Received 16 January 2024, accepted 9 February 2024, date of publication 19 February 2024, date of current version 23 February 2024.
*Digital Object Identifier 10.1109/ACCESS.2024.3367606*

SURVEY

# Application of Example-Based Explainable Artificial Intelligence (XAI) for Analysis and Interpretation of Medical Imaging: A Systematic Review

**MIGUEL FONTES**[1,2], **JOÃO DALLYSON SOUSA DE ALMEIDA**[3], **AND ANTÓNIO CUNHA**[1,2]
[1]Institute for Systems and Computer Engineering, Technology and Science, 4200-465 Porto, Portugal
[2]School of Sciences and Technology, University of Trás-os-Montes and Alto Douro, 5000-801 Vila Real, Portugal
[3]Applied Computing Group, Federal University of Maranhão (UFMA), São Luís, Maranhão 65080-805, Brazil

Corresponding author: Miguel Fontes (miguel.f.fontes@inesctec.pt)

National Funds finance this work through the Portuguese funding agency, FCT—Fundacão para a Ciência e a Tecnologia, within project PTDC/EEI-EEE/5557/2020. Co-funded by the European Union (grant number 101095359) and supported by the UK Research and Innovation (grant number 10058099). However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency (HaDEA).

**ABSTRACT** Explainable Artificial Intelligence (XAI) is an area of growing interest, particularly in medical imaging, where example-based techniques show great potential. This paper is a systematic review of recent example-based XAI techniques, a promising approach that remains relatively unexplored in clinical practice and medical image analysis. A selection and analysis of recent studies using example-based XAI techniques for interpreting medical images was carried out. Several approaches were examined, highlighting how each contributes to increasing accuracy, transparency, and usability in medical applications. These techniques were compared and discussed in detail, considering their advantages and limitations in the context of medical imaging, with a focus on improving the integration of these technologies into clinical practice and medical decision-making. The review also pointed out gaps in current research, suggesting directions for future investigations. The need to develop XAI methods that are not only technically efficient but also ethically responsible and adaptable to the needs of healthcare professionals was emphasised. Thus, the paper sought to establish a solid foundation for understanding and advancing example-based XAI techniques in medical imaging, promoting a more integrated and patient-centred approach to medicine.

**INDEX TERMS** Explainable artificial intelligence, XAI, example-based, explanations by example, medical imaging.

## I. INTRODUCTION

Integrating deep learning (DL) technologies into medical image analysis has represented a fundamental advance in modern medicine. These sophisticated algorithms have been instrumental in detecting intricate patterns in various diagnostic images [1]. Their ability to challenge and, in some cases, surpass the diagnostic accuracy of human experts marked a significant leap forward in medical diagnosis. However, the actual effectiveness of these complex systems has often been overshadowed by their opaque, ''black box'' nature. This lack of transparency has raised serious concerns about their reliability and acceptability, especially in the high-risk domain of healthcare [2].

In this scenario, Explainable Artificial Intelligence (XAI) has emerged as an innovative key, meeting the growing demand for transparency and reliability in AI-powered systems. XAI endeavoured to illuminate the inner workings of AI models, offering crucial insights into their decision-making processes. This transparency was not only a matter of ethical importance but also of practical significance, especially in clinical environments where important decisions

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

often depended on the guidance provided by these systems [2].

The example-based techniques within the XAI domain were particularly notable for their user-friendly and intuitive nature. By providing explanations based on specific examples, these methods made DL models more understandable to a broader audience. This included healthcare professionals who did not have specialised skills in programming or computer vision. This example-based approach could democratise the understanding of DL models, making them more accessible and transparent to a broader spectrum of users, including doctors, patients and other healthcare stakeholders [3].

The paper focused on exploring and consolidating the most promising example-based XAI methods adapted for medical imaging. The aim was to provide a comprehensive overview of current strategies, highlighting their strengths and limitations. This endeavour has contributed to the ongoing discourse and research within the field, aimed at charting a detailed landscape of available example-based XAI solutions. Its potential to improve the practice of medicine and the processes involved in clinical decision-making was emphasised.

The analysis aims to enhance comprehension and application of these technologies, which are necessary practical tools for improving health outcomes and patient care. Advancing the knowledge and use of XAI, based on examples in medical imaging, ensures that these technologies are improved in their capabilities and aligned with the core values of trust, transparency, and reliability in healthcare. The alignment was crucial for integrating AI into medical diagnostics and instilling confidence in healthcare professionals and patients regarding these emerging technological tools.

## II. RELATED WORK

Section II is structured in three subsections, addressing different aspects of example-based XAI in medical image analysis. Firstly, section II. An analysis of current research into example-based XAI, providing an overview of existing contributions. Next, section II-B identifies shortcomings and opportunities for future research in this area. Finally, section II-C discusses this review's importance and specific contributions, emphasising its role in advancing the understanding of example-based XAI applied to medical image analysis.

### A. ANALYSIS OF EXISTING SYSTEMATIC REVIEWS

To uncover the panorama of example-based XAI applied to medical image analysis, the review focused on dissecting and understanding the contributions other authors had presented in the literature. The decision to concentrate on works on example-based XAI was strategic, grounded in the inherent capacity of these techniques to offer intuitive explanations crucial for clinical application. As illustrated in Table 1, this review categorises and summarises the different techniques of example-based XAI discussed in various reviews. By analysing existing studies, the authors sought to capture the full spectrum of approaches proposed to increase the clarity and reliability of DL systems, paving the way for their more effective integration in various fields, particularly medical imaging.

