Automated Machine Learning in Medical Imaging

Imagiology

University of Minho

Abstract. Automated machine learning (AutoML) has emerged as a promising tool for developing and optimizing machine learning models in medical imaging. In this systematic review, we analyzed three articles that used AutoML for different medical imaging tasks. The first article proposed a T-AutoML method for lesion segmentation in 3D medical images that leverages transformer modules to optimize deep learning configurations. The T-AutoML method outperformed other existing methods in large-scale lesion segmentation datasets and demonstrated effective transferability to different datasets. The second article introduced StrokeViT with AutoML for brain stroke classification, which improved slice-wise accuracy and patient-wise prediction using a combination of CNN and transformer architectures. The proposed architecture has the potential to be generalized for volumetric scans aiding in the diagnosis of various diseases. Finally, the third article evaluated the effectiveness of generic AutoML tools for computational pathology and found that they can generate image classifiers that are comparable in performance to those reported in the literature. These tools provide an efficient way to eliminate the need for manual search of neural architectures and hyperparameter optimization. Overall, AutoML has the potential to revolutionize medical imaging by improving diagnostic accuracy and reducing the burden on radiologists. However, there is still a need for further research on reducing searching and training time given certain computational budgets and expand the applications of AutoML to other medical imaging tasks. The results of this systematic review provide valuable insights into the current state of AutoML in medical imaging and highlight opportunities for future research and development.

Keywords: Automated Machine Learning · AutoML · Medical Imaging.

1 Introduction

Automated Machine Learning (AutoML) is a process of automating the end-to-end process of applying machine learning to real-world problems. AutoML aims to simplify the machine learning process and make it more accessible to non-experts by automating various aspects of it, including data preparation, feature engineering, model selection, hyperparameter tuning, and model deployment.The AutoML process typically involves the following steps: Data pre-processing; Feature engineering; Model selection; Hyperparameter tuning.

Automated Machine Learning (AutoML) has become increasingly popular in medical imaging as it can help healthcare providers to quickly and accurately analyse large amounts of medical images, leading to faster and more accurate diagnosis and treatment of diseases.

Here are some of the keyways AutoML is being used in medical imaging:

- **Image classification**: AutoML algorithms can automatically classify medical images into different categories based on their features, such as detecting cancerous tumors or classifying different types of lung disease. This can help healthcare providers to quickly identify potential health risks and develop appropriate treatment plans;
- **Object detection**: AutoML algorithms can automatically detect and segment specific objects within medical images, such as tumors or blood vessels. This can help healthcare providers to identify the location, size, and shape of these objects, which is critical for accurate diagnosis and treatment planning;
- **Image segmentation**: AutoML algorithms can automatically segment medical images into different regions of interest, such as organs or tissues. This can help healthcare providers to identify abnormalities or diseases within these regions, leading to earlier diagnosis and better treatment outcomes;
- **Image registration**: AutoML algorithms can automatically align and register multiple medical images, such as CT or MRI scans, to create a single, unified image. This can help healthcare providers to better visualize and analyze complex structures within the body, such as blood vessels or tumors;
- **Disease diagnosis and prediction**: AutoML algorithms can automatically analyze medical images to identify patterns and features that are associated with specific diseases, such as Alzheimer's or cancer. This can help healthcare providers to diagnose these diseases earlier and predict their progression more accurately.

Overall, AutoML in medical imaging can help healthcare providers to save time and improve accuracy in the diagnosis and treatment of diseases. It can also help to democratize healthcare by making medical imaging analysis more accessible to healthcare providers with less specialized training in machine learning.

In recent years, there has been significant progress in the development of AutoML algorithms for medical imaging applications, so in this article, we will explore the current state of AutoML in medical imaging, highlighting its potential benefits, challenges, and limitations. We will also discuss some of the latest advances in AutoML techniques and applications in medical imaging, as well as the future directions of this rapidly evolving field [1].

In general, the use of machine learning models could improve patient safety, improve quality of care, and reduce healthcare costs. Specifically, automated machine learning has the capability to improve lesion segmentation in medical imaging, using a new automated machine learning algorithm, the TAutoML, which was validated on several large-scale public lesion segmentation datasets and achieved state-of-the-art performance [2]; improve human-computer interaction (HCI) in stroke diagnosis using CT scans, by integrating convolutional neural networks (CNNs), vision transformers (ViT), and automated machine learning (AutoML) to obtain both slice-level and patient-wise predictions [3]; and optimize CNN architectures and parametrizations with AutoML tools, allowing developers to focus on harder-to-automate tasks such as data curation [4].

