Proceedings of CONF-SEML 2026 Symposium: Multimodal Data Acquisition: Applications in Physiological and Behavioral Research
DOI: 10.54254/2755-2721/2026.AD33660

Research on Medical Image Classification Using AutoML: A Case Study of Baidu EasyDL

Zihan Tang

School of Computer Science and Engineering, University of Electronic Science and Technology of China, China
coolharrison231@gmail.com

Abstract. With the rapid development of deep learning techniques within medical image analysis, convolutional neural networks (CNNs) that play a crucial role in disease detection and auxiliary diagnosis are emerging. However, the traditional deep learning methods are extremely dependent on human expertise when it comes to model construction and hyperparameter adjustment which makes the development cycle rather long the technical threshold quite high and the resource consumption also quite large. In order to enhance the modeling efficiency and reduce the dependence on expertise automated machine learning (AutoML) has indeed become a crucial direction in medical AI research. It is able to automatically handle the key processes, like the things such as model selection, neural architecture search and hyperparameter optimization. Taking the classification of brain tumor magnetic resonance imaging images as an example this article by means of the Baidu EasyDL platform constructs an AutoML model based on the public datasets from Kaggle. A systematic comparative analysis is carried out on traditional self-defined CNN models. It turns out that AutoML has struck a fairly good balance between efficiency and performance, which has greatly reduced human intervention as well as model design costs and at the same time has maintained a relatively high classification accuracy. However, this research also points out the existing limitations in the interpretability of the model and the problems in overfitting control.

Keywords: AutoML, Medical Image Classification, Brain Tumor Detection, Convolutional Neural Networks, Intelligent Diagnosis

1. Introduction

Computer vision, which is a tool widely used in fields such as medicine and biology, can bring about substantial improvements in the efficiency and effectiveness within those domains. For example, systems based on deep learning have attained performances that are similar to those of human experts on specific tasks, such as detecting diabetic retinopathy in fundus photos or identifying lung nodules in CT scans and the like. The foundation of these advancements usually resides in convolutional neural networks (CNNs) [1] which can automatically learn those hierarchical features. However, even though advanced architectures like ResNet and EfficientNet have been developed, these models face great difficulties when manually adjusting hyperparameters—for example, having

DOI: 10.54254/2755-2721/2026.AD33660

to determine the optimal learning rate, batch size, network depth, and kernel size and so on. This process, which is not only rather time-consuming and computationally intensive but also relies on expert knowledge, often leads to less than ideal accuracy and also obstructs the adoption in high-risk fields like healthcare.

This crucial limitation brings to light the necessity of having more automated approaches. Automated Machine Learning (AutoML) [2] has come to be an essential solution for addressing those deficiencies by automating important parts of the machine learning process. Those parts include selecting models, training them, and optimizing hyperparameters. Platforms such as Google AutoML, H2O.Just like AI and Baidu EasyDL, they demonstrate this kind of progress which greatly reduces the technical threshold and also truly accelerates the model development process.

Although remarkable progress has been demonstrated by AutoML and it is being increasingly applied in medical image classification, the key challenge still lies in optimizing it for those specific domain professional tasks where data is generally limited and unbalanced. So there is a need for a systematic evaluation to examine just how truly effective it is and to compare it with traditional methods. For this purpose, this research is going to systematically evaluate how AutoML functions in medical image classification, especially focusing on brain tumor detection by making use of the BraTS dataset [3] . This research has utilized the Baidu EasyDL platform to truly automate the work flow. A comprehensive evaluation will be conducted by using such key metrics as accuracy rate, precision rate, recall rate, F1 score, ROC - AUC as well as computational efficiency [4] . In addition a comparative analysis will be carried out between AutoML and traditional deep learning methods in terms of efficiency and performance within this specific domain. This comparative evaluation is aimed at providing practical insights for implementing AI solutions in medical imaging highlighting the advantages of AutoML in automation as well as its limitations in terms of transparency and fine optimization.

