Artificial Intelligence In Medicine 156 (2024) 102949

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

A systematic review of few-shot learning in medical imaging

Eva Pachetti [a,b,*], Sara Colantonio [a]

[a] Institute of Information Science and Technologies "Alessandro Faedo", National Research Council of Italy (ISTI-CNR), via Giuseppe Moruzzi 1, Pisa, 56124, PI, Italy

[b] Department of Information Engineering, University of Pisa, via Girolamo Caruso 16, Pisa, 56122, PI, Italy

ARTICLE INFO

Keywords:
Few-shot learning
Medical imaging
Systematic review

ABSTRACT

The lack of annotated medical images limits the performance of deep learning models, which usually need large-scale labelled datasets. Few-shot learning techniques can reduce data scarcity issues and enhance medical image analysis speed and robustness. This systematic review gives a comprehensive overview of few-shot learning methods for medical image analysis, aiming to establish a standard methodological pipeline for future research reference. With a particular emphasis on the role of meta-learning, we analysed 80 relevant articles published from 2018 to 2023, conducting a risk of bias assessment and extracting relevant information, especially regarding the employed learning techniques. From this, we delineated a comprehensive methodological pipeline shared among all studies. In addition, we performed a statistical analysis of the studies' results concerning the clinical task and the meta-learning method employed while also presenting supplemental information such as imaging modalities and model robustness evaluation techniques. We discussed the findings of our analysis, providing a deep insight into the limitations of the state-of-the-art methods and the most promising approaches. Drawing on our investigation, we yielded recommendations on potential future research directions aiming to bridge the gap between research and clinical practice.

Contents

* Corresponding author at: Institute of Information Science and Technologies "Alessandro Faedo", National Research Council of Italy (ISTI-CNR), via Giuseppe Moruzzi 1, Pisa, 56124, PI, Italy.
E-mail addresses: eva.pachetti@isti.cnr.it (E. Pachetti), sara.colantonio@isti.cnr.it (S. Colantonio).

https://doi.org/10.1016/j.artmed.2024.102949
Received 19 September 2023; Received in revised form 16 July 2024; Accepted 13 August 2024
Available online 16 August 2024

1. Introduction

1.1. Rationale

The demand for deep learning (DL) models that can generalize well and achieve high performance with limited data is constantly increasing. Few-Shot Learning (FSL) plays a crucial role in addressing this challenge by enabling models to learn from only a few examples, mimicking the way humans naturally learn. In contrast to the typical practice in DL, which involves pre-training models on large datasets and fine-tuning them on specific tasks, FSL allows models to learn effectively with minimal labelled examples. Among the most prominent models that have successfully addressed this limitation is GPT-3 [1]. Unlike traditional models, GPT-3 does not require fine-tuning on specific tasks. Instead, it leverages FSL during inference by being exposed, for each task, to a few demonstrations for conditioning without updating its parameters [1]. This approach allows GPT-3 to perform various tasks with just a few examples, showcasing the power of FSL in natural language processing.

FSL finds one of its most crucial applications in medical image analysis for several compelling reasons. Firstly, medical datasets are often limited in size due to privacy concerns, high data acquisition costs, and the laborious process of expert annotation. FSL enables models to achieve robust generalization with minimal labelled examples, making it possible to develop effective medical imaging solutions even with scarce data. Secondly, FSL alleviates the burden of manual annotation by requiring only a few annotated examples for each new task or medical condition. This capability streamlines the annotation process and supports clinicians in their time-consuming tasks. Moreover, FSL proves particularly valuable for handling rare medical conditions where acquiring sufficient data for traditional DL approaches may be impractical. Leveraging knowledge from more prevalent diseases, FSL empowers models to adapt to new and rare cases with limited examples. Furthermore, the medical field constantly encounters new diseases, conditions, and imaging modalities. FSL enables medical imaging models to swiftly adapt and learn from a few examples of these novel tasks, facilitating their seamless integration into clinical practice. Finally, FSL holds potential in personalized medicine, where models must rapidly adapt to analyse images from individual patients. With just a few examples from each patient, FSL allows the model to tailor its analysis based on specific patient characteristics, enhancing the precision of medical diagnoses and treatments.

Despite FSL finds applications also in non-medical fields, the medical domain presents unique challenges and requirements. Data scarcity, variability, and the necessity for expert annotations make FSL particularly valuable in medical applications. The critical nature of medical decisions demands high accuracy, interpretability, and stringent ethical considerations, as errors can have severe consequences. This contrasts with non-medical fields where data is generally more abundant, expertise requirements are lower, and ethical stakes vary. Moreover, medical data is highly heterogeneous and complex, requiring specialized knowledge to interpret. These unique characteristics highlight the complexity and high stakes of deploying FSL in the medical field, necessitating specialized approaches and interdisciplinary collaboration. These are the reasons why we believe that a systematic and comprehensive review of the current state-of-the-art (SOTA) in FSL for medical imaging is essential to provide researchers with information on the most promising FSL methods and helping to identify areas where applications have not been sufficiently explored and the most promising approaches that deserve further investigation. This knowledge would ultimately accelerate the integration of FSL approaches into real-world clinical practice.

1.2. Objectives

This systematic review aims to present a comprehensive overview of the SOTA in FSL techniques applied to medical imaging, ultimately providing the reader with a standard methodological pipeline shared across all the examined studies to be used as a reference for further research in this field. While many FSL reviews primarily concentrate on the broader domain of computer vision [2–4], existing reviews addressing FSL for medical imaging [5,6] offer only a limited scope within this domain. These reviews include only a modest number of papers, focusing solely on classification and segmentation outcomes, along with a restricted number of clinical applications. Furthermore, they lack statistical analysis of collected results, as well as a discussion on the risk of bias and applicability concerns. Overall, they do not provide an overview of the shared aspects among the investigated techniques.

With this work, we aim to collect and highlight all the studies that, in the authors' opinion, make substantial and genuine contributions to this domain. Specifically, we focus on the primary applications of DL in medical imaging, namely segmentation, classification, and registration. For each study, we provide readers with risk of bias and applicability analyses, enabling a critical evaluation of their actual scientific impact. The objective is to showcase innovative techniques with proven effectiveness and robustness and leverage them to establish a standardized methodological pipeline applicable to various medical

outcomes. This pipeline can serve as a starting point for researchers to develop effective approaches in this area. In addition to our primary objective, this study aims to provide the following insights:

- **Presentation of studies distribution by outcome.** We highlight the distribution of studies across three outcomes: segmentation, classification and registration.
- **Presentation of studies and results distribution by clinical task.** For each outcome, we analyse the distribution of studies and provide a statistical analysis of their results w.r.t. the clinical task investigated.
- **Presentation of studies and results distribution by meta-learning technique.** Throughout our analysis, we provide special attention to the meta-learning domain as the most utilized approach to tackle FSL issues. Specifically, for each outcome, we provide a distribution analysis of the meta-learning methods used, as well as a statistical analysis of the studies' results grouped by meta-learning technique.
- **Additional analyses.** In addition to the core analyses, we explore some secondary aspects, such as data usage information, the most commonly used imaging modalities, and the model robustness assessment methods employed. To not affect the manuscript's readability, we will provide such additional information solely in textual form, prioritizing visual representations only for the primary analyses.

In the following, we outline the structure of the manuscript:

- We begin with a theoretical introduction to FSL focusing on meta-learning, with a brief overview of Zero-shot Learning (ZSL). Right after, we delve into the most popular meta-learning methods for FSL, highlighting their potential application also to the ZSL domain.
- We describe the literature search methods we employed, including the eligibility criteria and the databases analysed. We also illustrate the main features extracted from each study and the synthesis methods applied.
- We show the results of our analyses w.r.t. the objectives of our work as well as the analyses performed regarding the risk of bias and applicability concerns.
- We discuss our key findings, including highlighting the limitations of current SOTA approaches and exploring potential avenues for future research to guide further exploration in this field.
- We summarize the key findings from our work and, based on them, we finally draw conclusions.

2. Theoretical background

FSL has been receiving significant attention, especially since the advent of meta-learning. This section delves into the theoretical foundations of FSL, focusing on meta-learning methodologies as the most popular approach to tackling FSL problems. By exploring meta-learning, we aim to shed light on its core concepts, techniques, and applications specific to the FSL domain. Furthermore, we offer a brief focus on ZSL and Generalized ZSL (GZSL) domains, highlighting the utilization of several meta-learning methodologies within this realm as well.

Meta-learning, a.k.a. *learning-to-learn*, is a powerful paradigm that empowers models to rapidly adapt and generalize to new tasks with minimal training examples. Unlike the traditional training scheme where models are trained on data, meta-learning operates on a higher level by training models on *tasks* or *episodes*. Thus, this form of training is often referred to as *episodic training*. During training, the meta-learning model is exposed to multiple episodes, each comprising a few examples of a specific task. As a result, the model acquires transferable knowledge and learns to identify common patterns. Consequently, when faced with a new episode during the testing phase, the model can efficiently leverage its acquired meta-knowledge to make accurate predictions, even with limited examples. The combination of FSL and meta-learning has shown remarkable results, especially where data availability is limited or when handling novel tasks. Below, we provide a more formal formulation of the meta-learning framework, as outlined in [7], which we will leverage to describe the following SOTA methods. The inner algorithm (f) solves the task i by updating the model parameters θ to θ_i': this phase is called *base learning*. During the *meta-learning* phase, an outer algorithm updates the model parameters θ across all the tasks according to an outer objective; the updating entity is regulated by a meta-step hyperparameter β. As pointed out by Hospedales et al. [7], several classic algorithms, such as hyperparameter optimization, can match this definition; however, what actually defines a modern meta-learning algorithm is the definition of an outer objective with the simultaneous optimization of the inner algorithm w.r.t. to this objective.

A meta-learning training procedure consists of a *meta-training* and a *meta-testing* stage. During meta-training, a set of *source* tasks is sampled from the distribution of the tasks $P(\tau)$. Each source task is composed by a *support* ($S = \{(x_j, y_j)\}_{j=1}^{k}$) and a *query* set ($Q = \{(\hat{x}_j, \hat{y}_j)\}_{j=1}^{k}$), which corresponds to training and validation data in a classical training paradigm, respectively. The goal is to minimize a loss function $\mathcal{L}$ on the query samples conditioned to the support set. During the meta-testing stage, several *target* tasks are sampled as well. In this phase, the base learner is trained on the previously unseen tasks by exploiting the *meta-knowledge* learned during the meta-training phase. To speak about FSL, the number of examples for each class within the support set should be typically less than 10. Fig. 1 illustrates the meta-learning training process based on the N-way K-shot paradigm in a generic context where the model's task involves classifying medical images according to the depicted organ.

In scenarios with no shared classes between training and test sets or, more generally, where test samples can be from both training (seen) and unseen classes, we talk about ZSL [8] and GZSL [9], respectively. For addressing ZSL/GZSL, the most popular approach is to transfer knowledge from seen classes to unseen classes by sharing attributes [10]. In this sense, ZSL/GZSL is often tackled using different methods w.r.t. FSL, like learning intermediate attribute classifiers [11,12] or learning a mixture of seen class proportions [13,14] or adopting a direct approach as learning compatibility functions [15,16]. Given that a comprehensive discussion of the commonly used methods for addressing ZSL is beyond the scope of this paper, we highlight, in the description of meta-learning for FSL approaches, the ones that have also found utility for ZSL/GZSL.

Since meta-knowledge can manifest in various forms, such as initial parameters, optimization strategy, and learning algorithm [7], we adopt the taxonomy proposed by [2] to categorize meta-learning algorithms for FSL into three categories: *Initialization-based*, *Metric learning-based*, and *Hallucination-based* methods. A representation of the followed taxonomy is provided Fig. 2. In the subsequent paragraphs, we provide an overview of the most renowned algorithms developed within each category.

2.1. Initialization-based methods

Initialization-based methods refer to a class of approaches that focus on learning effective initializations for model parameters, i.e. *learning to initialize*. The model learns to adjust its parameters or initialization to better adapt to each task during the meta-training phase. The goal is to find parameter initializations that can be readily fine-tuned with only a few examples from a new episode, facilitating rapid generalization. The following are some of the most relevant SOTA algorithms that belong to the category of initialization-based methods in meta-learning.

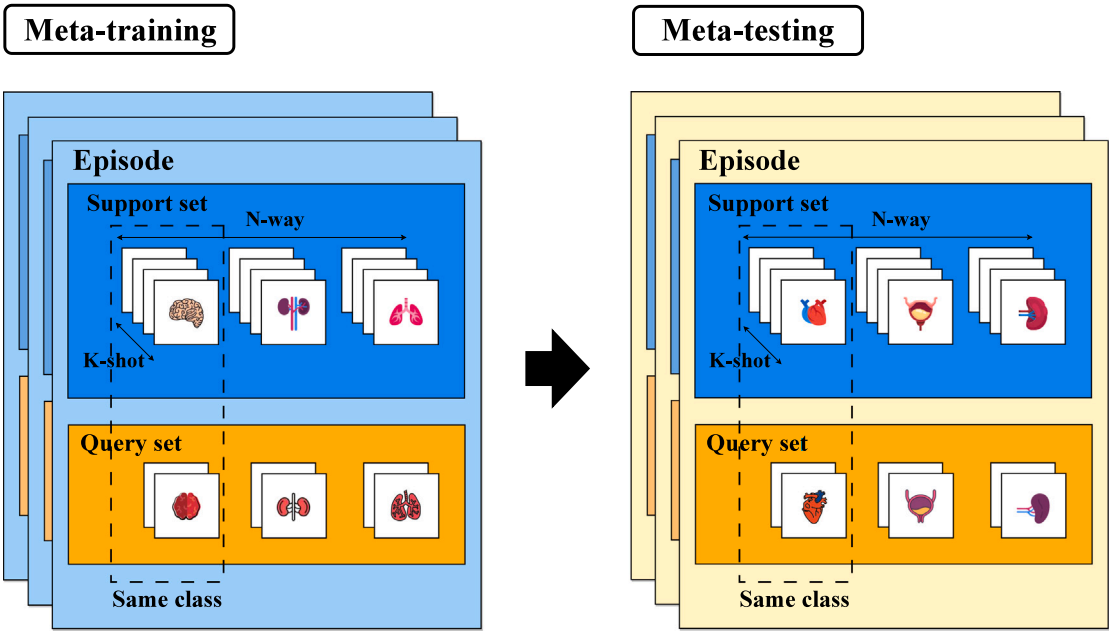


Fig. 1. N-way K-shot paradigm representation.

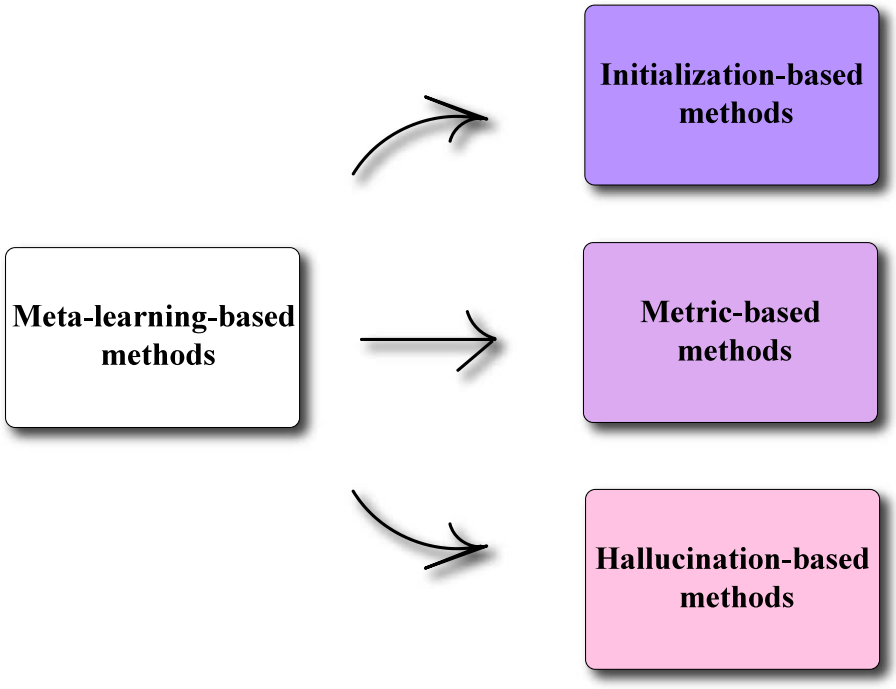


Fig. 2. Meta-learning methods taxonomy.

2.1.1. Model-agnostic meta-learning

In their paper, Finn et al. [17] present Model-Agnostic Meta-Learning (MAML), a meta-learning framework applicable to any model trained with gradient descent. The objective of MAML is to enable the model f_θ to adapt quickly to new tasks τ_i by finding the model parameters most sensitive to changes in the episode. In particular, the model's parameters are updated to θ_i' for a new task τ_i as follows:

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}\tau_i(f\theta) \tag{1}$$

where α is the step size of the gradient descent and $\mathcal{L}$ the loss function. The overall meta-objective is to minimize the loss across all tasks $P(\tau)$:

$$\min_\theta \sum_{\tau_i \sim P(\tau_i)} \mathcal{L}\tau_i(f\theta_i') \tag{2}$$

The model parameters are updated through stochastic gradient descent (SGD) as follows:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\tau_i \sim P(\tau_i)} \mathcal{L}\tau_i(f\theta_i') \tag{3}$$

Since computing gradients for both task and meta objectives can be computationally expensive, the authors also explored a first-order approximation (FOMAML) that omits the second derivatives. Surprisingly, their results showed that FOMAML performed almost as well

as the original MAML. A possible explanation for this observation is that certain ReLU neural networks are nearly linear locally, causing the second derivatives to be close to zero in practice.

In [18], the authors proposed exploiting MAML to perform GZSL. Specifically, they use the meta-learning framework to train a generative adversarial network conditioned on class attributes that can generate novel class samples. A key difference with classical MAML is that, for each task, the classes between the training and validation phases are disjoint.

2.1.2. Reptile

In their work Nichol, Achiam and Schulman [19] propose a variant of FOMAML called Reptile. Similar to MAML and FOMAML, Reptile updates the global parameters to create task-specific parameters. However, instead of following Eq. (3), Reptile uses the following update rule for N tasks:

$$\theta \leftarrow \theta + \beta \frac{1}{N} \sum_{i=1}^{N} (\theta_i' - \theta) \tag{4}$$

Here, the difference $(\theta_i' - \theta)$, instead of being updated towards θ, is treated as a gradient and can be utilized with an adaptive algorithm like Adam for the final update. This update rule is computationally more efficient compared to the complex second-order differentiation used in MAML. This efficiency makes Reptile easier to implement and can lead to faster training times.

2.1.3. Optimization as long short-term memory network cell update

In their work, Ravi and Larochelle [20] propose a meta-learning approach based on Long Short-Term Memory (LSTM) networks, aiming to learn an optimization algorithm for training another model in an FSL manner. The main idea stems from the observation that the parameter updating law in a generic gradient descent network is similar to the update equation of the cell state in an LSTM [21]:

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c_t} \tag{5}$$

where $f_t = 1$, $c_{t-1} = f_\theta$, $i_t = \alpha$, and $\widetilde{c_t} = -\nabla\theta\mathcal{L}$. Exploiting this relationship, the learning rate can be formulated as a function of the current parameter value θ, the current gradient $\nabla_\theta\mathcal{L}$, the current loss $\mathcal{L}$, and the previous learning rate α_{t-1}. By doing so, the meta-learner can effectively control the learning rate value, enabling the model to learn quickly. During training, while iterating on the episode's training set, the LSTM meta-learner receives the values $(\nabla_\theta\mathcal{L}\tau_i, \mathcal{L}\tau_i)$ from the model for each task τ_i. Subsequently, it generates the updated parameters θ_i' as its output. This process is repeated for a predefined number of steps, and at the end of these steps, the model's parameters are evaluated on the test set to compute the loss, which is then used for training the meta-learner.

2.1.4. Optimization with Markov decision process and reinforcement learning

In their paper, Li and Malik [22] propose a novel approach to learning an optimization algorithm using guided policy search through reinforcement learning in the form of a Markov decision process (MDP) [23]. The goal is to learn an optimization algorithm, represented by a *policy* π, that can efficiently update the current location in an iterative optimization process. The optimization algorithm under consideration performs updates to the current location using a step vector computed by a generic function π of the objective function, the current location, and past locations. Each value of π corresponds to a different optimization algorithm, so by learning π, one can effectively learn multiple optimization algorithms. However, learning a generic function π is challenging, so the authors restrict the dependence of π to the objective values and gradients evaluated at the present and past locations. Consequently, π can be modelled as a function that takes the objective values and gradients along the optimizer's trajectory and outputs the next step vector for the optimization.

The authors observe that executing an optimization algorithm can be seen as executing a policy in an MDP, where the current location serves as the state, the step vector as the action, and the transition probability is similar to the location update formula $(x^{(i)} \leftarrow x^{(i-1)} + \Delta x)$. The implemented policy corresponds to the choice of π used by the optimization algorithm. By searching over policies, they effectively explore a range of possible first-order optimization algorithms. To learn the policy π, they use reinforcement learning, with the speed of convergence serving as the cost function (policies that lead to slow convergence are penalized). Since π could be stochastic in general, the authors use a neural network to parameterize the mean of π. The current state in the MDP corresponds to the parameters of the neural network, and the system updates these parameters (takes an action from the policy) and receives a reward based on how the loss function changes.