Automated machine learning has proven to be useful in healthcare, but it requires a significant amount of effort from human experts as no algorithm can perform well on all problems. For this reason, in this article it is also analysed the improvements that are still necessary in AutomML as well as the limitations and challenges of using AutoML, such as the lack of expertise necessary to apply machine learning techniques to big data sources of healthcare researchers, the fact that data and human expertise are not readily available in healthcare settings, the need for large and diverse datasets, the difficulty in interpreting ML models, and the ethical considerations surrounding the use of ML in clinical practice [2].

2 Methods

This article is a systematic review of the literature on the use of automated machine learning (AutoML) in medical imaging. The review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

We searched for papers on the topic of automated machine learning (AutoML) in 9 electronic databases (MDPI, Science Direct, ARXIV, IMR, Springer Link, PubMed, Scopus, Web of Science, and IEEE Xplore) using a set of keywords, such as "AutoML", "Automated Machine Learning", "Machine Learning", between 2020 and 2023. The search strategy used a combination of medical subject headings and text words related to AutoML, medical imaging, and healthcare. The search was limited to articles published in English language. We excluded closed-source systems and only reviewed papers with available source code or links to project repositories. Papers with limited citation counts were also excluded.

Three reviewers independently screened the titles and abstracts of the articles identified in the search. The inclusion criteria were: (1) original research articles, (2) studies that used AutoML methods for medical image analysis, (3) studies that evaluated the performance of AutoML methods using appropriate metrics, and (4) studies that were published in peer-reviewed journals. Exclusion criteria were: (1) studies that did not use AutoML methods, (2) studies that did not report performance metrics.

A narrative synthesis was conducted to summarize the findings of the included studies. A qualitative analysis was performed to identify the strengths and limitations of the AutoML methods and to provide recommendations for future research.

3 Results

As it was mentioned previously, after a full article evaluation, a lot of studies were excluded for various reasons. Finally, three studies met the inclusion criteria and were included in the systematic review. The following paragraphs detail what results were obtained in each investigation, as well as the risk of bias and other relevant aspects like the sample size of experiments.

T-AutoML: Automated Machine Learning for Lesion Segmentation using Transformers in 3D Medical Imaging is the title of one of the reviewed studies. In this article it is exposed how a novel automated machine learning algorithm, called TAutoML, simultaneously searches for the best neural architecture and finds the best combination of hyperparameters and data augmentation strategies. The risk of bias in the T-AutoML study is considered low as it utilizes two public, large and diverse datasets – LiTs 2017 (contains 201 3D abdominal CT volumes) and MSD (contains 95 3D CT scans of lung lesion/tumor). LiTs was used as the training and validation dataset, while MSD was used to verify the transferability of the model trained on LiTs to another segmentation task, in this case, the segmentation of lung lesions in abdominal CT images [2].

T-AutoML achieved superior results compared to other methods for 3D medical imaging lesion segmentation. The evaluation of the model was made by calculating the Dice score per case for the LiTS challenge, and both the Dice score and normalized surface distance (NSD) for the lung task in the MSD challenge. The *Tables 2* and *3* and the *Figure 4* of the article demonstrate, with the respective values of Dice score and visual evidence, the superior performance of this method compared to others. For numerical references, the average Dice score for T-AutoML was 0.7533 and the average NSD was 0.7768. The researchers also migrated successfully the found solution from LiTS to MSD lung task without any additional search, which is an indicator that their method is transferable between different datasets and tasks, suggesting it has potential for use in a variety of 3D medical image segmentation tasks. They also compared three different types of predictors: accuracy-based predictor, regular relation predictor based on MLP, and relation and transformer based relation predictor. The results showed that the accuracy predictor was not able to produce a positive correlation between ground truth ranks and predicted ranks. However, both relation-based predictors were able to rank them in a positively correlated fashion. In particular, the transformer-based predictor showed more than 20% improvement over the MLP-based predictor due to the advanced capacities of transformer modules [2].

