithuania. He is a member of IEEE, and has authored numerous impactful scientific papers.

Samina Kausar is currently working as an Assistant Professor in University of Kotli Azad Jammu and Kashmir, Pakistan. She has received her Ph.D. degree from School of Computer Engineering and Science, Shanghai University, China. Her research interests are in the fields of big data, bioinformatics, computer networks, cloud computing, data mining, E-learning and machine learning algorithms.

Rizwan Rehman is an Assistant Professor from the Centre for Computer Science and Applications, Dibrugarh University, who specializes in Computer Programming, Machine Learning and Speech Processing. He has done his Ph.D. in Computer Science from Dibrugarh University. He has a teaching experience of 20 years.

Priyakshi Mahanta received her PhD in computer science and engineering from Tezpur University on 2017. She is working as an Assistant Professor in Centre for Computer science and application of Dibrugarh University from 2014. Her research interest include computational biology, machine learning.

Pranjal Kumar Bora pursued PhD degree from Dibrugarh University with specialization in computational Biology in 2021. Dr. Bora currently works as an Assistant Professor in the Centre for Computer Science and Applications, Dibrugarh University. His-research interests include Machine Learning, Graph Theory and Computational Biology. He published thirteen numbers of research papers in reputed journals over the years.

Ammar K. Almasri received the Ph.D. degree in Management Information Systems/Artificial Intelligence from Cyprus International University, Cyprus, in 2020. He is currently teaching at Al-Balqa Applied University since 2009, Amman, Jordan. His-research interests include Big Data analysis, Artificial Intelligence, and Deep Learning.

Rami S. Alkhawaldeh received his B.S. degree in Computer Information Systems from Yarmouk University in 2007, MSc. Degree in Computer Information Systems from the University **of Jordan** in 2010. Dr. Alkhawaldeh pursued his PhD degree in computing science from Glasgow University in 2017. The research interests include Artificial Intelligence, Deep and Machine Learning, Information Retrieval, VoIP, and Wireless networks.

Sadiq Hussain received the Ph.D. degree from Dibrugarh University, Assam, India in 2017. His-research interests include data mining, medical analytics, and machine learning. A paper titled, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges" was selected as the 2022 Best Survey Paper Award from Information Fusion journal (Elsevier).

Bilal Alatas received his B.S., M.S., and Ph.D. degrees from Firat University. He works as a Professor of Software Engineering at Firat University. His-research interests include artificial intelligence, data mining, social network analysis, and metaheuristic optimization. He is the editor of twelve journals five of which are indexed in SCI and reviewer of seventy SCI-indexed journals.

Afshin Shoeibi has been a Research Intern with the Biomedical Machine Learning Laboratory, Graduate School of Biomedical Engineering, UNSW Sydney, Australia, since 2022. Additionally, he is a Research Assistant with the University of Granada, Spain, and Macquarie Group, Australia. His-research interests include biomedical signal processing, fuzzy systems, computational neuroscience, deep learning, and VLSI for machine learning.

Hossein Moosaei received his PhD in Applied Mathematics in 2013. He is currently collaborating with Jan Evangelista Purkyně University. His-current research interests include machine learning, optimization and its applications, biomedical applications, numerical analysis, numerical linear algebra, and scientific computing. He has served as a member of the editorial board, guest editor, and reviewer for some international journals.

Milan Hladík is a professor and the head of the Department of Applied Mathematics of the Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic. He specializes in optimization, operations research, interval computation and linear algebra. He is a member of the editorial board of five international journals, including European Journal of Operational Research.

Saeid Nahavandi is currently Swinburne University of Technology's inaugural Associate Deputy Vice-Chancellor Research and Chief of Defence Innovation. He previously served as Pro-Vice-Chancellor (Defence Technologies) and Founding Director of the Institute for Intelligent Systems Research and Innovation, Deakin University. He is a Fellow of IEEE (FIEEE), and Engineers Australia (FIEAust).

Panos M. Pardalos is a world-renowned leader in Global Optimization, Mathematical Modeling, Energy Systems, Financial applications, and Data Sciences. He was awarded 2013 EURO Gold Medal prize bestowed by the Association for European Operational Research Societies. He has published over 600 journal papers and edited/ authored over 200 books. He has lectured and given invited keynote addresses worldwide in different countries.