2.1.5. Memory-augmented neural networks

In their paper, Santoro et al. [24] propose a solution to the FSL task using a differentiable version of Memory-augmented Neural Networks (MANNs) known as Neural Turing Machines (NTMs) [25]. An NTM consists of a controller, which can be a feed-forward network or a Long Short-Term Memory (LSTM) network, that interacts with an external memory module through reading and writing heads. The NTM's memory reading and writing operations are fast, making it suitable for meta-learning and few-shot predictions. In the proposed approach, the authors feed the model with an input while its label is provided a one-time step later. Specifically, at time step t, the model receives the input x_t and the label y_{t-1}. This approach prevents the model from simply learning to map the label to the output. To further ensure this, inputs and their corresponding labels are shuffled in each episode so that the model cannot learn the input sequence directly. During the training process, an external memory is utilized to store the input and its bounded label, which are always one time step apart. When an already-seen input shows up, the corresponding label is retrieved from the external memory, providing the input classification. The retrieval process is performed using a key k_t associated with the input x_t, produced by the controller and stored in a memory matrix M_t. Specifically, the cosine similarity between the key k_t and the memory matrix M_t contents is produced. The error signals from the prediction step are backpropagated to promote this binding strategy. We illustrate the described approach in Fig. 3.

2.2. Metric learning-based methods

The metric-learning-based category comprises all the algorithms that enable the model to *learn to compare*. The main idea is to train the model to understand the similarity between images, allowing it to classify a new instance based on its distance w.r.t the seen categories. Below, we report some of the most relevant SOTA metric-learning-based algorithms.

2.2.1. Siamese neural networks

Bromley et al. [26] first introduced Siamese Neural Networks for signature verification. In 2015, they were newly proposed by Koch, Zemel and Salakhutdinov [27], where they exploited Convolutional Neural Networks (CNNs) to perform one-shot image classification. A Siamese Network consists of two identical networks accepting different inputs and having bound weights to ensure that similar images are mapped close in the feature space. As the network undergoes training, it learns to differentiate between pairs of images that belong to the same class and those that belong to different classes. In the inference phase, a test image is compared with one image per novel class and a similarity score is computed. The network then assigns the highest probability to the pair with the highest score. Because the model is trained on an extensive set of training classes, it becomes proficient at general data discrimination during the training process.

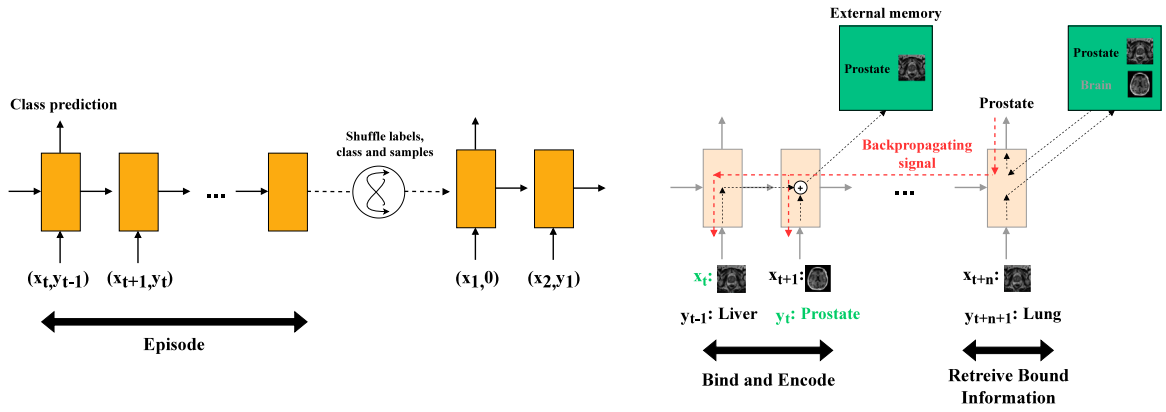


Fig. 3. Illustration of meta-learning with MANN approach.

2.2.2. Triplet networks

Triplet Networks, introduced by Hoffer et al. [28], were inspired by Siamese Networks and share the same architectural criterion. Here, the model is composed of three identical networks, having shared parameters, which are trained by triplets composed of anchor, positive and negative samples (positive examples belong to the same class as the anchor, while negative belongs to a different class). The network outputs the L_2 distances between the anchor and the positive and negative examples. The objective is to classify which between the positive and negative examples belongs to the same class as the anchor. During inference time, the model is fed with two inputs and assesses whether they belong to the same class by applying a threshold to the distance in the embedding space.

Triplet networks have also been employed in the ZSL and GZSL domains. Specifically, in [29], the authors propose a Dual-Triplet Network (DTNet) that employs two triplet networks for learning visual-semantic mapping: one focuses on negative attribute features, and the other on the negative visual features. Zero-shot classification is performed with a nearest-neighbour approach on the attribute feature, searching for the closest to the test sample.

2.2.3. Matching networks

Matching Networks proposed by Vinyals et al. [30], differently from Siamese and Triple Networks, can work in a multi-class way instead of in a pair-wise one. Matching Networks aim to map a support set to a classifier, which, given a query example, can produce a probability distribution of the output according to the following equation:

$$P(\hat{y}_j|\hat{x}_j) = \sum_{j=1}^{k} a(\hat{x}_j, x_j) y_j \tag{6}$$

where a acts as an attention mechanism. In the simplest implementation, a consists of computing a softmax over the cosine distance. At each iteration, a training episode is constructed, composed of a support and a query set. Based on the support set, the network provides the query label and the error is minimized.

2.2.4. Prototypical networks

Prototypical Network, proposed by Snell, Swersky, and Zemel [31], compute a representation or *prototype* of each class using an embedding function with trainable parameters. Given a class c, the prototypes are computed by averaging the embeddings of the support samples belonging to each class:

$$p_c = \frac{1}{|S_c|} \sum_{(x_j, y_j) \in S_c} f_\theta(x_j) \tag{7}$$

Given a generic distance function d, the prototypical network provides an output distribution based on the distance between the query embeddings and the prototypes of each class:

$$P(\hat{y}_j = c|\hat{x}_j) = \frac{exp(-d(f_\theta(\hat{x}_j), p_c))}{\sum_{c'} exp(-d(f_\theta(\hat{x}_j), p_{c'}))} \tag{8}$$

As for Matching Networks, training episodes are built by sampling a set of classes from the training set and choosing two groups of examples for each class as the support and query set, respectively. While in the original paper on Matching Networks, cosine distance was used as a distance function, here, the authors employ the negative squared Euclidean distance (greater distances provide smaller values). As pointed out by the authors, while prototypical networks differ from matching networks in a few-shot scenario, One-Shot Learning (OSL) makes them equivalent.

It is also possible to use this architecture for ZSL. Here, instead of having training points, we have a class meta-data vector for each class, which can be already known or learned, for example, from raw text [32]. Here, the prototype becomes an embedding of the meta-data vector.

2.2.5. Relation networks

Relation Networks were introduced by Santoro et al. in their paper [33], and they were initially employed in the FSL and ZSL domains in [34]. In contrast to Matching and Prototypical Networks, which use predefined distance functions, a relation network is trained end-to-end, including the metric to compare support and query embeddings. This part of the network is called *relation module*. In a one-shot setting, embeddings from support and query samples are first produced and concatenated in depth through an operator $Z(\cdot,\cdot)$. Concatenated embeddings are provided to the relation module g_ϕ, which outputs a scalar representing the similarity between the support and query embeddings:

$$r = g_\phi(Z(f_\theta(x_j), f_\theta(\hat{x}_j))) \tag{9}$$

For a generic FSL, the class feature map is calculated by summing all embedding module outputs from each sample in the training set. The class-level feature map is then combined with the query image feature map as in the one-shot scenario.

Relation Networks can be employed in a ZSL as well. In this case, a semantic class embedding vector is provided for each class. Since support and query vectors belong to different modalities (attributes and images, respectively), two embedding modules are employed. The relation module instead works as before.

2.3. Hallucination-based methods

The hallucination-based methods directly address the scarcity of data by *learning to augment*. These methods focus on generating additional data to overcome the limitations of the available dataset. In the following, we describe in detail the most prominent hallucination-based methods.

2.3.1. Hallucinating with intra-class analogies

Harihan and Girshick [35] propose to exploit intra-class analogies to augment the dataset when few examples are available. Their framework employs a learner, two training and a testing phase. In the first training phase, known as *representation learning* phase, the learner is fed with several base classes (C_{base}), for which a lot of examples are available for each class. The learner uses these data to set the parameters of its feature extractor. During the second phase (*low-shot* phase), the learner needs to distinguish a set of classes, both base and novel ones. For the novel classes, the learner has access only to a few examples, while for the base classes, it has access to the same dataset used for learning the feature extractor. During the test phase, the model predicts labels from both classes. For the categories with few examples, the idea is to hallucinate additional data using the many examples seen for the base classes to improve the model's performance. The goal is to learn a transformation that maps two images belonging to the same base class (e.g., bird on grass and bird on the sky) and apply this transformation to a novel class image. To achieve this, a function G is trained that takes the concatenated feature vectors of three examples and outputs a "hallucinated" feature vector. As G, they exploited an MLP with three fully connected layers.

2.3.2. Classificator and hallucinator end-to-end model

Wang et al. [36] further deepened the previously described method by combining a generator of "hallucinated" examples, with a meta-learning framework, by optimizing the two models jointly. The "hallucinator" G takes as input an example x, a noise vector z and produces a hallucinated example as the output according to the hallucinator parameters θ_G. During meta-testing, several hallucinated examples are computed by sampling from the initial training set S_{train}, producing a new training set S^G_{train}. The final training set S^{aug}_{train} is obtained by combining the two datasets. This dataset is then used to train the classification algorithm. During the meta-training phase, the hallucinator is trained jointly with the classification algorithm, exploiting a meta-learning paradigm. From the set of all classes, m classes are sampled, specifically n examples for one. The generator G is exploited to produce additional n augmented examples to add to the training set. This new dataset is employed to train the classification algorithm. This training process is agnostic w.r.t. specific meta-learning algorithm used.

After categorizing and describing the main meta-learning methods for FSL in the literature, the following chapter outlines the methods used for searching, selecting, and analysing SOTA works in the field of FSL for medical image analysis.

3. Methods

3.1. Study design

We conducted a systematic review in accordance with the "Preferred reporting items for systematic reviews and meta-analyses" (PRISMA) 2020 checklist [37]. The primary objective of this review is to establish a comprehensive methodological pipeline shared across all examined studies, serving as a reference for future research in this domain. Additionally, it aims to analyse the distribution of studies across three key outcomes: segmentation, classification, and registration, as well as w.r.t. the clinical tasks investigated and the meta-learning methods employed. Furthermore, it seeks to conduct a statistical analysis of the performance of the studies and provide additional insights, such as the imaging modality analysed.

3.2. Eligibility criteria

We established the inclusion criteria for paper selection based on three primary aspects:

- **Implementation of FSL techniques**: We selected papers that claimed to implement FSL in their work.
- **Application in medical imaging domain**: We considered papers that performed at least one experiment applied to the medical imaging domain.
- **Low data usage in training**: We included only papers that demonstrated using a small amount of data during training. In particular, we considered all the studies that employed a maximum of 20 training examples per class.

In addition, during the selection process, we excluded abstracts, non-peer-reviewed papers, papers written in languages other than English, and papers deemed to have significant theoretical errors. Furthermore, we did not include papers dealing with few-shot domain adaptation methods (FSDA), as [38–40]. FSDA, as highlighted by Li et al. [40], FSL focuses on adapting pre-trained models to perform well on novel tasks with limited training examples, whereas FSDA involves adapting models across different domains. Therefore, we considered FSDA papers outside the scope of this systematic review. By applying these inclusion and exclusion criteria, we aimed to ensure the selection of relevant and high-quality papers that specifically addressed the application of FSL techniques in medical imaging with limited training data.

3.3. Information sources

We searched for papers using the following databases:

- Web of science
- Scopus
- IEEE Xplore
- ACM Digital Library

To ensure comprehensive coverage and include recent studies in our analysis, we performed a two-step search, the first on September 7, 2022, and the second on January 25, 2023. In cases where we did not have full access to the papers, we took advantage of the Network Inter-Library Document Exchange (NILDE) platform, a web-based Document Delivery service through which we requested access to the missing PDF files, enabling us to obtain the complete papers for inclusion in our review.

3.4. Research strategies

For each of the mentioned databases, we listed the queries used in the study search in Table 1.

3.5. Selection process

During the review process, a single reviewer examined each record, including titles, abstracts, and any accompanying reports obtained during the research. No machine learning algorithms were employed to aid in eliminating records or to streamline the screening process. Additionally, no crowdsourcing or pre-screened datasets were employed for the records screening.

3.6. Data collection process

For data collection, a single reviewer was responsible for collecting the relevant information from each report. No automation processes were employed for the data collection process. During the review, all articles were examined in their original language. The selection of articles was based on the predefined eligibility criteria described above. No software or automated tools were used to extract data from the figures or graphical representations in the articles. Finally, the data collection process entailed a manual analysis of the articles to extract the pertinent information for the review.

Table 1
Research queries employed for each database.

Database	Query
Web of Science	(TS = (“few-shot”) OR TS = (“low-shot”) OR TS = (“one-shot”) OR TS = (“zero-shot”)) AND (TS = (“medical imag*”)) AND (TS = (“classif*”) OR TS = (“segment*”) OR TS = (“regist*”))
Scopus	TITLE-ABS-KEY (few-shot) OR TITLE-ABS-KEY (low-shot) OR TITLE-ABS-KEY (one-shot) OR TITLE-ABS-KEY (zero-shot) AND TITLE-ABS-KEY (medical imaging) OR TITLE-ABS-KEY (medical image) OR TITLE-ABS-KEY (medical images) AND TITLE-ABS-KEY (classif*) OR TITLE-ABS-KEY (segment*) OR TITLE-ABS-KEY (regist*)
IEEE Xplore	(((“Abstract”:“few-shot” OR “Abstract”:“low-shot” OR “Abstract”:“one-shot” OR “Abstract”:“zero-shot”) AND “Abstract”:“medical imag*” AND (“Abstract”:classification OR “Abstract”:segmentation OR “Abstract”:registration) OR (“Document Title”:“few-shot” OR “Document Title”:“low-shot” OR “Document Title”:“one-shot” OR “Document Title”:“zero-shot”) AND “Document Title”:“medical imag*” AND (“Document Title”:classification OR “Document Title”:segmentation OR “Document Title”:registration) OR (“Author Keywords”:“few-shot” OR ““Author Keywords””:“low-shot” OR ““Author Keywords””:“one-shot” OR ““Author Keywords””:“ero-shot”) AND ““Author Keywords””:“medical imag*” AND (““Author Keywords””:classif* OR ““Author Keywords””:segment* OR ““Author Keywords””:regist*)))
ACM Digital Library	“(Abstract:(““few-shot”” OR ““low-shot”” OR ““one-shot”” OR ““zero-shot””) OR Keyword:(““few-shot”” OR ““low-shot”” OR ““one-shot”” OR ““zero-shot””) OR Title:(““few-shot”” OR ““low-shot”” OR ““one-shot”” OR ““zero-shot””)) AND (Abstract:(““medical imaging”” OR ““medical images”” OR ““medical image””) OR Title:(““medical imaging”” OR ““medical images”” OR ““medical image””) OR Keyword:(““medical imaging”” OR ““medical images”” OR ““medical image””)) AND (Title:(classif* OR segment* OR regist*) OR Abstract:(classif* OR segment* OR regist*) OR Keyword:(classif* OR segment* OR regist*))”

3.7. Data item

In our study, we examined three primary outcomes: segmentation, classification, and registration. All of the reviewed studies were compatible with these three outcome domains. We did not alter or introduce any changes to the outcome domains or their significance in the review. Likewise, we did not modify the selection processes within these eligible outcome domains. Beyond the three outcomes previously mentioned, we also explored data pertaining to the utilization of FSL, OSL, and ZSL techniques, as well as their applications within the field of medical imaging.

3.8. Assessment of bias in studies

To evaluate the potential risk of bias (ROB) or concerns regarding applicability in each study, we utilized the PROBAST tool [41], designed for assessing the quality of diagnostic accuracy studies. For each outcome, we created a table denoting studies with low risk or concerns using a green checkmark symbol ✓ and those with high risk or concerns using a red cross symbol ✗.

3.9. Effect measures

In the segmentation studies included in our review, we evaluated the performance using two popular metrics: the Dice score and the Intersection over Union (IoU). When investigating studies that addressed a classification task, we considered Accuracy, F1-score, Recall, and the Area Under the Receiver Operating Characteristic (AUROC) metrics. Finally, in the registration domain, we examined the Dice score as well as the Average Landmark Distance (ALD) and the Target Registration Error (TRE).

3.10. Synthesis methods

In our systematic review, we structured the key information of each study within dedicated tables for each outcome category. The tables included the following information: first author, year of publication, the algorithm or framework used, the number of training data, the best performance achieved by the model, and whether the study utilized the meta-learning paradigm. To provide a visual summary of the results, we used forest plots. We generated these plots by grouping the studies based on the anatomical structure investigated and the meta-learning method employed in each outcome. We created separate forest plots for each performance metric (accuracy, AUROC, etc.), considering, in each study, the highest performance achieved (across various experiments and image modalities). In each forest plot, we reported the mean and the 95% confidence interval (CI) across all the studies within the corresponding group, whether organized by clinical task or meta-learning algorithm. It is important to note that we did not conduct a meta-analysis of the collected results. This is because the studies included in our review encompassed various clinical applications, making direct comparisons between the results inappropriate. Therefore, the forest plots served as a visual representation of the individual study findings rather than a quantitative synthesis of the data. Finally, for each study, we presented a table summarizing the main learning strategies employed w.r.t. three key phases: **Pre-training**, **Training**, and **Data augmentation**. Based on this, we identified a standard methodological pipeline shared among the examined studies.

4. Results

4.1. Study selection

In Fig. 4, we summarize the data selection flow using the PRISMA diagram. In total, we retrieved 314 studies and included 80 studies in the final analysis.

4.2. Studies characteristics

In this section, we present the findings resulting from our analysis of the selected research papers. Fig. 5 displays the distribution of studies across the three investigated outcomes. Below, we present the results of our analysis grouped by outcome. It is worth noting that several studies, namely [42–45], and [46], are included multiple times, as they address various outcomes simultaneously.

4.2.1. Segmentation

We selected 50 relevant studies, each focusing on medical segmentation as its primary task. All information extracted from the selected studies is provided in Table 2. In addition, for each study, we present ROB and the applicability analyses in Table 3. The selected studies implemented different strategies to perform FSL for segmentation purposes. For example, Khadka et al. [47] leveraged an initialization-based meta-learning approach, specifically the implicit MAML algorithm, to segment skin lesions and colonoscopy images. They evaluated the generalizability of their model by testing it on both the same and a different task w.r.t. the one used during training. Besides meta-learning algorithms, several studies leveraged other types of learning frameworks. For example, Khaled et al. [48] employed a semi-supervised approach leveraging a multi-stage Generative Adversarial Network (GAN) that performs a coarse-to-fine segmentation of brain tissues. To improve the robustness and generalization capabilities of the model, it is also usual to employ a pre-training step and data augmentation techniques. For example, Xu and Niethammer [46] leveraged a pre-trained network and a joint registration-segmentation network training. Specifically, the

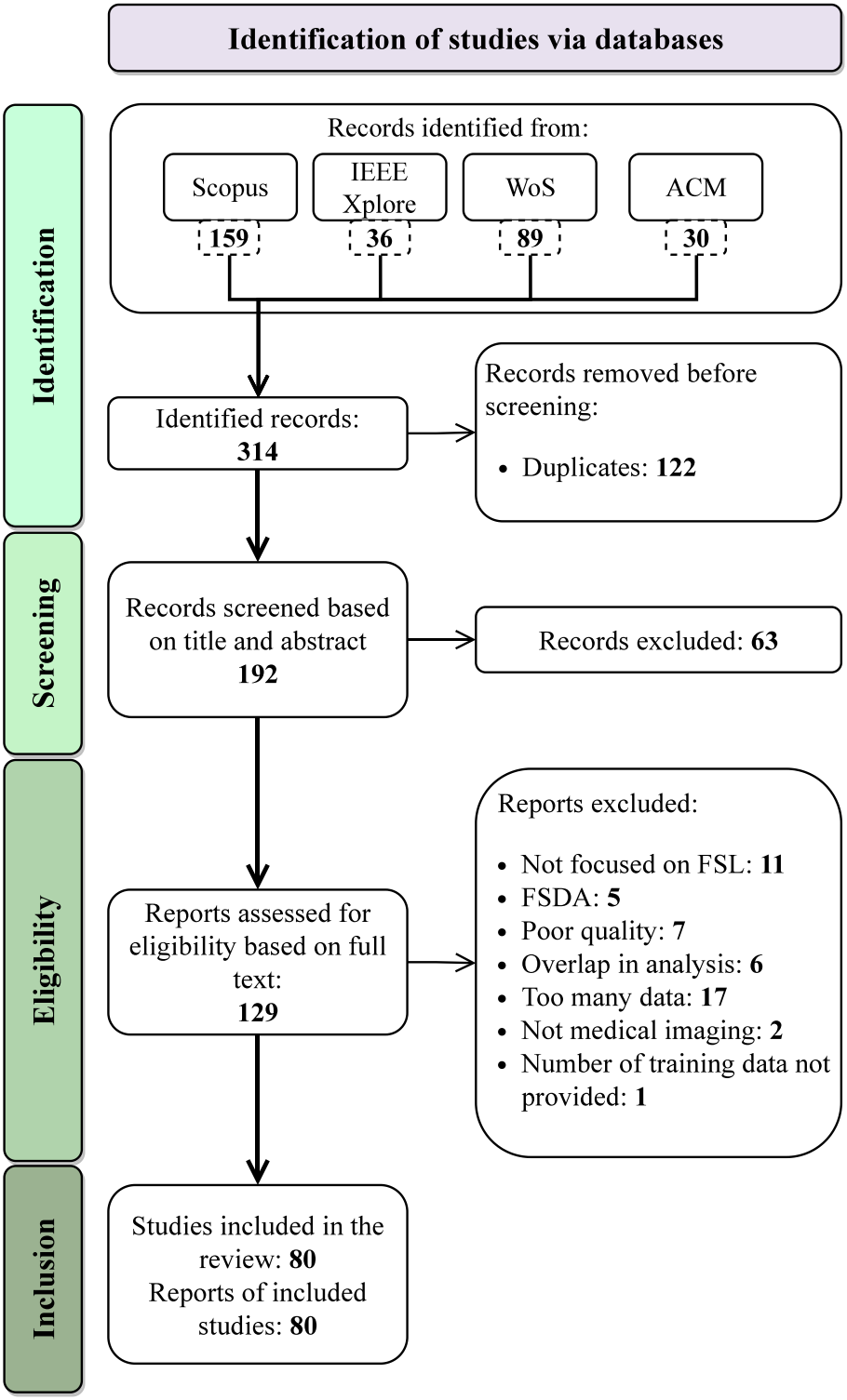


Fig. 4. PRISMA flow diagram.

registration network provides a realistic form of data augmentation, an approach also used in other FSL segmentation studies.