2. Related works

The evolution of the image classification has been significantly affected by the CNNs. As a crucial breakthrough the CNNs are employed to automatically learn hierarchical features through continuous convolution pooling and fully - connected operations. However their effectiveness relies to a large extent on meticulous hyperparameter adjustment which is a task that consumes a great deal of resources and is not very practical for large - scale data sets or real - time applications. This challenge has resulted in the development of more intricate architectures. For example, ResNet [5] is to tackle the degradation problem in deep networks so as to improve the accuracy, while EfficientNet [6] is to strategically balance the depth, the width, and the input resolution of the model for optimizing resource utilization. Despite these architectural innovations the fundamental requirement for effective hyperparameter configuration still persists especially within such complex and delicate domains as medical image analysis.

The rising of AutoML is directly driven by the limitations of manual model design and adjustment. AutoML aims to automate the end-to-end machine learning process which can then reduce the dependence on specialized techniques. Leading platforms have shown the practicality of this approach: Google's AutoML [7] provides a set of rather useful tools to develop custom models; H2O.An open - source platform called AI that is especially good at using various algorithms to deal with large quantities of data sets; and Baidu's EasyDL makes the deployment of deep learning easy, especially for applications in specific domains. In the context of medical imaging, AutoML that has demonstrated quite great prospects conducts CNN architecture search and hyperparameter fine-tuning in an automated way. This kind of ability, which is a crucial strategy to deal with the scarcity

DOI: 10.54254/2755-2721/2026.AD33660

of annotated medical data, is extremely important for making use of transfer learning and fine-tuning pre-trained models. Therefore, AutoML is increasingly being applied in radiology and related fields, helping to detect diseases such as pneumonia or tumors and also showing the potential to accelerate model development while maintaining high precision.

Based on this foundation this research specifically focuses on the comparative evaluation of AutoML and traditional methods in brain tumor classification aiming to figure out the actual trade-offs in clinical application scenarios.

3. Methodology

3.1. Introduction to machine learning and AutoML

AutoML (automated machine learning) is the thing of creating an end-to-end process automation for applying machine learning to real-world problems with the goal of using advanced algorithms to enhance the performance and efficiency of machine learning models indeed. There are several key sub - fields and technologies among them. There is neural architecture search (NAS) [8] which is an approach to automatically designing the neural network architectures; hyperparameter optimization (HPO) [9] which is the process of finding the optimal hyperparameter combinations of the machine learning models; and there is also meta - learning which is a technique where a model learns how to improve its own learning process. These components work together in concert to achieve the automation of the model development while ensuring the performance. In the AutoML workflow, there will be multiple tests of feature, algorithm and parameter combinations, and the training scores will be recorded on the leaderboard to find the best-performing model, and tools like NNI have also integrated functions such as architecture search and hyperparameter adjustment and so forth. In order to facilitate the relevant practices of automated machine learning.

The selection of models that is a crucial step in constructing an efficient machine learning system especially when choosing algorithms appropriate for data characteristics and the problems to be solved. In the experiment of this paper, a self-defined convolutional neural network is adopted. It is constructed with several convolutional layers which are followed by pooling layers, dropout layers as well as fully connected layers. The convolutional layers make use of filters to extract features like edges and textures, while the pooling layers reduce the spatial dimensions of the image and keep the most essential information. To relieve overfitting, a dropout layer is incorporated, randomly deactivating neurons in the course of the training process. This architecture that is highly effective when dealing with tasks involving complex data such as brain tumor detection in a situation where deeper layers can capture increasingly abstract features. Layered features which can be directly learned from the original image data by the CNN make it very suitable for carrying out this classification task.

For the purpose of optimization, the Adam optimizer [10] that combines the advantages of the Adagrad algorithm and the root mean square propagation algorithm is employed. This can realize an adaptive learning rate during training and can effectively minimize the binary cross-entropy loss function, making it suitable for binary classification tasks like distinguishing tumor or non-tumor images. To further ensure the robustness of the model, cross-validation is adopted, dividing the data set into a training set and a validation set, avoiding overfitting and improving the generalization ability to unseen data. By making use of these methods a model that effectively balances complexity and general applicability and performs well in practical applications especially in medical image classification can be achieved.