**TABLE 1. Techniques covered in the different systematic reviews.**

| Paper | [3] | [4] | [5] | [6] | [7] |
|---|---|---|---|---|---|
| Prototype | ✓ | ✓ | ✓ | ✓ | ✓ |
| Triplet Networks | ✓ | ✓ | | | |
| Influence functions | ✓ | | | | |
| Latent space examples | ✓ | | | | |
| CBR/CBIR | | | ✓ | | |
| Counterfactual explanations | | | ✓ | ✓ | ✓ |
| Criticisms | | | | ✓ | |
| Anchors | | | | | ✓ |
| Contrastive explanations | | | | | ✓ |
| Integrated Gradients | | | | | ✓ |

The paper [3] emphasises methods such as triple networks, influence functions, prototypes, and latent space examples. The triples network approach, in particular, is highlighted for its effective use in learning useful representations and explaining the model's reasoning by inspecting neighbours in the embedded representation.

The survey [4] discusses triple networks, ProtoPNet, and xDNN innovations. Triplet networks, with their example-based learning mechanism, are highlighted for their ability to bring the anchor closer to positive samples and further away from negative ones, providing a framework for explanations through similarities. ProtoPNet and xDNN, incorporating layers of prototypes, are illustrated as means for an intuitive "this looks like that" interpretation, promoting a bridge between the characteristics visualised in the medical images and the model outputs.

In a review [5], Case-based Reasoning/Content-Based Image Retrieval (CBR/CBIR), Counterfactual Explanations, and Prototypes are explored as methods that use example-based logic to make medical image classification systems more transparent. CBR and CBIR are recognised for their efficiency in retrieving and comparing images that best align with a query image, contributing to interpretation and confidence in the medical diagnosis.

The paper [6] also addresses Prototypes, Criticisms, and Counterfactuals, highlighting the significant role of prototypes in class representation and how criticisms can help clarify model limitations. At the same time, counterfactuals can illuminate how minimal changes can affect model output.

Finally, the study [7] reviews techniques such as Anchors, Counterfactual explanations, Contrastive Explanations, and Integrated Gradients. Anchors are particularly notable for establishing If-Then rules that identify crucial features influencing predictions. At the same time, Counterfactual explanations make it possible to understand which feature changes can modify the model's results.

### B. GAPS IN THE LITERATURE

A careful analysis of existing systematic reviews reveals several critical shortcomings when studying Example-based XAI.

Firstly, it can be observed that many of the reviews tend to approach Example-based XAI superficially, not penetrating sufficiently into the complexity or practical implications of these techniques. This superficial approach often fails to capture the essence and full potential of Example-based XAI in making informed clinical decisions and interpreting complex models.

Furthermore, there is a notable lack of consensus in the literature about which techniques should be classified under the Example-based XAI spectrum. This lack of uniformity confuses understanding of the field as a whole and makes it difficult to compare and evaluate the effectiveness of the proposed techniques.

Finally, another significant gap is the redundancies between the techniques presented as Example-based XAI. Many studies include approaches that, despite being labelled with different nomenclatures, show little differentiation in methodology or applicability, leading to a feeling of repetition and a potential dilution of the practical impact of Example-based XAI in the health area. Identifying and eliminating these redundancies is essential to distilling a set of genuinely distinct and effective techniques, promoting fundamental advances in artificial intelligence that healthcare professionals can confidently understand and apply.

### C. JUSTIFICATION AND CONTRIBUTIONS OF THE CURRENT REVIEW

The realisation of this systematic review and the expected contributions reflect a strategic move to overcome the limitations in the study of Example-based XAI, mainly applied to medical imaging. The current review aims to provide an in-depth analysis of the most recent and advanced techniques in the field of Example-based XAI, focusing on identifying and synthesising the methods that genuinely advance the interpretability of medical images.

This review is intended to establish a framework for understanding Example-based XAI techniques, creating a consensus on which methodologies are most effective and how they can be implemented to improve the analysis of medical images. The aim is to provide a compilation of the most pertinent innovations, enabling researchers and healthcare professionals to employ artificial intelligence methods with greater transparency and reliability. This will directly impact the quality of medical image interpretation, contributing to more accurate diagnoses and improved patient care.

## III. METHODOLOGY

This methodological section is dedicated to explaining the approach adopted for the systematic review, which is essential to ensure the reliability and replicability of the results.

In section III-A, the research questions that served as a beacon for the study are set out, directing the analysis and delimiting its scope. Next, in section III-B, the databases selected for the research are specified, with a view to a complete and representative exploration of the existing literature. The search strategy is detailed in section III-C, highlighting the search strings and selection steps, carefully chosen to capture as many relevant studies as possible. Finally, in section III-D, the characteristics to be observed that are fundamental to understanding and analysing the data collected are described, thus ensuring the robustness of the findings.

### A. RESEARCH QUESTIONS

The methodology of this systematic review focuses on three essential research questions that guide the analysis, each motivated by a desire to advance and make contributions to the field of XAI:

**RQ1:** What are the different existing example-based XAI methods, and what are their main advantages and limitations?

**Motivation:** Unraveling the different methods of Example-based XAI is the first step towards transforming how AI decisions are interacted with and trusted, equipping them with knowledge to maximise efficiency and transparency.

