



Few-shot classification of ultrasound breast cancer images using meta-learning algorithms

Gültekin Işık¹ · İshak Paçal¹

Received: 8 February 2023 / Accepted: 25 March 2024 / Published online: 18 April 2024
© The Author(s) 2024

Abstract

Medical datasets often have a skewed class distribution and a lack of high-quality annotated images. However, deep learning methods require a large amount of labeled data for classification. In this study, we present a few-shot learning approach for the classification of ultrasound breast cancer images using meta-learning methods. We used prototypical networks and model agnostic meta-learning (MAML) algorithms as meta-learning methods. The breast ultrasound images (BUSI) dataset, which has three classes and is difficult to use in meta-learning, was used for meta-testing in a cross-domain approach along with other datasets for meta-training. Our proposed approach yielded an accuracy range of 0.882–0.889, achieved by implementing the ResNet50 backbone with ProtoNet in a 10-shot setting. These results represent a significant improvement ranging from 6.27 to 7.10% over the baseline accuracy of 0.831. The results showed that ProtoNet outperformed the MAML method for all k-shot settings. In addition, the use of ResNet models as the backbone network for feature extraction was found to be more successful than the use of a four-layer convolutional model. Our proposed method is the first attempt to apply meta-learning for few-shot classification in the BUSI dataset while providing higher accuracy compared to deep learning methods for medical images with small-scale datasets and few classes. The methodology used in this study can be adapted to other datasets with similar problems.

Keywords Few-shot classification · Meta-learning · Prototypical networks · Breast cancer · ResNet

1 Introduction

Accurately classifying diseases from limited data is a major challenge in medical imaging. However, this task is challenged by limited data availability, a common issue in medical settings. Medical datasets are often constrained by privacy concerns, data acquisition costs, and the difficulty of obtaining labeled samples, resulting in small-scale datasets [1]. This scarcity of data presents a significant obstacle to traditional machine learning and deep learning methods, which typically require large volumes of labeled data for effective model training [2]. The problem of few-shot classification, where the goal is to classify with minimal examples, becomes particularly pertinent in this context. Breast cancer detection is a formidable challenge

due to the scarcity of comprehensive and diverse datasets, which hinders the development of accurate classification models. Traditional machine learning and deep learning approaches rely heavily on abundant labeled data for effective model training. However, the paucity of labeled samples in medical imaging, particularly in breast cancer datasets, limits the effectiveness of these methods. This limitation, compounded by class imbalance within the data, prompts the exploration of innovative strategies such as few-shot learning. The few-shot learning paradigm offers a promising solution [3, 4] by facilitating accurate classification even with minimally labeled samples, thus addressing the pressing need for robust and adaptive models in breast cancer diagnosis.

Metastatic cancers, particularly breast cancer, pose a substantial threat. Arising in mammary ducts or glands, it swiftly spreads [5], emphasizing the need for early detection, which has proven effective in reducing related deaths by 40% [6]. Ultrasound (US) imaging, a painless and comfortable real-time technique, is employed in breast

✉ Gültekin Işık
gultekin.isik@igdir.edu.tr

¹ Department of Computer Engineering, Igdir University, Igdir, Türkiye

cancer diagnosis [7]. It is a widely favored method in the initial assessment of breast cancer, surpassing alternatives like mammography and biopsy.

Meta-learning has emerged as a compelling approach to address the constraints of small datasets. This technique equips models with the ability to rapidly adapt to new tasks with minimal training samples, making it a promising tool for enhancing classification accuracy. In this paper, we delve into the application of meta-learning in the specific domain of breast cancer classification using medical images, aiming to leverage this approach to not only enhance classification accuracy but also contribute to more efficient and reliable early detection of metastatic breast cancers, thereby improving patient outcomes.

Computer-aided diagnosis systems (CADx) are frequently used in the early diagnosis of breast cancer as well as in the early diagnosis of many cancers. In recent years, deep learning-based CAD systems have given promising results in the diagnosis of many cancers. Previous CAD systems often included classical machine learning approaches, but these approaches lacked the ability to generalize. As deep learning-based approaches overcome this problem, studies are presented in the diagnosis of breast cancer from US images. Since there is only one public dataset in the literature containing US images (BUSI) [8], a competent CAD system has not been presented. Deep learning approaches offer more effective results when trained with large-scale data [9].

Meta-learning is a machine learning subfield that equips models with the ability to quickly adapt to new tasks based on prior experience, even with limited data. In other words, meta-learning aims to improve the ability of a model to learn new skills or adapt to new environments, by using the knowledge it has acquired from previous tasks.

Meta-learning finds applications across various domains, including image and speech recognition, natural language processing, robotics, and healthcare [10]. It enhances machine learning models for tasks like disease diagnosis, treatment planning, and drug discovery [11].

Meta-learning in cancer diagnosis helps to optimize model performance through hyperparameter tuning or by exploiting prior experience [12]. Ouyang et al. [12] introduced SSL-ALPNet, a self-supervised framework for medical images that uses superpixel-based pseudo-labels for annotation-free training. They improved the segmentation accuracy by incorporating an adaptive local prototype pooling module into the prototype networks. Sun et al. [13] proposed a fast learning model for medical image segmentation, using episodic training with a global correlation module. This method focuses on capturing spatial consistency, which is crucial in medical images, and shows improved discriminative ability with few training images.

Feng et al. [14] introduced interactive few-shot learning (IFSL) for medical image segmentation, easing annotation burdens. Their medical prior-based few-shot learning network (MPNet) uses a minimal set of annotated support images to guide query image segmentation without pre-training. The interactive learning-based test time optimization algorithm (IL-TTOA) optimizes and fortifies MPNet interactively, a pioneering aspect in few-shot segmentation. Singh et al. [15] proposed “MetaMed,” leveraging meta-learning for rare disease adaptation in medical image classification. Validated on Pap smear, BreakHis, and ISIC 2018 datasets, they used advanced augmentation techniques to combat overfitting, aiming to enhance few-shot learning. Hansen et al. [16] offered a novel approach for few-shot medical image segmentation, focusing on anomaly detection instead of background modeling. Their method, using a single foreground prototype and self-supervision with 3D structures, surpassed prior segmentation methods, tested on MRI datasets for abdominal organ and cardiac segmentation.

In a recent study [17], researchers developed a breast cancer classification model using meta-learning and multiple convolutional neural networks (CNNs). They improved accuracy by incorporating meta-learning, transfer learning, and data augmentation while focusing on the BUSI dataset. Within a meta-learning framework, various CNN models like Inception V3, ResNet50, and DenseNet121 were employed. The study evaluated both individual CNN models and their proposed meta-model’s performance in classifying benign and malignant breast lesions, achieving an accuracy of 0.90.