The paper named *StrokeViT with AutoML for brain stroke classification* proposes a novel approach for stroke classification. The proposed architecture improves upon this by providing slice-level predictions as well as patient-wise predictions, making the outcome more clinically relevant and closer to real-world scenarios. The proposed architecture achieved an accuracy of 87% for single slice-level prediction and an accuracy of 92% for patient-wise prediction, outperforming the state-of-the-art models by 12.9%. The risk of bias in the study is low to moderate. While the authors state that they used a large dataset, it is unclear how representative it is of the general population and whether any selection bias exists. Additionally, the study only reports results for a single institution, so it is unknown if the results would generalize to other settings. However, the study's use of AutoML to optimize the performance of the StrokeViT model is a strength, as it reduces the risk of bias introduced by human judgment. Overall, the study provides promising results, but further research is needed to determine the generalizability of the StrokeViT model to different populations and settings [3].

Evaluating generic AutoML tools for computational pathology is the title of an article that evaluates the performance of generic tools for neural network architecture search and hyperparameter optimization in image analysis tasks. More precisely, the article compares two generic AutoML tools, AutoGluon and AutoML Vision. As an evaluation of two AutoML tools, the study did not involve human subjects or patient data, so there was no risk of bias related to these factors. However, the evaluation relied on a limited set of classification tasks for histological images and did not consider other image analysis tasks or domains. Additionally, the study was limited to two specific AutoML tools and did not include other available tools or approaches for neural network architecture search and hyperparameter optimization. Therefore, there is a risk of bias in the generalizability of the findings to other use cases or AutoML tools [4].

Overall, the study suggests that generic CNN (Convolutional Neural Networks) architectures and AutoML tools can be a viable alternative to manual optimization, allowing developers to focus on more challenging tasks. Both AutoGluon and AutoML Vision were able to generate CNNs within the specified time budgets, with AutoML Vision finishing within a similar time as AutoGluon [4].

By comparing the performance of classifiers obtained from different presets, the authors can determine the effectiveness of the different configurations and select the best one for a particular use case. *Figure 1* of the article shows the classifier performance for different presets. For AutoGluon, preset 2 generally produced the best-performing classifiers. For tissue classification, the classifiers produced AUROCs that were even closer to 1 than the reference results from literature (Area Under the Receiver Operating Characteristic Curve - a higher AUROC indicates better model performance, with a value of 1 indicating perfect classification). For mutation prediction, the AutoGluon-generated classifiers resulted in slightly higher or lower AUROCs depending on the dataset. For grading, the values were slightly higher or lower than the literature reference for pathologists. For AutoML Vision, among the presets, none was generally superior to the others, and even two presets failed to generate a meaningful classifier in one case each. For the tissue classification use case, the classifiers generated by AutoML Vision resulted in AUROCs even closer to 1 than the AutoGluon-generated classifiers. The AUROCs for the mutation prediction use case were in the same range as the literature results and AutoGluon-generated classifiers. However, values for the grading use case were slightly lower than for the AutoGluon-generated classifiers, and in the same range as the reference results [4].

The study also conducted hyperparameter optimization in AutoGluon and found that it led to slightly better classifier performance compared to preset 2 for some tasks. However, the optimization did not substantially improve performance for other tasks. AutoML Vision did not report whether hyperparameter optimization took place during the respective run [4].

The reproducibility analysis of selected tasks in the study showed that there was notable variability in the performance metrics, including cases where the first AutoML run performed worse than the six repetitions for some tasks. *Figure 3* of the article represents precisely the variability of AutoML results. The classifiers generated by AutoML Vision were more variable than the AutoGluon-generated ones [4].

4 Discussion

The three articles discussed provide valuable insights into the application and potential of automated machine learning (AutoML) in medical image analysis and highlight the strengths and weaknesses of different AutoML tools and architectures. The first article investigates the use of AutoML with transformers for lesion segmentation in 3D medical imaging and suggests that AutoML search may be more effective for challenging tasks. The second article proposes a StrokeViT with AutoML approach for classifying brain stroke in CT scans, which achieves superior performance over state-of-the-art models. The third article evaluates the performance of two popular AutoML tools, AutoGluon and AutoML Vision, for computational pathology and highlights the importance of reproducibility in deep learning [2] [3] [4].

In more detail, the first article discusses a study on the use of AutoML for lesion segmentation in 3D medical imaging using transformers. The study shows the potential of AutoML using transformers for lesion segmentation in 3D medical imaging and provides insights into the best predictor for ranking data points in the validation set. The authors conducted ablation studies to investigate their choice of predictor, using three different types to rank validation data points: accuracy-based, regular relation-based MLP, and relation and transformer-based predictors. The results suggested that the accuracy predictor, due to fewer ground truth points, produced poor validation performance and overfitting. On the other hand, the relation-based predictors, which leverage pairwise information from the same training pool, performed better in terms of ranking performance. In particular, the transformer-based predictor showed significant improvement due to the advanced capacities of transformer modules. Overall, the study demonstrates the importance of carefully selecting the predictor type when utilizing AutoML in medical image analysis [2].

The article also highlights that the search for organ segmentation may not be able to generate good architecture or training configurations for lesion segmentation even in the same dataset. This is because organ body segmentation is much simpler than lesion segmentation, and organ segmentation models typically converge much faster than lesion segmentation models. Therefore, challenging tasks such as lesion segmentation are more helpful for AutoML search in order to transfer the searched results to other applications [2].

The second article, *StrokeViT with AutoML for brain stroke classification*, presents an innovative approach to classify brain stroke in CT scans using a combination of ResNet50 and Vision Transformer (ViT) with AutoML. The use of AutoML for patient-wise prediction and stroke-specific selection features such as the longest sequence of hemorrhage and infarct slices, the total number of hemorrhage and infarct slices, and the initial slice of hemorrhage and infarct prediction helps mitigate class imbalance and overfitting issues in small datasets [3].

The study shows that the slice-wise prediction results are balanced by AutoML's patient-wise prediction, which removes anomaly and out-of-distribution predictions. Additionally, the architecture provides real-time prediction outcomes in a sequence of CT scans, which helps in visualizing the actual scenario and ruling out anomalies in the prediction outcomes [3].

The proposed architecture's combination of ResNet50 and Vision Transformer shows that Vision Transformer can perform better than CNN models when self-attention is applied to the features extracted from the CNN architecture. However, the self-attention mechanism lacks the inductive bias evident in the CNN models, which limits its performance in low-data scenarios, such as medical image datasets. The authors addressed this limitation by training the Vision Transformer on the features extracted by ResNet50. Overall the StrokeViT with AutoML approach is a new way of classifying brain stroke in CT scans, which has shown promising results, outperforming existing methods by a significant margin, indicating that it could be a valuable tool in the diagnosis and treatment of strokes [3].

Finally the third article discusses the evaluation of generic AutoML tools for computational pathology. The authors compared the performance of two popular AutoML tools, AutoGluon and AutoML Vision. AutoGluon is an open-source tool that provides more flexibility in setting network architectures and parametrizing hyperparameter optimization, but comes with higher usage complexity. On the other hand, AutoML Vision is a black-box tool that offers only limited choices to the user, but it is easier to use. The authors also highlight that AutoGluon can be used on premises, which allows for confidentiality and privacy of training data. In contrast, cloud-based tools like AutoML Vision may not comply with local data protection regulations [4].

The results suggest that both AutoGluon and AutoML Vision produce classifiers that are generally on par with literature results. However, the authors observed two cases where AutoML Vision produced classifiers that predicted a constant single class for the entire test set, which suggests an obviously failed training [4].

The article also discusses the issue of reproducibility in deep learning, highlighting the non-deterministic nature of training and evaluation of CNN-based classifiers. The authors suggest that non-determinism poses a problem for hyperparameter optimization, making the objective function noisy, regardless of whether algorithmic or manual optimization is used. The authors also note that variability at the tile level can obfuscate aggregated performance metrics, indicating that underlying diagnostic tasks are robust with respect to misclassifications of part of the data for separate cases [4].

These articles contribute to the growing body of research that explores the potential of AutoML in medical image analysis. The first article demonstrates the potential of AutoML with transformers for lesion segmentation, which has significant implications for clinical applications. The results of the StrokeViT with AutoML approach are particularly promising, as the proposed architecture achieves a high accuracy rate while mitigating the class imbalance and overfitting issues in small datasets. The findings of the third article also offer insights into the strengths and weaknesses of popular AutoML tools and highlight the need for reproducibility in deep learning research [2] [3] [4].