**RQ2:** In which areas of medical imaging are these methods applied, and how do they behave in terms of performance in these applications?

**Motivation:** Identifying the areas in medical imaging where XAI methods excel and understanding their performance characteristics are crucial for developing reliable and interpretable AI tools for clinical decision support. Our knowledge will drive innovations to enhance diagnostic accuracy and ultimately improve patient outcomes.

**RQ3:** What are the future challenges and opportunities for the field of example-based XAI?

**Motivation:** Exploring the challenges and prospects of Example-based XAI is not just an academic necessity; it is a quest to anticipate the future of AI, ensuring its responsible development and alignment with human values.

Each question reflects the dedication to deepening knowledge in the field of XAI, intending to contribute to a future in which artificial intelligence is more understandable, reliable, and beneficial to all.

### B. DATABASES

A set of widely recognised databases was used for the systematic review, including ScienceDirect, IEEE Xplore, ACM Digital Library, arXiv, Springer Link, Nature, and Google Scholar. These were selected for their breadth and depth of coverage in the field of interest, providing various research materials relevant to the study of XAI.

### C. RESEARCH STRATEGY

The search strategy adopted in this systematic review followed a meticulous approach, outlined in four main steps, to identify relevant studies discussing Example-based XAI in medical imaging.

**Step 1:** The search began with the application of a specific search string: (''Medical Image'') AND (''Explainable AI'' OR ''XAI'') AND (''Example-based'' OR ''Explanations by example''), restricting the results to papers published between 2021 and 2023, in English, to ensure current data pertinent to the focus of the study.

**Step 2:** This was followed by removing duplicates, which is essential to avoid overlaps and ensure clarity in the data analysis.

**Steps 3 and 4:** The next two steps focused on qualitatively selecting papers. Firstly, the papers were filtered by analysing the titles and abstracts, discarding those not strictly aligned with using Example-based XAI techniques in medical imaging. Next, an evaluation of the complete texts made it possible to exclude any study that, although tangentially related, did not directly contribute to the scope of the review. This process was carried out in such a way as to avoid redundancies, ensuring that only papers with direct contributions to the topic were included in the subsequent analysis.

This structured approach created a cohesive data set focused on understanding the applicability and challenges of Example-based XAI in medical imaging.

### D. EXTRACTION OF STUDY CHARACTERISTICS

In the systematic review process, carefully extracting the specific characteristics of each paper is fundamental to uncovering the current state of the use of Example-based XAI techniques in medical imaging. Each study is analysed to identify the Example-based XAI technique employed, focusing on its application and what particularities it presents when used on medical datasets. In addition, attention is paid to the type of medical image data in which the technique is applied to understand the adaptations and applications depending on the nature of the data. Finally, it is discerned whether the examples used to explain the decisions of AI algorithms are based on real cases or whether they are constructed synthetically. The source of the examples is critical to appreciating the validity and applicability of the explanations generated, with real examples being potentially more valuable for their authenticity. In contrast, synthetic ones can offer control and consistency for testing specific hypotheses. Gathering this information is essential for drawing up an overview of current practices and identifying trends and gaps in the application of Example-based XAI in medical imaging.

## IV. RESULTS

This section details the results obtained through the search strategy adopted, covering the selection and concise synthesis of the relevant studies. The search methodology was designed to capture a representative set of publications within the scope of Example-based XAI in medical imaging. Each paper was rigorously assessed, highlighting the research objective, the techniques used, the main findings and their respective implications for the field. This analysis provides a solid basis for understanding the current state and emerging trends in the application of XAI techniques to medical imaging.

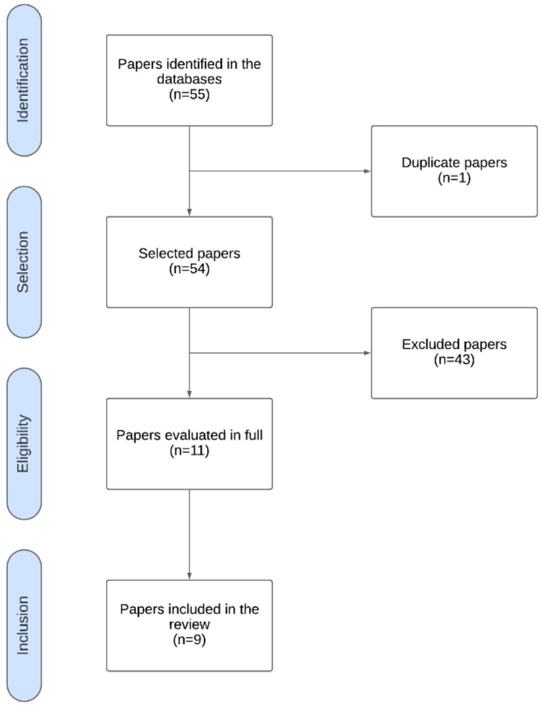


**FIGURE 1.** Flow diagram of the selection of the papers.

Figure 1 illustrates the process of selecting studies for the systematic review. Initially, 55 papers were identified in the various databases consulted using a specific search strategy. After removing one duplicate paper, an initial screening was carried out based on titles and abstracts, which resulted in the exclusion of 43 papers that did not align with the use of Example-based XAI techniques or the application to medical images. A detailed analysis of the full content of the remaining 11 papers led to the exclusion of more papers that did not fall within the specific scope of the review. Finally, 9 papers were considered strictly relevant to the analysis's focus and included for qualitative and quantitative synthesis.