In literature, it has been observed that there are many studies in the field of meta-learning for medical image analysis, particularly for cancer diagnosis, and these studies have yielded successful results. In particular, the use of ultrasound imaging for breast cancer diagnosis is a non-invasive and cost-effective tool for early diagnosis. However, the limited availability of publicly available datasets of ultrasound images of breast cancer and the insufficiency of the BUSI dataset for deep learning techniques necessitate the exploration of alternative methods. In this regard, meta-learning has emerged as a popular method that offers more effective and successful results. In this study, we used meta-learning algorithms that require less data than deep learning architectures for the diagnosis of breast cancer using ultrasound images, which is one of the deadliest types of cancer. The novel aspects of our study that differentiate it from other studies are as follows:

This study presents an approach for medical image classification using meta-learning in situations with limited data. The standard few-shot classification with meta-learning approach involves using two datasets with the same distribution as the source (base classes) and target

(novel classes) domains, with a portion of the dataset being designated as base classes for training and the rest as novel classes. In simpler terms, base classes represent the categories of data used for initial model training, while novel classes are categories introduced during testing, which the model has not encountered during its initial training. However, the challenge arose when working with the breast ultrasound images (BUSI) dataset in this study, as it consisted of only three classes, making it unsuitable for the standard few-shot classification approach. To overcome this limitation, we devised a cross-domain strategy. In this approach, we leveraged other datasets for the meta-training phase, where the model learned to perform few-shot classification tasks. Meanwhile, we reserved the BUSI dataset exclusively for the meta-testing phase, where we evaluated the model's performance on the specific medical image classification task at hand.

This innovative cross-domain approach effectively tackled the twin challenges of limited data and the low-class problem in the BUSI dataset. The term 'low-class problem' pertains to situations where the dataset contains a restricted number of distinct classes. As far as our knowledge extends, this marks the pioneering attempt to apply meta-learning for few-shot classification on the BUSI dataset, employing a cross-domain methodology to address its unique characteristics and challenges.

2 Few-shot classification scheme

In conventional classification tasks, training and test sets comprise distinct samples from identical classes. The classes used for testing are those previously encountered during training. Conversely, in few-shot learning during testing, instances from new, unseen classes are classified [18]. There's a clear demarcation between the classes in the training and test sets. During the test or inference phase, there exist only a limited number of labeled data points for each class. Few-shot learning consists of two main phases: an initial training phase, which aims to make a model adaptable to new tasks, and a subsequent adaptation phase, where the already trained model is fine-tuned to perform these new tasks. This learning method focuses mainly on classifying rare classes during the meta-testing phase, while the meta-training phase involves training the model to achieve adaptability.

The objective of few-shot classification is to train a model that can promptly adapt to a new classification task using a limited number of observations. During the training stage, a model f_θ is learned using a training algorithm on a dataset $D^{\text{train}} = \{(x_i, y_i)\}_{i=1}^{|D^{\text{train}}|}$. Each element (x_i, y_i) in the D^{train} represents i th example with its corresponding label,

where $x_i \in \mathbb{R}^D$, and the label y_i are both drawn from a set of N_c classes. During the subsequent adaptation stage, a set of tasks T for few-shot classification is formulated using the test dataset D^{test} , where the classes differ from those in D^{train} . Every task T_i is made up of a support set S , which is employed for adapting the model, and a query set Q , which is employed for evaluating the model, and has identical label categories as that of S . When there are N categories in the support set S , and each category has K examples, T_i is known as an N -way K -shot task. The query set Q has N categories and each category contains q samples. The objective is to categorize all $N \times q$ samples in Q into their respective N categories [18, 19]. The support and query sets are defined in Eqs. 1 and 2 formulas, respectively.

$$S = \{(x_i, y_i)\}_{i=1}^m, m = N \times K \quad (1)$$

$$Q = \{(x_j, y_j)\}_{j=1}^z, z = N \times q \quad (2)$$

where $x \in \mathbb{R}^D$ is the D -dimensional input vector, y is the class label of the input sample and q is the number of queries from each class. During training, the model's predictions on the query sets are compared with the ground truth labels and loss is calculated. In the testing phase, metrics such as accuracy and precision of the model are calculated on the query set with the information obtained from the support set.

The aim of few-shot learning is to learn a generalizable classification model f_θ for K samples from N unseen classes. Transfer learning and data augmentation techniques are also generally used for few-shot learning [19]. While these methods have their merits, our study specifically employs meta-learning algorithms for few-shot classification on breast images to address the challenges posed by limited data conditions in medical image classification scenarios. The distinct advantages of meta-learning, including rapid adaptation to new tasks with limited data and learning from related tasks, are exploited to develop a robust and generalizable few-shot classification model. The results obtained by our meta-learning approach are systematically compared with those obtained by the conventional transfer learning technique.

2.1 Meta-learning approaches

The goal of meta-learning is to develop models that can learn from a small amount of data and generalize to new tasks with similar characteristics. Meta-learning, akin to transfer learning, optimizes models for multiple sub-tasks during training. Instead of fine-tuning a single sub-task, meta-learning optimizes the model for success across numerous sub-tasks by segmenting the training phase into multiple subsets, referred to as episodes or tasks. This type of training is called episodic training [20–22] (Fig. 1).

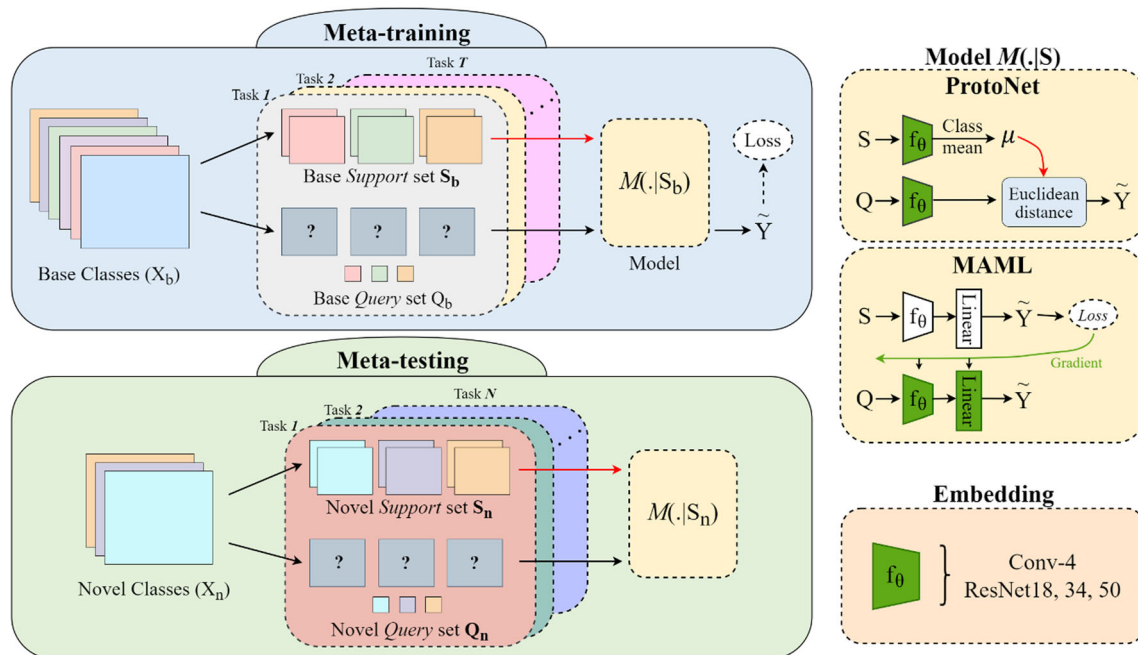


Fig. 1 For few-shot classification, the dataset is partitioned into meta-training and meta-testing and the classes in these parts do not overlap. Meta-training and meta-testing should have the same conditions. In both parts, the examples are organized into many tasks (episodes) and put into the classification process. Base classes generally refer to datasets with many examples, while novel classes refer to datasets with few and unseen examples to be classified. In ProtoNet, the embedding space (f_θ) is learnt from meta-training tasks according to

the distance metric. In meta-testing, the embedding vector of the instances is obtained using f_θ and the query set is classified based on the distance metric. In MAML, the parameter space learnt in meta-training is updated in meta-testing and the query set is classified accordingly. Thus, it can be said that ProtoNet learns to compare and MAML learns to fine-tune. Conv-4, ResNet18, ResNet34, and ResNet50 networks were used to obtain the embedding space. Figure adapted from [31]

In the episodic training process, the test process is mimicked with the query set used. Because the instances of the query set in each episode are different from the support instances. In each episode, the few-shot learning model performs N-way K-shot classification using distinct support and query sets. Each episode consists of new support and query sets and contains K and q randomly selected examples from N classes for these sets, respectively. In the test phase, the performance of the examples in the query set is measured using the knowledge of the support set.

