Algorithm Selection with Priority Order for Instances

Zhamilya Saparova
Department of Computer Science
Nazarbayev University
Astana, Kazakhstan
zhamilya.saparova@nu.edu.kz

Martin Lukac
Department of Computer Science
Nazarbayev University
Astana, Kazakhstan
and
Department of Networks and
Computer Architecture
Hiroshima City University
Hiroshima, Japan
malu@hiroshima-cu.ac.jp

Abstract

Reliability in medical image diagnostics is a required trait for any artificial system. Currently, most approaches rely on highly trained and specific models to leverage the feature quality learned from a particular type of medium, such as X-rays, NMR, PET scans and others. While this approach aligns with the standard human expert perspective, it also limits artificial systems to the representations learned from the dataset distribution. To gain a better understanding of how different media affect specific tasks, we explore task-specific feature transfer between domains. In this work, we propose the possibility of merging features from various areas to harness feature transfer in outlier cases. For this purpose, we develop an Algorithm Selection (AS) method that chooses algorithms trained on different sets of medical images and for different classification tasks. The AS system is then applied to a different classification task. The AS represents a set of methods that, given a problem and a range of existing algorithms, selects the best algorithm on a case-by-case basis. The results demonstrate the advantages of incorporating algorithms from different tasks and datasets in a supervised manner. By considering algorithms trained on diverse datasets, we can effectively capture outliers that might otherwise be neglected by more specific algorithms.

1 Introduction

Medical diagnostics is becoming more and more automated due to the usage of machine learning approaches. Learning deep models from larger and larger datasets allows to minimize the occurrence of outlier situations. However due to privacy concerns, medical data is often available only in small amount which prevents often an effective training of very deep models. In addition the data is also obtained from localized sources, minimizing the exposure for unexpected symptoms. Finally, medical data is also highly noisy, often contaminated with aperture artefacts and other medical artefacts. As a result, medical diagnostics in general achieves locally good results but the employed methods are very sensitive even small changes in the expected input.

In order to address this problem, in the area of medical image classification, several approaches have been proposed in the literature. Aurna et al. (2) introduced a two-stage feature ensemble approach using deep Convolutional Neural Networks (CNN) to classify three types of brain tumors (meningioma, glioma, pituitary) as well as normal brain images. Ullah et al. (11), on the other hand, utilized pre-trained transfer learning (TL) classifiers to accurately identify and detect glioma, meningioma, and pituitary brain tumors. They conducted evaluations on a baseline brain tumor

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

classification (MRI) dataset, considering nine different pre-trained TL algorithms. Another study by Razaa et al. (8) presented a transfer learning-based predictor method that relied on EfficientNet. Their research involved an ablation study using the "Chest CT-Scan images Dataset," with the goal of mapping images to their respective classes: normal, adenocarcinoma, large cell carcinoma, and squamous cell carcinoma.

Most of the algorithms used for medical image classification are trained on purpose on a specific dataset for a specific task. The feature and sample distribution therefore corresponds to knowledge in the target specific domain. However, knowledge from different datasets and different tasks could be used for improve knowledge about outlier samples and specific cases.

In addition because ultimately the different learning problems are represented by some features, we are interested to determine if the most accurate algorithms trained on different datasets and for different type of classification can be combined on a completely new task and provide overall improvement.

Thus to address the issue of data and algorithm specificity in medical images, we propose employing meta-learning and merging specific detectors in a supervised manner. The approach begins with a set of algorithms trained for a particular classification task using a specific dataset. These algorithms are then evaluated on a new dataset and a different classification task. Finally, the data obtained from this evaluation process is utilized to construct an algorithm selection framework that identifies the best algorithm on a case-by-case basis.

Algorithm Selection (AS) can also be seen as an optimization strategy: given a problem and a set of algorithms trained on different data or using different algorithmic biases, select the algorithm on a case-by-case basis to leverage maximum accuracy. AS has been successfully applied to decision-making problems where meta-information is available, particularly for problems with structured input information like logic formulae or problems specified using discrete features (4). However, when dealing with real-world information such as images or sound, AS has been used with varying degrees of success. To address some of the challenges encountered when no meta-information is available, the authors of (6) and (10) introduced the concept of selecting selectors based on data distillation or algorithm properties. In (6) the authors build a set of selectors using data distillation and then used an algorithm selector to select the best algorithm selector. In (10), the authors build an algorithm selector selector directly by evaluation algorithm selector algorithmic bias.