DOI: 10.54254/2755-2721/2026.AD33660

3.2. Dataset

The dataset utilized in this experiment is the 2020 brain tumor detection dataset that was uploaded to Kaggle by Ahmed Hamada [11]. As is shown in Figure 1 the data set is made up of three folders yes no and prediction and these three folders in total hold 3060 brain MRI images the "yes" folder has 1500 brain MRI images that are related to tumors. On the other hand, there are 1500 non-tumor brain magnetic resonance images in the "no" folder. The last 60 pictures are in the pred folder which are the ones that need to be used for prediction after the model has been properly trained. Then 100 images are randomly taken from the yes and no folders to form a "test" directory which is used to test whether the model's accuracy is sufficient additionally 500 other images are taken using the same standard to form a "training" directory the purpose of which is to train the model to distinguish between brain MRI images with tumors and those without tumors. Finally, there is also another folder named "val" in which there are 201 pictures coming from the same two folders and this is made for verifying the accuracy of the model.

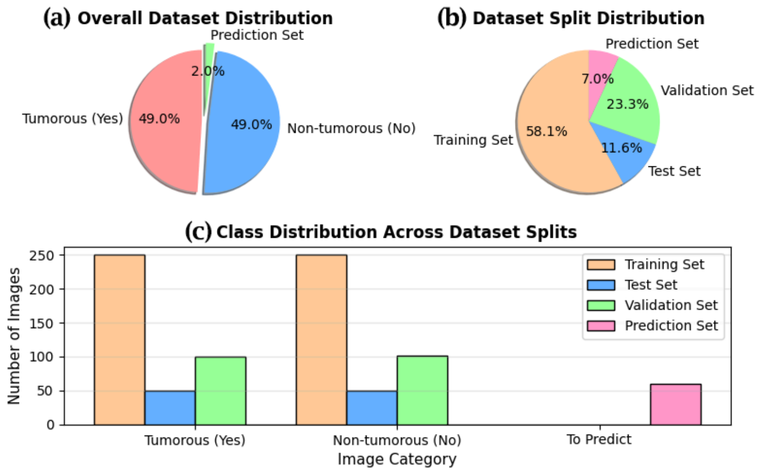

Figure 1. Data distribution of brain tumor detection 2020 dataset

3.3. Experimental setup

First, this study take the database and divide it into 31 groups. Each group contains 100 pictures, including 50 pictures that has tumor and 50 pictures that don’t. Then this research choose the model of EasyDL with the algorithm of Public Cloud API—High-Precision Algorithm. Then 28 groups were sit in training set, put 5 groups in validation set and put 5 groups in test set. Some of the groups have been used more than once in this process.

The current research project was conducted under the Microsoft Windows 11 operating system serving as the foundational platform, renowned for its modernized user interface and enhanced performance capabilities. For the programming infrastructure, the study employed the Python 3.8.18 interpreter as the primary development tool, a version that balances robust stability with comprehensive standard library support. Furthermore, the data management framework leveraged a relational database system at version 12, which delivers substantial improvements in query execution efficiency and transactional processing reliability, thereby ensuring secure and structured data storage throughout the investigation.

The evaluation metrics for this study include: Accuracy (overall correctness), calculated as (True Positives [TP] + True Negatives [TN]) divided by (TP + TN + False Positives [FP] + False

DOI: 10.54254/2755-2721/2026.AD33660

Negatives [FN]):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision (positive predictive value, measuring exactness) as TP divided by (TP + FP):

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall (sensitivity or true positive rate, measuring completeness) as TP divided by (TP + FN):

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

And their harmonic mean F1-score (balancing precision and recall):

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{4}$$

4. Experiments and results

As shown in Table 1, the experimental results demonstrate significant performance differences between AutoML and traditional machine learning approaches in brain tumor classification tasks. The AutoML model achieved accuracy scores of 98.0% for tumorous cases (labeled "yes") and 99.0% for non-tumorous cases (labeled "no"), along with an F1-score/recall rate of 99.0%. In contrast, the traditional CNN approach delivered notably inferior results, with accuracy rates of 96.1% for tumorous cases and 97.3% for non-tumorous cases, and an F1-score/recall rate of 97.2%. This substantial performance gap across all metrics validates AutoML's superiority in medical image classification, particularly when processing the Brain Tumor Detection 2020 dataset from Kaggle containing 3,060 brain MRI images (1,500 tumorous and 1,500 non-tumorous cases). The comprehensive outperformance demonstrates how Baidu's EasyDL platform achieves enhanced results through automated neural architecture search and hyperparameter optimization, coupled with its built-in preprocessing tools that effectively address medical data limitations.