Here, we present the findings of our segmentation papers analysis.

Clinical task. The segmentation studies investigated various anatomical structures and regions, as well as specific diseases such as polyps or tumours. Here's a breakdown of the papers categorized w.r.t. the clinical task addressed. Eighteen papers (36%) focused on liver segmentation; 18 studies (36%) concentrated on kidney segmentation; 17 papers (34%) centred around spleen segmentation; three papers (6%) pertained to psoas segmentation; four (8%) investigated a prostate segmentation task; three works (6%) involved bladder segmentation; four papers (8%) dealt with breast segmentation; one paper (2%) addressed colon segmentation; six (12%) concerned stomach segmentation; 12 (24%) were dedicated to brain segmentation; 14 papers (28%) revolved around heart segmentation; three (6%) involved pancreas segmentation; three (6%) pertained to cell segmentation; two papers (4%) considered to lung segmentation; one (2%) focused on eye

Table 2
FSL studies for medical image segmentation. AA = Ascending Aorta; CZ = Central Zone; IC = Intraretinal Cyst; LA = Left Atrium; LV = Left Ventricle; MAD = Mean Across Diseases; MAS = Mean Across Structures; MYO = Myocardium; PA = Pulmonary Artery; PZ = Peripheral Zone; RA = Right Atrium; RV = Right Ventricle; TZ = Transitional Zone.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
1	Blendowski, Nickisch, and Heinrich [49]	Siamese Network + SSL	1-shot 9-shot	Dice: 0.853 (Liver) 0.657 (Spleen) 0.663 (Kidney) 0.656 (Psoas)	None
2	Chan et al. [50]	Res2-UNeXT + Data augmentation with Daemons registration algorithm	8-shot	IoU: 0.943 (Cells)	None
3	Chen et al. [51]	Adversarial data augmentation framework (Advchain)	1-shot 3-shot 11-shot	Dice: 0.844 (LV) 0.647 (RV) 0.812 (MYO) 0.572 (Prostate, PZ) 0.845 (Prostate, CZ)	None
4	Cui et al. [52]	MRE-Net (Distance metric-learning + U-net)	1-shot 7-shot	Dice: 0.781 (Spleen) 0.774 (Kidney) 0.522 (Gallbladder) 0.568 (Esophagus) 0.597 (Stomach) 0.613 (Pancreas) 0.820 (Brain, MAS)	Metric learning
5	Ding, Wangbin et al. [53]	Registration model + Similarity model + Patch label fusion	20-shot	Dice: 0.817 (MYO)	Metric learning
6	Ding, Yu and Yang [54]	Registration network + Generative network + Segmentation network	1-shot	Dice: 0.851 (Brain)	Hallucination
7	Farshad et al. [55]	MetaMedSeg (Reptile-based with task weighting)	15-shot	IoU: 0.683 (Heart) 0.583 (Spleen) 0.227 (Prostate, PZ) 0.483 (Prostate, TZ)	Initialization
8	Feng et al. [56]	Medical prior-based FSL network + Interactive learning-based test time optimization algorithm	10-shot	Dice: 0.569 (Breast) 0.584 (Kidney) 0.751 (Liver) 0.675 (Stomach)	Metric learning
9	Gama, Oliveira and dos Santos [57]	Weakly-supervised segmentation learning	1-shot 5-shot 10-shot 20-shot	IoU: 0.870 (Lungs) 0.790 (Heart) 0.800 (Mandible) 0.870 (Breast)	Initialization
10	Gama et al. [58]	ProtoSeg	1-shot 5-shot 10-shot 20-shot	IoU: 0.800 (Heart) 0.920 (Lungs) 0.720 (Breast) 0.400 (Mandible)	Metric learning
11	Guo, Odu and Pedrosa [59]	Cascaded U-net + 3D augmentation	From 1-shot to 6-shot	Dice: 0.910 (Kidney)	None
12	Hansen et al. [60]	Anomaly detection-inspired model + SSL	1-shot 2-shot 3-shot	Dice: 0.875 (LV) 0.773 (RV) 0.624 (MYO) 0.833 (Kidney) 0.759 (Spleen) 0.808 (Liver)	Metric learning
13	He et al. [42]	Deep complementary joint model (Segmentation model + Pixel-wise discriminator + Registration model)	4-shot	Dice: 0.970 (AA) 0.920 (LA) 0.950 (LV) 0.870 (MYO) 0.800 (PA) 0.800 (RA) 0.810 (RV)	None

(*continued on next page*)

Table 2 (*continued*).

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
14	He et al. [43]	Knowledge consistency constraint strategy + Space-style sampling program + Mix misalignment regularization	1-shot 5-shot	Dice: 0.911 (Heart, MAS) 0.872 (Brain, MAS)	Hallucination
15	Jenssen et al. [61]	Self-guided anomaly detection-inspired model	1-shot	Dice: 0.840 (LV) 0.585 (MYO) 0.697 (RV)	Metric learning
16	Joyce and Kozerke [62]	Anatomical model + SSL	1-shot 3-shot 10-shot	Dice: 0.630 (Heart)	None
17	Khadka et al. [47]	Implicit MAML + Attention U-Net	5-shot 10-shot 20-shot	Dice: 0.833 (Skin, nevus)	Initialization
18	Khaled, Han and Ghaleb [48]	Multi-stage GAN	5-shot 10-shot	Dice: 0.940 (Brain, MAS)	None
19	Khandelwal and Yushkevich [63]	Gradient-based meta-learning domain generalization + 3D U-Net + Fine-tuning	2-shot 4-shot 6-shot	Dice: 0.823 (Spine, MAS)	Initialization
20	Kim et al. [64]	VGG16 + Bidirectional gated recurrent unit + U-Net + Fine-tuning	5-shot	Dice: 0.905 (Spleen) 0.900 (Kidney) 0.887 (Liver) 0.771 (Bladder)	Metric learning
21	Li et al. [65]	3D U-Net + Prototypical learning + Image alignment module	1-shot	Dice: 0.417 (Prostate, MAS)	Metric learning
22	Lu et al. [66]	Contour transformer network (ResNet-50 + Graph convolutional network blocks)	1-shot	IoU: 0.973 (Knee) 0.948 (Lung) 0.970 (Phalanx) 0.973 (Hip)	None
23	Lu and Ye [67]	TractSeg + Knowledge transfer with warmup	1-shot 5-shot	Dice: 0.812 (Brain, WM)	None
24	Ma et al. [68]	Segmentation network + Zero-shot segmentation network + Spatial Context Attention module	0-shot	Dice: 0.882 (Brain, tumour)	None
25	Niu et al. [69]	Conditioner + Segmenter + Symmetrical supervision mechanism + Transformer-based global feature alignment module	1-shot	Dice: 0.870 (LV) 0.815 (Kidney) 0.738 (Spleen) 0.729 (Liver)	Metric learning
26	Ouyang et al. [70]	Self-supervised adaptive local prototype pooling network	1-shot 5-shot	Dice: 0.862 (Kidney) 0.757 (Spleen) 0.821 (Liver) 0.870 (LV) 0.721 (MYO) 0.860 (RV)	Metric learning
27	Pham et al. [71]	Few-sample-fitting	1-shot to 20-shot	Dice: 0.990 (Femur)	None
28	Pham, Dovletov and Pauli [72]	3D U-Net + Imitating encoder + Prior encoder + Joint decoder	1-shot	Dice: 0.776 (Liver)	None
29	Roy et al. [73]	Conditioner arm + Segmenter arm + Channel squeeze & Spatial excitation blocks	1-shot	Dice: 0.700 (Liver) 0.607 (Spleen) 0.464 (Kidney) 0.499 (Psoas)	None
30	Roychowdhury et al. [44]	Echo state network + augmented U-Net	5-shot	Dice: 0.640 (Eye, IC)	None
31	Rutter, Lagergren and Flores [74]	CNN for boundary optimization	1-shot 3-shot 5-shot	Dice: 0.931 (Cells)	None

(*continued on next page*)

Table 2 (*continued*).

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
32	Shen et al. [75]	Large deformation diffeomorphic metric mapping model + Sample transformations + Interpolation	1-shot	Dice: 0.883 (Knee)	None
33	Shen et al. [76]	VGG-16 + Poisson learning + Spatial consistency calibration	1-shot	Dice: 0.619 (Skin, MAD) 0.610 (Liver) 0.536 (Kidney) 0.529 (Spleen)	None
34	Shi et al. [45]	Joint Registration and Segmentation Self-training Framework (JRSS)	5-shot	Dice: 0.795 (Brain, MAS) 0.753 (Abdomen, MAS)	None
35	Sun et al. [77]	2-branch CNN + Spatial squeeze excite module + Global correlation module + Discriminative embedding module	1-shot	Dice: 0.495 (Liver) 0.606 (Spleen) 0.830 (Kidney)	Metric learning
36	Tang et al. [78]	Recurrent prototypical networks (U-Net + Contex relation encoder + Prototypical network)	1-shot	Dice: 0.788 (Spleen) 0.851 (Kidney) 0.819 (Liver)	Metric learning
37	Tomar et al. [79]	Generative style transfer (Appearance model + Style encoder + Flow model + Flow adversarial autoencoder)	1-shot	Dice: 0.835 (Brain, MAS)	None
38	Wang et al. [80]	Label transfer network (Atlas-based segmentation + Forward-backward correspondence)	1-shot	Dice: 0.823 (Brain, MAS)	None
39	Wang et al. [81]	Siamese model and individual-difference- aware model (Encoders + Forward-backward consistency)	1-shot 5-shot	Dice: 0.862 (Brain, MAS) 0.803 (Spleen) 0.884 (Kidney) 0.916 (Liver) 0.684 (Stomach) 0.511 (Pancreas) 0.485 (Doudenum) 0.519 (Esophagus)	None
40	Wang et al. [82]	V-Net + Init-crop + Self-down + Self-crop	4-shot	Dice: 0.937 (LV) 0.890 (RV) 0.872 (LA) 0.909 (RA) 0.831 (MYO) 0.943 (AA) 0.798 (PA)	None
41	Wang, Zhou and Zheng [83]	Prototype learning + Self-reference + Contrastive learning	1-shot	Dice: 0.756 (Liver) 0.737 (Spleen) 0.842 (Kidney)	Metric learning
42	Wang et al. [84]	Alternating union network (Image sub-network + Label sub-network)	1-shot	Dice: 0.873 (LV) 0.637 (MYO) 0.720 (RV)	None
43	Wu, Xiao and Liang [85]	Dual contrastive learning + Anatomical auxiliary supervision + Constrained iterative prediction module	1-shot	Dice: 0.699 (Liver) 0.838 (Kidney) 0.749 (Spleen)	None
44	Wu et al. [86]	Self-learning + One-shot learning	1-shot	Dice: 0.850 (Spleen) 0.930 (Liver)	None
45	Xu and Niethammer [46]	DeepAtlas (Semi-supervised learning + Segmentation network + Registration network)	1-shot 5-shot 10-shot	Dice: 0.892 (Knee, MAS) 0.612 (Brain)	None

(*continued on next page*)

Table 2 (*continued*).

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
46	Yu et al. [87]	Location-sensitive local prototype network	1-shot	Dice: 0.793 (Liver) 0.733 (Spleen) 0.765 (Kidney) 0.524 (Psoas)	Metric learning
47	Yuan, Esteva and Xu [88]	MetaHistoSeg (U-Net + MAML)	8-shot	IoU: 0.326 (Cells) 0.682 (Cells nuclei) 0.557 (Gland) 0.632 (Colon, tumour)	Initialization
48	Zhao et al. [89]	Spatial and appearance transform models + Semi-supervised learning + Supervised learning	1-shot	Dice: 0.815 (Brain, MAS)	None
49	Zhao et al. [90]	Meta-hallucinator	1-shot 4-shot	Dice: 0.756 (AA) 0.751 (LA) 0.823 (LV) 0.696 (MYO)	Initialization and Hallucination
50	Zhou et al. [91]	OrganNet (3 encoders + Pyramid reasoning modules)	1-shot	Dice: 0.891 (Spleen) 0.860 (Kidney) 0.770 (Aorta) 0.728 (Pancreas) 0.826 (Stomach)	None

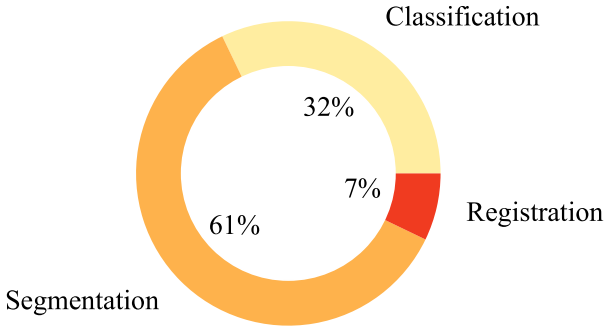


Fig. 5. Studies distribution by outcome.

segmentation; two papers (4%) involved mandible segmentation; one (2%) addressed duodenum segmentation; two papers (4%) dealt with skin segmentation; three papers (6%) considered to knee segmentation; one (2%) concerned phalanx segmentation; one (2%) dealt with hip segmentation; one paper (2%) was dedicated to spine segmentation. Fig. 6(a) illustrates the distribution of segmentation studies among the five most popular clinical tasks.

Meta-learning method. Out of the 50 studies we selected in the realm of FSL for medical image segmentation, the distribution of their meta-learning methods is as follows: six studies (12%) leveraged initialization-based methods; 14 studies (28%) utilized metric learning-based techniques; three studies (6%) employed hallucination-based methods; one study (2%) combined both initialization-based and hallucination-based methods. The remaining 28 studies (56%) did not incorporate any meta-learning technique. For a visual representation of the distribution of studies in terms of the meta-learning approach employed, refer to Fig. 6(b).

K-shot. According to the k-shot configuration used, we found that 15 studies (30%) utilized k-shot training with k ranging from 2 to 20; 14 studies (28%) performed both OSL and FSL; 20 works (40%) exclusively focused on OSL and one paper (2%) employed ZSL.

Image modality. In the following, we provide the distribution of studies in terms of the imaging modalities utilized: 26 (52%) used CT images; 30 papers (60%) utilized MRI; four (8%) relied on X-ray images; two (4%) involved dermoscopic images; one paper (2%) made use of endoscopic images; one (2%) used histopathology images; two (4%) employed microscopic images; one paper (2%) utilized OCT images.

Model evaluation. For evaluating the models' robustness, the selected studies used different evaluation techniques: 21 studies (42%) exclusively conducted ablation studies; 11 studies (22%) utilized both ablation studies and cross-validation; five studies (10%) relied solely on cross-validation; 13 studies (26%) did not employ any specific model evaluation technique.

Statistical analysis. Fig. 7 displays forest plots that summarize the performance of FSL models (mean and 95% CI) for each clinical task considering both Dice score and IoU metrics. Conversely, Fig. 8 depicts the models' results grouped by the meta-learning method used.

Standard pipeline. Table 4 details the learning frameworks employed in each step of the standard pipeline by the segmentation studies analysed in this review. In the following, we explore the distribution of the studies based on the techniques employed for each pipeline phase: two out of 50 studies (4%) employed meta-learning for pre-training; two studies (4%) utilized self-supervised learning and 13 studies (26%) relied on supervised learning. The majority, 33 out of 50 studies (66%), did not employ any pre-training stage. For their main training stage, 20 studies (40%) utilized meta-learning methods; 12 (24%) employed semi-supervised approaches; four studies (8%) employed self-supervised methods; 16 studies (32%) used traditional supervised techniques; one study (2%) employed a zero-shot learning method. Finally, concerning the data augmentation techniques, 16 studies (32%) exploited classical data augmentation techniques; five studies (10%) utilized generative methods for data augmentation; nine studies (18%) relied on registration-based augmentation. The remaining 24 out of 50 studies (48%) did not employ data augmentation.

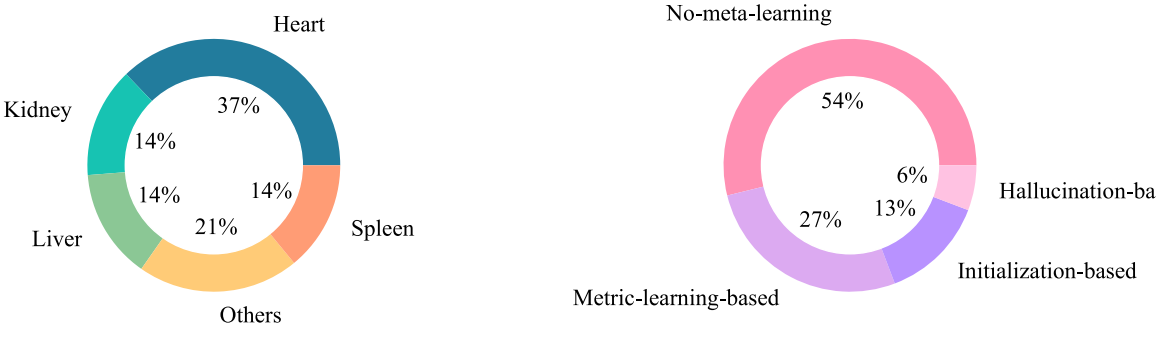
4.2.2. Classification

We identified 27 relevant studies, each focusing on medical classification as its primary task. To enhance clarity and facilitate easy reference, we present all the information extracted from the selected studies in Table 5. In addition, we provide information concerning ROB and the applicability concerns in Table 6.

As for segmentation, different strategies are present in SOTA for addressing FSL classification in the medical imaging domain. Regarding the meta-learning techniques, for example, Dai et al. [92] introduced

Table 3
ROB of FSL studies for medical image segmentation.

Study ID	Pub. ref.	Risk of bias					Applicability			
		Part.	Pred	Out.	Analysis	Overall	Part.	Pred.	Out.	Overall
1	Blendowski, Nickisch, and Heinrich [49]	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	Chan et al. [50]	✓	✓	✓	✗	✗	✓	✓	✓	✓
3	Chen et al. [51]	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	Cui et al. [52]	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	Ding, Wangbin et al. [53]	✓	✓	✓	✗	✗	✓	✓	✓	✓
6	Ding, Yu and Yang [54]	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	Farshad et al. [55]	✓	✓	✓	✗	✗	✓	✓	✓	✓
8	Feng et al. [56]	✓	✓	✓	✓	✓	✓	✓	✓	✓
9	Gama, Oliveira and dos Santos [57]	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	Gama et al. [58]	✓	✓	✓	✓	✓	✓	✓	✓	✓
11	Guo, Odu and Pedrosa [59]	✓	✓	✓	✗	✗	✓	✓	✓	✓
12	Hansen et al. [60]	✓	✓	✓	✓	✓	✓	✓	✓	✓
13	He et al. [42]	✓	✓	✓	✗	✗	✓	✓	✓	✓
14	He et al. [43]	✓	✓	✓	✓	✓	✓	✓	✓	✓
15	Jenssen et al. [61]	✓	✓	✓	✗	✗	✓	✓	✓	✓
16	Joyce and Kozerke [62]	✓	✓	✓	✗	✗	✓	✓	✓	✓
17	Khadka et al. [47]	✓	✓	✓	✓	✓	✓	✓	✓	✓
18	Khaled, Han and Ghaleb [48]	✓	✓	✓	✓	✓	✓	✓	✓	✓
19	Khandelwal and Yushkevich [63]	✓	✓	✓	✗	✗	✓	✓	✓	✓
20	Kim et al. [64]	✓	✓	✓	✓	✓	✓	✓	✓	✓
21	Li et al. [65]	✓	✓	✓	✓	✓	✓	✓	✓	✓
22	Lu et al. [66]	✓	✓	✓	✓	✓	✓	✓	✓	✓
23	Lu and Ye [67]	✓	✓	✓	✓	✓	✓	✓	✓	✓
24	Ma et al. [68]	✓	✓	✓	✓	✓	✓	✓	✓	✓
25	Niu et al. [69]	✓	✓	✓	✓	✓	✓	✓	✓	✓
26	Ouyang et al. [70]	✓	✓	✓	✓	✓	✓	✓	✓	✓
27	Pham et al. [71]	✓	✓	✓	✗	✗	✓	✓	✓	✓
28	Pham, Dovletov and Pauli [72]	✓	✓	✓	✓	✓	✓	✓	✓	✓
29	Roy et al. [73]	✓	✓	✓	✓	✓	✓	✓	✓	✓
30	Roychowdhury et al. [44]	✓	✓	✓	✗	✗	✓	✓	✓	✓
31	Rutter, Lagergren and Flores [74]	✓	✓	✓	✓	✓	✓	✓	✓	✓
32	Shen et al. [75]	✓	✓	✓	✓	✓	✓	✓	✓	✓
33	Shen et al. [76]	✓	✓	✓	✓	✓	✓	✓	✓	✓
34	Shi et al. [45]	✓	✓	✓	✓	✓	✓	✓	✓	✓
35	Sun et al. [77]	✓	✓	✓	✓	✓	✓	✓	✓	✓
36	Tang et al. [78]	✓	✓	✓	✓	✓	✓	✓	✓	✓
37	Tomar et al. [79]	✓	✓	✓	✓	✓	✓	✓	✓	✓
38	Wang et al. [80]	✓	✓	✓	✓	✓	✓	✓	✓	✓
39	Wang et al. [81]	✓	✓	✓	✓	✓	✓	✓	✓	✓
40	Wang et al. [82]	✓	✓	✓	✓	✓	✓	✓	✓	✓
41	Wang, Zhou and Zheng [83]	✓	✓	✓	✓	✓	✓	✓	✓	✓
42	Wang et al. [84]	✓	✓	✓	✓	✓	✓	✓	✓	✓
43	Wu, Xiao and Liang [85]	✓	✓	✓	✓	✓	✓	✓	✓	✓
44	Wu et al. [86]	✓	✓	✓	✓	✓	✓	✓	✓	✓
45	Xu and Niethammer [46]	✓	✓	✓	✓	✓	✓	✓	✓	✓
46	Yu et al. [87]	✓	✓	✓	✓	✓	✓	✓	✓	✓
47	Yuan, Esteva and Xu [88].	✓	✓	✓	✓	✓	✓	✓	✓	✓
48	Zhao et al. [89]	✓	✓	✓	✓	✓	✓	✓	✓	✓
49	Zhao et al. [90]	✓	✓	✓	✓	✓	✓	✓	✓	✓
50	Zhou et al. [91].	✓	✓	✓	✓	✓	✓	✓	✓	✓



(a) Grouped by clinical task (b) Grouped by meta-learning method

Fig. 6. Segmentation studies distribution grouped by clinical task and meta-learning method.