Despite the valuable insights provided by these articles, there are limitations to the evidence included in the review. For instance, the studies presented in these articles may not be representative of all the possible use cases of AutoML in medical image analysis, and there may be other factors that influence the performance of AutoML tools, such as the small sample sizes and lack of external validation on other datasets of the training dataset [2] [3] [4].

In summary, AutoML has a high potential to increase the accuracy and efficiency of medical image analysis. However, more research is needed to explore the generalizability of AutoML approaches in different medical imaging contexts. So, future research should focus on addressing these limitations by conducting large-scale studies with external validation on other datasets and transparently reporting the review processes used. Additionally, future research should investigate the use of AutoML in other medical image analysis tasks and compare the performance of different AutoML tools [2] [3] [4].

5 Conclusion

In recent years, automated machine learning (AutoML) has emerged as a promising tool for developing and optimizing machine learning models in medical imaging. AutoML offers the potential to significantly reduce development effort and improve performance, while also allowing researchers and clinicians to focus on harder-to-automate tasks like data curation.

In conclusion, this systematic review analysed three articles on the topic of automated machine learning (AutoML) in medical imaging. The T-AutoML method proposed in the first article presents a promising approach to optimize deep learning configurations for lesion segmentation in 3D medical images. By leveraging the advanced capacity of transformer modules, the proposed method can predict the relation between different training configurations and neural networks. The experiments showed that the T-AutoML method outperformed other existing methods in the literature, achieving the most advanced performance in large-scale lesion segmentation datasets. Additionally, the proposed methods have been shown to be effectively transferable to different datasets. Future work could investigate how to further reduce searching and training time given certain computational budgets. Overall, the T-AutoML method presents a valuable contribution to the field of automated machine learning in medical imaging [2].

The second article introduced *StrokeViT with AutoML* for brain stroke classification, which improved slice-wise accuracy and patient-wise prediction using a combination of CNN and transformer architectures. This approach demonstrated significant improvements in slice-wise accuracy and patient-wise predictions, which are clinically relevant and closer to the real-world scenario. The proposed architecture has the potential to be generalized for volumetric scans aiding in the diagnosis of various diseases. Overall, this paper presents a promising direction in using AutoML for medical image analysis and has the potential to contribute to the development of computer-aided diagnosis systems in the future [3].

Finally, the third article evaluated the effectiveness of generic AutoML tools for computational pathology and found that generic AutoML tools such as AutoGluon and AutoML Vision can generate image classifiers for use cases in computational pathology that are comparable in performance to those reported in the literature. These tools provide an efficient way to eliminate the need for manual search of neural architectures and hyperparameter optimization, thus significantly reducing development effort and allowing developers to focus on more complex tasks. However, the authors note that further investigation is needed to determine the impact of actual AutoML features, as AutoGluon relied on a generic preset for CNN architecture and training hyperparameters. Overall, the use of AutoML tools in computational pathology has the potential to improve the speed and accuracy of image analysis, ultimately leading to better patient outcomes [4].

Overall, these articles demonstrate the potential of AutoML in medical imaging to improve accuracy, efficiency, and clinical relevance. However, there is still a need for further research on reducing searching and training time given certain computational budgets and expand the applications of AutoML to other medical imaging tasks.

In summary, AutoML has the potential to revolutionize medical imaging by improving diagnostic accuracy and reducing the burden on radiologists. The results of this systematic review provide valuable insights into the current state of AutoML in medical imaging and highlight opportunities for future research and development.

References

1. Waring, J., Lindvall, C., Umeton, R.: Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artificial Intelligence in Medicine **104**, 101822 (2020). https://doi.org/10.1016/j.artmed.2020.101822
2. Yang, D., Myronenko, A., Wang, X., Xu, Z., Roth, H., Xu, D.: T-AutoML: Automated Machine Learning for Lesion Segmentation Using Transformers in 3D Medical Imaging. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 3962-3974) (2021).
3. Raj, R., Mathew, J., Kannath, S.K., Rajan, J.: StrokeViT with AutoML for brain stroke classification. Engineering Applications of Artificial Intelligence **119**, 105772 (2023). https://doi.org/10.1016/j.engappai.2022.105772
4. Schwen, L.O., Schacherer, D., Geißler, C., Homeyer, A.: Evaluating generic AutoML tools for computational pathology. Informatics in Medicine Unlocked **29**, 100853 (2022). https://doi.org/10.1016/j.imu.2022.100853