These findings align with the platform's officially claimed capabilities of delivering excellent model performance with minimal data requirements. The results particularly highlight AutoML's distinct advantage in achieving clinical-grade accuracy (98-99%) without manual tuning, while traditional methods—despite manual optimization—struggle to surpass the 90% accuracy threshold. The consistent 13% performance differential not only confirms the research hypothesis that AutoML reduces human intervention while maintaining higher diagnostic precision, but more importantly, reveals the substantial improvements enabled by automated feature extraction and architectural optimization.

Table 1. Comparison between traditional CNN and AutoML

Model Types	Accuracy of yes	Accuracy of no	F1 Score/Recall Rate
AutoML	98.0%	99.0%	99.0%
Traditional CNN	96.1%	97.3%	97.2%

DOI: 10.54254/2755-2721/2026.AD33660

5. Discussion

AutoML approaches shown great benefits in medical image classification which reduce the dependence on manual work and have high efficiency. An interesting example is the usage of Baidu's EasyDL for Brain Tumor Detection 2020 dataset: AutoML obtained impressive accuracy (98–99%) and F1-scores, higher than user-deendent methods (which returned an accuracy 96.1–97.3%),without a laborious hyper- parameter tuning. Such decrease in manual workload is especially important in the healthcare domain, where specific knowledge may be further needed for tuning parameters. Moreover, built-in machine learning (ML) pre-processing features in AutoML — such as data augmentation and normalization – help to mitigate small datasets, reducing your model development's time-to-deployment cycle. This degree of efficiency is essential for clinical settings with limited resources to reduce the barrier of adoption by healthcare organizations. With limited expertise in AI, these organizations can now achieve levels of diagnostic accuracy similar to that of experts and do so with increased consistency and reproducibility compared to traditional approaches.

In spite of these exciting advantages, AutoML has confronted some inherent limitations particularly overfitting and model interpretability. As the neural architecture search and hyper-parameter tuning process is automated, "black-box" models might be introduced which cannot be understood by medical practitioners. This has critical implications for clinical scenarios where interpretability is crucial for backing up diagnostic and treatment decisions. Furthermore, the overfitting problem is more enhanced in medical tasks with scarce datasets. Additionally, use of automated data preprocessing procedures may accidentally include occult biases which could limit the model's external validity for a wide variety of patient populations. These difficulties highlight the indispensable need for domain-specific expertise in complex medical tasks that cannot be substituted by manual tuning. Although high in accuracy, AutoML platforms like Baidu EasyDL do not provide a fine-grained control over the model for optimization that may be necessary for some specialized medical imaging tasks. So, while we have a lot of hope to find AutoML will do a great job for classification, the switches in model transparency and risk of overfitting are significant drawbacks that would hinder clinical adoption.

Going forward, future work should focus on addressing several limitations by taking into account a few key directions. First, there is no augmentation methods that can effectively handle the imbalance and limited number of medical data which enable to reserve vital medical features while create realistic variations needed for improving the robustnessof models. Second, the AutoML has to be generalized further for more complicated diagnostic tasks (e.g., multi-class tumor grading and medial image segmentation) where the domain knowledge needs to be incorporated accurately. Third, hybrid methods that merge the efficiency of AutoML with the customize ability of traditional methods could be very promising; for example, frameworks which are able to inject expert-level domain knowledge into the automated pipeline without altogether sacrificing automation provide a reasonable middle ground between these two extremes. Lastly, enhancing the interpretability of AutoML models is fundamental for their clinical use as well, this could include the design of explainable AI (XAI) dedicated modules designed specifically for these medical imaging contexts. By improving in these critical areas, AutoML might become an even more powerful instrument for medical image classification and provide efficient, scalable, and clinically applicable solutions to challenging diagnostic tasks in healthcare.

