

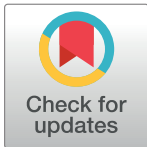
EDUCATION

Ten quick tips for computational analysis of medical images

Davide Chicco^{1*}, Rakesh Shiradkar²¹ Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada,² Department of Biomedical Engineering, Emory University, Atlanta, Georgia, United States of America* davidechicco@davidechicco.it

Abstract

Medical imaging is a great asset for modern medicine, since it allows physicians to spatially interrogate a disease site, resulting in precise intervention for diagnosis and treatment, and to observe particular aspect of patients' conditions that otherwise would not be noticeable. Computational analysis of medical images, moreover, can allow the discovery of disease patterns and correlations among cohorts of patients with the same disease, thus suggesting common causes or providing useful information for better therapies and cures. Machine learning and deep learning applied to medical images, in particular, have produced new, unprecedented results that can pave the way to advanced frontiers of medical discoveries. While computational analysis of medical images has become easier, however, the possibility to make mistakes or generate inflated or misleading results has become easier, too, hindering reproducibility and deployment. In this article, we provide ten quick tips to perform computational analysis of medical images avoiding common mistakes and pitfalls that we noticed in multiple studies in the past. We believe our ten guidelines, if taken into practice, can help the computational–medical imaging community to perform better scientific research that eventually can have a positive impact on the lives of patients worldwide.



OPEN ACCESS

Citation: Chicco D, Shiradkar R (2023) Ten quick tips for computational analysis of medical images. *PLoS Comput Biol* 19(1): e1010778. <https://doi.org/10.1371/journal.pcbi.1010778>

Editor: Patricia M. Palagi, SIB Swiss Institute of Bioinformatics, SWITZERLAND

Published: January 5, 2023

Copyright: © 2023 Chicco, Shiradkar. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Medical images are useful data elements that provide visual and spatial information regarding the condition of a particular organ or tissue of a patient. Medical images supply a representation of a health condition to medical doctors, who can interpret them and made better decisions regarding treatments or therapies. Since the 1970s, new biomedical engineering technologies have allowed the medical community to develop novel medical imaging modalities to evaluate and assess the medical conditions of patients. To mention a few, these image modalities include X-ray radiography, magnetic resonance imaging (MRI), positron emission tomography (PET), computed tomography (CT) scan, and single-photon emission computed tomography (SPECT). Since 2010s, new computational analyses of medical images have been made possible through machine learning [1–3]. In the meanwhile, open access medical images have become available online on several data repositories. While performing computational analyses of medical images has become easier, making mistakes during these analyses has become easier, too.

Following the examples of the debates about best practices for machine learning in bioinformatics [4–9], best practices for machine learning in health informatics [10–12], best practices for computational statistics [13,14], and best practices for pathway enrichment analysis [15–17], we decided to present here ten quick tips to perform computational analyses of medical images avoiding common mistakes and pitfalls that we noticed in multiple studies on this topic. In the biomedical literature, some studies already reported guidelines for machine learning applications to medical image analysis [18–21]: Most of those articles, although useful, are limited to computational intelligence.

With this article, instead, we propose ten quick tips not only related to data mining and pattern recognition, but to any computational analysis of health imaging data, including segmentation, coregistration, quality improvement, and classification. In the past, the PLOS Computational Biology Quick Tips series published no articles on guidelines for medical images and released only a study on scripts for neuroimaging [22]; we fill the gap by presenting this study. We designed these simple recommendations for beginners, but we believe they should be kept in mind by experts, too.

Tip 1: Before starting, make sure you have enough computational resources

As simple as it might seem, before starting any computational analysis, you have to make sure you have enough computational resources, both in terms of hard disk drives and random-access memory (RAM) slots. For other biomedical informatics fields, such as bioinformatics or analysis of electronic health records, one might tend to forget about this issue, because the data files are rarely so large to need special treatments.

A recent study one of us authors published, for example, involved the machine learning analysis of data of patients with inflammatory bowel disease [23]: The dataset analyzed was contained in a comma-separated values (CSV) file of 13 kilobytes size, previously released openly on FigShare [24]. Another study involved the usage of microarray gene expression data of patients with heart failure [25], whose dataset was originally released on Gene Expression Omnibus (GEO) as a TAR archive of CEL files of 1.8 gigabytes [26]. For the computational analysis done on both these datasets, a personal computer with limited resources such as Dell Latitude E5420 with Intel Core i5-2520M central processing unit (CPU) @ 2.50 gigahertz processor and 5.7 gigabytes RAM was sufficient, but it would be unsuitable for computational analyses of larger datasets, such as medical images.

Medical image datasets, in fact, have much larger sizes. The collection of magnetic resonance imaging scans for glioblastoma patients from the University of Pennsylvania Health System on the Cancer Imaging Archive (TCIA) [27], for example, contains Digital Imaging and Communications in Medicine (DICOM) files of 139.4 gigabytes. And the dataset of the OASIS brain project of neuroimaging scans weights approximately 114 gigabytes [28,29], just to name another one. With datasets of these sizes, most common laptop would be insufficient to perform any scientific analysis in reasonable time. To be more precise, not only medical images' data require larger computational resources, but also bioinformatics data (such as single-cell RNA-seq [30], for example) could easily need more than hundreds of gigabytes of disk spaces, when thousands or millions of cells are involved in the studies.

Scientific analysis of large datasets like single-cell RNA-seq data or medical images necessitates of high performance computing systems able to process a fair number of medical images in few minutes [31,32]. Moreover, computational resource allocation depends also on which computational techniques you plan to use for your scientific analysis: deep learning models (for example, convolutional neural networks; [33]) would require more RAM memory and disk space than simpler machine learning methods (such as linear regression) or traditional

statistics tests (for example, Mann–Whitney U test; [34]), of course. So, in addition to the data size, you need to consider which methods you would like to employ.

Therefore, here is our first simple tip: Before embarking in an image analysis project, make sure you enough computational resources to perform the analysis. If necessary, go and talk to the computational resources department of your institute or company. Usage of high-performance computing (HPC) [35] or graphics processing units (GPUs) [36] can be beneficial for the analysis.