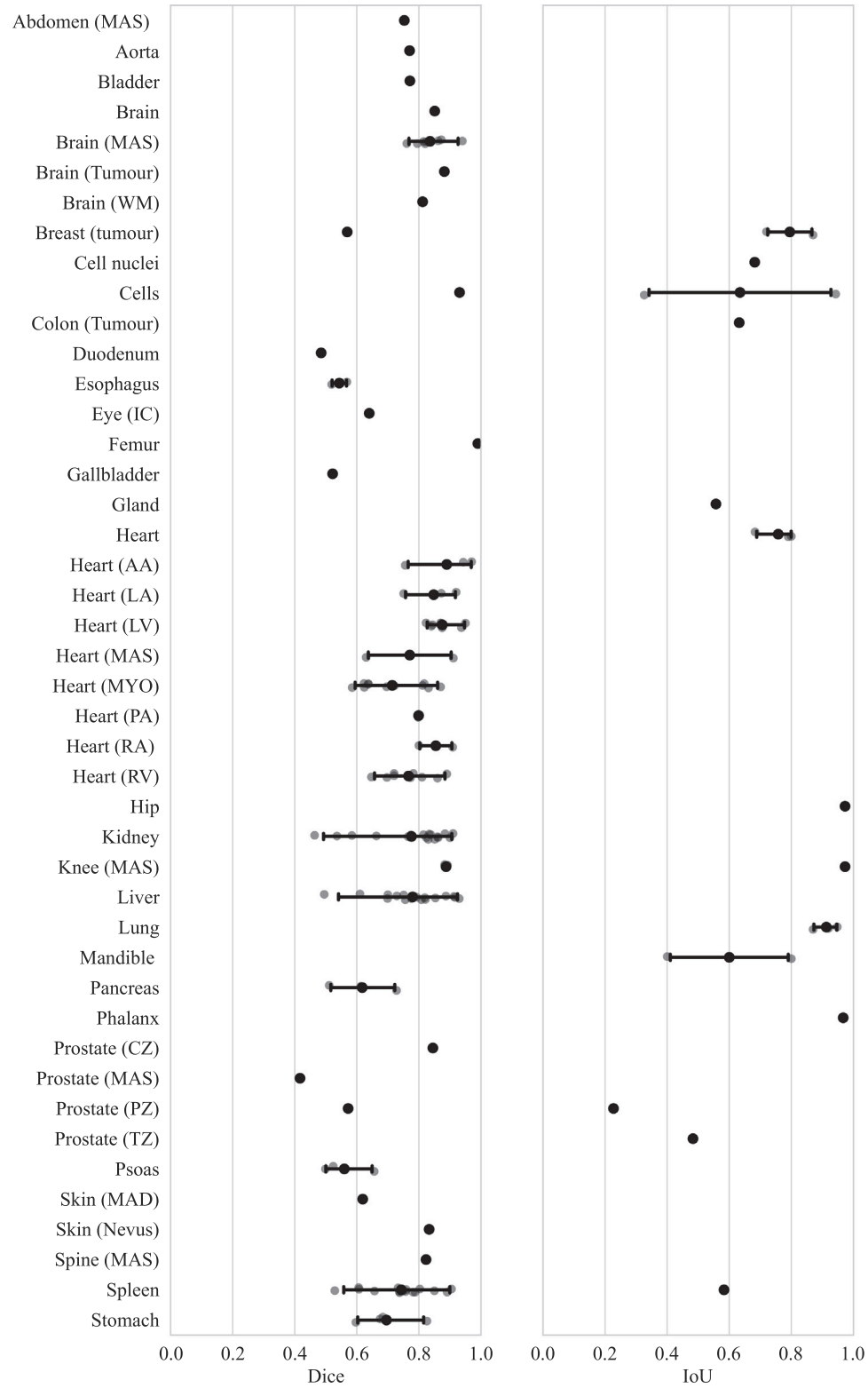


Fig. 7. Forest plot of segmentation studies performance based on Dice and IoU metrics. Studies are grouped by the clinical task investigated. AA = Ascending Aorta; CZ = Central Zone; IC = Intraretinal Cyst; LA = Left Atrium; LV = Left Ventricle; MAD = Mean Across Diseases; MAS = Mean Across Structures; MYO = Myocardium; PA = Pulmonary Artery; PZ = Peripheral Zone; RA = Right Atrium; RV = Right Ventricle; TZ = Transitional Zone; WM = White Matter.

PFEMed, i.e., an episodic-trained model that leverages a dual-encoder structure and a prior-guided Variational Autoencoder (VAE) to enhance the robustness of the target features w.r.t. classify query samples. We observed that meta-training can also serve as a pre-training step before fine-tuning on the downstream task. This approach was adopted by Maicas et al. [93], who proposed to meta-train the model on a wide selection of tasks and fine-tuning the pre-trained model on the target task using a classical fully-supervised approach. In terms of other learning techniques, Mahapatra et al. [94] addressed GZSL by exploiting an SSL to generate anchor vectors for seen and unseen classes. Anchor vectors

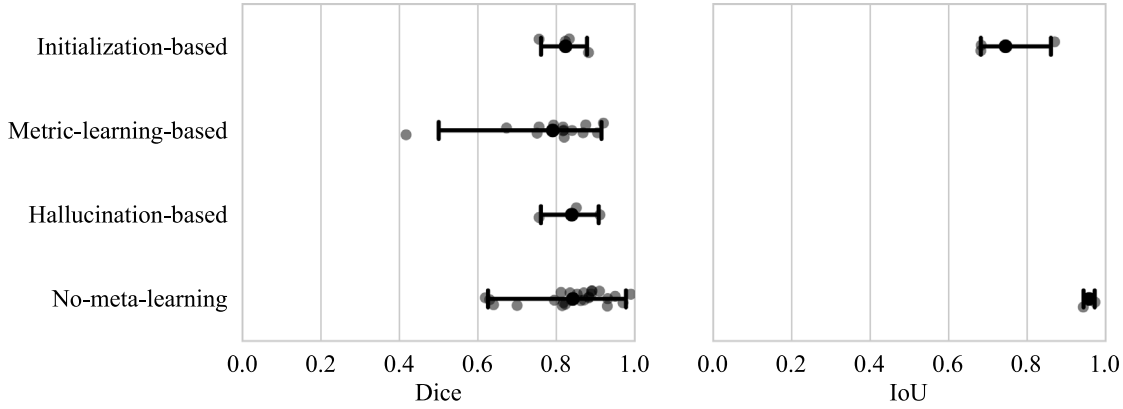
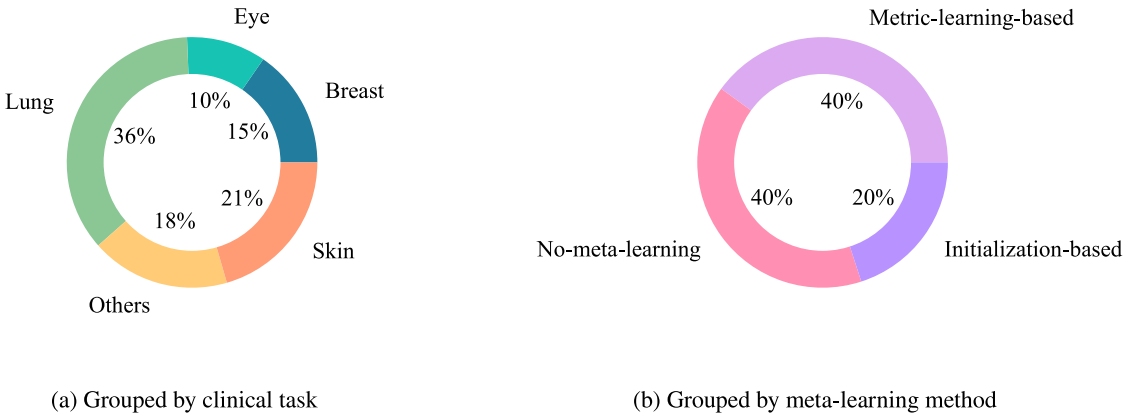


Fig. 8. Forest plot of segmentation studies performance based on Dice and IoU metrics. Studies are grouped by the meta-learning method employed.

Table 4
Standard pipeline steps adopted by segmentation studies.

Study ID	Pub. ref.	Pre-training	Training	Data augmentation
1	Blendowski, Nickisch, and Heinrich [49]	Self-supervised	None	None
2	Chan et al. [50]	None	Supervised	Registration-based
3	Chen et al. [51]	None	Supervised	Generative
4	Cui et al. [52]	None	Meta	Classical
5	Ding, Wangbin et al. [53]	None	Meta	Classical
6	Ding, Yu and Yang [54]	None	Semi-supervised and meta	Generative and registration-based
7	Farshad et al. [55]	Meta	Supervised	None
8	Feng et al. [56]	None	Meta	None
9	Gama, Oliveira and dos Santos [57]	None	Meta	None
10	Gama et al. [58]	None	Meta	None
11	Guo, Odu and Pedrosa [59]	None	Supervised	Classical
12	Hansen et al. [60]	Supervised	Self-supervised and meta	None
13	He et al. [42]	None	Supervised	Registration-based
14	He et al. [43]	None	Meta	Registration-based
15	Jenssen et al. [61]	Supervised	Self-supervised and meta	None
16	Joyce and Kozerke [62]	None	Self-supervised	Classical
17	Khadka et al. [47]	Supervised	Meta	None
18	Khaled, Han and Ghaleb [48]	None	Semi-supervised	None
19	Khandelwal and Yushkevich [63]	None	Meta	Classical
20	Kim et al. [64]	Meta	Meta	Classical
21	Li et al. [65]	None	Meta	Classical
22	Lu et al. [66]	Supervised	Semi-supervised	None
23	Lu and Ye [67]	Supervised	Supervised	None
24	Ma et al. [68]	None	Zero-shot	None
25	Niu et al. [69]	None	Meta	None
26	Ouyang et al. [70]	Supervised	Self-supervised and meta	None
27	Pham et al. [71]	None	Supervised	Classical
28	Pham, Dovletov and Pauli [72]	None	Supervised	None
29	Roy et al. [73]	None	Supervised	None
30	Roychowdhury et al. [44]	None	Supervised	Classical
31	Rutter, Lagergren and Flores [74]	None	Semi-supervised	Classical
32	Shen et al. [75]	None	Semi-supervised	Registration-based
33	Shen et al. [76]	Supervised	Semi-supervised	None
34	Shi et al. [45]	Supervised	Semi-supervised	Registration-based
35	Sun et al. [77]	Supervised	Meta	None
36	Tang et al. [78]	None	Meta	None
37	Tomar et al. [79]	Self-supervised	Supervised	Generative and registration-based
38	Wang et al. [80]	None	Semi-supervised	None
39	Wang et al. [81]	None	Supervised	Classical
40	Wang et al. [82]	None	Semi-supervised	Classical
41	Wang, Zhou and Zheng [83]	Supervised	Meta	None
42	Wang et al. [84]	None	Supervised	None
43	Wu, Xiao and Liang [85]	None	Supervised	Classical
44	Wu et al. [86]	None	Semi-supervised	None
45	Xu and Niethammer [46]	Supervised	Semi-supervised	Classical and registration-based
46	Yu et al. [87]	Supervised	Meta	Classical
47	Yuan, Esteva and Xu [88]	None	Supervised and meta	Classical
48	Zhao et al. [89]	None	Semi-supervised	Generative and registration-based
49	Zhao et al. [90]	None	Meta	Classical and generative
50	Zhou et al. [91]	Supervised	Supervised	None



(a) Grouped by clinical task (b) Grouped by meta-learning method

Fig. 9. Classification studies distribution grouped by clinical task and meta-learning method.

of the seen class samples were used to get SSL-based loss terms for the unseen class samples clustering stage. As a second step, they performed feature generation and used synthesized and actual features of unseen and seen classes to train a classifier.

Here, we present the findings of our classification papers analysis.

Clinical task. The classification papers covered a wide range of anatomical structures and regions, as well as various diseases. In the following, we provide a breakdown of articles categorized by the clinical task investigated: two out of 27 studies (7%) performed brain image classification, focusing on different types of tumours and MRI contrast types; six studies (22%) addressed breast image classification, with four concentrating on breast tumours and two on breast metastases involving nearby lymph nodes; two studies (7%) investigated cell image classification; two studies (7%) focused on cervix image classification; three studies (11%) pertained to colon image classification; four studies (15%) were dedicated to fundus eye image classification, with 2 investigating different diseases; one study (4%) dealt with liver disease classification; 11 studies (41%) involved lung image classification; one study (4%) concerned pancreas image classification; one study (4%) classified prostate tumour images; 7 studies (26%) addressed skin image classification, covering different diseases; one study (4%) investigated esophagus image classification; one study (4%) focused on stomach image classification. Note that [116] is not included in this analysis as it did not specify which anatomical structures were part of their study. An illustration of the classification studies distribution in terms of the five most popular clinical task investigated is provided in Fig. 9(a).

Meta-learning method. In the context of classification studies employing FSL, the distribution of meta-learning methods is as follows: six out of 27 studies (22%) utilized initialization-based methods; 10 studies (37%) opted for metric-learning-based algorithms; the remaining 11 studies (41%) did not incorporate any meta-learning technique. A representation of the classification studies distribution in terms of the meta-learning methods employed is provided in Fig. 9(b).

K-shot. Among the 27 selected studies in the classification domain using FSL, the training configurations were distributed as follows: 13 studies (48%) employed k-shot training with k ranging from 2 to 20; six studies (22%) utilized both OSL and FSL; one study (4%) used both FSL and ZSL; five studies (19%) exclusively performed OSL and two studies (7%) solely employed ZSL.

Image modality. In the following, we provide the distribution of studies in terms of the imaging modalities utilized: three studies out of 27 (11%) used CT images; three studies (11%) employed MRI images; seven studies (26%) utilized dermoscopic images; 11 studies (41%) relied on X-ray images; three studies (11%) involved fundus images; two studies (7%) made use of microscopic images; nine studies (33%) investigated histopathological images; one study (4%) utilized endoscopy images; one study (4%) involved cytological images and one study (4%) used OCT images.

Model evaluation. For assessing the robustness of the models, various evaluation techniques were employed: nine studies (33%) utilized ablation studies; one study (4%) conducted both ablation studies and cross-validation; one study (4%) solely relied on cross-validation; two studies (7%) repeated experiments multiple times for evaluation. The remaining 14 studies (52%) did not employ any specific model evaluation technique.

Statistical analysis. Fig. 10 displays forest plots that summarize the performance of FSL models (mean and 95% CI) for each clinical task considering both AUROC score and accuracy metrics. Conversely, Fig. 11 depicts the models' results grouped by the meta-learning method used. Note that the results of [95,116] were not included in the forest plots since they provide average results across different clinical tasks.

Standard pipeline. Table 7 outlines the learning frameworks employed in each step of the standard pipeline by the classification studies analysed in this review. In the following, we explore the distribution of the studies based on the techniques employed for each pipeline phase: one out of 27 studies (4%) employed a meta-learning algorithm for pre-training; 13 studies (48%) employed classical supervised pre-training; one study (4%) used unsupervised pre-training; 12 studies (44%) did not employ any pre-training stage. For training, fifteen studies (56%) utilized meta-learning; one study (4%) employed semi-supervised training; one study (4%) employed self-supervised training; nine studies (33%) used traditional supervised training; two studies (7%) employed zero-shot learning methods. Finally, concerning the data augmentation techniques, 10 out of 27 studies (37%) relied on classical data augmentation techniques; two studies (7%) utilized generative methods for data augmentation. The remaining 15 studies did not employ data augmentation.

4.2.3. Registration

We included six relevant studies, each focusing on medical registration as its primary task. Table 8 summarizes all the information extracted from the selected studies. In addition, we provide information concerning ROB and the applicability of each study in Table 9. Unlike segmentation and classification, medical image registration approaches for FSL generally do not employ meta-learning techniques but typically rely on an unsupervised learning framework where the model is trained until convergence using only the pair of images to be registered. An example of this is provided by Fechter et al. [118], who proposed a OSL registration method for periodic motion tracking that combines a U-Net architecture trained in a supervised way with a coarse-to-fine approach and a differential spatial transformer module. Two other

Table 5
FSL studies for medical image classification. CTC = Contrast-type Classification; DR = Diabetic Retinopath; IC = Intraretinal Cyst; LN = Limph Nodes; MAD = Mean Across Diseases; MAO = Mean Across Organs; MAS = Mean Across Structures; MAT = Mean Across Tumours.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
51	Ali et al. [95]	Prototypical network	5-shot	Accuracy: 0.906 (Endoscopic images, MAO)	Metric learning
52	Cai, Hu, and Zheng [96]	Prototypical network + Attention module (CBAM)	20-shot	Accuracy: 0.924 (Brain, MAT)	Metric learning
53	Cai et al. [97]	Pre-Moco diagnosis network (Pre-training+ Contrastive learning)	1-shot 5-shot 10-shot 20-shot	Accuracy: 0.832 (Skin, MAD) 0.675 (Eye, MAD)	Metric learning
54	Cano and Cruz-Roa [98]	Siamese Neural Network	1-shot	Accuracy: 0.908 (Breast, MAT)	Metric learning
55	Chen et al. [99]	2D CNN ranking + 2D CNN classification + Heatmap for segmentation	2-shot	AUROC: 0.883 (Breast, LN metastases)	None
56	Chou et al. [100]	Siamese Neural Network (Triple encoder + Triple loss)	1-shot	Accuracy: 0.986 (Brain, CTC)	None
57	Dai et al. [92]	Prior guided feature enhancement for few-shot medical image classification	3-shot 5-shot 10-shot	Accuracy: 0.851 (Brain, MAT) 0.960 (Skin, MAT) 0.803 (Cervix, MAT)	Metric learning
58	Huang, Huang and Tang [101]	One-shot anomaly detection framework	1-shot	AUROC: 0.961 (Eye, MAD) 0.955 (Lung, COVID)	None
59	Jiang et al. [102]	Autoencoder + Metric learner + Task learner (Transfer learning phase + Meta-learning phase)	1-shot 5-shot 10-shot	Accuracy: 0.762 (Cells, MAS) 0.762 (Colon, MAD) 0.506 (Lungs, MAD)	Metric learning
60	Jin et al. [103]	ViT-L/16 + ResNet50 + Metric-learning	1-shot 5-shot 8-shot	Accuracy: 0.346 (Lungs, MAD)	Metric learning
61	Mahapatra, Ge and Reyes [94]	Self-supervised clustering based generalized zero-shot learning	0-shot	Accuracy: 0.921 (Breast, LN metastases) 0.909 (Lungs, MAD) 0.942 (Eye, DR) 0.911 (Prostate, tumour)	None
62	Maicas et al. [93]	Pre and post-hoc diagnosis and interpretation + 3D DenseNet	4-shot	AUROC: 0,910 (Breast, tumour)	Initialization
63	Mohan et al. [104]	Siamese Network + Classifier	1-shot	Accuracy: 0.930 (Lung, COVID and Pneumonia)	None
64	Moukheiber et al. [105]	DeepVoro multi-label ensemble	5-shot 10-shot	AUROC: 0.679 (Lung, MAD)	Initialization-based
65	Naren, Zhu and Wang [106]	8 block VGG + MAML++	1-shot to 5-shot	Accuracy: 0.857 (Lung, COVID)	Initialization
66	Ouahab, Ben-Ahmed and Fernandez-Maloigne [107]	Self-attention augmented MAML	3-shot 5-shot	Accuracy: 0.819 (Skin, MAD) 0.703 (Lungs, MAD) AUROC: 0.843 (Skin, MAD) 0.734 (Lungs, MAD)	Initialization
67	Paul, Tang and Summers [108]	DenseNet-121 (feature extractor) + Autoencoder ensemble (classificator)	5-shot	F1-score: 0.440 (Lung, MAD) Recall: 0.490 (Lung, MAD)	None

(*continued on next page*)

Table 5 (*continued*).