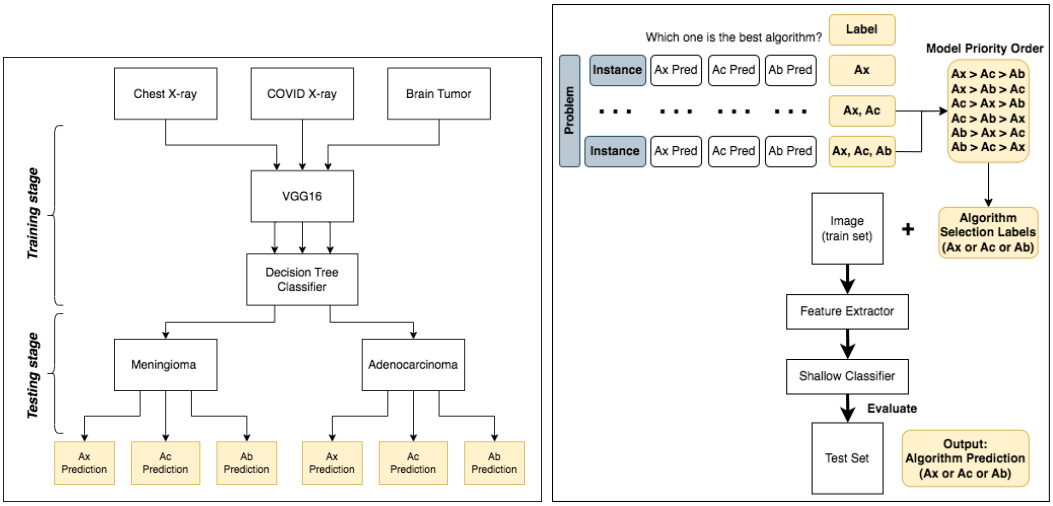
2 Proposed Method

The method starts by taking a set of pre-trained, or locally trained (trained on specific distinct datasets and task) algorithms. Figure 1a shows the overall picture of the training procedure. We train three different algorithms each on a different dataset for the classification task. The first algorithm $\mathcal{A}_X$ is trained on the Chest X-ray (7), the algorithm $\mathcal{A}_C$ is trained on Covid X-ray (9) and the algorithm $\mathcal{A}_B$ is trained Brain tumor (5) dataset. Each algorithm is classifying samples into two labels $\mathcal{L} = \{Healthy, Not - healthy\}$. The training stage is shown in Figure 1a, *Training Stage*.

Once the classification algorithms are trained, we proceed to evaluate each of them on two different classification tasks. The first task involves meningioma classification (3), while the second task focuses on adenocarcinoma classification (1) (Figure 1a, *Testing Stage*). To ensure balanced representation, all datasets have been prepared by sampling from under-represented classes. Furthermore, to enable seamless combination and utilization in different classification tasks, all tasks have been transformed into binary classification.

Then, we built the Algorithm Selection using the training set of the two target tasks: classification of meningioma and adenocarcinoma. However, because there are no meta-features available from the images, the AS uses the features from VGG16 pre-trained on ImageNet. The overview of the method is shown in Figure 1b.

Note that when all three algorithms were trained, we performed a verification of an ideal algorithm selection with a selection accuracy of 100%. Under such conditions, the accuracy of both the classification of meningioma and adenocarcinoma attained 100% of accuracy.



(a) The training of the individual Algorithms (b) Overview of the Algorithm Selection framework

Figure 1: The Algorithm selection framework (a) training and (b) processing methodology.

One of the problems in the multi-algorithm classification is that in many cases more than one algorithm can generate the correct result. The problem for training a meta-classifier with multiple correct choices is that the result often suffers from algorithm and data bias.

As it is in our case, the amount of training data is in the order of few thousands samples. When training the meta classifier, each sample where more than one algorithm is correct, the choice of the target must be made. Random sampling is a possible choice but is inconsistent across trials. Therefore, in order to deal with this specific problem we implemented an priority based algorithm selection.

Instead of training thus a specific selection, we propose to train the algorithm selection based on algorithm selection priority. The general idea is that every time multiple algorithms have same result, we evaluate different priorities to be used. For instance, let the priority be $\mathcal{A}_X > \mathcal{A}_B > \mathcal{A}_C$ and if the results of processing an input image of $\mathcal{A}_B$ and of$\mathcal{A}_C$ algorithms had the same correct result, the $\mathcal{A}_B$ would be selected. As a result we construct for six different algorithm selectors. Each trained using a different priority for training sample selection.

3 Experiments and Results

3.1 Datasets

In this research, five distinct 2D image datasets are employed, specifically chest X-ray (7), brain tumor (5), COVID X-ray (9), meningioma (3), and adenocarcinoma (1). The fundamental characteristics of each dataset are presented in Table 1. Category 1 denotes images associated with pathological conditions, while Category 2 represents data indicative of a healthy state.

Dataset Name	Sample Number	Category 1	Category 2
Chest X-ray	5216	3875	1341
COVID X-ray	1204	536	668
Brain Tumor	4600	2513	2087
Meningioma	1437	937	500
Adenocarcinoma	517	315	202

Table 1: The overview of some properties of used Datasets.

The data pre-processing phase commences with the classification of images related to viral infections through a binary labeling scheme, in which the label "0" is assigned to images portraying virus-infected states, while "1" is designated for images representing healthy conditions. The dataset

undergoes normalization and appropriate reshaping procedures. To ensure balanced representation, all datasets have been prepared by sampling from under-represented classes.

3.2 Experiments

The results of the evaluation of the whole system are shown in Table 2. The various orders of algorithm selection are shown in the first column of the table. Columns two to four show the results on evaluation dataset when classifying the meningioma and columns five to seven show the results of classifying adenocarcinoma. The shown results use logistic regression classifier for selecting the algorithm because it experimentally showed the best results.