Tip 2: Before starting, make sure you have enough medical images for your study

Presentation of disease patterns on medical imaging is often subtle, and radiologists require significant training to identify them. Use of machine learning algorithms to automate the task of detection and classification of disease patterns on medical imaging is often difficult. Therefore, a sufficient number of studies that capture different presentations of disease patterns under varying conditions including scanner differences, extent of disease, and subtypes of disease is required to train reliable and reproducible machine learning models.

Unfortunately, a number of machine learning studies that have been published in the past were just proofs of concept and preliminary studies. They have often used small datasets and have demonstrated the promise of the machine learning potential; however, such studies have often failed when validated on external datasets [37]. As the machine learning field has matured, one of the important considerations for publishing high-impact studies is to demonstrate the reproducibility of these methods [38].

While we cannot provide a general rule of thumb for the sample size required for given study, we recommend that power analysis and sample size calculations are conducted prior to conducting experiments. Several methods and software programs are available for sample size calculations such as MedCalc [39], Power Analysis and Sample Size (PASS) [40], R sampler and pwr packages [41,42], and IBM Statistical Package for the Social Sciences (SPSS) [43], which can help determine sample size for a given study.

Another aspect to keep in mind for binary classification and regression analysis tasks is class imbalance: When dataset has skewed class proportions, with one class being much larger than the others, some problems may arise, which might compromise the prediction task [5]. In these cases, we advise using techniques for class imbalance data handling [44], such as data weighting [45], oversampling [46], undersampling [44], and data augmentation [47]. The data imbalance issue, moreover, can be noticed and handled only by using metrics that take it into account, such as the Matthews correlation coefficient (MCC) for binary classifications and the coefficient of determination (R^2) for regression analyses (Tip #8).

Tip 3: Take care of noise and apply image harmonization

Noise. Inexperienced users and beginners often do not know about it, but medical images always come with noise. Always. Noise is an invisible issue: You might think it is not there, but it is. And repositories, which provide medical images, do not mention it, because it is implied: No repository presents their data by writing something like “breast cancer screening mammograms images with noise,” for example.

The bad news is that noise is there, anyway. The good news is there are methods for handling them, called denoising techniques [48]. Specific denoising techniques are suitable for specific medical images, such as X-ray radiography images [49–51], CT scan images [52–54], magnetic resonance images (MRI) [55–57], ultrasound images [58,59], positron emission

tomography (PET) images [60], just to mention a few. Our recommendation, simply, is not to analyze the raw dataset as it comes, but always remember to perform a denoising step first.

Harmonization. Data harmonization implies minimizing biases and variations that might arise on account of differences in scan parameters, post-processing protocol, and site-specific aspects [61–63].

When datasets from multiple institutions and hospitals are involved, harmonization becomes crucial to ensure that studies belonging to a specific class have similar intensity ranges and patterns. There are several open source tools available that can assist in checking the presence of batch effects arising due to scanner and site-specific variations. MRQy [64] and LAB-QA2GO [65], for example, are some recently developed tools with the aim of identifying batch effects and assist in quality control. There are several methods available for data harmonization.

One of the simplest methods is histogram matching, where intensity distributions of scans from multiple batches (scanners/sites) are aligned, rendering the processed scans to share a similar appearance [66]. Some of the more recent and advanced techniques use deep learning for harmonizing datasets [67]. In certain cases, despite the use of image harmonization tools, batch effects continue to persist and in such cases, one might make a call to develop and validate machine learning or computational image analytic approaches that are specific to a particular site or scanner.

Additionally, we can consider using computational imaging approaches that are independent of image intensity measurements such as shape and volume based features [68].

Tip 4: Only use open source programming languages and software platforms

When one decides to start a project involving computational analyses of medical images, they face an important question: Which programming language and software platforms should I use for this study? The answer is clear and straight for us: a fast, open source one like Python, R, C, C++, or Java.

Using an open source programming language brings several benefits and advantages. First, the programmer would be able to easily and freely share their software code with collaborators, and even reuse it after the project has finished, even if they moved to a different institution or company. Moreover, using an open permissive software license would allow the reproducibility of the computational analysis [69]. We also suggest to keep an eye on the TIOBE Programming Community Index [70] to get updates on the most used programming languages worldwide.

Since Python is not only the most popular programming language in machine learning and data science, but also the most used programming language in any field [70], we recommend all the beginners and students to start their projects with it. Moreover, the software packages' catalogue of Python include some popular, effective software libraries for deep learning, such as TensorFlow [71] and PyTorch [72].

Avoid proprietary software: Proprietary software programs cannot be shared openly and restrict their usage only to users who have their expensive licenses. Moreover, what is the point of spending your money or your institution's money for a license when free, open source alternatives are available? Scientific researchers often work in public universities and research centres funded by taxpayers' money. In this context, we suggest to keep in mind the quote of the economist Bruno Leoni: "Money of everybody should not be treated like money of nobody." At the end of the project, we also advise releasing publicly your software code on public repositories such as GitLab or GitHub, to enhance reproducibility [69], and facilitate the detection of possible mistakes in the analyses [73].

Tip 5: Identify features specific to your scientific problem

Computationally derived features provide alternate representations that enhance specific attributes and patterns for specific tasks. For instance, feature representations have been used for coregistration [74], segmentation [75], and classification [68].

Tasks using medical imaging data typically take advantage of handcrafted features from the diseased regions of interest to train on. Pyradiomics is a popular open source radiomic package that can serve as a general start to extract computational features from radiology imaging data [76]. This library includes intensity statistics, first order, second order, co-occurrence-based features. This can be a starting point for several classification tasks where we would like to quantify lesion characteristics using quantitative features.

However, depending on the clinical problem and disease, the starting set of features can be different. For instance, if we are planning to quantify the shape of a lesion, then volume and shape-based features such as fractals [77]. In case, lesion edge and gradients are the region where differential characteristics of less and more aggressive lesions are observed, then we might start off with edge-based features [78]. As deep learning methods are becoming popular, they are also being used as feature extractors [79]. Features from deep learning networks can be integrated with other clinical features, which may not be possible when using end-to-end deep learning networks.

Deep learning features can also be used in conjunction with simpler machine learning classifiers such as Naïve Bayes or logistic regression. These pieces of advice are useful not only for machine learning but also for segmentation [80] and coregistration [81] tasks.

Tip 6: Reduce overfitting through feature selection and dimensionality reduction

Classification is one of the most common tasks in computational imaging, and one of its important challenges is to avoid overfitting for the models to be generalizable and reproducible [82]. As mentioned earlier, open source feature extraction libraries such as Pyradiomics [76] provide a large initial set of features.

When these techniques are used to train machine learning models, overfitting occurs, and this problem is commonly known as the curse of dimensionality [83]. To avoid this problem, we need to reduce the set of features we work with. As a rule of thumb, use less than 10% of the sample size as the number of features for classification problem. For machine learning and deep learning tasks, imaging data are first converted to a feature vectors on which feature selection and dimensionality reduction techniques are then employed.

The feature vector usually includes shape features, color histogram features, color moment features, and texture features [79]. Often the success of a machine learning application depends on how accurately one can quantify input data (image or volume or text) in terms of quantitative features. To get to a reduced feature set, there are basically two methods one can adopt: feature pruning and dimensionality reduction.

Feature pruning. Feature pruning approaches involve discarding those features that are not important for the classification problem.

One way to approach this is to identify features that show significant differences between different classes using a *t* test and only retain those. Another way to do this is to remove features that are highly correlated with each other and essentially are duplicates of each other. Correlation coefficients such as Pearson's correlation coefficient (PCC) can be used to set threshold (for example, $PCC > 0.9$) and discard features that are highly correlated. Lastly, feature selection methods (for example, minimum redundancy maximum relevance (MRMR);

[84]) and feature ranking methods (for example, feature importance in nonlinear embedding (FINE); [85]) can be used to identify a subset of features that can be used for machine learning.

Dimensionality reduction. Dimensionality reduction approaches essentially identify characteristics of the features in the high dimensional space and determine approximations in the low dimensional space [86,87].

A popular dimensionality reduction method is principal component analysis where eigenvectors pertaining to the top eigenvalues are used to obtain feature characteristics [88]. Machine learning models can then be trained on these low dimensional feature vectors. Recently, embedding approaches are being widely used for dimensionality reduction.

Uniform manifold approximation and projection (UMAP) [89] and t-distributed stochastic neighbor embedding (t-SNE) [90] are popular embedding approaches for visualization used in a wide variety of machine learning applications. Choice of feature selection and dimensionality reduction depends on whether one cares about the interpretability of the features. Feature selection retains the original features that can then be overlaid over medical imaging data on a pixel-wise basis for visualization and interpretation.

However, if all one cares is a good classification performance, dimensionality reduction methods provide an efficient means.

Tip 7: Split your dataset into training set and test set, and look for a validation cohort dataset online

Dataset split. Splitting the dataset into a training set and a test set, by making sure that no data element is shared between the two, is a key pillar of any supervised machine learning analysis [5]. We would like to reaffirm this concept and enforce it by stating clearly that the same practice should be employed also in any studies, even the ones not involving a computational intelligence phase.

Even for probabilistic projects or unsupervised machine learning studies, in fact, splitting the dataset into two different subsets is a good idea for validation purposes. In a cluster analysis applied to medical images, for example, one can split the dataset into 60% randomly selected images for the first set and using the remaining 40% for the second set, applying the computational method to the former one, then to the latter one, and comparing their results. If the researcher observes the same scientific results on both the subsets, they can consider the results more consistent.

Of course, this split should not be performed only once: The researcher should repeat the same analysis at least 100 times (if they have the sufficient computational resources, as explained in Tip #1), by splitting the dataset in different ways each time. If, at the end of 100 executions, a scientific result always shows up on both the subsets, it can be considered truthful.

Validation cohort search. A key component of solid scientific studies is the repetition of the application of same method on a second dataset, with the same features and a different origin, usually called validation cohort. Often, however, it is difficult to find an alternative dataset where to repeat a computational analysis.

Even if it is difficult to find the right resource, we suggest the readership to look for an alternative dataset on public web repositories of image data such as the Cancer Imaging Archive (TCIA) [27], Open Access Series of Imaging Studies (OASIS) [28,29], and OpenfMRI [91,92], or on general data repositories such as Re3data [93], Google Datasets Search [94], Kaggle [95], and the University of California Irvine Machine Learning Repository [96].

We know that a bit of luck is fundamental to find a dataset having the same disease and of the same features as the one used in the primary analysis, but it never hurts to try. Of course,

obtaining the same results on a different dataset by using the same method employed on the primary dataset would make the study much more robust and valid.

Tip 8: Use multiple evaluation metrics to assess your results, not only one

A common mistake we see in multiple health informatics studies is reporting the results measured with only one rate that usually is the most common rate of the scientific field. This approach has a hidden problem: Since no statistical rate is perfect, the results measured only with one single rate are going to hide some drawbacks and bad news. For this reason, we recommend anyone performing any analysis of medical images to employ multiple statistical rates, and not only one.

A study whose results are based only on receiver operating characteristic (ROC) area under the curve (AUC), for example, would hide the performance of the binary classification measured as precision ($PPV = TP/(TP + FP)$) and negative predictive value ($NPV = TN/(TN + FN)$), where TP are true positives, FP are false positives, TN are true negatives, and FN are false negatives.

For binary classifications, we suggest to always include at least the MCC [97], sensitivity, specificity, precision, and negative predictive value, and to base the rankings of the methods on the MCC [98–102]. Never base all your results on ROC AUC [103–107]. ROC AUC, accuracy, F1 score, and most common statistical rates, in fact, might illude clinicians and researchers when analyzing classification dataset with skewed class proportions. The MCC, on the other hand, is the only statistic that generates an informative, truthful outcome for binary classifications [98–102]. Regarding regression analyses, we recommend to always include at least coefficient of determination (R-squared or R^2), symmetric mean absolute percentage error (SMAPE), mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and build the rankings of the methods on R-squared [108].

For clustering analysis, always include Davies–Bouldin index [109], Dunn index [110], and silhouette coefficient [111]. Depending on your scientific goal, considering results measured through mutual information, adjusted mutual information, Rand index, adjusted Rand index, homogeneity, completeness, and V-measure could be a good idea. When doing statistical and probabilistic analyses, always include at least adjusted p -value, nominal p -value, q -value, and z -score, if available, and rank the methods based on the former one [112].

Of course, particular statistics can be utilized based on the scientific goals of a study (for three-dimensional medical image segmentation [113], for example), but we would like to reaffirm our take-home message here: Never use a singular rate, always employ multiple statistics.

Tip 9: Ask a radiologist or a medical doctor to review your medical image analysis results

After months of work, multiple executions of your software, your analysis is done, and you are collecting the results, making them ready for the manuscript. You feel like you did your homework, you reached the mountaintop, and now you just need to present your methods and results in an article draft. Before writing your paper, however, there is a final step of your analysis that we suggest you to do: Show your results to a radiologist or a medical doctor, and ask them their honest feedback about them.

Are your results something they would expect to see for patients of that particular disease? Do the results make sense? How do your medical results relate to the existing knowledge in the biomedical literature? These are all pertinent questions you might ask your medical collaborator. Their external, agnostic feedback would be pivotal for your study: If they approved your

results, you would rest assured about the quality of your analysis; otherwise, if they noticed something wrong, it would be much better to hear the bad news from them informally and confidentially before submitting the article than having your article rejected by a scientific journal later on.

Additionally, when analyzing patients' data with computational intelligence, researchers should investigate the interpretability of their statistical models, by opening the black box to explain what the model does. Frameworks for explainable artificial intelligence (XAI) can be employed [114]. Van der Velden and colleagues [115] write about interpretable AI: "Something can be considered a good explanation if it gives insight into how a neural network came to its decision and/or can make the decision understandable."

Interpretation of the behaviors of machine learning models applied to medical images can then be assessed through an application-grounded evaluation, where a domain expert can inspect the behavior of the machine learning model, compare the actual results with their expected results, and eventually say if the interpretation is correct [115]. XAI is a broad theme and its discussion goes beyond the scope of this study; however, we advise the readership considering the importance of interpretability of computational intelligence models applied to medical images [18,116–120].

Tip 10: If possible, release your dataset publicly online

Earlier, we recommended to look for validation cohort datasets on public online repositories (Tip #7). Here, we follow up on that piece of advice to giving back to the online communities rather than only receiving: After finishing your medical image analysis, if you are authorized, we recommend you to publicly share online your dataset on repositories such as Kaggle [95], the University of California Irvine Machine Learning Repository [96], FigShare [121], or Zenodo [122], following the findability, accessibility, interoperability, and reusability (FAIR) principles [123].

Of course, to do so, you first need to verify you have all the authorizations from the ethical committee of your institutions or hospital and the authorizations from the patients or the patients' families to use their data. If you have them, releasing your dataset online can actually bring multiple advantages for your study.

First of all, it would allow the reproducibility of your analysis: Anyone around the world would be able to repeat your analysis and reach the same conclusions. Moreover, they would be able to find any potential mistakes, which is a fundamental aspect of scientific research involving patients. Additionally, users around the world would be able to reuse your data and generate different, additional results, building on your original outcomes. This aspect would also increase the impact of your study, in terms of article citations and in other ways.

If you also have the chance to decide on which journal to submit your article, we advocate to select an open access one: If published, your article will be readable by way more people in the world than if it was published in a restricted-access journal. An updated list of open access journals of health informatics can be found on Scimago Journal Ranking [124].

Conclusions

Medical imaging is an efficient and useful tool that can show particular aspects of patients' organs and tissues that would otherwise would be unnoticeable with traditional medical techniques. Through medical images, physicians can infer new information about the conditions of a patient and, therefore, design a better therapy.

Computational analysis of medical images can highlight particular trends among those images, indicating some correlations between patients that otherwise would be unobservable.

Looking at images does not only allow the observer to see specific conditions; more importantly, it facilitates the understanding of a phenomenon.

As the computer vision expert Tomaso Poggio once wrote [125], it is not by chance that in the English language people use the term “I see” as a synonym of “I understand.” While computational tools for medical image analysis have become more common, making mistakes have become easier as well. With these quick tips, we report a list of best practices for computational analyses of medical images that can help users avoid common mistakes and wrongdoings. We believe our ten recommendations can help researchers generate better and more reliable results, ultimately leading to better treatments and cures for patients.

References

1. Al-Galal SAY, Alshaiikhli IFT, Abdulrazzaq M. MRI brain tumor medical images analysis using deep learning techniques: a systematic review. *Health and Technology*. 2021; 11(2):267–282.
2. Maksoud EAA, Barakat S, Elmogy M. Medical images analysis based on multilabel classification. *Machine Learning in Bio-Signal Analysis and Diagnostic Imaging*. Elsevier; 2019. p. 209–245.
3. Farouk R, Aziz MA, Habib S. Medical images analysis based on fractal dimension and wavelet transform. *Journal of Computer Science Approaches*. 2016; 2(1).
4. Domingos P. A few useful things to know about machine learning. *Communications of the ACM*. 2012; 55(10):78–87.
5. Chicco D. Ten quick tips for machine learning in computational biology. *BioData Mining*. 2017; 10(1):1–17.
6. Jones DT. Setting the standards for machine learning in biology. *Nature Reviews Molecular Cell Biology*. 2019; 20(11):659–660. <https://doi.org/10.1038/s41580-019-0176-5> PMID: 31548714
7. Walsh I, Fishman D, Garcia-Gasulla D, Titma T, Pollastri G, Harrow J, et al. DOME: Recommendations for supervised machine learning validation in biology. *Nature Methods*. 2021; 18(10):1122–1127. <https://doi.org/10.1038/s41592-021-01205-4> PMID: 34316068
8. Whalen S, Schreiber J, Noble WS, Pollard KS. Navigating the pitfalls of applying machine learning in genomics. *Nature Reviews Genetics*. 2021; 23:169–181. <https://doi.org/10.1038/s41576-021-00434-9> PMID: 34837041
9. Lee BD, Gitter A, Greene CS, Raschka S, Maguire F, Titus AJ, et al. Ten quick tips for deep learning in biology. *PLoS Computational Biology*. 2022; 18(3):e1009803. <https://doi.org/10.1371/journal.pcbi.1009803> PMID: 35324884
10. Cho SM, Austin PC, Ross HJ, Abdel-Qadir H, Chicco D, Tomlinson G, et al. Machine learning compared with conventional statistical models for predicting myocardial infarction readmission and mortality: a systematic review. *Canadian Journal of Cardiology*. 2021; 37(8):1207–1214. <https://doi.org/10.1016/j.cjca.2021.02.020> PMID: 33677098
11. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *International Journal of Medical Informatics*. 2021; 153:104510. <https://doi.org/10.1016/j.ijmedinf.2021.104510> PMID: 34108105
12. Chicco D, Jurman G. The ABC recommendations for validation of supervised machine learning results in biomedical sciences. *Frontiers in Big Data*. 2022; 5(979465):1–5. <https://doi.org/10.3389/fdata.2022.979465> PMID: 36238654
13. Makin TR, de Xivry JJO. Science forum: ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *eLife*. 2019; 8:e48175.
14. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers EJ, Berk R, et al. Redefine statistical significance. *Nature Human Behaviour*. 2018; 2(1):6–10. <https://doi.org/10.1038/s41562-017-0189-z> PMID: 30980045
15. Mubeen S, Tom Kodamullil A, Hofmann-Apitius M, Domingo-Fernández D. On the influence of several factors on pathway enrichment analysis. *Briefings in Bioinformatics*. 2022; 23(3):bbac143. <https://doi.org/10.1093/bib/bbac143> PMID: 35453140
16. Wieder C, Frainay C, Poupin N, et al. Pathway analysis in metabolomics: recommendations for the use of over-representation analysis. *PLoS Computational Biology*. 2021; 17(9):e1009105. <https://doi.org/10.1371/journal.pcbi.1009105> PMID: 34492007
17. Chicco D, Agapito G. Nine quick tips for pathway enrichment analysis. *PLoS Computational Biology*. 2022; 18(8):1010348. <https://doi.org/10.1371/journal.pcbi.1010348> PMID: 35951505

18. Jin W, Li X, Fatehi M, Hamarneh G. Guidelines and evaluation for clinical explainable AI on medical image analysis. arXiv:220210553 [Preprint]. 2022.
19. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. *npj Digital Medicine*. 2022; 5(1):1–8.
20. Block KT. Subtle pitfalls in the search for faster medical imaging. *Proceedings of the National Academy of Sciences*. 2022; 119(17):e2203040119. <https://doi.org/10.1073/pnas.2203040119> PMID: 35452309
21. Guillermo M, Pengo T, Sanders MA. Imaging methods are vastly underreported in biomedical research. *eLife*. 2020; 9:e55133. <https://doi.org/10.7554/eLife.55133> PMID: 32780019
22. Van Vliet M. Seven quick tips for analysis scripts in neuroimaging. *PLoS Computational Biology*. 2020; 16(3):e1007358. <https://doi.org/10.1371/journal.pcbi.1007358> PMID: 32214316
23. Chicco D, Jurman G. Arterial disease computational prediction and health record feature ranking among patients diagnosed with inflammatory bowel disease. *IEEE Access*. 2021; 9:78648–78657.
24. Le Gall G, Kirchgessner J, Bejaoui M, Landman C, Nion-Larmurier I, Bourrier A, et al. Clinical activity is an independent risk factor of ischemic heart and cerebrovascular arterial disease in patients with inflammatory bowel disease. *PLoS ONE*. 2018; 13(8):e0201991. <https://doi.org/10.1371/journal.pone.0201991> PMID: 30169521
25. Chicco D, Oneto L. An enhanced Random Forests approach to predict heart failure from small imbalanced gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2020; 18(6):2759–2765.
26. Maciejak A, Kiliszek M, Michalak M, Tulacz D, Opolski G, Matlak K, et al. Gene expression profiling reveals potential prognostic biomarkers associated with the progression of heart failure. *Genome Medicine*. 2015; 7(1):1–15.
27. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of Digital Imaging*. 2013; 26(6):1045–1057. <https://doi.org/10.1007/s10278-013-9622-7> PMID: 23884657
28. OASIS. Open Access Series of Imaging Studies; 2022. Available from: <http://www.oasis-brains.org/> [cited 2022 Aug 2].
29. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of Cognitive Neuroscience*. 2007; 19(9):1498–1507. <https://doi.org/10.1162/jocn.2007.19.9.1498> PMID: 17714011
30. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology*. 2019; 15(6):e8746. <https://doi.org/10.15252/msb.20188746> PMID: 31217225
31. Kikinis R, Warfield S, Westin CF. High performance computing (HPC) in medical image analysis (MIA) at the surgical planning laboratory (SPL). *Proceedings of Supercomputing ASIA 2023 –the 3rd High Performance Computing Asia Conference & Exhibition*. Citeseer; 1998. p. 1–15.
32. Gulo CA, Sementille AC, Tavares JMR. Techniques of medical image processing and analysis accelerated by high-performance computing: a systematic literature review. *Journal of Real-Time Image Processing*. 2019; 16(6):1891–1908.
33. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. *Pattern Recognition*. 2018; 77:354–377.
34. MacFarland TW, Yates JM. Mann-Whitney U test. *Introduction to nonparametric statistics for the biological sciences using R*. Springer; 2016. p. 103–132.
35. Alnasir JJ. Fifteen quick tips for success with HPC, ie, responsibly BASHing that Linux cluster. *PLoS Computational Biology*. 2021; 17(8):e1009207. <https://doi.org/10.1371/journal.pcbi.1009207> PMID: 34351904
36. Bizzego A, Bussola N, Chierici M, Maggio V, Francescato M, Cima L, et al. Evaluating reproducibility of AI algorithms in digital pathology with DAPPER. *PLoS Computational Biology*. 2019; 15(3):e1006269. <https://doi.org/10.1371/journal.pcbi.1006269> PMID: 30917113
37. Balki I, Amirabadi A, Levman J, Martel AL, Emersic Z, Meden B, et al. Sample-size determination methodologies for machine learning in medical imaging research: a systematic review. *Canadian Association of Radiologists Journal*. 2019; 70(4):344–353. <https://doi.org/10.1016/j.carj.2019.06.002> PMID: 31522841
38. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. *JAMA*. 2020; 323(4):305–306. <https://doi.org/10.1001/jama.2019.20866> PMID: 31904799
39. Schoonjans F, Zalata A, Depuydt C, Comhaire F. MedCalc: a new computer program for medical statistics. *Computer Methods and Programs in Biomedicine*. 1995; 48(3):257–262. [https://doi.org/10.1016/0169-2607\(95\)01703-8](https://doi.org/10.1016/0169-2607(95)01703-8) PMID: 8925653

40. PASS. Sample Size & Power; 2022. Available from: <https://www.ncss.com/software/pass/> [cited 2022 Aug 24].
41. Baldassarro M. sampler R package; 2021. Available from: <https://cran.r-project.org/web/packages/sampler/> [cited 2022 Aug 24].
42. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A, et al. pwr R package; 2018. Available from: <https://cran.r-project.org/web/packages/pwr/> [cited 2022 Aug 24].
43. Field A. Discovering statistics using IBM SPSS statistics. SAGE; 2013.
44. He H, Garcia EA. Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*. 2009; 21(9):1263–1284.
45. Anand A, Pugalenthi G, Fogel GB, Suganthan P. An approach for classification of highly imbalanced data using weighting and undersampling. *Amino Acids*. 2010; 39(5):1385–1391. <https://doi.org/10.1007/s00726-010-0595-2> PMID: 20411285
46. Gosain A, Sardana S. Handling class imbalance problem using oversampling techniques: a review. In: *Proceedings of ICACCI 2017—the 2017 International Conference on Advances in Computing, Communications and Informatics*. IEEE; 2017. p. 79–85.
47. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. *AMIA Annual Symposium Proceedings*. vol. 2017. American Medical Informatics Association; 2017. p. 979. PMID: 29854165
48. Goel N, Yadav A, Singh BM. Medical image processing: a review. *Proceedings of CIPECH 2016—the 2nd International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity*. 2016. p. 57–62.
49. Lee D, Choi S, Kim HJ. Performance evaluation of image denoising developed using convolutional denoising autoencoders in chest radiography. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*. 2018; 884:97–104.
50. Mredhula L, Dorairangasamy M. An extensive review of significant researches on medical image denoising techniques. *International Journal of Computer Applications*. 2013; 64(14).
51. Sun Y, Liu X, Cong P, Li L, Zhao Z. Digital radiography image denoising using a generative adversarial network. *Journal of X-ray Science and Technology*. 2018; 26(4):523–534. <https://doi.org/10.3233/XST-17356> PMID: 29889095
52. Mohammadi S, Leventouri T. A study of wavelet-based denoising and a new shrinkage function for low-dose CT scans. *Biomedical Physics & Engineering Express*. 2019; 5(3):035018.
53. Diwakar M, Kumar M. A review on CT image noise and its denoising. *Biomedical Signal Processing and Control*. 2018; 42:73–88.
54. Gajera B, Kapil SR, Ziaei D, Mangalagiri J, Siegel E, Chapman D. CT-scan denoising using a charbonnier loss generative adversarial network. *IEEE Access*. 2021; 9:84093–84109.
55. Heunis S, Lamerichs R, Zinger S, Caballero-Gaudes C, Jansen JF, Aldenkamp B, et al. Quality and denoising in real-time functional magnetic resonance imaging neurofeedback: a methods review. *Human Brain Mapping*. 2020; 41(12):3439–3467. <https://doi.org/10.1002/hbm.25010> PMID: 32333624
56. Bhujle HV, Vadavadagi BH. NLM based magnetic resonance image denoising—A review. *Biomedical Signal Processing and Control*. 2019; 47:252–261.
57. Mohan J, Krishnaveni V, Guo Y. A survey on the magnetic resonance image denoising methods. *Biomedical Signal Processing and Control*. 2014; 9:56–69.
58. Ragesh N, Anil A, Rajesh R. Digital image denoising in medical ultrasound images: a survey. *Proceedings of AIML-11—the ICGST International Conference on Artificial Intelligence and Machine Learning*. vol. 12; 2011. p. 14.
59. Sagheer SVM, George SN. A review on medical image denoising algorithms. *Biomedical Signal Processing and Control*. 2020; 61:102036.
60. Gong K, Guan J, Liu CC, Qi J. PET image denoising using a deep neural network through fine tuning. *IEEE Transactions on Radiation and Plasma Medical Sciences*. 2018; 3(2):153–161. <https://doi.org/10.1109/TRPMS.2018.2877644> PMID: 32754674
61. Li XT, Huang RY. Standardization of imaging methods for machine learning in neuro-oncology. *Neuro-Oncology*. *Advances*. 2020; 2(Supplement 4):iv49–iv55. <https://doi.org/10.1093/neoajnl/vdaa054> PMID: 33521640
62. Papadimitroulas P, Brocki L, Chung NC, Marchadour W, Vermet F, Gaubert L, et al. Artificial intelligence: deep learning in oncological radiomics and challenges of interpretability and data harmonization. *Physica Medica*. 2021; 83:108–121. <https://doi.org/10.1016/j.ejmp.2021.03.009> PMID: 33765601

63. Zhu AH, Moyer DC, Nir TM, Thompson PM, Jahanshad N. Challenges and opportunities in dMRI data harmonization. In: Proceedings of MICCAI 2019 –the 22nd International Conference on Medical Image Computing and Computer-Assisted Intervention, Computational Diffusion MRI Workshop. Springer; 2019. p. 157–172.
64. Sadri AR, Janowczyk A, Zhou R, Verma R, Beig N, Antunes J, et al. MRQy—An open-source tool for quality control of MR imaging data. *Medical Physics*. 2020; 47(12):6029–6038.
65. Vogelbacher C, Bopp MH, Schuster V, Herholz P, Jansen A, Sommer J. LAB-QA2GO: a free, easy-to-use toolbox for the quality assessment of magnetic resonance imaging data. *Frontiers in Neuroscience*. 2019; 13:688. <https://doi.org/10.3389/fnins.2019.00688> PMID: 31333406
66. Nyúl LG, Udupa JK. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine*. 1999; 42(6):1072–1081. [https://doi.org/10.1002/\(sici\)1522-2594\(199912\)42:6<1072::aid-mrm11>3.0.co;2-m](https://doi.org/10.1002/(sici)1522-2594(199912)42:6<1072::aid-mrm11>3.0.co;2-m) PMID: 10571928
67. Bashyam VM, Doshi J, Erus G, Srinivasan D, Abdulkadir A, Singh A, et al. Deep generative medical image harmonization for improving cross-site generalization in deep learning predictors. *Journal of Magnetic Resonance Imaging*. 2021; 55(3):908–916. <https://doi.org/10.1002/jmri.27908> PMID: 34564904
68. Shiradkar R, Ghose S, Mahran A, Li L, Hubbard I, Fu P, et al. Prostate surface distension and tumor texture descriptors from pre-treatment MRI are associated with biochemical recurrence following radical prostatectomy: preliminary findings. *Frontiers in Oncology*. 2022. p. 2055. <https://doi.org/10.3389/fonc.2022.841801> PMID: 35669420
69. Cadwallader L, Mac Gabhann F, Papin J, Pitzer VE. Advancing code sharing in the computational biology community. *PLoS Computational Biology*. 2022; 18(6):e1010193. <https://doi.org/10.1371/journal.pcbi.1010193> PMID: 35653366
70. TIOBE. TIOBE Index for July 2022; 2022. <https://www.tiobe.com/tiobe-index/> URL visited on 2nd August 2022.
71. Pang B, Nijkamp E, Wu YN. Deep learning with TensorFlow: a review. *Journal of Educational and Behavioral Statistics*. 2020; 45(2):227–248.
72. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing systems*. 2019; 32.
73. Ioannidis JP. Why most published research findings are false. *PLOS Medicine*. 2005; 2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124> PMID: 16060722
74. Li L, Pahwa S, Penzias G, Rusu M, Gollamudi J, Viswanath S, et al. Co-registration of ex vivo surgical histopathology and in vivo T2 weighted MRI of the prostate via multi-scale spectral embedding representation. *Scientific Reports*. 2017; 7(1):1–12.
75. Wu H, Li X, Cheng KT. Exploring feature representation learning for semi-supervised medical image segmentation. *ArXiv*. 2021;arXiv:2111. 10989:1–10.
76. Van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*. 2017; 77(21):e104–e107. <https://doi.org/10.1158/0008-5472.CAN-17-0339> PMID: 29092951
77. Rong Q, Thangaraj C, Easwaramoorthy D, He S. Multifractal based image processing for estimating the complexity of COVID-19 dynamics. *The European Physical Journal Special Topics*. 2021; 230(21):3947–3954. <https://doi.org/10.1140/epjs/s11734-021-00336-1> PMID: 34815830
78. Alilou M, Prasanna P, Bera K, Gupta A, Rajiah P, Yang M, et al. A novel nodule edge sharpness radiomic biomarker improves performance of lung-RADS for distinguishing adenocarcinomas from granulomas on non-contrast CT scans. *Cancers*. 2021; 13(11):2781. <https://doi.org/10.3390/cancers13112781> PMID: 34205005
79. Lai Z, Deng H. Medical image classification based on deep features extracted by deep model and statistical feature fusion with multilayer perceptron. *Computational Intelligence and Neuroscience*. 2018; 2018:1–13.
80. Olabbarriaga SD, Smeulders AW. Interaction in the segmentation of medical images: a survey. *Medical Image Analysis*. 2001; 5(2):127–142. [https://doi.org/10.1016/s1361-8415\(00\)00041-4](https://doi.org/10.1016/s1361-8415(00)00041-4) PMID: 11516707
81. Oliveira FP, Tavares JMR. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*. 2014; 17(2):73–93. <https://doi.org/10.1080/10255842.2012.670855> PMID: 22435355
82. Mwangi B, Tian TS, Soares JC. A review of feature reduction techniques in neuroimaging. *Neuroinformatics*. 2014; 12(2):229–244. <https://doi.org/10.1007/s12021-013-9204-3> PMID: 24013948
83. Debie E, Shafi K. Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*. 2019; 22(2):519–536.