Table 2 provides a detailed summary of the Example-based XAI techniques discussed in the papers selected for this review. It highlights the specific approaches used and the variations and contexts in which they were applied within the medical imaging domain. This table overviews current methodologies, allowing a direct comparative view between the different studies and their contributions to XAI in medical applications.

Two recent studies illustrate significant advances in prototype learning, a method that aims to improve the transparency and explainability of machine learning models. The first paper [8] highlights the ExPeRT (Explainable Prototype-based model for Regression using optimal Transport) architecture in the context of explainable regression. This innovation uses a convolutional neural network to extract

**TABLE 2. Different XAI Example-based techniques. Acronyms used: Ultrasound (US); Magnetic Resonance Imaging (MRI); X-Ray (XR).**

| Technique | Paper | Type of Image | Type of Example |
|---|---|---|---|
| Prototypes | [8] | US and MRI | Real |
| | [9] | Dermatoscopic | Real |
| CBIR/CBR | [10] | XR and Dermatoscopic | Real |
| | [11] | N.A. | Real/Synthetic |
| GAN's | [12] | Histological images | Synthetic |
| | [13] | XR, MRI, US, etc. | Synthetic |
| SCOPe | [14] | XR and Dermatoscopic | Synthetic |
| TraCE | [15] | XR | Synthetic |
| CapsNet | [16] | Endoscopy | Real |

features from images, complementing a fully connected neural network that determines distances in a latent space. The model has been trained with representative prototypes to facilitate accurate inferences based on their proximity to new samples. Applied to medical images, such as MRI scans and fetal ultrasounds, ExPeRT obtained results demonstrating state-of-the-art brain age prediction accuracy and providing advanced explainability, promoting trust and understanding among medical professionals.

The second piece of research [9] introduces a distinctive model for skin cancer diagnosis. Using convolutional neural networks, the model is based on class prototypes to offer clear interpretations of the decisions made by the algorithm. Through a framework that contrasts dermoscopy images with prototypes, measuring similarities through cosine similarity, the model demonstrated superior performance compared to other explainable approaches. Tested on the ISIC 2019 dataset, which encompasses various skin lesions, it proved particularly effective when integrating local prototypes. It increased diagnostic accuracy and enriched the field of XAI by offering detailed and accessible justifications for the diagnoses made.

Two studies about CBR/CBIR techniques were highlighted. The study [10] introduces X-MIR (EXplainable Medical Image Retrieval, a pioneering technique for explainable retrieval of medical images that integrates deep neural networks and deep metric learning. This innovative approach structures latent representations of medical images clustered, significantly improving the distinction between pathologies. The technique uses Euclidean and cosine distances to compare query images with an extensive database, identifying similar images with high precision. With its successful application to chest X-rays of COVID-19 patients and images of skin lesions, X-MIR has demonstrated its effectiveness in different clinical contexts, promoting highly accurate and interpretable medical image retrieval and reinforcing the role of XAI in medicine.

The research [11] explores the potential of "twin systems" in XAI, combining the processing power of DL models with the clarity and adaptability of CBR. The genius of this method lies in its dual system: a DL model, responsible for efficiency and accuracy, and a CBR model, which acts as an explanatory agent. The explanatory strategies, including factual, counterfactual, and semi-factual explanations, allow the CBR system to explain the DL model's decisions. Tested in a range of practical scenarios, from agriculture to medical diagnosis, this methodology has not only clarified the logic behind automatic predictions and diagnoses. Still, it has also promoted the generation of new training data to improve the generalizability and reduce the bias of DL models. This research exemplifies the continuous advance in the field of CBR/CBIR and underlines the importance of developing AI tools that are transparent and reliable for professionals in various fields.

Two recent studies on Generative Adversarial Networks (GANs) to create explanatory examples stood out for their significant advances in digital pathology and medicine. In work [12], a conditional Generative Adversarial Network (cGAN) based on the StyleGAN2 architecture was specialised to produce detailed histological images. The structure consists of a generator, responsible for creating realistic images, and a discriminator, which evaluates the authenticity of these images. cGAN was adequately trained with histological images covering various types of cancer. The results were remarkable, with the generation of images that experts judged to align with biological expectations and display peculiar characteristics of tumour subtypes. The images generated not only improved pathologists' training in identifying rare cancer subtypes but also enriched XAI by serving as examples to clarify the model's decisions, illustrating histological correlations with specific tumour subtypes and increasing the interpretability of histopathology results.

In parallel, the study [13] represents a milestone in using GANs to develop XAI within the medical domain. DeepSynthBody not only tackles the problem of limited data in the healthcare sector by generating synthetic data using advanced GANs but also integrates a crucial explanatory dimension. By implementing the "Explainable DeepSynth AI and DeepSynth XAI" layer, the model provides previously unattainable transparency, using the generated synthetic data as explanatory examples that clarify the reasoning behind the machine learning models. This exemplary XAI methodology promotes user confidence in AI systems. It addresses ethical concerns related to data privacy by replacing sensitive information from actual patients with synthetic data in explanatory processes. DeepSynthBody thus paves the way for future XAI research in healthcare, underlining the importance of clarity and transparency in applications that handle critical and confidential data.