Observe, that the results from classification of meningioma resulted in accuracy higher than the state of the art reported in (2) being 99.13%. For adenocarcinoma, the accuracy got very close to the state of the art. The currently reported single algorithm accuracy was described in (8) with maximum accuracy of 94%.

Of course, one must consider that our results are based on binary classification, while the compared state of the art is used for multi-label classification. However, our purpose was different: we studied if it is possible to use the different sources of information to improve algorithm selection. According to these preliminary results, one can observe that when the task is similar but using a different medium the meta-selector can efficiently adapt. The reason is most likely due to the fact that the dataset are similar, the features are descriptive enough.

However it also poses the question about task and algorithm specificity. In particular, what is the general requirement for algorithms being able to process any input data and simply change the task by the final set of computational components?

4 Conclusion

In this paper we presented a set of experiments aimed at leveraging inter-data set features for a task of classification. We trained three algorithms, each on a different dataset. The algorithm selection was then trained again on a different dataset so that we can evaluate the selection of algorithms that extract different features for different classification. The most interesting result is that the perfect algorithm selection results in 100% for both datasets used for training the algorithm selection while none of the used algorithms have been trained on it. This would indeed indicate that the various sources of information in these datasets can be used in a complementary manner without compromising the accuracy of the result.

5 Acknowledgments

This work was partially supported by the Faculty-development competitive research grants program with number Faculty-development competitive research grants program 20122022FD4109 and by the Erasmus+ CBHE ACeSYRI grant with reg. num. 610166-EPP-1-2019-1-SK-EPPKA2-CBHE-JP.

Priorities	Meningioma Evaluation			Adenocarcinoma Evaluation		
	VGG16	VGG19	ResNet50	VGG16	VGG19	ResNet50
$\mathcal{A}_X > \mathcal{A}_C > \mathcal{A}_B$	0.982	0.968	0.795	0.724	0.833	0.586
$\mathcal{A}_X > \mathcal{A}_B > \mathcal{A}_C$	**0.995**	0.973	0.836	0.655	0.632	0.54
$\mathcal{A}_C > \mathcal{A}_X > \mathcal{A}_B$	0.977	0.973	0.782	0.92	0.879	0.667
$\mathcal{A}_C > \mathcal{A}_B > \mathcal{A}_X$	0.977	0.973	0.845	**0.925**	0.897	0.684
$\mathcal{A}_B > \mathcal{A}_X > \mathcal{A}_C$	0.982	0.955	0.845	0.69	0.621	0.586
$\mathcal{A}_B > \mathcal{A}_C > \mathcal{A}_X$	0.964	0.968	0.85	0.776	0.828	0.563

Table 2: Results of the various orders of algorithm selection on both evaluation datasets.

References

[1] Anon, 2023a. Chest CT-Scan images Dataset. Available: Chest CT-scan images dataset. https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images.

[2] Aurna, N. F., Yousuf, M. A., Taher, K. A., Azad, A. K. M., & Moni, M. A. (2022). A classification of MRI brain tumor based on two-stage feature-level ensemble of deep CNN models. Computational Biology and Medicine, 146, 105539. DOI: 10.1016/j.compbiomed.2022.105539.

[3] Bhuvaji, S., Kadam, A., Bhumkar, A., Dedge, S. & Kanchan, S. Brain Tumor Classification - MRI. Kaggle. Retrieved from https://www.kaggle.com/datasets/sartajbhuvaji/brain-tumor-classification-mri.

[4] Bischl, B. et al. (2016) 'ASlib: A Benchmark Library for algorithm selection', Artificial Intelligence, 237, pp. 41–58. doi:10.1016/j.artint.2016.04.003.

[5] Hamada, A. Brain Tumor Detection Dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-detection?resource=download&select=no.

[6] Lukac, M. and Abdiyeva, K. and Bayanov, A. and Li, Albina, and Izbassarova, N. and Gabidolla, M. and Kameyama, M., Selecting Algorithms without meta-features, ICPR 2020, 1st International Workshop on Industrial Machine Learning, 2020

[7] Mooney, P. T. Chest X-ray Pneumonia. Kaggle. Retrieved from https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[8] Raza, R., Zulfiqar, F., Khan, M. O., Arif, M., Alvi, A., Iftikhar, M. A., & Alam, T. (2023). Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images. Engineering Applications of Artificial Intelligence, 126(Part B), 106902.

[9] Siddhartha, A. (2021). COVID CXR Image Dataset - Research. Kaggle.Retrieved from https://www.kaggle.com/datasets/sid321axn/covid-cxr-image-dataset-research.

[10] Tornede, A., Gehring, L., Tornede, T. et al. Algorithm selection on a meta level. Mach Learn 112, 1253–1286 (2023). https://doi.org/10.1007/s10994-022-06161-4

[11] Ullah, N., Khan, J. A., Khan, M. S., Khan, W., Hassan, I., Obayya, M., Negm, N., & Salama, A. S. (2022). An Effective Approach to Detect and Identify Brain Tumors Using Transfer Learning. Applied Sciences, 12(11), 5645. https://doi.org/10.3390/app12115645