There are generally 3 different approaches in meta-learning: metric-based [20, 23–25], model-based [26, 27], and optimization-based [28–30]. In our study, we employed two prominent meta-learning techniques, prototypical networks (ProtoNets) [24] and model agnostic meta-learning (MAML) [29], to enhance the accuracy of medical image classification, particularly in the context of breast cancer diagnosis. We chose ProtoNets and MAML for their effectiveness in few-shot learning scenarios. ProtoNets uses metric-based meta-learning to generate prototypes for each class, which facilitates fast classification. Meanwhile, MAML, an optimization-based approach, enables rapid adaptation to new tasks by fine-tuning model

parameters. Both techniques excel in scenarios with limited labeled data.

2.2 Prototypical networks (ProtoNet)

Prototypical networks are a metric-based meta-learning algorithm that works similarly to the nearest neighbor method. In metric-based approaches, an embedding space (feature vectors) is learnt in which instances belonging to the same class are close to each other according to certain metrics. This approach is based on similarity or distance metrics. Using these metrics, the similarity or distance of the instances in the query set to the instances in the support set is calculated. Accordingly, the mean of the feature vectors of all instances in the support set is computed. Given a new query instance x is classified according to its distance or similarity to these means (Fig. 2). Here, differentiable distance functions such as Euclidean distance [24] or cosine distance [20] can be used to calculate distances (d). In the ProtoNet approach, a neural network model (f_θ), known as the backbone, transforms support set samples into feature vectors. These feature vectors are used to calculate prototypes, which represent the mean feature vectors of instances within the same class (c) (Eq. 3). In

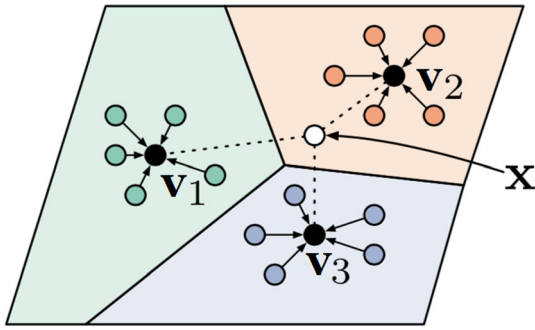


Fig. 2 The instances of the support set (S) and the instance x (query) are mapped into the embedding space by passing through the backbone model f_θ . The prototype of each class (v_c) is calculated and the instance x is assigned to one of the support set prototypes according to the distance metric d

Eq. 3, S_c denotes the instances of class c in the support set and v_c denotes the prototype of class c (Fig. 2).

The examples in the query set can be classified according to the prototypes calculated using the support set. For this purpose, the feature vectors of the instances in the query set (x) should be extracted with the embedding function f_θ . Then, these feature vectors and the prototype of each class (v_c) are passed through the distance function (d) and probability distributions are calculated. A small distance between the two (i.e. high similarity) means that the probability value is high. The softmax function used for probability distributions is given in Eq. 4.

$$v_c = \frac{1}{|S_c|} \sum_{(x_i) \in S_c} f_\theta(x_i) \quad (3)$$

$$p(y = c|x) = \frac{\exp(-d(f_\theta(x), v_c))}{\sum_{c' \in C} \exp(-d(f_\theta(x), v_{c'}))} \quad (4)$$

Here, training is performed based on the cross entropy error of the query set instances in meta-training. The parameters of the f_θ backbone network are updated to minimize the loss values.

Briefly, ProtoNets work by learning a prototype representation for each class in a way that minimizes the distance between the prototype and support examples (training images). In our application, ProtoNets were adapted to the medical image classification task by utilizing them to generate compact and discriminative representations of breast cancer lesions. These representations were then used to classify new, previously unseen lesions as normal, benign, or malignant.

2.3 Model agnostic meta-learning (MAML)

MAML, one of the optimization-based approaches, allows the θ parameters of the model to adapt quickly and easily to

new tasks (classes). The initial θ parameters of the model are adapted to new tasks with only a few examples in one or a few gradient steps. The difference from ProtoNet is that the model is fine-tuned for each episode in training and prediction is made by repeating the same process in testing. A diagram of the MAML algorithm is given in Fig. 3.

The meta-training and meta-testing phases consist of episodes (tasks). The dataset for meta-training consists of episodes in the form of $D_{\text{train}} = \{E_1, E_2, \dots, E_n\}$ and each episode consists of a support and query set in the form of $E_i = \{S_i, Q_i\}$. The MAML algorithm shown in Fig. 3 basically performs the computations for each episode in two stages:

$$f_\theta(S_i) = \bar{Y}_i^S, \quad \theta_i = \theta - \alpha \nabla_{\theta} L(S_i, \bar{Y}_i^S) \quad (5)$$

$$f_{\theta_i}(Q_i) = \bar{Y}_i^Q, \quad \theta = \theta_i - \beta \nabla_{\theta} L(Q_i, \bar{Y}_i^Q) \quad (6)$$

Initially, the parameters θ are random values. α and β are hyperparameters indicating the step size. In the first stage, the θ values for each episode are fine-tuned with the samples of the support set (Eq. 5). In the second stage, the θ parameters are updated using the samples of the query set (Eq. 6). These stages are repeated for each episode and the θ parameters are updated by the stochastic gradient descent (SGD) method according to the derivative of the total Loss (L) value. Here, L is calculated by comparing the model predictions (\bar{Y}) with the ground truth.

At meta-testing stage, the learnt θ parameters are used as a starting point for the given support and query sets. The fine-tuning process in training with new tasks is repeated here and new parameters (θ_{test}) are obtained. Predictions are made using the new parameters.

In our context, MAML was applied to fine-tune the CNNs used for image classification. By initializing the CNNs with parameters optimized for rapid adaptation to the task of breast cancer classification, MAML enabled our models to perform effectively even with limited training data.