Two papers offer techniques for creating synthetic examples that dispense with GANs. The first study introduces the SCOPe technique (Synthesizing Counterfactuals by Optimizing Pre-Images) [14], marking a significant advance in the explainability of deep classifiers. This approach is distinguished by its ability to generate accurate counterfactual explanations and maintain the original identity of the analysed input. The SCOPe consists of a trio of elements: an image generator based on Deep Image Priors (DIP) or Implicit Neural Representations (INRs), an OOD detector to validate the adherence of synthetic data to the distribution of actual data, and a classifier responsible for evaluating the explanations generated. Tests on a variety of datasets, such as CelebA, CheXpert, and ISIC2018, revealed that SCOPe not only preserves the identity of the inputs but also generates counterfactuals with semantic coherence, making it easier to interpret the decisions of DL models in critical situations.

The second study describes the development of TraCE (Training Calibration-based Explainers) [15], a methodology designed to provide interpretable, high-confidence counterfactual explanations applied to medical imaging. TraCE is built on a three-phase architecture consisting of a convolutional autoencoder that condenses images into a compact latent space, a predictive model that provides interpretations and evaluates uncertainties in predictions. A counterfactual optimisation algorithm focused on calibrating uncertainties. In addition, an auxiliary interval model is integrated to increase precision in the calibration of uncertainties. When applied to a large set of 30,000 chest X-ray images from the RSNA pneumonia detection challenge, TraCE demonstrated an exceptional ability to identify pathologies and clarify the relationship between patient characteristics and the severity of the condition, promoting greater clarity in medical decisions based on DL models, and reinforcing the critical role of AI in evidence-based medicine.

In conclusion, the results section of this systematic review in the XAI example-based domain focusing on medical images shows that the study [16] emerges as an important milestone. This research stands out for its innovative application of the CapsNet (Capsule Network) architecture in diagnosing gastrointestinal pathologies. It is mainly distinguished by its advanced use of visualisations of feature clusters derived from the outputs of class capsules. Through the dynamic routing mechanism peculiar to CapsNet, the clusters are evaluated in terms of their formation, separability, and density, providing an in-depth analysis of the model's discriminative potential. The study shows that the clusters obtained are more distinct and separable than those of conventional models, underlining the effectiveness of CapsNet in differentiating between diseases. This ability to visualise and interpret the model's decision-making bases reinforces the importance of clusters for post-hoc interpretability, strengthening the credibility of capsule networks as a diagnostic tool and advancing the use of XAI methods in critical areas such as health. This study emphasises the need for transparency and understanding of AI models, which are as essential as the accuracy of their predictions.

## V. DISCUSSION

Section V discusses four main aspects of example-based XAI techniques for analysing medical images. In section V-A, different XAI approaches were compared, highlighting their clarity and diagnostic accuracy advantages. In section V-B, the effectiveness and consistency of these techniques were evaluated. Section V-C. discussed the resources required to implement each XAI technique. Finally, section V-D. they discussed and developed the research questions. This section provided a clear overview of XAI techniques, emphasising their uses and challenges.

### A. COMPARISON OF TECHNIQUES

CBR/CBIR techniques, such as DeepCBR and X-MIR, are noted for their strong tendency towards transparency and interpretability. DeepCBR, in particular, effectively simplifies the inherent complexity of DL by providing diverse explanations, ranging from the factual to the counterfactual [11]. On the other hand, X-MIR uses deep neural networks and deep metric learning to optimise the retrieval of medical images, excelling in differentiating between various pathologies, including cases of COVID-19 and skin lesions [10].

On the other hand, Prototypes techniques, applied in contexts such as skin cancer diagnosis and explainable regression, focus on diagnostic accuracy and explanatory details. These approaches, which include using cosine similarity measures and integrating local prototypes, balance comprehensive explainability with high diagnostic efficiency [9]. The ExPeRT model exemplifies this methodology, employing convolutional neural networks for detailed latent analysis [8].

GAN-based techniques such as cGANs and DeepSynthBody provide a global perspective by generating synthetic images. cGANs excel at generating hypotheses and visual insights into image characteristics, while DeepSynthBody distinguishes itself by integrating explainability into all stages of the generation process, reinforcing confidence in model results [12], [13]. In contrast, approaches such as SCOPe and TraCE, which do not rely on GANs, focus on creating high-quality counterfactuals, demonstrating versatility and skill in identifying shortcuts in complex models [14], [15].

Furthermore, CapsNet deserves to be highlighted due to its ability to model spatial variability in images and its robustness in reduced data sets. The ability to reconstruct input images through the decoder layer is a significant advantage, particularly in clinical contexts where accuracy and clarity are essential [16].

Specific objectives and application context should guide the selection of XAI techniques. Choosing between methods that emphasise transparency, diagnostic accuracy, or data flexibility is crucial to drive the practical application of machine learning in medicine, ensuring reliable and transparent results.