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
68	Paul et al. [109]	DenseNet + Vanilla autoencoder	5-shot	F1-score: 0.470 (Lungs, MAD) AUROC: 0.647 (Lungs, MAD)	None
69	Paul et al. [110]	DenseNet + MVSE network + Self-training	0-shot	Recall: 0.454 (Lungs, MAD)	None
30	Roychowdhury et al. [44]	Echo state network (ParESN) + Target label selection algorithm (TLSA)	5-shot	Accuracy: 0.970 (Eye, IC)	None
70	Singh et al. [111]	MetaMed	3-shot 5-shot 10-shot	Accuracy: 0.864 (Breast, MAD) 0.843 (Skin, MAD) 0.934 (Cervix, MAT)	Initialization
71	Vetil et al. [112]	VAE + Distribution learning	0-shot 15-shot	AUROC: 0.789 (Pancreas)	None
72	Xiao et al. [113]	CNN feature extractor + classification prototype + similarity module + rectified corruption function	5-shot 10-shot	Accuracy: 0.874 (Skin, MAD)	Metric learning
73	Yan et al. [114]	Siamese-Prototypical Network	1-shot 5-shot	Accuracy: 0.686 (Skin, MAD) 0.608 (Liver, MAD) 0.626 (Colon, MAD)	Metric learning
74	Yarlagadda et al. [115]	Region proposal network + Inception-ResNet-v2 + Memory module with regional maximum activation of convolutions global descriptors	1-shot	Accuracy: 0.946 (Cells)	None
75	Zhang, Cui and Ren [116]	MAML	1-shot 3-shot 5-shot	Accuracy: 0.788 (VQA-RAD, MAS) 0.614 (PathVQA, MAS)	Initialization
76	Zhu et al. [117]	Query-relative loss + Adaptive hard margin + Prototypical Network/ Matching Network	1-shot 5-shot	Accuracy: 0.719 (Skin, MAD)	Metric learning

Table 6
ROB of FSL studies for medical image classification.

Study ID	Pub. ref.	Risk of bias					Applicability			
		Part.	Pred.	Out.	Analysis	Overall	Part.	Pred.	Out.	Overall
51	Ali et al. [95]	✓	✓	✓	✗	✗	✓	✓	✓	✓
52	Cai, Hu, and Zheng [96]	✓	✓	✓	✓	✓	✓	✓	✓	✓
53	Cai et al. [97]	✓	✓	✓	✓	✓	✓	✓	✓	✓
54	Cano and Cruz-Roa [98]	✓	✓	✓	✗	✗	✓	✓	✓	✓
55	Chen et al. [2]	✓	✓	✓	✗	✗	✓	✓	✓	✓
56	Chou et al. [100]	✓	✓	✓	✗	✗	✓	✓	✓	✓
57	Dai et al. [92]	✓	✓	✓	✓	✓	✓	✓	✓	✓
58	Huang, Huang and Tang [101]	✓	✓	✓	✓	✓	✓	✓	✓	✓
59	Jiang et al. [102]	✓	✓	✓	✓	✓	✓	✓	✓	✓
60	Jin et al. [103]	✓	✓	✓	✓	✓	✓	✓	✓	✓
61	Mahapatra, Ge and Reyes [94]	✓	✓	✓	✓	✓	✓	✓	✓	✓
62	Maicas et al. [93]	✓	✓	✓	✓	✓	✓	✓	✓	✓
63	Mohan et al. [104]	✓	✓	✓	✗	✗	✓	✓	✓	✓
64	Moukheiber et al. [105]	✓	✓	✓	✓	✓	✓	✓	✓	✓
65	Naren, Zhu and Wang [106]	✓	✓	✓	✓	✓	✓	✓	✓	✓
66	Ouahab, Ben-Ahmed and Fernandez-Maloigne [107]	✓	✓	✓	✓	✓	✓	✓	✓	✓
67	Paul, Tang and Summers [108]	✓	✓	✓	✓	✓	✓	✓	✓	✓
68	Paul et al. [109]	✓	✓	✓	✓	✓	✓	✓	✓	✓
69	Paul et al. [110]	✓	✓	✓	✓	✓	✓	✓	✓	✓
30	Roychowdhury et al. [44]	✓	✓	✓	✗	✗	✓	✓	✓	✓
70	Singh et al. [111]	✓	✓	✓	✓	✓	✓	✓	✓	✓
71	Vetil et al. [112]	✓	✓	✓	✓	✓	✓	✓	✓	✓
72	Xiao et al. [113]	✓	✓	✓	✓	✓	✓	✓	✓	✓
73	Yan et al. [114]	✓	✓	✓	✓	✓	✓	✓	✓	✓
74	Yarlagadda et al. [115]	✓	✓	✓	✗	✗	✓	✓	✓	✓
75	Zhang, Cui and Ren [116]	✓	✓	✓	✗	✗	✓	✓	✓	✓
76	Zhu et al. [117]	✓	✓	✓	✓	✓	✓	✓	✓	✓

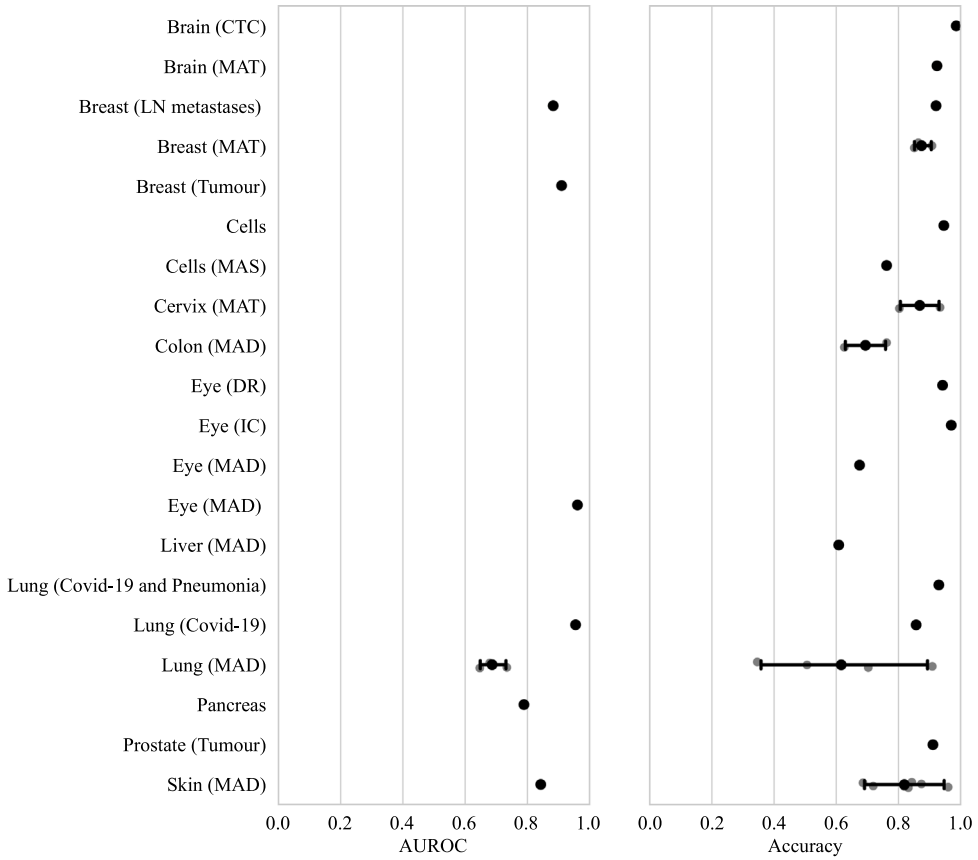


Fig. 10. Forest plot of classification studies performance based on AUROC and Accuracy metrics. Studies are grouped by the clinical task investigated. CTC = Contrast-type Classification; DR = Diabetic Retinopathy; IC = Intraretinal Cyst; LN = Limph Nodes; MAD = Mean Across Diseases; MAS = Mean Across Structures; MAT = Mean Across Tumours.

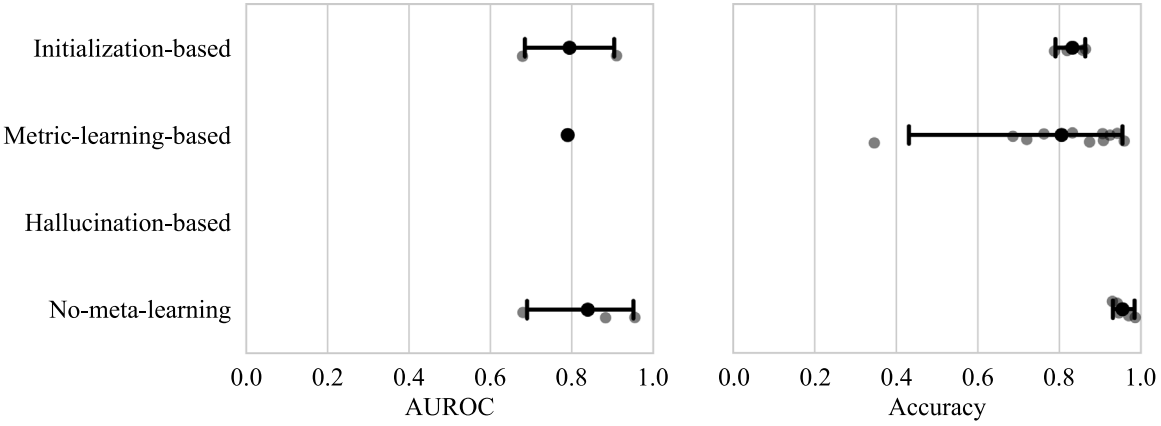


Fig. 11. Forest plot of classification studies performance based on AUROC and Accuracy metrics. Studies are grouped by the meta-learning method employed.

works, instead, leveraged joint training between segmentation and registration networks and provided weak supervision to the registration network to improve performance. For example, Shi et al. [45] employed a segmentation module to predict pseudo-labels on unlabelled data to provide weak supervision for the registration network.

Here, we present the findings derived from our comprehensive analysis of the registration papers.

Clinical task. The selected registration papers investigate a range of anatomical regions. Here's the breakdown of studies w.r.t. the clinical task addressed: three out of 6 studies (50%) explored brain registration; one study (17%) focused on the registration of knee bones and cartilages; three studies (50%) delved into heart registration; three studies (50%) concentrated on lung registration; one study (17%) pertained to abdominal registration; one study (17%) dealt with cervical vertebra registration. We represent the registration studies distribution w.r.t. clinical task in Fig. 12(a).

Meta-learning method. In the registration studies domain, all of the selected papers (100%) did not employ the meta-learning paradigm, as illustrated in Fig. 12(b).

K-shot. Among the six selected studies, the distribution of k-shot configurations is as follows: two studies (33%) solely employed FSL; three studies (50%) investigated solely OSL; one study (17%) performed both FSL and OSL.

Image modality. In the context of registration studies, the distribution of imaging modalities used among the selected papers is as follows: four

Table 7
Standard pipeline steps adopted by classification studies.

Study ID	Pub. ref.	Pre-training	Training	Data augmentation
51	Ali et al. [95]	Supervised	Meta	None
52	Cai, Hu, and Zheng [96]	None	Meta	Classical
53	Cai et al. [97]	Supervised	Meta	Classical
54	Cano and Cruz-Roa [98]	None	Meta	None
55	Chen et al. [99]	Unsupervised	Supervised	None
56	Chou et al. [100]	None	Supervised	None
57	Dai et al. [92]	Supervised	Meta	None
58	Huang, Huang and Tang [101]	None	Supervised	Generative
59	Jiang et al. [102]	Supervised	Meta	Classical
60	Jin et al. [103]	None	Meta	Classical
61	Mahapatra, Ge and Reyes [94]	Supervised	Zero-shot self-supervised and supervised	Generative
62	Maicas et al. [93]	Meta	Supervised	None
63	Mohan et al. [104]	Supervised	Supervised	Classical
64	Moukheiber et al. [105]	Supervised	Meta	None
65	Naren, Zhu and Wang [106]	None	Meta	None
66	Ouahab, Ben-Ahmed and Fernandez-Maloigne [107]	Supervised	Meta	Classical
67	Paul, Tang and Summers [108]	Supervised	Supervised	None
68	Paul et al. [110]	Supervised	Zero-shot and semi-supervised	None
69	Paul et al. [109]	Supervised	Supervised	None
30	Roychowdhury et al. [44]	None	Supervised	Classical
70	Singh et al. [111]	None	Meta	Classical
71	Vétil et al. [112]	None	Zero-shot and supervised	Classical
72	Xiao et al. [113]	None	Meta	None
73	Yan et al. [114]	Supervised	Meta	Classical
74	Yarlagadda et al. [115]	Supervised	Supervised	None
75	Zhang, Cui and Ren [116]	None	Meta	None
76	Zhu et al. [117]	None	Meta	None

Table 8
FSL studies for medical image registration. MAS = Mean Across Structures.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
77	Fechter and Baltas [118]	U-net + Differential spatial transformer module	1-shot	ALD: 1.49 (Lungs) Dice: 0.860 (Heart)	None
78	Ferrante et al. [119]	U-net + Unsupervised learning	1-shot	Dice: 0.920 (Heart) 0.890 (Lungs)	None
79	He et al. [120]	Perception-correspondence registration	5-shot	Dice: 0.857 (Heart, MAS) 0.867 (Cervical vertebra, MAS) 0.800 (Brain, MAS)	None
34	Shi et al. [45]	Joint registration and segmentation self-training framework	5-shot	Dice: 0.759 (Brain, MAS) 0.539 (Abdomen, MAS)	None
45	Xu and Niethammer [46]	Semi-supervised learning + Segmentation network + Registration network	1-shot 5-shot 10-shot	Dice: 0.759 (Brain, MAS) 0.539 (Abdomen, MAS)	None
80	Zhang et al. [121]	CNN + Spatial transformer + similarity loss + smooth loss + cyclic loss	1-shot	TRE: 1.03 (Lung)	None

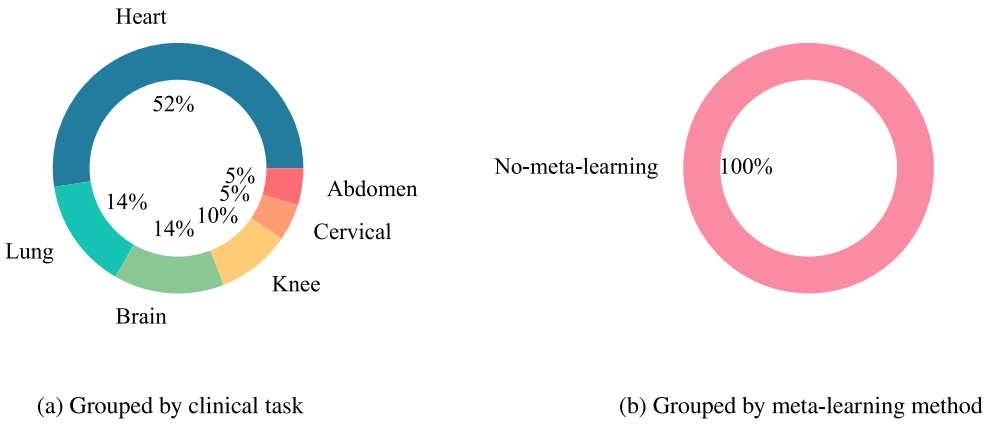
Table 9
ROB of FSL studies for medical image registration.

Study ID	Pub. ref.	Risk of bias					Applicability			
		Part.	Pred	Out.	Analysis	Overall	Part.	Pred.	Out.	Overall
77	Fechter, Baltas [118]	✓	✓	✓	✓	✓	✓	✓	✓	✓
78	Ferrante et al. [119].	✓	✓	✓	✓	✓	✓	✓	✓	✓
79	He et al. [120]	✓	✓	✓	✓	✓	✓	✓	✓	✓
34	Shi et al. [45]	✓	✓	✓	✓	✓	✓	✓	✓	✓
45	Xu and Niethammer [46]	✓	✓	✓	✓	✓	✓	✓	✓	✓
80	Zhang et al. [121]	✓	✓	✓	✓	✓	✓	✓	✓	✓

out of six studies (67%) employ CT acquisitions; five out of six studies (83%) utilize MRI images; one out of six studies (17%) involves X-ray images.

Model evaluation. To investigate the models' robustness, several evaluation techniques were employed: two studies (33%) utilized only ablation studies and one study (17%) used cross-validation. The remaining studies (50%) did not employ any specific model evaluation technique.

Statistical analysis. Fig. 13 displays forest plots that summarize the performance of FSL models (mean and 95% CI) for each clinical task



(a) Grouped by clinical task (b) Grouped by meta-learning method

Fig. 12. Registration studies distribution grouped by clinical task and meta-learning method.

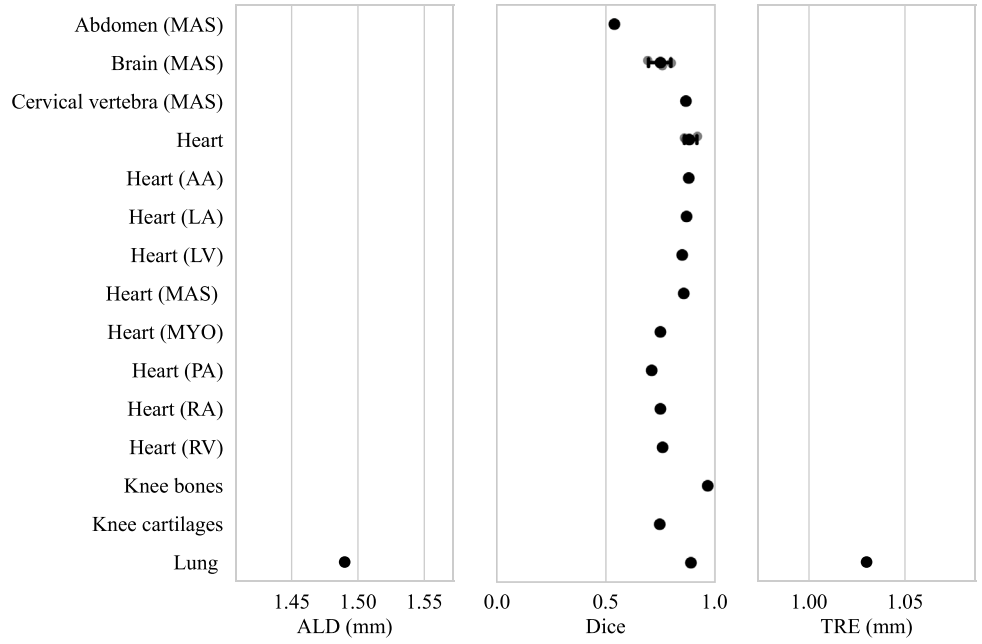


Fig. 13. Forest plot of registration studies performance based on ALD, Dice, and TRE metrics. Studies are grouped by the clinical task investigated. AA = Ascending Aorta; LA = Left Atrium; LV = Left Ventricle; MAS = Mean Across Structures; MYO = Myocardium; PA = Pulmonary Artery; RA = Right Atrium; RV = Right Ventricle.

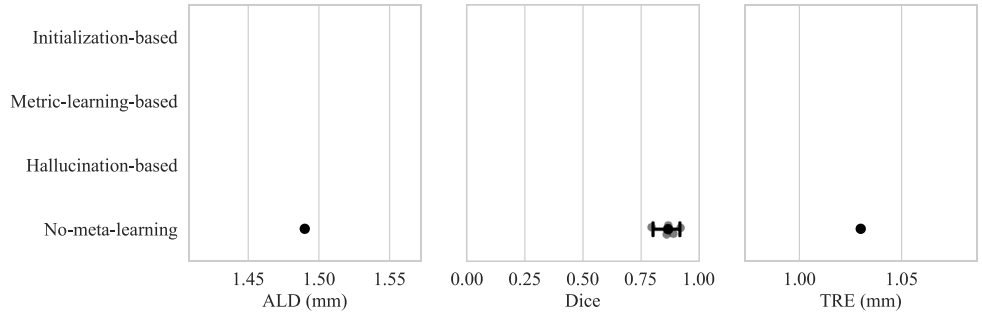


Fig. 14. Forest plot of registration studies performance based on ALD, Dice and TRE metrics. Studies are grouped by the meta-learning method employed.

considering ALD, Dice score and TRE metrics. Conversely, Fig. 14 depicts the models' results grouped by the meta-learning method used.

Standard pipeline. Table 10 outlines the learning frameworks employed in each step of the standard pipeline by the registration studies analysed in this review. In the following, we explore the distribution of the studies based on the techniques employed for each pipeline phase: two out of six studies (33%) utilized a classical supervised approach in both pre-training and training steps. The remaining four studies (67%) did not employ any pre-training and adopted an unsupervised approach during training. Two out of six studies (33%) used classical data augmentation techniques. The other four studies (67%) did not exploit data augmentation.

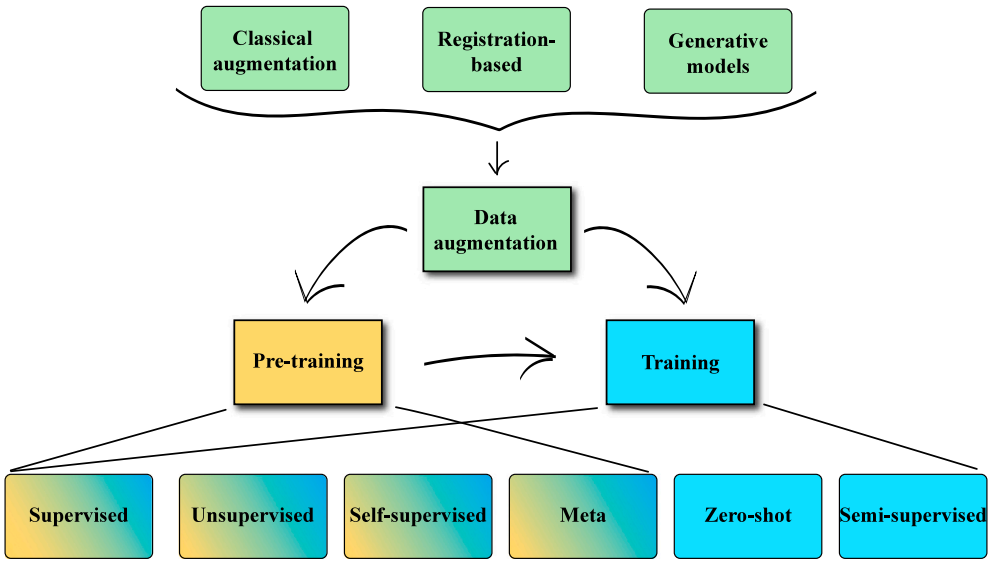


Fig. 15. An illustration of the proposed standard pipeline. Learning frameworks suitable for both pre-training and training phases are highlighted in shaded boxes. For optimal clarity, please view this figure in colour.

Table 10
Standard pipeline steps adopted by registration studies.

Study ID	Pub. ref.	Pre-training	Training	Data augmentation
77	Fechter, Baltas [118]	None	Unsupervised	None
78	Ferrante et al. [119]	None	Unsupervised	None
79	He et al. [120]	None	Unsupervised	Classical
34	Shi et al. [45]	Supervised	Supervised	None
45	Xu and Niethammer [46]	Supervised	Supervised	Classical
80	Zhang et al. [121]	None	Unsupervised	None

4.3. Standard pipeline

Fig. 15 presents the proposed standardized pipeline for FSL in medical imaging. This pipeline was derived from a comprehensive analysis of all the selected studies across classification, segmentation, and registration tasks.

5. Discussion

This review analysed 80 FSL studies applied to the medical imaging domain. We organized the studies into three main categories based on their primary outcome: segmentation, classification, and registration. Within each category, we extracted relevant information from each study, including the utilized algorithm or pipeline, employed meta-learning methods, the amount of labelled training data, and the highest performance achieved. Additionally, we conducted a ROB and applicability analysis using the PROBAST method for each study. We further performed a statistical analysis, grouping the studies by both the investigated clinical task and the employed meta-learning techniques. Furthermore, we analysed the learning frameworks used in each study across three key phases: pre-training, training, and data augmentation. Based on this comprehensive analysis, we defined a general pipeline encompassing the techniques shared across the studies and relevant to the three targeted outcomes.

In the following paragraphs, we discuss the results of our analysis based on the objectives outline provided in Section 1.2. First, we examine the distribution of studies across various outcomes. Next, we discuss the results of our statistical analysis as well as additional aspects like data usage, imaging modalities, and model robustness assessment methods. Later, we provide insights into our main finding, i.e. the definition of a standard pipeline. Based on this comprehensive analysis, we offer the reader insights into potential future research directions. This includes highlighting techniques that show promise due to their effectiveness and identifying areas where further investigation is warranted. To conclude, we highlight the main takeaways from our research and identify any limitations in our analysis.

Studies distribution by outcome. As shown in Fig. 5, segmentation tasks dominate the landscape of FSL studies in medical imaging, accounting for 61% of the reviewed studies. Classification tasks follow at 32%, and registration tasks make up the remaining 7%. In the following paragraphs, we delve deeper into how these studies are distributed, considering both the clinical task and the employed meta-learning methods.

Studies and results distribution by clinical task. Fig. 6(a) offers an insightful overview of the distribution of clinical tasks related to the segmentation outcome. Notably, the heart emerges as the most extensively investigated anatomical structure, comprising 34% of the studies. Following closely behind are the kidney, spleen, and liver, each accounting for 13% of the research. The brain also features significantly, representing 10% of the studies. As for classification, Fig. 9(a) shows that the lungs take the lead, constituting the primary focus in 36% of the research. The skin follows closely with 21%, while the breast and eye account for 15% and 10%, respectively. Lastly, according to Fig. 12(a), among registration tasks, the heart emerges as the most commonly investigated, representing 52% of the studies focus, followed by lungs and brain, accounting for 14% of the studies and the knee for 10%. Finally, the cervical vertebra and abdomen represent the main application in 5% of the studies.

Moving to our statistical analysis of results, Fig. 7 provides insights into the performance of segmentation tasks in terms of the Dice score and IoU across various anatomical structures. Notably, femur segmentation demonstrates the highest Dice score, although it is worth mentioning that only one study addresses this task, making the result only partially reliable. In contrast, AA and LV segmentation exhibit consistently good average results across multiple peer studies, achieving Dice scores of 0.89 and 0.88, respectively. The worst-performing segmentation task appears to be the prostate, with a mean Dice score of 0.42 across different structures. As for the IoU metric, hip, knee, and phalanx segmentation provide the best results. It is worth noting that the present findings are based on a limited pool of studies, potentially impacting their generalizability. On the other hand, lung segmentation demonstrates a high IoU of 0.91, with a small CI, across several studies,

indicating robust and consistent performance. Conversely, prostate segmentation consistently yields lower IoU scores, with a 0.23 IoU for the segmentation of the peripheral zone, indicating room for improvement in this specific task. In terms of the classification outcome, even though Fig. 10 indicates that classifying brain images into different contrast types exhibits the highest accuracy, it is essential to acknowledge that this specific task is considered relatively straightforward. In contrast, the more challenging task of skin lesion classification achieves a high mean accuracy (0.82) across a considerable number of studies. Conversely, liver segmentation demonstrates the lowest average performance, yielding an accuracy of 0.61 across various diseases. Concerning the AUROC metric, eye image classification achieves the highest performance, scoring 0.96. Conversely, lung segmentation exhibits the lowest average performance (0.69 AUROC across various diseases). However, it is noteworthy that this task involved an ample number of studies. Finally, it is worth mentioning that a few studies evaluated the classification performance of different lung diseases employing the F1-score and Recall metrics. However, the results in both cases were poor, scoring below 0.5. Finally, concerning the registration task, Fig. 13 suggests that FSL achieves its highest registration performance in terms of Dice score for knee bones. However, cautious interpretation is advised due to data from a single study. Conversely, whole heart image registration demonstrates a consistently high mean Dice score (0.88) with a narrow CI, reflecting findings from a larger number of studies. Regarding the ALD and TRE metrics, it is noteworthy that they were exclusively employed for lung image registration.

Studies and results distributions by meta-learning technique. Fig. 6(b) reveals that most studies (55%) did not utilize any meta-learning method for segmentation. Among those that did, meta-learning approaches (26%) were the most popular, followed by initialization-based (13%) and hallucination-based methods (6%). Shifting the focus on classification tasks, Fig. 9(b) reveals that 40% of classification studies did not leverage meta-learning algorithms. However, of those who did (60%), metric learning-based methods were the most common (40%), followed by initialization-based methods (20%). Notably, no studies used hallucination-based methods for classification. Finally, in terms of registration, we found that none of the registration studies employed meta-learning approaches.

Our statistical analysis of segmentation studies performance (Fig. 8) reveals interesting findings. Both no-meta-learning methods and hallucination-based methods achieve the highest mean Dice scores (0.84). However, no-meta-learning methods exhibit a wider CI, likely due to their application in a higher number of studies. Metric-learning approaches, the most common meta-learning method in segmentation, show slightly lower performance with a mean Dice score of 0.79 and a larger CI. Concerning the IoU metric, this was only employed for evaluating initialization-based and no-meta-learning methods. In this case, no-meta-learning methods significantly outperform the others, achieving a notably better score with a smaller CI. Analysing classification tasks (Fig. 11), we observe that no-meta-learning studies consistently achieve the highest accuracy (with a narrow CI). Conversely, metric-learning methods exhibit lower performance (mean accuracy of 0.81 and a larger CI). Regarding AUROC, no-meta-learning algorithms again yield the best average score (0.84). Interestingly, both initialization-based and metric-learning methods have a mean AUROC of 0.79, although metric-learning methods have a wider CI. Finally, Fig. 14 depicts the mean performance metrics (ALD, Dice, and TRE) for registration tasks, all utilizing non-meta-learning approaches.

Additional analyses. Our analysis of the amount of labelled training data employed reveals that in segmentation studies, the majority (58%) incorporate one or more labelled data samples, with a significant portion (40%) utilizing OSL and only 2% employing ZSL. A similar pattern emerges in classification studies, where 70% involve at least one labelled sample, 19% solely rely on OSL, and 7% utilize ZSL. However, registration studies diverge, with the primary approach being OSL (50%). The remaining studies are split between using multiple labelled examples (40%) and ZSL (17%).

In terms of imaging modalities, MRI emerges as the dominant modality across segmentation and registration studies, accounting for 60% and 83% of studies, respectively. In contrast, classification studies primarily rely on X-ray imaging, with 41% of studies utilizing this modality.

Finally, considering the model's robustness evaluation, 74% of the segmentation studies incorporate some form of model robustness evaluation. This typically involves conducting ablation studies and or employing cross-validation techniques. Conversely, only half of the classification and registration studies utilize robustness assessment methods. Ablation studies are the most common approach in these domains.

Standard pipeline discussion and recommendations. As a primary objective, this work aimed to define a standard pipeline reflecting the shared methods employed across all the studies. Building upon the findings from Fig. 15, in the following paragraphs, we elaborate on the specific arrangement of various learning frameworks within the three pipeline elements: pre-training, training, and data augmentation.

Among the analysed studies that utilize pre-training, we identified four main learning paradigms: supervised, unsupervised, self-supervised, and meta-learning. The most common approach, namely supervised learning, involves leveraging pre-trained convolutional backbones on large datasets for transfer learning, as seen in the works of Hansen et al. [60] and Ali et al. [95]. For unsupervised learning, Chen et al. [99] is the only example. Specifically, they employed an unsupervised cell ranking as a pre-training step for classifying sentinel breast cancer lymph node images. Self-supervised learning is used as a pre-training phase in segmentation studies alone, as exemplified by Tomar et al. [79]. Here, they pre-trained a style encoder in a self-supervised manner using a volumetric contrastive loss to make it learn clustering similar styled images. The so-trained model is then jointly optimized during the training phase with an appearance and a flow model by minimizing a shared loss. Finally, we found that meta-learning is also used as a pre-training step, as for Maicas et al. [93], who meta-pre-trained their backbone on various tasks before fine-tuning on the task of interest using a fully-supervised approach.

Overall, our analysis revealed that only a few studies employ some pre-training strategy, with supervised learning being the most popular approach. However, while transfer learning offers advantages, several studies highlighted the potential of SSL in outperforming supervised pre-training [122,123]. Ericsson et al. [122] demonstrated that leading SSL methods usually surpass supervised pre-training as a source of knowledge transfer. Specifically in the medical domain, Zhao et al. [123] highlighted that SSL models can exceed baseline models using only 1% labelled data compared to training from scratch. This efficiency makes SSL particularly valuable in FSL. Additionally, SSL approaches have been shown to learn colour-invariant features [122], a potentially beneficial feature in medical imaging where data is often grayscale. In terms of meta-learning, despite its primary use in downstream task training, its ability to improve generalization can also provide a strong foundation for pre-training, as demonstrated by Maicas et al. [93]. Given these findings, we recommend that future FSL research delve deeper into the potential of SSL as a pre-training strategy. Building on the success of SSL in surpassing supervised approaches and achieving high accuracy with minimal labelled data, FSL research can explore how to best leverage SSL for few-shot tasks. This includes investigating which SSL objectives are most beneficial for FSL settings and how to best integrate this pre-training with fine-tuning techniques. Additionally, inspired by Maicas et al. [93], a promising direction is to leverage the generalization strength of meta-learning for robust model pre-training, followed by fine-tuning with appropriate techniques. This could involve designing new meta-learning algorithms tailored to extract meaningful features from a large pretext dataset and leveraging this knowledge for downstream FSL tasks.

For the main training phase, a wider range of learning frameworks compared to pre-training have been utilized. Indeed, while all four

pre-training paradigms (supervised, unsupervised, self-supervised, and meta-learning) are still present, semi-supervised and ZSL approaches are also employed. Meta-learning emerges as the dominant choice, particularly for segmentation and classification. Interestingly, registration studies lean solely on unsupervised or weakly-supervised learning, potentially due to the task nature, which is often addressed by training the model until convergence, mimicking classical registration algorithms. However, research by Kanter et al. [124] and Park et al. [125] demonstrates the potential of meta-learning for image registration outside the FSL context. This suggests that future research should explore leveraging meta-learning frameworks to enhance registration models in FSL settings, mirroring its success in segmentation and classification tasks. Potential approaches include using the MAML method to learn adaptable initializations for registration models and combining meta-learning's ability to learn from a few labelled examples with the strengths of existing unsupervised or weakly supervised learning registration methods that utilize unlabelled data.

Beyond meta-learning, supervised learning remains the second most employed approach for segmentation and classification tasks. This framework utilizes various strategies to enhance model performance and generalizability. In segmentation studies, data augmentation is a popular choice for addressing FSL in a classical supervised way, often involving jointly trained segmentation and registration models where the registration output serves as realistic additional data (e.g., Chan et al. [50]; He et al. [42]). Other studies, instead, leverage classical data augmentation methods (Guo et al. [59]). Additionally, some works focus on improving the generalization capabilities of the model itself, like Roy et al. [73], who proposed a two-armed architecture with a "conditioner arm" processing labelled slices to generate task-specific representations to be provided to a "segmenter arm" performing segmentation on query images. Similar to segmentation, classification studies often employ data augmentation techniques to achieve good performance with supervised learning frameworks [44,101], or rely on robust pre-training with unsupervised or meta-learning approaches, as seen in the works of Chen et al. [99] and Maicas et al. [93], respectively. Unlike segmentation and classification tasks, which primarily rely on supervised learning, only two registration studies employ a supervised approach, i.e., Shi et al. [45] and Xu and Niethammer [46], who leverage pseudo-labels generated by the segmentation network to train the registration network in a weakly-supervised manner. On the other hand, the most popular learning framework among registration studies involves an unsupervised method where the model is trained until convergence using only the two images to be registered. Interestingly, no segmentation or classification study in our analysis employed unsupervised learning in their training phase.

The rest of the segmentation and classification studies rely on SSL, ZSL, and semi-supervised learning techniques. These approaches address the data scarcity by leveraging additional information from unlabelled data or by encoding distinguishing properties of objects to be transferred to unseen classes. For instance, Paul et al. [110] leveraged semi-supervised learning and GZSL by proposing a multi-view semantic embedding network that maps a feature embedding extracted from an X-ray image to a semantic signature corresponding to the disease present in the image itself and combines it with a self-training approach to improve diagnosis on unseen classes. Regarding the SSL domain, we found that, in segmentation studies, the main thread is to couple SSL with a meta-learning framework. An example of this approach is provided by Hansen et al. [60], who propose a detection-inspired-network trained self-supervised end-to-end in an episodic manner. On the other hand, in classification, only Mahapatra et al. [94] leveraged SSL, coupling it with supervised learning for GZSL.

While supervised learning remains prevalent in FSL, our analysis highlights the potential of SSL for achieving similar or higher results with less labelled data (e.g., [122,123]). Following this, FSL research should delve deeper into two key areas: exploring hybrid approaches that combine SSL pre-training with meta-learning, leveraging the strengths of both for rapid adaptation in few-shot tasks (inspired by promising results in segmentation) and investigating semi-supervised learning, which has shown success both independently (e.g., [126,127]) and when combined with meta-learning (e.g., [128–130]). By focusing on these less data-hungry learning paradigms, FSL research can unlock significant advancements while utilizing readily available unlabelled data better.

Finally, speaking about data augmentation techniques, our analysis shows that three data-augmentations varieties are employed among FSL for medical imaging studies: classical, registration-based and generative augmentation. Classical data augmentation techniques often regard, among others, geometric image transformations, random erasing, and colour jitter. In addition, some studies provided a custom interpretation of classical augmentation, as Guo et al. [59], who proposed a 3D augmentation of MRI images, which becomes the key to performing kidney segmentation in a supervised way using a limited number of data. Another popular type of data augmentation consists of leveraging registration models jointly trained with segmentation ones and using the output of the registration model as a realistic form of data augmentation to improve segmentation performance [42,43,45,46]. Related to registration-based augmentations are also the works of Chan et al. [50] and Shen et al. [75]. In the first one, the authors leverage the Daemons registration algorithm to generate additional images and labels to train a segmentation network. In the second one, the authors propose a fluid-based augmentation method which generates anatomically meaningful images via interpolation from the geodesic subspace underlying the provided samples.

The last data augmentation category we identified among the examined studies is generative-based. This category considers all the studies that propose data generation schemes for data augmentation. For example, Huang et al. [101] proposed the AugPaste framework, where lesion patches are processed with data augmentation operations and the transformed lesion patches are randomly pasted to the normal samples via MixUp, yielding synthetic anomalous samples. In addition, several segmentation studies combined generative and registration-based approaches by employing image registration w.r.t. an atlas to gain shape deformations and intensity changes between the two images, which are then exploited to synthesize additional images [46,54,79].

Apart from classical data augmentation techniques, which are widely used across various research domains due to their proven ability to enhance model performance by introducing diversity, two typical research strands can be identified in the field of FSL. Firstly, using images generated by registration models and traditional registration algorithms as data augmentation techniques for segmentation models stands out. This approach produces augmented images that are more realistic and informative for the final task. Secondly, more sophisticated techniques, such as AugPaste and methods involving deformations and intensity variations to create synthetic examples, offer tailored examples for specific tasks. These techniques provide more informative content for the model training process. Therefore, it is recommended to prioritize these advanced forms of data augmentation to expand the dataset. They have been shown to yield excellent results, such as 0.961 AUROC for the classification of eye diseases [101], 0.851 Dice for brain segmentation [54], and 0.911 mean Dice for heart structure segmentation [43].

Closing remarks. Given that the meta-learning framework is the prevailing method for addressing FSL, we devoted a significant portion of our analysis to it. Our investigation revealed variations in the adoption and success rates of different meta-learning techniques. Specifically, methods based on metric-learning have garnered substantial attention, while hallucination-based techniques still require further exploration despite showing promising results. Furthermore, we showed that, on average, non-meta-learning methods outperform meta-learning approaches. Several factors may contribute to this behaviour. We argue that one of the most likely reasons lies in the inherent nature of meta-learning algorithms. Indeed, meta-learning algorithms are designed to

learn a generalized approach to solving tasks by training on diverse tasks. Consequently, there often exists a class mismatch between the tasks seen during training and those encountered during testing. In other words, models are evaluated on tasks involving classes that were never seen during training. Conversely, in classical supervised frameworks, for example, models are specifically trained to excel at a particular task, and the limited availability of data is usually mitigated by leveraging robust augmentation strategies and/or unlabelled data, as in self-supervised and semi-supervised approaches. Therefore, to interpret these results effectively, one must go beyond simple average performance metrics and consider the actual generalizability of the model across different tasks, particularly when additional data are not employed.

In terms of the clinical task investigated, our analysis highlights that the heart, abdomen, and lungs have been the primary areas of focus in the examined studies. This is likely due to the availability of well-established benchmark datasets such as CHAOS [131], MS-CMRSeg [132], and NIH Chest X-ray [133]. Nevertheless, there is an unexplored potential for researchers to delve into relatively less-explored medical applications, including the prostate and digestive organs. These areas remain critical in clinical practice, and the adoption of FSL techniques could open new avenues for their utilization in the clinical domain.

Moreover, our observations reveal that certain studies, especially those focusing on classification and registration tasks, may lack comprehensive model investigation analyses. This gap in research practices could result in incomplete and unreliable performance assessments. Additionally, we identified issues pertaining to the reporting of ROB in some studies. Indeed, many of these studies lack clarity in explaining their approach to addressing FSL tasks, even when asserting the utilization of reduced amounts of labelled data.

In light of our findings, we encourage future researchers in the field to consider the following actions:

- Focus the research on the development and improvement of meta-learning methods, as this is the framework that guarantees better generalizability for the model. Specifically, evaluate the combined use of meta-learning and SSL as this showed promising results.
- Prefer the use of meta-learning, SSL and semi-supervised learning approaches w.r.t. fully-supervised as several studies demonstrated their success in overreaching fully-supervised methods with much less data.
- Consider the use of meta-learning methods for performing registration tasks, as all the investigated studies at the moment rely on unsupervised and weakly-supervised approaches.
- Among the meta-learning methods, explore the hallucination-based methods in more depth, as they are the least investigated despite being promising.
- Expand the scope of medical applications investigated beyond heart, abdomen and lung anatomical structures.
- Prioritize thorough model validation and comprehensive analyses to facilitate fair comparisons and the practical implementation of FSL models in clinical settings.

Limitations. Our study includes some limitations. Primarily, in our research, we focused on multidisciplinary and computer science-related databases, omitting medical databases. This may have affected the comprehensiveness of our literature review. Additionally, we did not encompass FSL papers addressing the image reconstruction task; instead, we concentrated on segmentation, classification, and registration studies.

Still valid in 2024? Since our systematic research covers papers up to January 2023, we wanted to assess whether the considerations made were still valid in the latest papers published. We reviewed several recent papers in the field, specifically, those we believe have had a significant impact on FSL for medical imaging literature. In the FSL for medical image segmentation domain, Cheng et al. [134] propose leveraging a prototypical meta-learning approach for medical image segmentation (abdomen, heart, and prostate), generating multiple representative descriptors instead of a single prototype to comprehensively represent the commonalities of the class distribution. Leng et al. [135] propose Self-Sampling Meta Segment Anything Model (SAM), an optimized SAM [136] version for few-shot medical image segmentation. They redesign the backbone by removing the original SAM's prompt encoder and mask decoder, employing a self-sampling prompt encoder and Flexible Mask Attention Decoder. Additionally, they embed a meta-learning-based optimizer layered on top based on MAML++. Their approach is evaluated on abdomen segmentation in CT and MRI images. Both [134,135] did not leverage data augmentation or pre-training steps. Another segmentation method built on the SAM model was proposed by Xie et al. [137], who suggest a fine-tuning strategy for SAM by reformulating SAM's mask decoder to accept few-shot embeddings as prompts. They use classical data augmentation techniques to augment the training set, do not leverage pre-training, and perform classical fully supervised training. Their approach is evaluated by segmenting bones (tibia, femur, and vertebrae) and heart elements.

Concerning the classification task, Ansari et al. [138] propose integrating meta-learning, specifically MAML, into a ViT while also leveraging extensive classical data augmentation techniques (such as noise injection and masking) to increase the training dataset size. They evaluated their approach to several medical tasks, including classifying breast cancer from histopathology images, cervical cells from Pap smear slides, dermoscopic skin lesion images, and retina images to classify eye disorders. Besides proposing new approaches, other works apply existing meta-learning frameworks to specific clinical tasks. For example, Maia et al. [139] apply both Siamese Neural Networks and Prototypical Networks to classify histopathology images of oral cancer. Similarly, Icsik et al. [140] apply and compare Prototypical Networks and MAML for ultrasound breast cancer images. Both studies do not use any data augmentation or pre-training techniques.

Analysing the above-mentioned papers, we find clear consistency with the assessments made in our systematic review. Meta-learning remains the most employed framework for addressing FSL, with metric-learning-based methods (most notably Prototypical Networks) and initialization-based methods (most notably MAML) prevalent in segmentation and classification tasks. In segmentation, the most investigated anatomical structures continue to be the abdomen and heart, while in classification, the focus remains on breast, skin, and eye conditions. Unfortunately, we did not find any relevant new research in the domain of FSL for medical image registration. Compared to our findings, the main difference observed in the new papers is the extensive use of the SAM, a powerful agnostic segmentation model that is gaining significant success in the DL literature, with various fine-tuning solutions tailored to specific tasks. We suggest readers monitor this trend, as it offers promising solutions in the FSL medical image segmentation domain. Specifically, it would be valuable to investigate using SAM as a base architecture and performing meta-training with a prototypical approach since only MAML has been investigated so far. Additionally, exploring pre-training SAM leveraging SSL on an unlabelled dataset could further enhance performance.

6. Conclusions

This work presented a comprehensive analysis of SOTA FSL techniques for the medical imaging domain. We extracted the most relevant information from each study and performed a statistical analysis of the studies' results w.r.t. both the clinical task investigated and the meta-learning algorithm employed. Our work culminated with designing a standard methodological pipeline shared across all the studies. Our findings not only reveal the limitations of existing SOTA methods but also pave the way for future advancements. We advocate for

prioritizing approaches that enhance model generalizability and reduce reliance on labelled data, such as self-supervised, semi-supervised, and meta-learning techniques. Additionally, we encourage investigating the potential of hallucination-based methods for FSL in medical imaging. Furthermore, we emphasize the need to expand research across diverse clinical domains as well as employ techniques to assess the robustness and generalizability potential of developed models. These recommendations aim to guide researchers towards achieving significant advancements in this domain and ultimately bridge the gap between FSL and real-world clinical applications.

CRediT authorship contribution statement

Eva Pachetti: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Sara Colantonio:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

None declared.

Acknowledgements

This study was funded by the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement No 952159 (ProCAncer-I) and by the Regional Project PAR FAS Tuscany-NAVIGATOR. The funders had no role in the study design, data collection, analysis, interpretation, or manuscript writing.

References

[1] Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, et al. Language models are few-shot learners. Adv Neural Inf Process Syst 2020;33:1877–901, URL https://splab.sdu.edu.cn/GPT3.pdf.

[2] Chen W-Y, Liu Y-C, Kira Z, Wang Y-CF, Huang J-B. A closer look at few-shot classification. 2019, arXiv preprint arXiv:1904.04232.

[3] Rezaei M, Shahidi M. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. Intell-Based Med 2020;3:100005. http://dx.doi.org/10.1016/j.ibmed.2020.100005, Publisher: Elsevier.

[4] Wang Y, Yao Q, Kwok JT, Ni LM. Generalizing from a few examples: A survey on few-shot learning. ACM Comput Surv (csur) 2020;53(3):1–34. http://dx.doi.org/10.1145/3386252, Publisher: ACM New York, NY, USA.

[5] Kotia J, Kotwal A, Bharti R, Mangrulkar R. Few shot learning for medical imaging. In: Machine learning algorithms for industrial applications. Cham: Springer International Publishing; 2021, p. 107–32. http://dx.doi.org/10.1007/978-3-030-50641-4_7.

[6] Nayem J, Hasan SS, Amina N, Das B, Ali MS, Ahsan MM, Raman S. Few shot learning for medical imaging: A comparative analysis of methodologies and formal mathematical framework. 2023, arXiv preprint arXiv:2305.04401.

[7] Hospedales T, Antoniou A, Micaelli P, Storkey A. Meta-learning in neural networks: A survey. IEEE Trans Pattern Anal Mach Intell 2021;44(9):5149–69. http://dx.doi.org/10.1109/TPAMI.2021.3079209, Publisher: IEEE.

[8] Xian Y, Schiele B, Akata Z. Zero-shot learning - the good, the bad and the ugly. In: Proceedings of the IEEE conference on computer vision and pattern recognition. CVPR, 2017, http://dx.doi.org/10.1109/CVPR.2017.328.

[9] Pourpanah F, Abdar M, Luo Y, Zhou X, Wang R, Lim CP, Wang X-Z, Wu QMJ. A review of generalized zero-shot learning methods. IEEE Trans Pattern Anal Mach Intell 2023;45(4):4051–70. http://dx.doi.org/10.1109/TPAMI.2022.3191696.

[10] Sun X, Gu J, Sun H. Research progress of zero-shot learning. Appl Intell 2021;51:3600–14. http://dx.doi.org/10.1007/s10489-020-02075-7.

[11] Rohrbach M, Stark M, Schiele B. Evaluating knowledge transfer and zero-shot learning in a large-scale setting. In: CVPR 2011. IEEE; 2011, p. 1641–8. http://dx.doi.org/10.1109/CVPR.2011.5995627.

[12] Lampert CH, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization. IEEE Trans Pattern Anal Mach Intell 2013;36(3):453–65. http://dx.doi.org/10.1109/TPAMI.2013.140.

[13] Zhang Z, Saligrama V. Zero-shot learning via semantic similarity embedding. In: Proceedings of the IEEE international conference on computer vision. ICCV, 2015, http://dx.doi.org/10.1109/ICCV.2015.474.

[14] Changpinyo S, Chao W-L, Gong B, Sha F. Synthesized classifiers for zero-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. CVPR, 2016, http://dx.doi.org/10.1109/CVPR.2016.575.

[15] Palatucci M, Pomerleau D, Hinton GE, Mitchell TM. Zero-shot learning with semantic output codes. In: Bengio Y, Schuurmans D, Lafferty J, Williams C, Culotta A, editors. Advances in neural information processing systems. Vol. 22, Curran Associates, Inc.; 2009, URL https://proceedings.neurips.cc/paper_files/paper/2009/file/1543843a4723ed2ab08e18053ae6dc5b-Paper.pdf.

[16] Akata Z, Reed S, Walter D, Lee H, Schiele B. Evaluation of output embeddings for fine-grained image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 2927–36. http://dx.doi.org/10.1109/CVPR.2015.7298911.

[17] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: International conference on machine learning. PMLR; 2017, p. 1126–35, URL http://proceedings.mlr.press/v70/finn17a/finn17a.pdf.

[18] Verma VK, Brahma D, Rai P. Meta-learning for generalized zero-shot learning. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 34, 2020, p. 6062–9. http://dx.doi.org/10.1609/aaai.v34i04.6069.

[19] Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms. 2018, arXiv preprint arXiv:1803.02999.

[20] Ravi S, Larochelle H. Optimization as a model for few-shot learning. In: International conference on learning representations. 2016, URL https://openreview.net/forum?id=rJY0-Kcll.

[21] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80. http://dx.doi.org/10.1007/978-3-642-24797-2_4, Publisher: MIT press.

[22] Li K, Malik J. Learning to optimize. 2016, arXiv preprint arXiv:1606.01885.

[23] Bellman R. A Markovian decision process. J Math Mech 1957;679–84, URL https://www.jstor.org/stable/24900506.

[24] Santoro A, Bartunov S, Botvinick M, Wierstra D, Lillicrap T. Meta-learning with memory-augmented neural networks. In: International conference on machine learning. PMLR; 2016, p. 1842–50, URL http://proceedings.mlr.press/v48/santoro16.pdf.

[25] Graves A, Wayne G, Danihelka I. Neural turing machines. 2014, arXiv preprint arXiv:1410.5401.

[26] Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a" siamese" time delay neural network. Adv Neural Inf Process Syst 1993;6. URL https://proceedings.neurips.cc/paper_files/paper/1993/file/288cc0ff022877bd3df94bc9360b9c5d-Paper.pdf.

[27] Koch G, Zemel R, Salakhutdinov R, et al. Siamese neural networks for one-shot image recognition. In: ICML deep learning workshop. Vol. 2, Lille; 2015, URL https://www.cs.utoronto.ca/~rsalakhu/papers/oneshot1.pdf.

[28] Hoffer E, Ailon N. Deep metric learning using triplet network. In: International workshop on similarity-based pattern recognition. Springer; 2015, p. 84–92. http://dx.doi.org/10.1007/978-3-319-24261-3_7.

[29] Ji Z, Wang H, Pang Y, Shao L. Dual triplet network for image zero-shot learning. Neurocomputing 2020;373:90–7. http://dx.doi.org/10.1016/j.neucom.2019.09.062, URL https://www.sciencedirect.com/science/article/pii/S092523121931330X.

[30] Vinyals O, Blundell C, Lillicrap T, Wierstra D, et al. Matching networks for one shot learning. Adv Neural Inf Process Syst 2016;29. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/90e1357833654983612fb05e3ec9148c-Paper.pdf.

[31] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. Adv Neural Inf Process Syst 2017;30. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf.

[32] Elhoseiny M, Saleh B, Elgammal A. Write a classifier: Zero-shot learning using purely textual descriptions. In: Proceedings of the IEEE international conference on computer vision. 2013, p. 2584–91. http://dx.doi.org/10.1109/ICCV.2013.321.

[33] Santoro A, Raposo D, Barrett DG, Malinowski M, Pascanu R, Battaglia P, Lillicrap T. A simple neural network module for relational reasoning. Adv Neural Inf Process Syst 2017;30. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/e6acf4b0f69f6f6e60e9a815938aa1ff-Paper.pdf.

[34] Sung F, Yang Y, Zhang L, Xiang T, Torr PH, Hospedales TM. Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 1199–208. http://dx.doi.org/10.1109/CVPR.2018.00131.

[35] Hariharan B, Girshick R. Low-shot visual recognition by shrinking and hallucinating features. In: Proceedings of the IEEE international conference on computer vision. 2017, p. 3018–27. http://dx.doi.org/10.1109/ICCV.2017.328.

[36] Wang Y-X, Girshick R, Hebert M, Hariharan B. Low-shot learning from imaginary data. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7278–86. http://dx.doi.org/10.1109/CVPR.2018.00760.

[37] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021;88:105906. http://dx.doi.org/10.1136/bmj.n71, Publisher: Elsevier.

[38] Gu M, Vesal S, Kosti R, Maier A. Few-shot unsupervised domain adaptation for multi-modal cardiac image segmentation. In: Bildverarbeitung für die medizin 2022: proceedings, german workshop on medical image computing, Heidelberg, June 26-28, 2022. Springer; 2022, p. 20–5. http://dx.doi.org/10.1007/978-3-658-36932-3_5.

[39] Keaton MR, Zaveri RJ, Doretto G. CellTranspose: Few-shot domain adaptation for cellular instance segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2023, p. 455–66. http://dx.doi.org/10.1109/WACV56688.2023.00053.

[40] Li S, Sui X, Fu J, Fu H, Luo X, Feng Y, Xu X, Liu Y, Ting DS, Goh RSM. Few-shot domain adaptation with polymorphic transformers. In: International conference on medical image computing and computer-assisted intervention. Springer; 2021, p. 330–40. http://dx.doi.org/10.1007/978-3-030-87196-3_31.

[41] Wolff RF, Moons KG, Riley RD, Whiting PF, Westwood M, Collins GS, Reitsma JB, Kleijnen J, Mallett S, Group P. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Internal Med 2019;170(1):51–8. http://dx.doi.org/10.7326/M18-1376.

[42] He Y, Li T, Yang G, Kong Y, Chen Y, Shu H, Coatrieux J-L, Dillenseger J-L, Li S. Deep complementary joint model for complex scene registration and few-shot segmentation on medical images. In: European conference on computer vision. Springer; 2020, p. 770–86. http://dx.doi.org/10.1007/978-3-030-58523-5_45.

[43] He Y, Ge R, Qi X, Chen Y, Wu J, Coatrieux J-L, Yang G, Li S. Learning better registration to learn better few-shot medical image segmentation: Authenticity, diversity, and robustness. IEEE Trans Neural Netw Learn Syst 2022. http://dx.doi.org/10.1109/TNNLS.2022.3190452, Publisher: IEEE.

[44] Roychowdhury S. Few shot learning framework to reduce inter-observer variability in medical images. In: 2020 25th international conference on pattern recognition. ICPR, IEEE; 2021, p. 4581–8. http://dx.doi.org/10.1109/ICPR48806.2021.9412620.

[45] Shi H, Lu L, Yin M, Zhong C, Yang F. Joint few-shot registration and segmentation self-training of 3D medical images. Biomed Signal Process Control 2023;80:104294. http://dx.doi.org/10.1016/j.bspc.2022.10429, Publisher: Elsevier.

[46] Xu Z, Niethammer M. DeepAtlas: Joint semi-supervised learning of image registration and segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22. Springer; 2019, p. 420–9. http://dx.doi.org/10.1007/978-3-030-32245-8_47.

[47] Khadka R, Jha D, Hicks S, Thambawita V, Riegler MA, Ali S, Halvorsen P. Meta-learning with implicit gradients in a few-shot setting for medical image segmentation. Comput Biol Med 2022;143:105227. http://dx.doi.org/10.1016/j.compbiomed.2022.105227, Publisher: Elsevier.

[48] Khaled A, Han J-J, Ghaleb TA. Multi-model medical image segmentation using multi-stage generative adversarial networks. IEEE Access 2022;10:28590–9. http://dx.doi.org/10.1109/ACCESS.2022.3158342, Publisher: IEEE.

[49] Blendowski M, Nickisch H, Heinrich MP. How to learn from unlabeled volume data: Self-supervised 3d context feature learning. In: International conference on medical image computing and computer-assisted intervention. Springer; 2019, p. 649–57. http://dx.doi.org/10.1007/978-3-030-32226-7_72.

[50] Chan S, Huang C, Bai C, Ding W, Chen S. Res2-UNeXt: a novel deep learning framework for few-shot cell image segmentation. Multimedia Tools Appl 2022;81(10):13275–88. http://dx.doi.org/10.1007/s11042-021-10536-5, Publisher: Springer.

[51] Chen C, Qin C, Ouyang C, Li Z, Wang S, Qiu H, Chen L, Tarroni G, Bai W, Rueckert D. Enhancing MR image segmentation with realistic adversarial data augmentation. Med Image Anal 2022;82:102597. http://dx.doi.org/10.1016/j.media.2022.102597, Publisher: Elsevier.

[52] Cui H, Wei D, Ma K, Gu S, Zheng Y. A unified framework for generalized low-shot medical image segmentation with scarce data. IEEE Trans Med Imaging 2020;40(10):2656–71. http://dx.doi.org/10.1109/TMI.2020.3045775, Publisher: IEEE.

[53] Ding W, Li L, Zhuang X, Huang L. Cross-modality multi-atlas segmentation using deep neural networks. In: International conference on medical image computing and computer-assisted intervention. Springer; 2020, p. 233–42. http://dx.doi.org/10.1007/978-3-030-59716-0_23.

[54] Ding Y, Yu X, Yang Y. Modeling the probabilistic distribution of unlabeled data for one-shot medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 35, 2021, p. 1246–54. http://dx.doi.org/10.1609/aaai.v35i2.16212, Issue: 2.

[55] Farshad A, Makarevich A, Belagiannis V, Navab N. MetaMedSeg: Volumetric meta-learning for few-shot organ segmentation. In: Domain adaptation and representation transfer: 4th MICCAI workshop, DART 2022, held in conjunction with MICCAI 2022, Singapore, September 22, 2022, proceedings. Springer; 2022, p. 45–55. http://dx.doi.org/10.1007/978-3-031-16852-9_5.

[56] Feng R, Zheng X, Gao T, Chen J, Wang W, Chen DZ, Wu J. Interactive few-shot learning: Limited supervision, better medical image segmentation. IEEE Trans Med Imaging 2021;40(10):2575–88. http://dx.doi.org/10.1109/TMI.2021.3060551, Publisher: IEEE.

[57] Gama PH, Oliveira H, dos Santos JA. Learning to segment medical images from few-shot sparse labels. In: 2021 34th SIBGRAPI conference on graphics, patterns and images. SIBGRAPI, IEEE; 2021, p. 89–96. http://dx.doi.org/10.1109/SIBGRAPI54419.2021.00021.

[58] Gama PHT, Oliveira HN, Marcato J, Dos Santos J. Weakly supervised few-shot segmentation via meta-learning. IEEE Trans Multimed 2022. http://dx.doi.org/10.1109/TMM.2022.3162951, Publisher: IEEE.

[59] Guo J, Odu A, Pedrosa I. Deep learning kidney segmentation with very limited training data using a cascaded convolution neural network. PLoS One 2022;17(5):e0267753. http://dx.doi.org/10.1371/journal.pone.0267753, Publisher: Public Library of Science San Francisco, CA USA.

[60] Hansen S, Gautam S, Jenssen R, Kampffmeyer M. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. Med Image Anal 2022;78:102385. http://dx.doi.org/10.1016/j.media.2022.102385, Publisher: Elsevier.

[61] Jenssen R, Hansen S, Gautam S, Kampffmeyer M. A self-guided anomaly detection-inspired few-shot segmentation network. 2022, URL https://hdl.handle.net/10037/28499.

[62] Joyce T, Kozerke S. 3D medical image synthesis by factorised representation and deformable model learning. In: International workshop on simulation and synthesis in medical imaging. Springer; 2019, p. 110–9. http://dx.doi.org/10.1007/978-3-030-32778-1_12.

[63] Khandelwal P, Yushkevich P. Domain generalizer: A few-shot meta learning framework for domain generalization in medical imaging. In: Domain adaptation and representation transfer, and distributed and collaborative learning. Springer; 2020, p. 73–84. http://dx.doi.org/10.1007/978-3-030-60548-3_8.

[64] Kim S, An S, Chikontwe P, Park SH. Bidirectional rnn-based few shot learning for 3d medical image segmentation. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 35, 2021, p. 1808–16. http://dx.doi.org/10.1609/aaai.v35i3.16275, Issue: 3.

[65] Li Y, Fu Y, Yang Q, Min Z, Yan W, Huisman H, Barratt D, Prisacariu VA, Hu Y. Few-shot image segmentation for cross-institution male pelvic organs using registration-assisted prototypical learning. In: 2022 IEEE 19th international symposium on biomedical imaging. ISBI, IEEE; 2022, p. 1–5. http://dx.doi.org/10.1109/ISBI52829.2022.9761453.

[66] Lu Y, Zheng K, Li W, Wang Y, Harrison AP, Lin C, Wang S, Xiao J, Lu L, Kuo C-F, et al. Contour transformer network for one-shot segmentation of anatomical structures. IEEE Trans Med Imaging 2020;40(10):2672–84. http://dx.doi.org/10.1109/TMI.2020.3043375, Publisher: IEEE.

[67] Lu Q, Ye C. Knowledge transfer for few-shot segmentation of novel white matter tracts. In: International conference on information processing in medical imaging. Springer; 2021, p. 216–27. http://dx.doi.org/10.1007/978-3-030-78191-0_17.

[68] Ma S, Li X, Tang J, Guo F. A zero-shot method for 3d medical image segmentation. In: 2021 IEEE international conference on multimedia and expo. ICME, IEEE; 2021, p. 1–6. http://dx.doi.org/10.1109/ICME51207.2021.9428261.