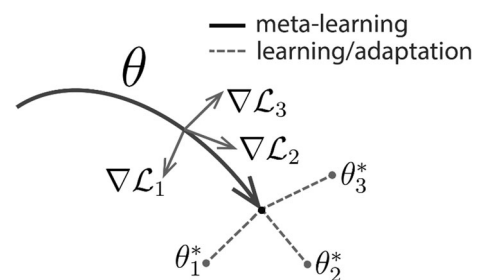


Fig. 3 MAML algorithm: The initial parameters of the model (θ) are fine-tuned for three different tasks ($\theta_1^*, \theta_2^*, \theta_3^*$) and adapted to the new tasks. Thus, the few-shot process (meta-testing) is actually imitated in the meta-training phase

By incorporating these meta-learning techniques into our methodology, we effectively addressed the challenges posed by small medical datasets, ultimately leading to improved classification accuracy. In the following sections, we provide further details on our experimental setup, datasets, and evaluation metrics, shedding light on how these techniques were applied in the context of our study.

3 Experiments

3.1 Datasets

Three datasets were used in this study: mini-ImageNet [19], BreakHis [32], and BUSI [8]. Of these, mini-ImageNet and BreakHis were used to obtain the few-shot model in the meta-training stage, while BUSI was used for few-shot classification in the meta-testing stage.

3.1.1 Mini-ImageNet

The mini-ImageNet dataset was proposed by Vinyals et al. [19] for few-shot learning. It is a dataset consisting of 100 ImageNet [33] classes and each class contains 600 images of size 84×84 . Base, validation and novel sets contain 64, 16 and 20 classes, respectively. In this study, the mini-ImageNet dataset was used in the meta-training stage to obtain the $M(\cdot|S_b)$ model. Therefore, 80 classes (base + validation) were used as the base and the remaining 20 classes (novel) as the validation set. Since this dataset is lightweight, it was possible to quickly obtain the results of many experiments.

3.1.2 BreakHis

The BreakHis dataset is a collection of 9109 microscopic images of breast tumor tissue taken from 82 individuals at different magnification levels (40x, 100x, 200x, and 400x). The images have dimensions of 700 pixels (height) and 460 pixels (width). The dataset has 8 different sub-classes of two main classes (malignant, benign), and examples are shown in Fig. 4. Here, as in mini-ImageNet, the BreakHis dataset was used in the meta-training stage to obtain the $M(\cdot|S_b)$ model. For this, 5 classes were used as the base and 3 classes as the validation purposes.

3.1.3 BUSI

The breast ultrasound images dataset (BUSI) which contains 780 images collected from 600 female patients between the ages of 25 and 75 in 2018, with a resolution of 500×500 pixels. These images are divided into three categories: normal (133), malignant (210), and benign

(437). Ground truth images are also available for use in image segmentation or detection. The BUSI dataset is an important tool for the classification of breast cancer in deep learning. Figure 5 illustrates some examples from the BUSI dataset. The BUSI dataset was used in the meta-testing stage to test the performance of the $M(\cdot|S_n)$ model obtained by fast adaptation. For this purpose, all 3 classes were used as the novel set.

In the context of meta-learning, splitting datasets into training and testing sets becomes challenging when dealing with datasets like BUSI, which have a small number of classes (normal, benign, and malignant). In conventional machine learning, we typically split the data into training and testing sets based on individual examples within each class, but in meta-learning, we need to allocate entire classes for training and testing, which can be problematic when dealing with a small number of classes.

The BUSI dataset, with its three classes, presents this challenge. To address it, we employed a cross-domain strategy that involved leveraging two distinct datasets for our meta-learning task. We used the BreakHis and mini-ImageNet datasets during the meta-training phase. In the meta-testing phase, we then applied our trained model to the actual BUSI dataset to evaluate its performance on the specific breast image analysis task we were interested in.

3.2 Used backbones

In few-shot learning models, backbone networks are used for feature-extracting functions. In this study, two backbone networks are used.

3.2.1 Conv-4

The configuration in [19, 23] was used for this backbone. There are four conv blocks in this architecture. Each conv block has 3×3 conv layers with 64 filters, batch normalization, ReLU nonlinearity, and 2×2 max-pooling.

3.2.2 ResNet architectures

ResNet (Residual Network) architectures were proposed by He et al. [34]. In this study, the proposed ResNet18, ResNet34, and ResNet50 architectures were used. These architectures differ from each other in terms of the number of layers, the number of conv layers in each residual block, and filter size.

ResNet18 is the simplest version of ResNet, and it consists of 18 layers, including the input and output layers. It has 11.69 million parameters. This architecture is relatively lightweight and fast to train, making it a good choice for tasks where computational resources are limited. ResNet34 is slightly more complex than ResNet18, and it

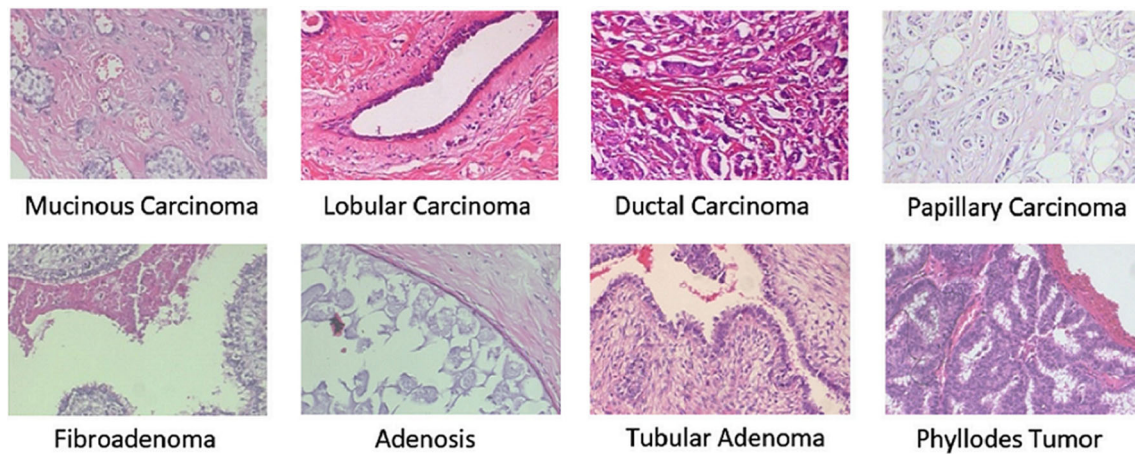


Fig. 4 Examples from 8 classes in the BreakHis dataset (images from [15])

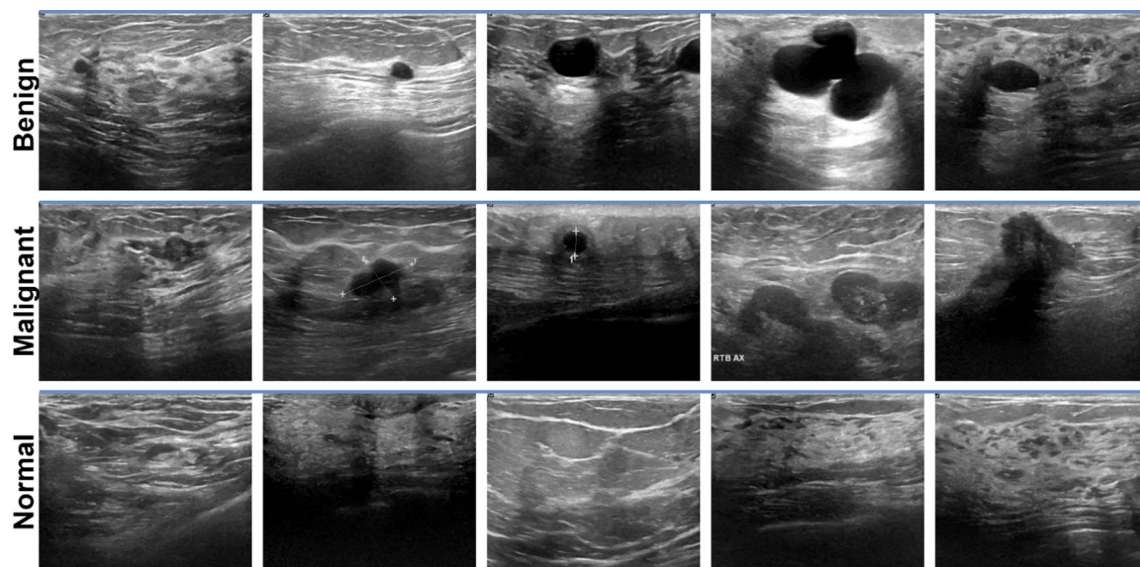


Fig. 5 Sample images from the breast ultrasound images (BUSI) dataset showing instances from each of the three classes: benign, malignant, and normal. “Benign” indicates an image displaying a

tumor, but one that is not caused by cancerous cells. “Malignant” describes an image showing a tumor with cancerous cells. “Normal” refers to an image that does not show any cancerous cells

has 34 layers, including the input and output layers. It has 21.8 million parameters. ResNet34 is more powerful and can learn more complex features than ResNet18. ResNet50 is the more complex version of ResNet, and it consists of 50 layers, including the input and output layers. It has 25.56 million parameters. ResNet50 is the most powerful of the three ResNet architectures and can learn the most

complex features. In Table 1, the input sizes of the used backbones are given.

ResNet has a deeper network structure than Conv-4. With skip connections, ResNet prevents gradient vanishing, allowing for smoother training and hierarchical feature learning. Its depth improves performance on limited datasets by capturing finer patterns. In addition, ResNet’s regularization effect minimizes overfitting, ideal for medical imaging tasks with limited data.

Table 1 Input image sizes of backbone networks

Backbone	Input image size
Conv-4	84×84
ResNet18, 34, 50	224×224

3.3 Experimental setup

In this study, all experiments were carried out on a personal computer with specific hardware specifications. The most recent version of Linux operating system, Ubuntu 22.04,

which is well-suited for deep learning platforms, was used. The computer was equipped with an Intel® Core™ i7-12700 K Processor, 64 GB of DDR5 RAM, and a NVIDIA RTX 3090 graphics card. The NVIDIA RTX 3090 graphics card has 10,496 CUDA cores, 328 tensor cores, and a 384-bit memory interface with 24 GB of GDDR6X memory. The programming language used was Python, and the PyTorch and NVIDIA CUDA Toolkit 11.7 frameworks were employed.

3.4 Evaluation

Evaluation or performance metrics are used to determine the generalization performance of the deep learning model. These metrics are commonly used in measuring the performance of the model and its ability to generalize. In this study, we employ a set of performance metrics including accuracy, sensitivity, specificity, and the F1-score for evaluating model performance. Accuracy signifies the proportion of accurate predictions among the total predictions made. Sensitivity measures the model's proficiency in detecting positive samples among the actual positive instances. Specificity quantifies the model's aptitude in recognizing negative samples within the entire pool of actual negatives. The F1-score serves as a consolidated metric, taking into account the model's capacity to accurately identify positive samples while also minimizing the occurrence of false positives (Eq. 7). In Eqs. 7–10, we use the following abbreviations to denote prediction outcomes: TP represents true positives, TN stands for true negatives, FP indicates false positives, and FN represents false negatives.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Sensitivity(Recall)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

$$\text{F1 - score} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (10)$$

3.5 Experimental details

All experiments were performed using PyTorch [35] in combination with the torchmeta [36] library. The torchmeta library is a collection of data loaders for few-shot learning tasks.

The MAML algorithm requires a second-order derivative to calculate the gradients of the θ parameters. Due to the high input size and time problems of second-order derivative calculations, the first-order MAML (FOMAML)

algorithm was proposed by Finn et al. [28]. The performance of the proposed algorithm on the Mini-ImageNet dataset was very close to the MAML algorithm. The FOMAML algorithm was utilized in this study due to its faster nature.

Four different convolutional network architectures are utilized as the basis for the feature extraction function (f_θ) in the evaluation of our methods: Conv-4, ResNet18, ResNet34, and ResNet50. All networks are trained for 100 epochs using stochastic gradient descent, with an initial learning rate of 0.01 and a batch size of 16. The learning rate is scaled by 0.1 after every 20 epochs. In this study, the algorithm followed for few-shot classification of the three-class BUSI dataset using meta-learning has four steps:

1. *Defining the set of meta-training tasks:* These tasks represent different 3-way classification problems. In this step, mini-ImageNet and BreakHis datasets were used. The separation of the datasets into base, validation, and novel sets is described in Sect. 3.1.
2. *Training the meta-learner:* A meta-learner was trained on the meta-training tasks. The goal of the meta-learner is to predict the best combination for a given task. In this step, ProtoNet and FOMAML were trained as meta-learner.
3. *Evaluating the meta-learner on the set of meta-validation tasks:* The meta-learner was evaluated on a set of validation tasks to assess its performance and to get the hyperparameters that work best. In this case, the optimal number of episodes for the best accuracy was determined through the use of the validation set. The hyperparameters of ProtoNet and FOMAML networks are given in Tables 2 and 3, respectively.
4. *Using the meta-learner to solve the novel task:* Given a new few-shot classification task, the meta-learner was used to predict the best combination of model architecture, optimizer, and hyperparameters. A model was then trained using these predictions and used to make predictions on the new task. In this step, the BUSI dataset was fine-tuned for 3-way classification.

For N -way classification in meta-training, N was chosen as 3 in all experiments. There are 3 classes in each episode. For the number of samples in the support set (K -shot),

Table 2 Prototypical networks hyperparameters

Hyperparameter	Options
Backbone	ResNets, Conv-4
Learning rate (LR)	0.0001
LR decay rate	0.9

Adapted from [37]

Table 3 FOMAML hyperparameters

Hyperparameter	Options
Backbone	ResNets, Conv-4
Outer loop LR	0.0004
Outer loop LR decay rate	0.65
Inner loop LR	0.005
Inner loop steps	10

Adapted from [37]

$K = 3$ and for the query set $q = 15$. In the meta-testing stage, the BUSI dataset was used as novel classes. In the meta-testing, N was chosen as 3 and experiments were performed separately for K values 1, 2, 5, and 10.

To be clear, basically two experiments were conducted: (i) mini-ImageNet in meta-training stage and BUSI in meta-testing (mini-ImageNet \rightarrow BUSI), (ii) BreakHis in meta-training stage and BUSI in meta-testing (BreakHis \rightarrow BUSI).

4 Results

In addition to experiments with meta-learning algorithms, transfer learning was also used as a baseline model. Transfer learning is a machine learning approach that allows a pre-trained model to be reused for another task [38]. Transferring information from learned tasks to new tasks with few examples increases classification performance [39].

Here, a pre-trained ResNet18 network with ImageNet dataset was used. Then, the obtained model was fine-tuned on the BUSI dataset with the transfer learning approach. For this, the images were converted to the input size of the ResNet18 network. The hyperparameters are as follows: learning rate 0.001, minibatch size 16 and 300 epochs. This model was used as a baseline for comparison purposes and achieved an accuracy of 0.831.

Tables 4 and 5 present the accuracy achieved in the 3-way classification of the BUSI dataset using different backbones and two meta-learning algorithms for various shot values during meta-training with the mini-ImageNet

and BreakHis datasets, respectively. When examining the results, we can observe that across both ProtoNet and FOMAML, the choice of backbone architecture significantly impacts accuracy. As we move from 1-shot to 10-shot, accuracy consistently improves for all backbones and algorithms. This suggests that having more labeled examples (shots) for each class results in better classification performance. ProtoNet consistently outperforms FOMAML in terms of accuracy across all shot values and backbones. This indicates that, in this context, ProtoNet is more effective in learning a generalizable classification model for the BUSI dataset. The highest accuracy achieved in these experiments are 0.882 and 0.889, respectively, which corresponds to using the ResNet50 backbone with ProtoNet and a 10-shot setting. This configuration results in the most accurate classification of breast ultrasound images in the 3-way classification task.

To provide a clearer overview of the results, we have computed the average accuracy values for the different methods, and these consolidated results are presented in Fig. 6. It appears that ProtoNet outperforms FOMAML in accuracy, especially as seen in (Fig. 7, top). The implementation simplicity and remarkable performance of ProtoNet indicate its potential for accurately classifying the BUSI dataset with few examples.

During the meta-training stage, both methods utilized mini-ImageNet and BreakHis datasets. In ProtoNet, the performance obtained with the BreakHis dataset outperforms mini-ImageNet, especially after 5-shots. This is observed in the FOMAML method after 2-shots (Fig. 7, left). As the number of K increases, the BreakHis dataset becomes more successful than mini-ImageNet after a certain value. This may be due to the fact that the BreakHis and BUSI datasets are similar and at the same time fine-grained.

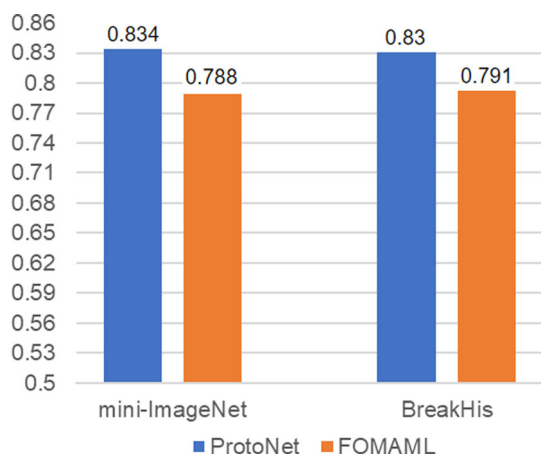
It has been established that a high degree of similarity between the source and target datasets enhances the accuracy of few-shot classification [23, 28, 30]. In general, deeper and more complex backbones, such as ResNet models, tend to yield higher accuracy compared to Conv-4, indicating that deeper models are better at capturing the underlying patterns and features in the BUSI dataset. The improved performance of ResNet models with higher complexity has led to longer training times and greater

Table 4 Accuracy rates for the 3-way classification of BUSI using mini-ImageNet dataset in meta-training

Backbone	ProtoNet				FOMAML			
	1-shot	2-shot	5-shot	10-shot	1-shot	2-shot	5-shot	10-shot
Conv-4	0.801	0.804	0.824	0.841	0.750	0.755	0.782	0.812
ResNet18	0.815	0.821	0.835	0.866	0.758	0.766	0.789	0.825
ResNet34	0.815	0.825	0.842	0.880	0.762	0.781	0.789	0.834
ResNet50	0.820	0.829	0.850	0.882	0.769	0.795	0.810	0.845

Table 5 Accuracy rates for the 3-way classification of BUSI using BreakHis dataset in meta-training

Backbone	ProtoNet				FOMAML			
	1-shot	2-shot	5-shot	10-shot	1-shot	2-shot	5-shot	10-shot
Conv-4	0.790	0.798	0.812	0.848	0.741	0.755	0.784	0.820
ResNet18	0.802	0.810	0.837	0.870	0.745	0.771	0.801	0.825
ResNet34	0.805	0.814	0.845	0.883	0.754	0.786	0.815	0.842
ResNet50	0.808	0.814	0.858	0.889	0.756	0.801	0.821	0.854

**Fig. 6** Average accuracy scores of the methods across the datasets

computational resource requirements. Thus, there needs to be a tradeoff between performance and cost. Efforts should be made to make these networks more efficient for low-powered devices, such as mobile phones.

Since accuracy consistently increases with higher K values, we calculated sensitivity (recall), specificity, and the F1-score using the maximum K value (10-shot) in Tables 6 and 7. Also, according to these tables, ProtoNet performs better than FOMAML in terms of sensitivity, specificity, and F1 score in the 3-way 10-shot classification of the BUSI dataset, regardless of the backbone and dataset used.

The comparison between meta-learning methods and the transfer learning approach utilizing pre-trained ResNet18 is evident in Fig. 7. For low values of K, transfer learning outperforms the other methods. It also generally outperformed the FOMAML method. However, when K is larger than 10, the meta-learning methods begin to surpass the

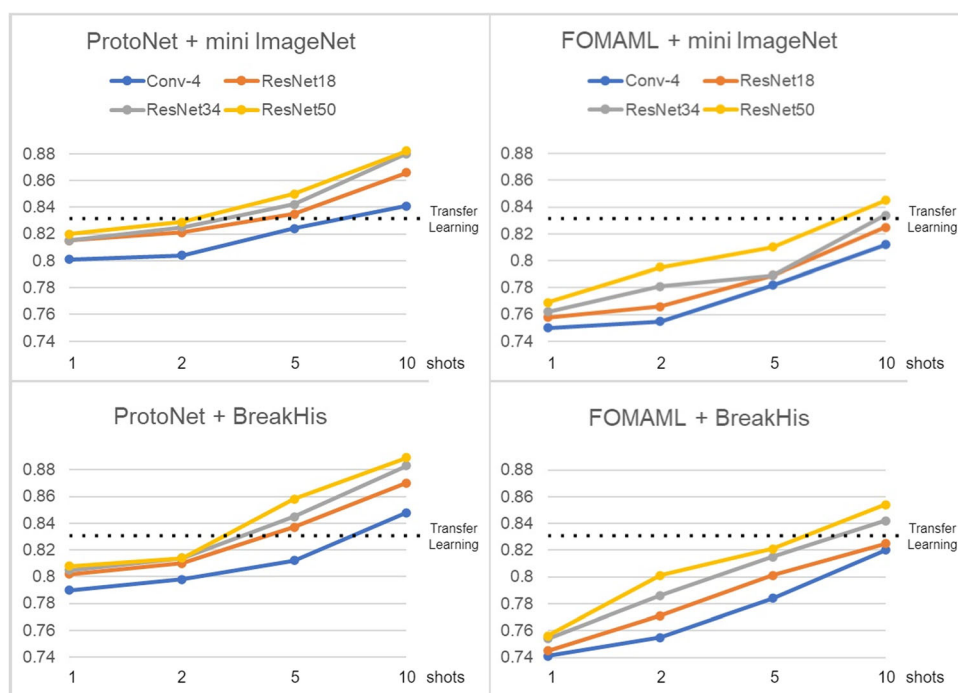
Fig. 7 The effectiveness of meta-learning methods with various configurations improves as the number of examples increases. Additionally, in some cases, the transfer learning approach outperforms meta-learning

Table 6 The sensitivity, specificity, and F1-score for the 3-way 10-shot classification of BUSI using the mini-ImageNet dataset during meta-training

Backbone	ProtoNet			FOMAML		
	Sensitivity	Specificity	F1-score	Sensitivity	Specificity	F1-score
Conv-4	0.820	0.855	0.838	0.781	0.840	0.825
ResNet18	0.832	0.850	0.855	0.790	0.840	0.832
ResNet34	0.875	0.889	0.861	0.810	0.855	0.846
ResNet50	0.880	0.899	0.872	0.815	0.872	0.855

Table 7 The sensitivity, specificity, and F1-score for the 3-way 10-shot classification of the BUSI using the BreakHis dataset during meta-training

Backbone	ProtoNet			FOMAML		
	Sensitivity	Specificity	F1-score	Sensitivity	Specificity	F1-score
Conv-4	0.835	0.855	0.848	0.815	0.840	0.829
ResNet18	0.842	0.855	0.855	0.825	0.845	0.845
ResNet34	0.885	0.898	0.877	0.825	0.865	0.850
ResNet50	0.888	0.910	0.892	0.845	0.879	0.870

transfer learning approach. If stronger models are used in place of ResNet18, transfer learning may perform better than certain meta-learning setups. Nevertheless, in the absence of pre-trained models required in transfer learning, meta-learning is a viable option [40].

5 Discussions

In the realm of medical applications, the integration of meta-learning techniques has witnessed significant traction, particularly in the domain of medical image segmentation. Noteworthy studies conducted by Sun et al. [13], Feng et al. [14], and Hansen et al. [16] exemplify the burgeoning interest in harnessing the potential of meta-learning to bolster the accuracy and efficacy of medical image segmentation tasks. These pioneering investigations have illuminated the promising prospects of employing meta-learning in medical image analysis.

Specifically, Singh et al. [15] made substantial strides by introducing a gradient-based Reptile meta-learning algorithm, thoughtfully coupled with innovative image augmentation strategies. Their meticulously crafted approach was synergistically fused with the Conv-4 backbone architecture, rendering impressive outcomes. Their diligent efforts resulted in an outstanding accuracy rate of 86.12% when confronted with the formidable challenges posed by the BreakHis dataset, a benchmark task that entails 10-shot classification. This accomplishment underscores the capability of meta-learning in substantially enhancing medical image analysis tasks, exemplifying its potential to excel in complex and demanding contexts.

In a parallel vein, Ali et al. [17] embarked on a mission to tackle breast cancer classification, particularly the binary discrimination between benign and malignant cases, using

the BUSI dataset. Their strategy hinged upon the utilization of ensemble learning, a technique that amalgamates the outputs of multiple convolutional neural networks (CNNs). The outcome of their collective endeavor yielded an impressive accuracy rate of 90%, underscoring the proficiency of ensemble learning in enhancing the classification accuracy of breast cancer lesions.

However, it is imperative to discern the nuanced distinctions between these laudable endeavors and our present study. While Ali et al. [17] focused on a binary classification task, our research takes a more intricate route by addressing the three-class classification problem. In doing so, we not only differentiate between benign and malignant lesions but also incorporate a 'normal' class into the classification schema. This augmentation of complexity necessitates a more nuanced evaluation, and direct comparisons between these studies may not be entirely equitable due to the inherent differences in the classification tasks.

In summary, our study reaffirms the promising potential of meta-learning techniques, mirroring the strides made in medical image analysis by researchers such as Singh et al. [15] and Ali et al. [17]. However, it distinguishes itself by undertaking a more intricate and challenging three-class classification task, further contributing to the growing body of knowledge in the field of medical image analysis.

6 Conclusion

An approach that utilizes meta-learning for medical image classification under limited data conditions as a few-shot learning task was presented in this research. The few-shot classification with meta-learning approach typically involves using two datasets from the same distribution as

the source and target domains. For training, a portion of the same dataset is used as the base classes while the remaining part is designated as the novel classes. However, in this study, the BUSI dataset, which has only three classes, could not be divided into base and novel sets. To overcome this issue, we used other datasets (mini-ImageNet and BreakHis) for meta-training and the BUSI dataset for meta-testing in a cross-domain approach. This not only addressed the problem of data scarcity but also solved the low-class problem of the BUSI dataset. The similarity between the BreakHis and BUSI datasets also contributed to the success of the few-shot classification. To the best of our knowledge, this study is the first attempt to apply few-shot classification using meta-learning on the BUSI dataset. The methodology we used can be adapted to other datasets with similar problems.

In the study, two meta-learning algorithms were used: metric-based prototypical networks (ProtoNet) and optimization-based first-order model agnostic meta-learning (FOMAML). The results of the experiments showed that ProtoNet outperformed FOMAML. Furthermore, the use of ResNet models as the backbone networks for feature extraction was more successful than using a four-layer convolutional model.

In our upcoming efforts, our focus extends to an in-depth exploration of various meta-learning methods and datasets to strengthen the adaptability and robustness of our approach. This includes an investigation of prominent meta-learning algorithms such as relation networks, matching networks, and gradient-based meta-learning. Our rationale lies in the search for more nuanced strategies to effectively navigate the intricacies of limited medical datasets. By exploring these alternative methodologies, we aim to fine-tune our approach and uncover innovative ways to address the inherent challenges posed by limited data availability in medical imaging. At the same time, our plan includes an extensive evaluation of various datasets that offer distinctive complexities and reflect different medical imaging scenarios.

In addition, our future endeavors include an in-depth exploration of data augmentation strategies to mitigate overfitting risks and enhance learning capabilities from sparse datasets. We aim to explore techniques such as Mixup, CutMix, and Random Erasing, among others, to increase the robustness of our models, especially in the face of highly imbalanced or limited data scenarios. This research seeks to exploit the potential of augmentation strategies to enrich the model's learning patterns, promote generalization, and mitigate the risks of overfitting. Additionally, within metric-based models, the exploration of various distance metrics such as cosine similarity, Euclidean distance, and Mahalanobis distance is critical. This strategic exploration aims to enhance the discriminative

power of our models and potentially guide advances in classification accuracy by optimizing the model's ability to decipher intricate patterns within medical images.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK).

Data availability The datasets analyzed during the current study are available in their own Kaggle repositories.

Declarations

Conflict of interest The authors did not receive support from any organization for the submitted work and declare they have no financial interests. Also, the authors have no conflicts of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Budd J, Miller BS, Manning EM, Lamos V, Zhuang M, Edelstein M, Rees G, Emery VC, Stevens MM, Keegan N, Short MJ, Pillay D, Manley E, Cox JJ, Heymann D, Johnson AM, McKendry RA (2020) Digital technologies in the public-health response to COVID-19. *Nat Med* 26(8):1183–1192
2. Zhang L, Tan J, Han D, Zhu H (2017) From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug Discov Today* 22(11):1680–1685
3. Feng Y, Chen J, Xie J, Zhang T, Lv H, Pan T (2022) Meta-learning as a promising approach for few-shot cross-domain fault diagnosis: algorithms, applications, and prospects. *Knowl-Based Syst* 235:107646
4. Ma J, Fong SH, Luo Y, Bakkenist CJ, Shen JP, Mourragui S, Wessels LF, Hafner M, Sharan R, Peng J, Ideker T (2021) Few-shot learning creates predictive models of drug response that translate from high-throughput screens to individual patients. *Nat Cancer* 2(2):233–244
5. Mishra AK, Roy P, Bandyopadhyay S, Das SK (2021) Breast ultrasound tumour classification: a machine learning—radiomics based approach. *Expert Syst* 38(7):1–12. <https://doi.org/10.1111/essy.12713>
6. Seely JM, Alhassan T (2018) Screening for breast cancer in 2018—what should we be doing today? *Seely Alhassan Curr Oncol*. <https://doi.org/10.3747/co.25.3770>
7. Zhang G, Zhao K, Hong Y, Qiu X, Zhang K, Wei B (2021) SHA-MTL: soft and hard attention multi-task learning for automated breast cancer ultrasound image segmentation and classification.

- Int J Comput Assist Radiol Surg 16(10):1719–1725. <https://doi.org/10.1007/s11548-021-02445-7>
8. Al-Dhabyani W, Gomaa M, Khalel H, Fahmy A (2020) Dataset of breast ultrasound images. Data Brief 28:104863. <https://doi.org/10.1016/j.dib.2019.104863>
 9. Gündüz MŞ, Işık G (2023) A new YOLO-based method for social distancing from real-time videos. Neural Comput Appl. <https://doi.org/10.1007/s00521-023-08556-3>
 10. Işık G (2023) Improving plant disease recognition through gradient-based few-shot learning with attention mechanisms. J Inst Sci Technol 13(3):1482–1495
 11. Zhang Y, Qiu L, Ren Y, Cheng Z, Li L, Yao S, Zhang C, Luo Z, Lu H (2022) A meta-learning approach to improving radiation response prediction in cancers. Comput Biol Med 150:106163
 12. Ouyang C, Biffi C, Chen C, Kart T, Qiu H, Rueckert D (2022) Self-supervised learning for few-shot medical image segmentation. IEEE Trans Med Imaging 41(7):1837–1848. <https://doi.org/10.1109/TMI.2022.3150682>
 13. Sun L, Li C, Ding X, Huang Y, Chen Z, Wang G, Yu Y, Paisley J (2022) Few-shot medical image segmentation using a global correlation network with discriminative embedding. Comput Biol Med. <https://doi.org/10.1016/j.combiomed.2021.105067>
 14. Feng R, Zheng X, Gao T, Chen J, Wang W, Chen DZ, Wu J (2021) Interactive few-shot learning: limited supervision, better medical image segmentation. IEEE Trans Med Imaging 40(10):2575–2588. <https://doi.org/10.1109/TMI.2021.3060551>
 15. Singh R, Bharti V, Purohit V, Kumar A, Singh AK, Singh SK (2021) MetaMed: few-shot medical image classification using gradient-based meta-learning. Pattern Recogn 120:108111
 16. Hansen S, Gautam S, Jenssen R, Kampffmeyer M (2022) Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Med Image Anal. <https://doi.org/10.1016/j.media.2022.102385>
 17. Ali MD, Saleem A, Elahi H, Khan MA, Khan MI, Yaqoob MM, Farooq Khattak U, Al-Rasheed A (2023) Breast cancer classification through meta-learning ensemble technique using convolution neural networks. Diagnostics 13(13):2242
 18. Hospedales TM, Antoniou A, Micaelli P, Storkey AJ (2021) Meta-learning in neural networks: a survey. IEEE Trans Pattern Anal Mach Intell 44(9):5149–5169
 19. Wang Y, Yao Q, Kwok JT, Ni LM (2020) Generalizing from a few examples: a survey on few-shot learning. ACM Comput Surv (csur) 53(3):1–34
 20. Vinyals O, Blundell C, Lillicrap T, Kavukcuoglu K, Wierstra D (2016) Matching networks for one shot learning. In Advances in neural information processing systems, pp. 3630–3638.
 21. Triantafyllou E, Zhu T, Dumoulin V, Lamblin P, Evci U, Xu K, Goroshin R, Gelada C, Swersky K, Manzagol PA, Larochelle H (2020) Meta-dataset: a dataset of datasets for learning to learn from few examples. In: International conference on learning representations.
 22. Laenen S, Bertinetto L (2021) On episodes, prototypical networks, and few-shot learning. Adv Neural Inf Process Syst 34:24581–24592
 23. Koch G, Zemel R, Salakhutdinov R (2015) Siamese neural networks for one-shot image recognition. In ICML deep learning workshop, vol. 2.
 24. Snell J, Swersky K, Zemel R (2017) Prototypical networks for few-shot learning. In: Advances in neural information processing systems, 30.
 25. Sung F, Yang Y, Zhang L, Xiang T, Torr PH, Hospedales TM (2018) Learning to compare: relation network for few-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1199–1208.
 26. Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T (2016) Meta-learning with memory-augmented neural networks. In: International conference on machine learning, pp. 1842–1850. PMLR.
 27. Munkhdalai T, Yu H (2017) Meta networks. In: International conference on machine learning, pp. 2554–2563. PMLR.
 28. Ravi S, Larochelle H (2016) Optimization as a model for few-shot learning. In: International conference on learning representations.
 29. Finn C, Abbeel P, Levine S (2017) Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning, pp. 1126–1135. PMLR.
 30. Nichol A, Achiam J, Schulman J (2018) On first-order meta-learning algorithms. arXiv preprint [arXiv:1803.02999](https://arxiv.org/abs/1803.02999).
 31. Chen WY, Liu YC, Kira Z, Wang YCF, Huang JB (2019) A closer look at few-shot classification. In: International conference on learning representations.
 32. Spanhol FA, Oliveira LS, Petitjean C, Heutte L (2015) A dataset for breast cancer histopathological image classification. IEEE Trans Biomed Eng 63(7):1455–1462
 33. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg AC, Fei-Fei L (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115(3):211–252
 34. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034.
 35. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Chintala S (2019) Pytorch: an imperative style, high-performance deep learning library. In: Advances in neural information processing systems, 32.
 36. Deleu T, Würfl T, Samiei M, Cohen JP, Bengio Y (2019) Torchmeta: a meta-learning library for PyTorch.
 37. Dumoulin V, Houlsby N, Evci U, Zhai X, Goroshin R, Gelly S, Larochelle H (2021) Comparing transfer and meta learning approaches on a unified few-shot classification benchmark. arXiv preprint [arXiv:2104.02638](https://arxiv.org/abs/2104.02638).
 38. Morid MA, Borjali A, Del Fiol G (2021) A scoping review of transfer learning research on medical image analysis using ImageNet. Comput Biol Med 128:104115
 39. Mehmood A, Yang S, Feng Z, Wang M, Ahmad AS, Khan R, Maqsood M, Yaqub M (2021) A transfer learning approach for early diagnosis of Alzheimer's disease on MRI images. Neuroscience 460:43–52
 40. Pons J, Serra J, Serra X (2019) Training neural audio classifiers with few data. In: ICASSP 2019–2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 16–20. IEEE.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.