### B. VALIDATION OF TECHNIQUES

Various explainability techniques in DL models exhibit their strengths in terms of validity and consistency. The X-MIR method uses saliency maps, evaluated by metrics such as insertion and deletion. Still, it shows variations in results between different algorithms, suggesting that consistency may depend on the specific method [10]. CapsNet and TraCE are notable for their intuitive visualisations and generation of reliable counterfactuals, respectively, offering detailed explanations in line with human interpretations [15], [16]. ExPeRT, through optimal transport pairing, and the Global and Local Explanations using the Prototypes model provide clear and coherent justifications. At the same time, SCOPe stands out for generating counterfactuals close to the data manifold [8], [9], [14]. DeepSynthBody raises the bar by integrating explainability into the synthetic data generation process, increasing confidence in machine learning solutions in the medical domain [13].

Choosing the most appropriate method depends on several factors, including the validity of the explanations, the specific application context, and clinical needs. Techniques such as DeepCBR, with the Twinning approach, and cGANs, in the generation of synthetic histology, demonstrate consistency in their explanations. However, there are variations depending on the specific strategy and context of the model [11], [12]. This variety of approaches emphasises the importance of careful analysis when selecting explainability techniques for medical imaging applications, with a view to effective and reliable implementations in contexts where precision and clarity are crucial.

### C. COMPUTATIONAL COST OF THE TECHNIQUES

Regarding the computational cost of the different DL techniques, it is crucial to consider the variations in the resources required for each method. For example, training models such as cGANs and DeepSynthBody, including convolutional neural networks and synthetic data generation, require significant computational resources [12], [13]. These methods can demand more time and resources, especially when generating large volumes of data or using complex models. However, optimisations such as hardware acceleration and parallelisation techniques can mitigate these costs.

On the other hand, the DeepCBR Twinning method, which maps DL models to a more transparent CBR model, can be applied to image classification problems. Still, there is no detailed information on its computational cost [11]. Similarly, SCOPe, Global and Local Explanations for Skin Cancer Diagnosis Using Prototypes, ExPeRT, and TraCE, despite their efficiency in generating explanations, do not have specific information about the computational cost detailed in the documents [8], [9], [14], [15]. For ExPeRT, using the Sinkhorn algorithm and entropic regularisation suggests a moderate computational cost [8]. Overall, the choice of a particular technique should consider its explanatory effectiveness, the computational resources available, and the environment in which it will be implemented, especially in cases of large data sets or hardware limitations.

### D. DISCUSSING AND DEVELOPING RESEARCH QUESTIONS

**RQ1:** Research into example-based XAI methods reveals a range of innovative approaches, each with unique characteristics. Within the scope of Prototype Learning, the ExPeRT architecture and the model for diagnosing skin cancer are remarkable [8], [9]. ExPeRT uses a convolutional neural network to extract features from images and a fully connected neural network to determine distances in a latent space. This results in accurate predictions and advanced explainability, especially in medical images such as MRIs and fetal ultrasounds. On the other hand, the model for diagnosing skin cancer relies on class prototypes to provide clear interpretations of decisions made by the algorithm, demonstrating superiority compared to other explainable approaches, especially when integrating local prototypes.

As for CBR and CBIR techniques, X-MIR and CBR "twin systems" are exemplary [10], [11]. X-MIR integrates deep neural networks and deep metric learning, significantly improving the distinction between different pathologies. On the other hand, the "twin-systems" method combines the efficiency of DL models with the clarity of CBR, offering detailed explanations of DL models' decisions.

Regarding GANs, cGAN's and DeepSynthBody are highlighted [12], [13]. The cGAN produces detailed histological images aligned with biological expectations, improving the training of pathologists. DeepSynthBody, in turn, addresses data limitations in medicine by generating synthetic data and integrating a crucial explanatory dimension in the process.

Finally, in counterfactual explanations, work on SCOPe and TraCE offers notable approaches [14], [15]. SCOPe generates accurate counterfactuals, maintaining the original identity of the input. At the same time, TraCE, when applied to a vast set of chest X-ray images, identifies pathologies and clarifies the relationship between patient characteristics and the severity of the condition, reinforcing the role of AI in evidence-based medicine.

**RQ2:** The applicability and performance of example-based XAI techniques are widely demonstrated in several areas of medicine. CapsNet has been applied in diagnosing gastrointestinal pathologies. It stands out for its ability to model spatial variability in images and generate intuitive visualisations of feature clusters, which are fundamental for the post hoc interpretation of model decisions. This advance reinforces the credibility of capsule networks as a diagnostic tool, underlining the importance of XAI methods in healthcare [16].

ExPeRT and the model for skin cancer diagnosis demonstrate high diagnostic accuracy, especially in the context of explainable regression and in the diagnosis of skin lesions. ExPeRT, by combining convolutional neural networks with fully connected neural networks, provides advanced explainability. At the same time, with its class prototype-based approach, the model for skin cancer diagnosis offers clear and detailed interpretations of algorithm decisions, highlighting the usefulness of these techniques in clinical applications [8], [9].

CBR/CBIR techniques, such as X-MIR and the "twin-systems" approach, show effectiveness in different clinical settings, including diagnosing COVID-19 and skin lesions. X-MIR integrates deep metric learning with deep neural networks to structure latent representations of medical images, promoting highly accurate and interpretable image retrieval [10], [11].