[69] Niu Y, Luo Z, Lian S, Li L, Li S, Song H. Symmetrical supervision with transformer for few-shot medical image segmentation. In: 2022 IEEE international conference on bioinformatics and biomedicine. BIBM, IEEE; 2022, p. 1683–7. http://dx.doi.org/10.1109/BIBM55620.2022.9995238.

[70] Ouyang C, Biffi C, Chen C, Kart T, Qiu H, Rueckert D. Self-supervised learning for few-shot medical image segmentation. IEEE Trans Med Imaging 2022. http://dx.doi.org/10.1109/TMI.2022.3150682, Publisher: IEEE.

[71] Pham DD, Lausen M, Dovletov G, Serong S, Landgraeber S, Jäger M, Pauli J. U-net in constraint few-shot settings. In: Bildverarbeitung für die medizin 2020. Springer; 2020, p. 280–5. http://dx.doi.org/10.1007/978-3-658-29267-6_62.

[72] Pham DD, Dovletov G, Pauli J. Using anatomical priors for deep 3D one-shot segmentation. In: BIOIMAGING. 2021, p. 174–81, URL https://www.scitepress.org/Papers/2021/103031/103031.pdf.

[73] Roy AG, Siddiqui S, Pölsterl S, Navab N, Wachinger C. 'Squeeze & excite'guided few-shot segmentation of volumetric images. Med Image Anal 2020;59:101587. http://dx.doi.org/10.1016/j.media.2019.101587, Publisher: Elsevier.

[74] Rutter EM, Lagergren JH, Flores KB. A convolutional neural network method for boundary optimization enables few-shot learning for biomedical image segmentation. In: Domain adaptation and representation transfer and medical image learning with less labels and imperfect data. Springer; 2019, p. 190–8. http://dx.doi.org/10.1007/978-3-030-33391-1_22.

[75] Shen Z, Xu Z, Olut S, Niethammer M. Anatomical data augmentation via fluid-based image registration. In: International conference on medical image computing and computer-assisted intervention. Springer; 2020, p. 318–28. http://dx.doi.org/10.1007/978-3-030-59716-0_31.

[76] Shen X, Zhang G, Lai H, Luo J, Lu J, Luo Y. PoissonSeg: Semi-supervised few-shot medical image segmentation via poisson learning. In: 2021 IEEE international conference on bioinformatics and biomedicine. BIBM, IEEE; 2021, p. 1513–8. http://dx.doi.org/10.1109/BIBM52615.2021.9669727.

[77] Sun L, Li C, Ding X, Huang Y, Chen Z, Wang G, Yu Y, Paisley J. Few-shot medical image segmentation using a global correlation network with discriminative embedding. Comput Biol Med 2022;140:105067. http://dx.doi.org/10.1016/j.compbiomed.2021.105067, Publisher: Elsevier.

[78] Tang H, Liu X, Sun S, Yan X, Xie X. Recurrent mask refinement for few-shot medical image segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 3918–28. http://dx.doi.org/10.1109/ICCV48922.2021.00389.

[79] Tomar D, Bozorgtabar B, Lortkipanidze M, Vray G, Rad MS, Thiran J-P. Self-supervised generative style transfer for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022, p. 1998–2008. http://dx.doi.org/10.1109/WACV51458.2022.00180.

[80] Wang S, Cao S, Wei D, Wang R, Ma K, Wang L, Meng D, Zheng Y. LT-Net: Label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 9162–71. http://dx.doi.org/10.1109/CVPR42600.2020.00918.

[81] Wang S, Cao S, Wei D, Xie C, Ma K, Wang L, Meng D, Zheng Y. Alternative baselines for low-shot 3D medical image segmentation—An atlas perspective. In: Proceedings of the AAAI conference on artificial intelligence. Vol. 35, 2021, p. 634–42. http://dx.doi.org/10.1609/aaai.v35i1.16143, Issue: 1.

[82] Wang W, Xia Q, Hu Z, Yan Z, Li Z, Wu Y, Huang N, Gao Y, Metaxas D, Zhang S. Few-shot learning by a cascaded framework with shape-constrained pseudo label assessment for whole heart segmentation. IEEE Trans Med Imaging 2021;40(10):2629–41. http://dx.doi.org/10.1109/TMI.2021.3053008, Publisher: IEEE.

[83] Wang R, Zhou Q, Zheng G. Few-shot medical image segmentation regularized with self-reference and contrastive learning. In: Medical image computing and computer assisted intervention–MICCAI 2022: 25th international conference, Singapore, September 18–22, 2022, proceedings, part IV. Springer; 2022, p. 514–23. http://dx.doi.org/10.1007/978-3-031-16440-8_49.

[84] Wang H, Li Q, Yuan Y, Zhang Z, Wang K, Zhang H. Inter-subject registration-based one-shot segmentation with alternating union network for cardiac MRI images. Med Image Anal 2022;79:102455. http://dx.doi.org/10.1016/j.media.2022.102455, Publisher: Elsevier.

[85] Wu H, Xiao F, Liang C. Dual contrastive learning with anatomical auxiliary supervision for few-shot medical image segmentation. In: Computer vision–ECCV 2022: 17th European conference, Tel aviv, Israel, October 23–27, 2022, proceedings, part XX. Springer; 2022, p. 417–34. http://dx.doi.org/10.1007/978-3-031-20044-1_24.

[86] Wu Y, Zheng B, Chen J, Chen DZ, Wu J. Self-learning and one-shot learning based single-slice annotation for 3D medical image segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2022: 25th international conference, Singapore, September 18–22, 2022, proceedings, part VIII. Springer; 2022, p. 244–54. http://dx.doi.org/10.1007/978-3-031-16452-1_24.

[87] Yu Q, Dang K, Tajbakhsh N, Terzopoulos D, Ding X. A location-sensitive local prototype network for few-shot medical image segmentation. In: 2021 IEEE 18th international symposium on biomedical imaging. ISBI, IEEE; 2021, p. 262–6. http://dx.doi.org/10.1109/ISBI48211.2021.9434008.

[88] Yuan Z, Esteva A, Xu R. MetaHistoSeg: A python framework for meta learning in histopathology image segmentation. In: Deep generative models, and data augmentation, labelling, and imperfections. Springer; 2021, p. 268–75. http://dx.doi.org/10.1007/978-3-030-88210-5_27.

[89] Zhao A, Balakrishnan G, Durand F, Guttag JV, Dalca AV. Data augmentation using learned transformations for one-shot medical image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019, p. 8543–53. http://dx.doi.org/10.1109/CVPR.2019.00874.

[90] Zhao Z, Zhou F, Zeng Z, Guan C, Zhou SK. Meta-hallucinator: Towards few-shot cross-modality cardiac image segmentation. In: Medical image computing and computer assisted intervention–MICCAI 2022: 25th international conference, Singapore, September 18–22, 2022, proceedings, part v. Springer; 2022, p. 128–39. http://dx.doi.org/10.1007/978-3-031-16443-9_13.

[91] Zhou H-Y, Liu H, Cao S, Wei D, Lu C, Yu Y, Ma K, Zheng Y. Generalized organ segmentation by imitating one-shot reasoning using anatomical correlation. In: International conference on information processing in medical imaging. Springer; 2021, p. 452–64. http://dx.doi.org/10.1007/978-3-030-78191-0_35.

[92] Dai Z, Yi J, Yan L, Xu Q, Hu L, Zhang Q, Li J, Wang G. PFEMed: Few-shot medical image classification using prior guided feature enhancement. Pattern Recognit 2023;134:109108. http://dx.doi.org/10.1016/j.patcog.2022.109108, Publisher: Elsevier.

[93] Maicas G, Bradley AP, Nascimento JC, Reid I, Carneiro G. Pre and post-hoc diagnosis and interpretation of malignancy from breast DCE-MRI. Med Image Anal 2019;58:101562. http://dx.doi.org/10.1016/j.media.2019.101562, Publisher: Elsevier.

[94] Mahapatra D, Ge Z, Reyes M. Self-supervised generalized zero shot learning for medical image classification using novel interpretable saliency maps. IEEE Trans Med Imaging 2022. http://dx.doi.org/10.1109/TMI.2022.3163232, Publisher: IEEE.

[95] Ali S, Bhattarai B, Kim T-K, Rittscher J. Additive angular margin for few shot learning to classify clinical endoscopy images. In: International workshop on machine learning in medical imaging. Springer; 2020, p. 494–503. http://dx.doi.org/10.1007/978-3-030-59861-7_50.

[96] Cai A, Hu W, Zheng J. Few-shot learning for medical image classification. In: International conference on artificial neural networks. Springer; 2020, p. 441–52. http://dx.doi.org/10.1007/978-3-030-61609-0_35.

[97] Cai A, Chen L, Chen Y, Fang J, Sun M, Chuan Z. Pre-MocoDiagnosis: Few-shot ophthalmic diseases recognition using contrastive learning. In: 2022 IEEE international conference on bioinformatics and biomedicine. BIBM, IEEE; 2022, p. 2059–66. http://dx.doi.org/10.1109/BIBM55620.2022.9994890.

[98] Cano F, Cruz-Roa A. An exploratory study of one-shot learning using Siamese convolutional neural network for histopathology image classification in breast cancer from few data examples. In: 15th international symposium on medical information processing and analysis. Vol. 11330, SPIE; 2020, p. 66–73. http://dx.doi.org/10.1117/12.2546488.

[99] Chen J, Jiao J, He S, Han G, Qin J. Few-shot breast cancer metastases classification via unsupervised cell ranking. IEEE/ACM Trans Comput Biol Bioinform 2019;18(5):1914–23. http://dx.doi.org/10.1109/TCBB.2019.2960019, Publisher: IEEE.

[100] Chou Y, Remedios SW, Butman JA, Pham DL. Automatic classification of MRI contrasts using a deep Siamese network and one-shot learning. In: Medical imaging 2022: image processing. Vol. 12032, SPIE; 2022, p. 110–4. http://dx.doi.org/10.1117/12.2613052.

[101] Huang W, Huang Y, Tang X. AugPaste: One-shot anomaly detection for medical images. In: Ophthalmic medical image analysis: 9th international workshop, OMIA 2022, held in conjunction with MICCAI 2022, Singapore, Singapore, September 22, 2022, proceedings. Springer; 2022, p. 1–11. http://dx.doi.org/10.1007/978-3-031-16525-2_1.

[102] Jiang H, Gao M, Li H, Jin R, Miao H, Liu J. Multi-learner based deep meta-learning for few-shot medical image classification. IEEE J Biomed Health Inf 2022;27(1):17–28. http://dx.doi.org/10.1109/JBHI.2022.3215147, Publisher: IEEE.

[103] Jin Y, Lu H, Zhu W, Yan K, Gao Z, Li Z. CTFC: A convolution and visual transformer based classifier for few-shot chest X-ray images. In: 2021 2nd international conference on artificial intelligence and computer engineering (ICAICE). IEEE; 2021, p. 616–22. http://dx.doi.org/10.1109/ICAICE54393.2021.00122.

[104] Mohan V. Detection of COVID-19 from chest X-ray images: A deep learning approach. In: 2021 ethics and explainability for responsible data science. EE-RDS, IEEE; 2021, p. 1–7. http://dx.doi.org/10.1109/EE-RDS53766.2021.9708594.

[105] Moukheiber D, Mahindre S, Moukheiber L, Moukheiber M, Wang S, Ma C, Shih G, Peng Y, Gao M. Few-shot learning geometric ensemble for multi-label classification of chest X-Rays. In: Data augmentation, labelling, and imperfections: second MICCAI workshop, DALI 2022, held in conjunction with MICCAI 2022, Singapore, September 22, 2022, proceedings. Springer; 2022, p. 112–22. http://dx.doi.org/10.1007/978-3-031-17027-0_12.

[106] Naren T, Zhu Y, Wang MD. COVID-19 diagnosis using model agnostic meta-learning on limited chest X-ray images. In: Proceedings of the 12th ACM conference on bioinformatics, computational biology, and health informatics. 2021, p. 1–9. http://dx.doi.org/10.1145/3459930.3469517.

[107] Ouahab A, Ben-Ahmed O, Fernandez-Maloigne C. A self-attentive meta-learning approach for image-based few-shot disease detection. In: Resource-efficient medical image analysis: first MICCAI workshop, REMIA 2022, Singapore, September 22, 2022, proceedings. Springer; 2022, p. 115–25. http://dx.doi.org/10.1007/978-3-031-16876-5_12.

[108] Paul A, Tang Y-X, Summers RM. Fast few-shot transfer learning for disease identification from chest x-ray images using autoencoder ensemble. In: Medical imaging 2020: computer-aided diagnosis. Vol. 11314, SPIE; 2020, p. 33–8. http://dx.doi.org/10.1117/12.2549060.

[109] Paul A, Shen TC, Peng Y, Lu Z, Summers RM. Learning few-shot chest X-ray diagnosis using images from the published scientific literature. In: 2021 IEEE 18th international symposium on biomedical imaging. ISBI, IEEE; 2021, p. 344–8. http://dx.doi.org/10.1109/ISBI48211.2021.9434059.

[110] Paul A, Shen TC, Lee S, Balachandar N, Peng Y, Lu Z, Summers RM. Generalized zero-shot chest x-ray diagnosis through trait-guided multi-view semantic embedding with self-training. IEEE Trans Med Imaging 2021;40(10):2642–55. http://dx.doi.org/10.1109/TMI.2021.3054817, Publisher: IEEE.

[111] Singh R, Bharti V, Purohit V, Kumar A, Singh AK, Singh SK. MetaMed: Few-shot medical image classification using gradient-based meta-learning. Pattern Recognit 2021;120:108111. http://dx.doi.org/10.1016/j.patcog.2021.108111, Publisher: Elsevier.

[112] Vétil R, Abi-Nader C, Bône A, Vullierme M-P, Rohé M-M, Gori P, Bloch I. Learning shape distributions from large databases of healthy organs: applications to zero-shot and few-shot abnormal pancreas detection. In: Medical image computing and computer assisted intervention–MICCAI 2022: 25th international conference, Singapore, September 18–22, 2022, proceedings, part II. Springer; 2022, p. 464–73. http://dx.doi.org/10.1007/978-3-031-16434-7_45.

[113] Xiao J, Xu H, Fang D, Cheng C, Gao H. Boosting and rectifying few-shot learning prototype network for skin lesion classification based on the internet of medical things. Wirel Netw 2021;1–15. http://dx.doi.org/10.1007/s11276-021-02713-z, Publisher: Springer.

[114] Yan J, Feng K, Zhao H, Sheng K. Siamese-prototypical network with data augmentation pre-training for few-shot medical image classification. In: 2022 2nd international conference on frontiers of electronics, information and computation technologies. ICFEICT, IEEE; 2022, p. 387–91. http://dx.doi.org/10.1109/ICFEICT57213.2022.00075.

[115] Yarlagadda DVK, Rao P, Rao D, Tawfik O. A system for one-shot learning of cervical cancer cell classification in histopathology images. In: Medical imaging 2019: digital pathology. Vol. 10956, SPIE; 2019, p. 216–21. http://dx.doi.org/10.1117/12.2512963.

[116] Zhang C, Cui Q, Ren S. Few-shot medical image classification with MAML based on dice loss. In: 2022 IEEE 2nd international conference on data science and computer application. ICDSCA, IEEE; 2022, p. 348–51. http://dx.doi.org/10.1109/ICDSCA56264.2022.9988390.

[117] Zhu W, Liao H, Li W, Li W, Luo J. Alleviating the incompatibility between cross entropy loss and episode training for few-shot skin disease classification. In: International conference on medical image computing and computer-assisted intervention. Springer; 2020, p. 330–9. http://dx.doi.org/10.1007/978-3-030-59725-2_32.

[118] Fechter T, Baltas D. One-shot learning for deformable medical image registration and periodic motion tracking. IEEE Trans Med Imaging 2020;39(7):2506–17. http://dx.doi.org/10.1109/TMI.2020.2972616, Publisher: IEEE.

[119] Ferrante E, Oktay O, Glocker B, Milone DH. On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains. In: International workshop on machine learning in medical imaging. Springer; 2018, p. 294–302. http://dx.doi.org/10.1007/978-3-030-00919-9_34.

[120] He Y, Li T, Ge R, Yang J, Kong Y, Zhu J, Shu H, Yang G, Li S. Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching. IEEE J Biomed Health Inf 2022;26(3):1177–87. http://dx.doi.org/10.1109/JBHI.2021.3095409, Publisher: IEEE.

[121] Zhang Y, Wu X, Gach HM, Li H, Yang D. GroupRegNet: a groupwise one-shot deep learning-based 4D image registration method. Phys Med Biol 2021;66(4):045030. http://dx.doi.org/10.1088/1361-6560/abd956, Publisher: IOP Publishing.

[122] Ericsson L, Gouk H, Hospedales TM. How well do self-supervised models transfer? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021, p. 5414–23. http://dx.doi.org/10.1109/CVPR46437.2021.00537.

[123] Zhao Z, Alzubaidi L, Zhang J, Duan Y, Gu Y. A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations. Expert Syst Appl 2024;242:122807. http://dx.doi.org/10.1016/j.eswa.2023.122807, URL https://www.sciencedirect.com/science/article/pii/S0957417423033092.

[124] Kanter F, Lellmann J. A flexible meta learning model for image registration. In: International conference on medical imaging with deep learning. PMLR; 2022, p. 638–52, URL https://proceedings.mlr.press/v172/kanter22a/kanter22a.pdf.

[125] Park H, Lee GM, Kim S, Ryu GH, Jeong A, Sagong M, Park SH. A meta-learning approach for medical image registration. In: 2022 IEEE 19th international symposium on biomedical imaging. ISBI, IEEE; 2022, p. 1–5. http://dx.doi.org/10.1109/ISBI52829.2022.9761512.

[126] Berthelot D, Carlini N, Goodfellow I, Papernot N, Oliver A, Raffel CA. MixMatch: A holistic approach to semi-supervised learning. In: Wallach H, Larochelle H, Beygelzimer A, d' Alché-Buc F, Fox E, Garnett R, editors. Advances in neural information processing systems. Vol. 32, Curran Associates, Inc.; 2019, URL https://proceedings.neurips.cc/paper_files/paper/2019/file/1cd138d0499a68f4bb72bee04bbec2d7-Paper.pdf.

[127] Rasmus A, Berglund M, Honkala M, Valpola H, Raiko T. Semi-supervised learning with ladder networks. Adv Neural Inf Process Syst 2015;28. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/378a063b8fdb1db941e34f4bde584c7d-Paper.pdf.

[128] Ren M, Triantafillou E, Ravi S, Snell J, Swersky K, Tenenbaum JB, Larochelle H, Zemel RS. Meta-learning for semi-supervised few-shot classification. 2018, arXiv preprint arXiv:1803.00676.

[129] Li X, Sun Q, Liu Y, Zhou Q, Zheng S, Chua T-S, Schiele B. Learning to self-train for semi-supervised few-shot classification. Adv Neural Inf Process Syst 2019;32. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/bf25356fd2a6e038f1a3a59c26687e80-Paper.pdf.

[130] Boney R, Ilin A. Semi-supervised few-shot learning with MAML. 2018, URL https://openreview.net/pdf?id=r1n5Osurf.

[131] Kavur AE, Gezer NS, Barış M, Aslan S, Conze P-H, Groza V, Pham DD, Chatterjee S, Ernst P, Özkan S, et al. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. Med Image Anal 2021;69:101950. http://dx.doi.org/10.1016/j.media.2020.101950, Publisher: Elsevier.

[132] Zhuang X. Multivariate mixture model for myocardial segmentation combining multi-source images. IEEE Trans Pattern Anal Mach Intell 2018;41(12):2933–46. http://dx.doi.org/10.1109/TPAMI.2018.2869576, Publisher: IEEE.

[133] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 2097–106. http://dx.doi.org/10.1109/CVPR.2017.369.

[134] Cheng Z, Wang S, Xin T, Zhou T, Zhang H, Shao L. Few-shot medical image segmentation via generating multiple representative descriptors. IEEE Trans Med Imaging 2024. http://dx.doi.org/10.1109/TMI.2024.3358295.

[135] Leng T, Zhang Y, Han K, Xie X. Self-sampling meta SAM: enhancing few-shot medical image segmentation with meta-learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024, p. 7925–35. http://dx.doi.org/10.1109/WACV57701.2024.00774.

[136] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, et al. Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. 2023, p. 4015–26. http://dx.doi.org/10.1109/ICCV51070.2023.00371.

[137] Xie W, Willems N, Patil S, Li Y, Kumar M. Sam fewshot finetuning for anatomical segmentation in medical images. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2024, p. 3253–61. http://dx.doi.org/10.1109/WACV57701.2024.00322.

[138] Ansari SA, Agrawal AP, Wajid MA, Wajid MS, Zafar A. MetaV: A pioneer in feature augmented meta-learning based vision transformer for medical image classification. Interdiscip Sci: Comput Life Sci 2024;1–20. http://dx.doi.org/10.1007/s12539-024-00630-1.

[139] Maia BMS, de Assis MCFR, de Lima LM, Rocha MB, Calente HG, Correa MLA, Camisasca DR, Krohling RA. Transformers, convolutional neural networks, and few-shot learning for classification of histopathological images of oral cancer. Expert Syst Appl 2024;241:122418. http://dx.doi.org/10.1016/j.eswa.2023.122418.

[140] Işık G, Paçal İ. Few-shot classification of ultrasound breast cancer images using meta-learning algorithms. Neural Comput Appl 2024;1–13. http://dx.doi.org/10.1007/s00521-024-09767-y.