84. Radovic M, Ghalwash M, Filipovic N, Obradovic Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics*. 2017; 18(1):1–14.
85. Ginsburg SB, Lee G, Ali S, Madabhushi A. Feature importance in nonlinear embeddings (FINE): applications in digital pathology. *IEEE Transactions on Medical Imaging*. 2015; 35(1):76–88. <https://doi.org/10.1109/TMI.2015.2456188> PMID: 26186772
86. Nguyen LH, Holmes S. Ten quick tips for effective dimensionality reduction. *PLoS Computational Biology*. 2019; 15(6):e1006907. <https://doi.org/10.1371/journal.pcbi.1006907> PMID: 31220072
87. Chicco D, Masseroli M. Software suite for gene and protein annotation prediction and similarity search. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 2014; 12(4):837–843.
88. Reddy GT, Reddy MPK, Lakshmana K, Kaluri R, Rajput DS, Srivastava G, et al. Analysis of dimensionality reduction techniques on big data. *IEEE Access*. 2020; 8:54776–54788.
89. McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*. 2018;arXiv:1802.03426:1–63.
90. Belkina AC, Ciccolella CO, Anno R, Halpert R, Spidlen J, Snyder-Cappione JE. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nature Communications*. 2019; 10(1):1–12.
91. Poldrack RA, Barch DM, Mitchell JP, Wager TD, Wagner AD, Devlin JT, et al. Toward open sharing of task-based fMRI data: the OpenfMRI project. *Frontiers in Neuroinformatics*. 2013; 7:12. <https://doi.org/10.3389/fninf.2013.00012> PMID: 23847528
92. Poldrack RA, Gorgolewski KJ. OpenfMRI: Open sharing of task fMRI data. *Neuroimage*. 2017; 144:259–261. <https://doi.org/10.1016/j.neuroimage.2015.05.073> PMID: 26048618
93. Re3data. Registry of research data repositories; 2022. Available from: <https://www.re3data.org/> [cited 2022 Jun 24].
94. Google. Google Dataset Search; 2022. Available from: <https://datasetsearch.research.google.com/> [cited 2022 Jul 29].
95. Kaggle. Kaggle datasets—Explore, analyze, and share quality data; 2022. Available from: <https://www.kaggle.com/datasets> [cited 2022 Jun 24].
96. University of California Irvine. Machine Learning Repository; 1987. Available from: <https://archive.ics.uci.edu/ml> [cited 2022 Jun 24].
97. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein. Structure*. 1975; 405(2):442–451.
98. Jurman G, Riccadonna S, Furlanello C. A comparison of MCC and CEN error measures in multi-class prediction. *PLOS ONE*. 2012; 7(8):e41882. <https://doi.org/10.1371/journal.pone.0041882> PMID: 22905111
99. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020; 21(1):6. <https://doi.org/10.1186/s12864-019-6413-7> PMID: 31898477
100. Chicco D, Tötsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Mining*. 2021; 14(1):1–22.
101. Chicco D, Starovoitov V, Jurman G. The benefits of the Matthews correlation coefficient (MCC) over the diagnostic odds ratio (DOR) in binary classification assessment. *IEEE Access*. 2021; 9:47112–47124.
102. Chicco D, Warrens MJ, Jurman G. The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. *IEEE Access*. 2021; 9:78368–78381.
103. Wald NJ, Bestwick JP. Is the area under an ROC curve a valid measure of the performance of a screening or diagnostic test? *Journal of Medical Screening*. 2014; 21(1):51–56. <https://doi.org/10.1177/0969141313517497> PMID: 24407586
104. Muschelli J. ROC and AUC with a binary predictor: a potentially misleading metric. *Journal of Classification*. 2020; 37(3):696–708. <https://doi.org/10.1007/s00357-019-09345-1> PMID: 33250548
105. Movahedi F, Padman R, Antaki JF. Limitations of receiver operating characteristic curve on imbalanced data: assist device mortality risk scores. *Journal of Thoracic and Cardiovascular Surgery*. 2021. <https://doi.org/10.1016/j.jtcvs.2021.07.041> PMID: 34446286
106. Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. *European Radiology*. 2015; 25(4):932–939. <https://doi.org/10.1007/s00330-014-3487-0> PMID: 25599932

107. Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*. 2008; 17(2):145–151.
108. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*. 2021; 7:e623. <https://doi.org/10.7717/peerj-cs.623> PMID: 34307865
109. Davies DL, Bouldin DW. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979; PAMI-1(2):224–227. PMID: 21868852
110. Dunn JC. Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*. 1974; 4(1):95–104.
111. Kaufman L, Rousseeuw PJ. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons; 2009.
112. Jafari M, Ansari-Pour N. Why, when and how to adjust your P values? *Cell Journal*. 2019; 20(4):604. <https://doi.org/10.22074/cellj.2019.5992> PMID: 30124010
113. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*. 2015; 15(1):1–28.
114. Doran D, Schulz S, Besold TR. What does explainable AI really mean? A new conceptualization of perspectives. *arXiv*. 2017;arXiv:1710.00794:1–8.
115. van der Velden BH, Kuijff HJ, Gilhuijs KG, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Medical Image Analysis*. 2022:102470. <https://doi.org/10.1016/j.media.2022.102470> PMID: 35576821
116. Bourdon P, Ahmed OB, Urruty T, Djemal K, Fernandez-Maloigne C. Explainable AI for medical imaging: knowledge matters. *Multi-Faceted Deep Learning*. Springer; 2021. p. 267–292.
117. Folke T, Yang SCH, Anderson S, Shafto P. Explainable AI for medical imaging explaining pneumothorax diagnoses with Bayesian teaching. *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III*. vol. 11746. SPIE; 2021. p. 644–664.
118. Jin W, Li X, Hamarneh G. Evaluating explainable AI on a multi-modal medical imaging task: can existing algorithms fulfill clinical requirements? *Association for the Advancement of Artificial Intelligence Conference (AAAI)*; 2022. p. 1–9.
119. Cabitza F, Campagner A, Malgieri G, Natali C, Schneeberger D, Stoeger K, et al. Quod erat demonstrandum?—Towards a typology of the concept of explanation for the design of explainable AI. *Expert Systems with Applications*. 2023; 213:118888.
120. Cabitza F, Campagner A, Sconfienza LM. As if sand were stone. New concepts and metrics to probe the ground on which to build trustable AI. *BMC Medical Informatics and Decision Making*. 2020; 20(1):1–21.
121. FigShare. Store, share, discover research; 2011. Available from: <https://www.figshare.com> [cited 2022 Jul 25].
122. Zenodo. Zenodo: research, shared; 2013. Available from: <https://www.zenodo.org> [cited 2022 Jul 25].
123. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. 2016; 3(1):1–9. <https://doi.org/10.1038/sdata.2016.18> PMID: 26978244
124. Scimago Journal Ranking. Health informatics open access journals; 2022. Available from: <https://www.scimagojr.com/journalrank.php?openaccess=true&type=j&category=2718> [cited 2022 Jun 26].
125. Poggio T, Fontana M. L'occhio e il cervello (in Italian). *Theoria*; 1991. p. 1–124.