GANs, including cGAN and DeepSynthBody, are crucial in generating synthetic images and integrating explainability into the data generation process. These methods improve the training of pathologists in identifying rare subtypes of cancer, promote user trust in AI systems, and address ethical concerns related to data privacy [12], [13].

**RQ3:** The future of example-based XAI in medical imaging presents promising challenges and opportunities. One of the main challenges is the need for more consistent approaches with lower computational costs. For example, despite its effectiveness, the X-MIR method presents variations in results, indicating a need for greater consistency [10]. Furthermore, techniques such as cGANs and DeepSynthBody, although innovative in generating synthetic data, require significant computational resources, highlighting the need for optimisations such as hardware acceleration and parallelisation techniques [12], [13].

Regarding opportunities, integration with other data modalities and developing hybrid methods are promising areas. For example, the "twin-systems" approach combines the power of DL models with the clarity of CBR, offering detailed explanations of DL model decisions. This demonstrates the potential of hybrid methods to provide more prosperous, more multifaceted explanations [11].

CapsNet effectively differentiates between diseases and in advanced visualisation of feature clusters. This ability to visualise and interpret the model's decision-making bases reinforces the importance of advances in intuitive visualisations and the generation of reliable counterfactuals [16].

Furthermore, adapting to new AI technologies is crucial to the future of XAI. Methods such as SCOPe and TraCE, which generate detailed and coherent explanations, illustrate the ability to create precise and coherent counterfactuals aligned with human interpretations, promoting greater clarity in medical decisions based on DL models [14], [15].

## VI. CONCLUSION

In conclusion, this systematic review highlights the growing importance of example-based XAI techniques in medical images, establishing a fundamental bridge between the complexity inherent in DL models and the pressing need for transparency and understandability in medicine. The methodologies explored, including CapsNet, GANs, and counterfactual approaches, reveal a field full of innovations and possibilities, paving the way for more reliable and accessible clinical applications.

The answers obtained concerning the research questions reflected the diversity and potential of these methods.

- **Q1:** What are the different existing example-based XAI methods, and what are their main advantages and limitations?
  **R:** Several example-based XAI methods that stood out for their accuracy and explanatory power in medical imaging were identified. These methods combined advanced techniques such as DL and CBR/CBIR with innovative approaches in GANs and counterfactual explanations.
- **Q2:** In which areas are these methods applied, and how do they behave in terms of performance in these applications?
  **R:** They have demonstrated efficacy in some medical areas, excelling in diagnostic accuracy and synthetic image generation, contributing to increasing confidence in AI systems and addressing ethical issues.
- **Q3:** What are the future challenges and opportunities for the field of example-based XAI?
  **R:** The future of XAI in medical imaging was identified as overcoming challenges such as the need for consistency and computational efficiency, seizing opportunities in hybrid methods and intuitive visualisations, and adapting to new AI technologies for more transparent and more coherent explanations.

These approaches not only elevate confidence in AI-supported decisions but also encourage greater involvement of healthcare professionals in the decision-making process, aiding in understanding the principles and limitations of AI models.

Moreover, the growing implementation of example-based XAI in medical imaging is pivotal in alleviating ethical concerns and enhancing the acceptance of AI systems among professionals and patients. By offering clear and intuitive explanations, these methods elucidate AI model outcomes, enabling a more critical and informed evaluation of their suggestions. Thus, example-based XAI enriches AI in healthcare and emphasises the importance of humanisation and ethics in digital medicine. This study contributes significantly to comprehending XAI techniques in medical imaging, laying a robust groundwork for future research and practical deployments in healthcare.

## ACKNOWLEDGMENT

National Funds finance this work through the Portuguese funding agency, FCT—Fundacão para a Ciência e a Tecnologia, within project PTDC/EEI-EEE/5557/2020. Co-funded by the European Union (grant number 101095359) and supported by the UK Research and Innovation (grant number 10058099). However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency (HaDEA).

## REFERENCES

[1] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, no. 1, pp. 221–248, Jun. 2017, doi: 10.1146/annurev-bioeng-071516-044442.

[2] A. Sohail, M. A. Fahmy, and U. A. Khan, "XAI hybrid multi-staged algorithm for routine & quantum boosted oncological medical imaging," *Comput. Part. Mech.*, vol. 10, no. 2, pp. 209–219, Apr. 2023, doi: 10.1007/s40571-022-00490-w.

[3] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102470, doi: 10.1016/j.media.2022.102470.

[4] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable AI techniques in healthcare," *Sensors*, vol. 23, no. 2, p. 634, Jan. 2023, doi: 10.3390/s23020634.

[5] C. Patrício, J. C. Neves, and L. F. Teixeira, "Explainable deep learning methods in medical image classification: A survey," *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1–41, Apr. 2024, doi: 10.1145/3625287.

[6] T. Clement, N. Kemmerzell, M. Abdelaal, and M. Amberg, "XAIR: A systematic metareview of explainable AI (XAI) aligned to the software development process," *Mach. Learn. Knowl. Extraction*, vol. 5, no. 1, pp. 78–108, Jan. 2023, doi: 10.3390/make5010006.

[7] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Jan. 2023, doi: 10.1145/3561048.

[8] L. S. Hesse, N. K. Dinsdale, and A. I. L. Namburete, "Prototype learning for explainable brain age prediction," 2023, *arXiv:2306.09858*.

[9] C. Santiago, M. Correia, M. R. Verdelho, A. Bissoto, and C. Barata, "Global and local explanations for skin cancer diagnosis using prototypes," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* (Lecture Notes in Computer Science), M. E. Celebi, M. S. Salekin, H. Kim, S. Albarqouni, C. Barata, A. Halpern, P. Tschandl, M. Combalia, Y. Liu, G. Zamzmi, J. Levy, H. Rangwala, A. Reinke, D. Wynn, B. Landman, W.-K. Jeong, Y. Shen, Z. Deng, S. Bakas, X. Li, C. Qin, N. Rieke, H. Roth, and D. Xu, Eds. Cham, switzerland: Springer, 2023, pp. 47–56, doi: 10.1007/978-3-031-47401-9_5.

[10] B. Hu, B. Vasu, and A. Hoogs, "X-MIR: EXplainable medical image retrieval," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 1544–1554. [Online]. Available: https://openaccess.thecvf.com/content/WACV2022/html/Hu_X-MIR_EXplainable_Medical_Image_Retrieval_WACV_2022_paper.html

[11] M. T. Keane, E. M. Kenny, M. Temraz, D. Greene, and B. Smyth, "Twin systems for DeepCBR: A menagerie of deep learning and case-based reasoning pairings for explanation and data augmentation," 2021, *arXiv:2104.14461*.

[12] J. M. Dolezal et al., "Deep learning generates synthetic cancer histology for explainability and education," *NPJ Precis. Oncol.*, vol. 7, no. 1, May 2023, Art. no. 1, doi: 10.1038/s41698-023-00399-4.

[13] V. Thambawita, S. A. Hicks, J. Isaksen, M. H. Stensen, T. B. Haugen, J. Kanters, S. Parasa, T. de Lange, H. D. Johansen, D. Johansen, H. L. Hammer, P. Halvorsen, and M. A. Riegler, "DeepSynthBody: The beginning of the end for data deficiency in medicine," in *Proc. Int. Conf. Appl. Artif. Intell. (ICAPAI)*, May 2021, pp. 1–8, doi: 10.1109/ICAPAI49758.2021.9462062.

[14] J. J. Thiagarajan, V. Narayanaswamy, D. Rajan, J. Liang, A. Chaudhari, and A. Spanias, "Designing counterfactual generators on-the-fly," Lawrence Livermore Nat. Lab. (LLNL), Livermore, CA, USA, Tech. Rep. LLNL-CONF-819407, Feb. 2021. Accessed: Feb. 19, 2024. [Online]. Available: https://www.osti.gov/biblio/1810656

[15] J. J. Thiagarajan, K. Thopalli, D. Rajan, and P. Turaga, "Training calibration-based counterfactual explainers for deep learning models in medical image analysis," *Sci. Rep.*, vol. 12, no. 1, Jan. 2022, Art. no. 1, doi: 10.1038/s41598-021-04529-5.

[16] M. A. Ayidzoe, Y. Yongbin, P. K. Mensah, J. Cai, and F. U. Bawah, "Visual interpretability of capsule network for medical image analysis," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 30, no. 3, pp. 978–995, Mar. 2022, doi: 10.55730/1300-0632.3822.



**MIGUEL FONTES** received the bachelor's degree in bioengineering from the University of Trás-os-Montes and Alto Douro, where he is currently pursuing the master's degree in biomedical engineering. He is a Researcher with the Computer Assisted Gastric Cancer Diagnosis, a project at Inesc Tec and the University of Trás-os-Montes e Alto Douro. His research interests include example-based XAI techniques and deep-learning methods for detecting gastric lesions.

**JOÃO DALLYSON SOUSA DE ALMEIDA** received the bachelor's degree in computer science and the master's and Ph.D. degrees in electrical engineering from the Federal University of Maranhão (UFMA), in 2007, 2010, and 2013, respectively. He is currently an Associate Professor with UFMA. He coordinates the Vision and Image Processing Laboratory (VipLab-UFMA). He has experience in computer science, working mainly on the following topics, such as image processing, machine learning, ophthalmic medical images, and time series.

**ANTÓNIO CUNHA** is currently a Ph.D. Senior Researcher and an Auxiliary Professor with the Department of Engineering, University of Trás-os-Montes and Alto Douro (UTAD). He participated as a member of seven funded research projects. Since 2015, he has been a member of the Centre for Biomedical Engineering Research (C-BER), INESC TEC, where one of the final goals is to create a CAD system based on DL approaches to assist clinical doctors in different biomedical image-related diagnoses/screening tasks. Thus, he has plenty of experience in medical imaging analysis and student supervision. In the last five years, he supervised two Ph.D. students and 22 M.Sc. students and published 17 journal articles and 34 Scopus conference papers in this area. His research interests include medical image analysis, bio-image analysis, computer vision, machine learning, and artificial intelligence, particularly in computer-aided diagnosis applied in several imaging modalities, such as computed tomography of the lung and endoscopic videos.

He was a part of the organization committee HCIST-International Conference on Health and Social Care Information Systems and Technologies, from 2013 to 2015 and from 2020 to 2023, the Organization Chair, in 2012, and an Advisory Board, from 2016 to 2023. He is the General Chair of the MobiHealth2022 International Conference. He is a referee for several international journals and conferences and participated yearly in organizing several international scientific events.