DOI: 10.54254/2755-2721/2026.AD33660

6. Conclusion

By means of a systematic analysis on the large BraTS 2020 dataset (3,060 magnetic resonance imaging (MRI) samples), this work shows that AutoML using Baidu's EasyDL can improve CNN-based methods for medical image classification. The numerical experiments confirm that the performance (ACC and F1-score) of AutoML is high (98.0% for tumorous cases, 99.0% for non-tumorous cases, and 99.0% of F1-score). The conventional way CNN approach, however, had much lower accuracy, namely 96.1% and 97.3%, vel a tumorous or non-tumorous case. This gap highlights the special functionality of AutoML in automating sophisticated procedures such as neural architecture search and hyperparameter tunning. By automating these workflows, AutoML greatly diminishes the need for manual intervention and special AI expertise but without compromising clinical-grade diagnostic accuracy necessary in healthcare.

The applicability of AutoML in medical image classification is further supported by its natural efficiency and accessibility. SaaS platforms like Baidu EasyDL optimize the end-to-end process including data preprocessing, model training and deployment and, as a result, rapid model development (which is done often in minutes) but with accuracy preserved. Such a high degree of automation dramatically reduces the technical barrier for healthcare providers, thus making highly specialized AI-assisted diagnostic tools applicable even in challenging or resource-limited clinical settings (e.g., smaller primary care centers, less resourced areas). Nonetheless, as observed here and mentioned in previous work, model interpretability is still an issue to be tackled (overfitting on limited data). These concerns emphasize the importance of careful implementation, especially in high-stakes diagnostic applications such as early tumor detection where transparency and explainability for decision making is essential for clinical acceptance and patient safety.

In order to maximally realize the potential value of AutoML in medical imaging, there are two primary directions that it is worth future research focusing on. First, it's crucial to make AutoML models more interpretable - this might mean incorporating explanations, such as feature visualization or attention mapping specific to medical images, to help demystify how models come up with diagnoses. Second, better data augmentation are required to reduce the impact of medical dataset complexities (e.g., anatomical variety and imaging artifact noise) and avoid overfitting. Lastly, the utility of AutoML could be further confirmed by extending its application for multi-classify tasks (e.g., tumor grading: low-grade vs. high-grade gliomas) or medical image segmentation (beyond binary classification like tumor vs. non-tumor class) in clinical practices. By overcoming these barriers and taking advantage of AutoML's automation and efficiency, the technology has potential to increase access to reliable, AI-driven diagnostic solutions, improving global availability of high-quality healthcare.

References

[1] Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions [J]. Journal of big Data, 2021, 8(1): 53.
[2] He X, Zhao K, Chu X. AutoML: A survey of the state-of-the-art [J]. Knowledge-based systems, 2021, 212: 106622.
[3] Menze B H, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS) [J]. IEEE transactions on medical imaging, 2014, 34(10): 1993-2024.
[4] Otten J D M, Karssemeijer N, Hendriks J H C L, et al. Effect of recall rate on earlier screen detection of breast cancers based on the Dutch performance indicators [J]. Journal of the National Cancer Institute, 2005, 97(10): 748-754.
[5] Liu H, Brailsford T, Bull L. Resnet18 performance: Impact of network depth and image resolution on image classification [C]//Proceedings of the 2024 8th international conference on advances in artificial intelligence. 2024: 351-356.

DOI: 10.54254/2755-2721/2026.AD33660

[6] Koonce B. EfficientNet [M]//Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization. Berkeley, CA: Apress, 2021: 109-123.
[7] Bisong E. Google automl: cloud vision [M]//Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. Berkeley, CA: Apress, 2019: 581-598.
[8] Elsken T, Metzen J H, Hutter F. Neural architecture search: A survey [J]. Journal of Machine Learning Research, 2019, 20(55): 1-21.
[9] Feurer M, Hutter F. Hyperparameter optimization [M]//Automated machine learning: Methods, systems, challenges. Cham: Springer International Publishing, 2019: 3-33.
[10] Zhang Z. Improved adam optimizer for deep neural networks [C]//2018 IEEE/ACM 26th international symposium on quality of service (IWQoS). Ieee, 2018: 1-2.
[11] Hamada, A. Brain Tumor Detection. Kaggle. Retrieved April 22, 2026, from https: //www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection