A SYSTEMATIC REVIEW OF FEW-SHOT LEARNING IN MEDICAL IMAGING

PREPRINT SUBMITTED TO ARTIFICIAL INTELLIGENCE IN MEDICINE

Eva Pachetti[1,2] **Sara Colantonio** [1]

[1] Institute of Information Science and Technologies (ISTI),
National Research Council (CNR), Pisa, Italy
[2] Department of Information Engineering,
University of Pisa, Pisa, Italy
{eva.pachetti, sara.colantonio}@isti.cnr.it

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, especially with meta-learning. This systematic review gives a comprehensive overview of few-shot learning in medical imaging. We searched the literature systematically and selected 80 relevant articles published from 2018 to 2023. We clustered the articles based on medical outcomes, such as tumour segmentation, disease classification, and image registration; anatomical structure investigated (i.e. heart, lung, etc.); and the meta-learning method used. For each cluster, we examined the papers' distributions and the results provided by the state-of-the-art. In addition, we identified a generic pipeline shared among all the studies. The review shows that few-shot learning can overcome data scarcity in most outcomes and that meta-learning is a popular choice to perform few-shot learning because it can adapt to new tasks with few labelled samples. In addition, following meta-learning, supervised learning and semi-supervised learning stand out as the predominant techniques employed to tackle few-shot learning challenges in medical imaging and also best performing. Lastly, we observed that the primary application areas predominantly encompass cardiac, pulmonary, and abdominal domains. This systematic review aims to inspire further research to improve medical image analysis and patient care.

Keywords Few-shot learning · Medical imaging · Systematic review ·

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 analyze 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.

While existing reviews have primarily focused on FSL in computer vision as a whole [2, 3, 4], the ones specific to FSL in medical imaging have often focused on particular aspects such as Neural Architecture Search [5] or have examined only a subset of published studies [6]. In contrast, we believe that a comprehensive review in the field of FSL for medical imaging can provide a global understanding of the current state of the art (SOTA). Specifically, this review will discuss how FSL declines in segmentation, classification, and registration tasks when used in medical image analysis. These three applications within the FSL are all generally characterized by using a limited number of annotations in the training phase, whether labels in the case of classification or annotations on image data in the case of segmentation and registration. To address this challenge, some works, for example, cope with the lack of annotated data by exploiting a large amount of unannotated data in the pre-training phase [7, 8, 9]. Others, instead, generate data artificially [10]. However, one thing that most work share consists of implementing meta-learning techniques. Indeed, meta-learning presents one promising direction for FSL by extracting and propagating transferable knowledge from a set of tasks to avoid overfitting [2]. Therefore, we will analyze the methods used in the FSL for medical imaging literature by paying particular attention to whether meta-learning methods were applied. Researchers can use this as a guide in developing new techniques and exploring uncharted territory. In conclusion, we will offer the reader an overview of the most frequently employed approaches, aside from meta-learning, for tackling FSL in the medical imaging domain, and we will also propose a comprehensive pipeline that encompasses all the studies we have reviewed.

1.2 Objectives

This systematic review aims to provide a comprehensive overview of the SOTA in FSL techniques applied to medical imaging and to offer readers insight into the most valuable works in this field. Alongside the theoretical background, we aim to collect and highlight papers 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. The objective is to present innovative techniques that have demonstrated tangible results, catalyzing advancements for each outcome and, specifically, in each medical application. A particular emphasis is placed on meta-learning, as it is a common approach used to tackle FSL problems. Below, we provide a detailed breakdown of the specific objectives of this study:

- **Present a distribution of studies by outcome.** The aim is to highlight the distribution of studies across the three outcomes: segmentation, classification and registration. This analysis will provide insight into the emphasis placed on each task in the field of FSL for medical imaging.
- **Present a distribution of studies and their results based on the anatomical structures investigated.** For each outcome, we analyze the most commonly addressed tasks w.r.t. the anatomical structures investigated and examine the average performance achieved by the SOTA methods.
- **Offer an analysis of the distribution of studies and their results w.r.t. the meta-learning methods employed.** We provide a distribution analysis of the meta-learning methods used for each outcome. This analysis will reveal which meta-learning techniques are predominantly employed and highlight cases where meta-learning methods are not utilized. Additionally, for each meta-learning set of techniques, we present the average performance achieved by SOTA.
- **Provide distributions for training data, imaging modalities, and evaluations of robustness.** Further to the above analyses, we provide data usage information for each study, examine the most commonly used imaging modalities and explore the model robustness assessment methods employed by the reviewed studies.
- **Identify a standard pipeline among the studies.** In conclusion, we identify a generic pipeline shared among all the studies we reviewed. This pipeline illustrates the most frequently used methodologies across all studies that aim to conduct FSL. For each study, we explicitly indicate which elements of the pipeline are adopted and offer an indication of the prevalence of various techniques across all the reviewed studies.

By accomplishing these objectives, our systematic review aims to offer a comprehensive and up-to-date understanding of the current landscape of FSL in medical imaging analysis. This review will serve as a valuable resource, providing researchers and practitioners with an overview of the SOTA techniques and approaches in FSL applied to medical imaging. By synthesizing the existing literature and highlighting key findings, the review will facilitate progress in the field by identifying gaps, challenges, and opportunities for future research. Furthermore, our work will aid in identifying the best practices and effective methodologies in FSL for medical imaging tasks, enabling researchers and practitioners to make informed decisions when designing and implementing FSL-based solutions in their work.

In the following, we explain our manuscript's organization:

- We begin with a theoretical introduction to the concepts of FSL and meta-learning, followed by a discussion of the key SOTA works in the field of meta-learning for FSL.
- Next, we outline the methods employed to perform the literature search, including the eligibility criteria and the databases utilized. We also detail the key aspects examined for each work and the synthesis methods employed.
- We present the obtained results by providing a comprehensive overview of the main characteristics of each selected paper and reporting the analyses conducted according to the review objectives. Additionally, we present the results of the risk of bias assessment and applicability analysis for each study and draw a synthesis of the employed methods.
- Finally, we discuss the findings regarding each objective of the review and draw conclusions based on the evidence presented.

2 Theoretical background

FSL has been gaining significant attention, particularly with the rise of meta-learning. 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 [11]. The inner algorithm (f) solves the task i by updating the model parameters θ to ${\theta'}_i$: 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. [11], 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. Figure 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.

Meta-knowledge can manifest in various forms, such as initial parameters, optimization strategy, and learning algorithm [11]. Accordingly, we adopt the taxonomy proposed by [2] to categorize meta-learning algorithms for FSL into three categories: *Initialization-based* methods, *Metric learning-based* methods, and *Hallucination-based* methods. Figure 2 illustrates this taxonomy. In the subsequent paragraphs, we provide an overview of the most renowned algorithms developed within each category.

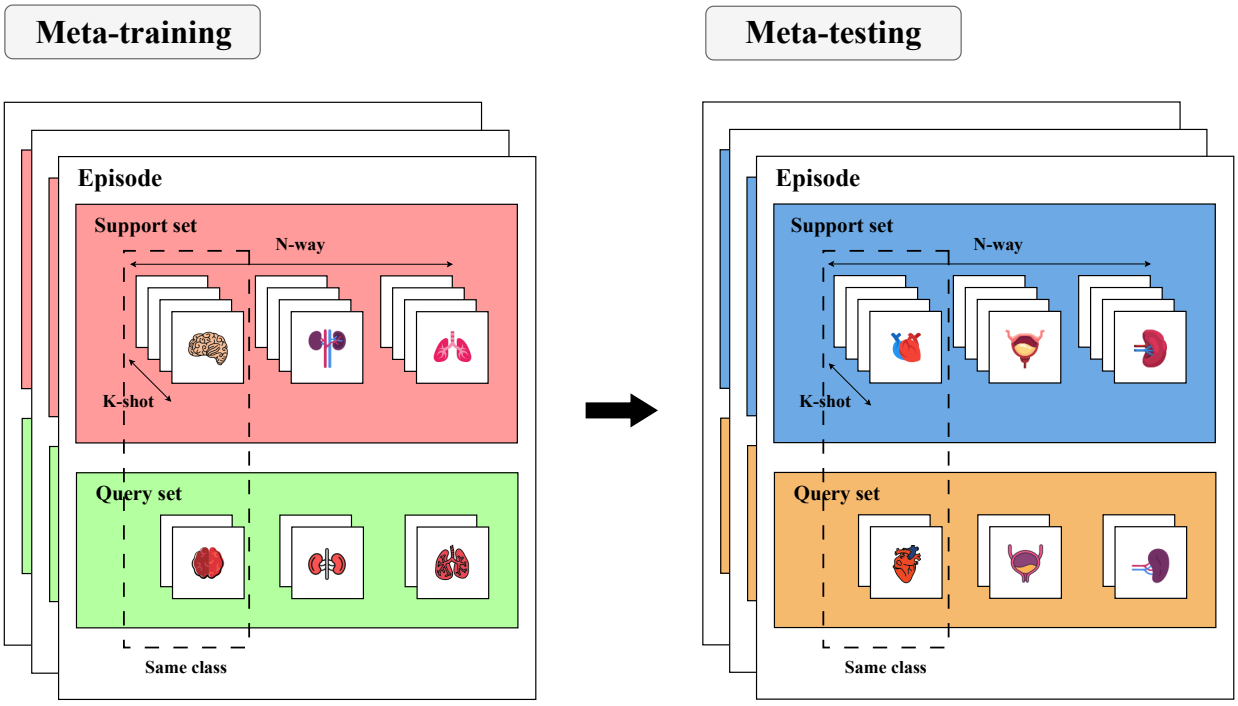

Figure 1: N-way K-shot paradigm representation.

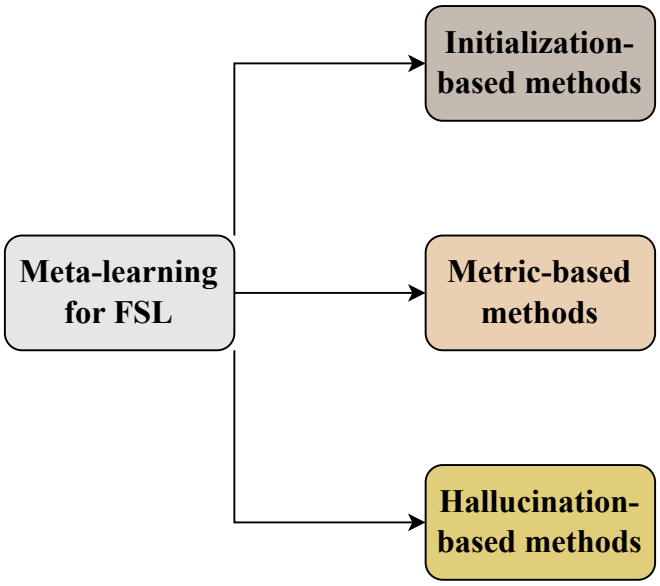

Figure 2: Meta-learning methods taxonomy.

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.

2.1.1 Model-Agnostic Meta-Learning

In their paper, Finn et al. [12] 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 $\theta_i^{'}$ 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.

2.1.2 Reptile

In their work Nichol, Achiam and Schulman [13] 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 Equation 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 [14] 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 [15]:

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

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 $\theta_i^{'}$ 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 [16] 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) [17]. 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. [18] propose a solution to the FSL task using a differentiable version of Memory-augmented Neural Networks (MANNs) known as Neural Turing Machines (NTMs). 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. It can store information for both short-term and long-term durations, making it capable of handling tasks with limited data. During training, the model is fed 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}, the label at the previous time step. 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. The external memory is utilized to store the input-label pairs discovered by the model during the training process. When a previously encountered input shows up again, the corresponding label is retrieved from the external memory, effectively making it a prediction for the current input. This 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. The retrieval is done by computing the cosine similarity between the key k_t and the contents of the memory matrix M_t. Once the label is retrieved, the error is backpropagated, and the model's weights are updated to improve the input-label binding strategy.

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. [19] first introduced Siamese Neural Networks for signature verification. In 2015, they were newly proposed by Koch, Zemel and Salakhutdinov [20], 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.

2.2.2 Triplet Networks

Triplet Networks, introduced by Hoffer et al. [21], 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.

2.2.3 Matching Networks

Matching Networks proposed by Vinyals et al. [22], 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 [23], 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 Zero-Shot Learning (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 ([24]). 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 [25], and they were initially employed in the FSL and ZSL domains in [26]. 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 [27] 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. [28] 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 computer 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 analyzing 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 [29]. The review has five main objectives. Firstly, it aims to analyze the distribution of studies among the three outcomes (segmentation, classification, and registration) in the field of FSL for medical imaging. Secondly, for each outcome, it examines the most commonly addressed tasks concerning the anatomical structures studied. Thirdly, it provides a distribution of the meta-learning methods used for both classification, segmentation and registration tasks by shedding light on prevalent meta-learning techniques and cases where meta-learning methods are not utilized. In addition, the review offers additional insights into the data usage, including the most commonly used imaging modalities, and explores the techniques of model robustness assessment employed in the reviewed studies. Finally, it offers an overview of the most commonly used methods among the selected studies while also outlining a general pipeline for conducting FSL in the field of medical imaging.

3.2 Eligilibity criteria

We established the inclusion criteria for paper selection in this systematic review 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 [30, 31, 32]. FSDA, as highlighted by Li et al. [32], 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 didn't 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, examinated 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 was employed for the data collection process. During the review, all articles were examinated 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.

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

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"":"zero-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*))"

Table 1: Research queries employed for each database.

domains or their significance in the review. Likewise, there were no modifications made to 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 [33], 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 commonly used metrics: the Dice score and the Intersection over Union (IoU). These metrics provide quantitative measures of the overlap between the predicted segmentation and the ground truth. For the classification outcome, we evaluated its effectiveness through various measures. One of the metrics employed was Accuracy, which determines the proportion of correctly classified samples. Additionally, we considered the F1-score and Recall. Moreover, we investigated the Area Under the Receiver Operating Characteristic (AUROC) curve as a performance metric, particularly for binary classification tasks. In the registration domain, we investigated different metrics to evaluate the effectiveness: the Dice score, the average landmark distance (ALD), and the target registration error (TRE). These metrics were used to quantify the performance of the models in their respective outcome domains, providing objective measures of effectiveness and allowing for comparisons between different approaches.

3.10 Synthesis methods

In our systematic review, we structured the results of each study within dedicated tables for each outcome category, including segmentation, classification, and registration. 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 organ 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. In conclusion, we furnished a comprehensive overview by creating a unified pipeline that encompasses all the papers reviewed within each outcome. For each outcome, we presented a table that delineated the specific elements of the core pipeline utilized by each study.

4 Results

4.1 Study selection

In Figure 3, we show the PRISMA diagram where we summarize the data selection flow. 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. Figure 4 displays the distribution of studies across the three primary outcomes, while Figure 5 illustrates the proposed unified pipeline for the various methods employed in performing FSL. Below, we present the results of our analysis grouped by the primary outcome: segmentation, classification, and registration. It's worth noting that several studies, namely [34], [35], [36], [37], and [38], are included multiple times, as they address multiple outcomes simultaneously.

4.2.1 Segmentation

We selected 50 relevant studies, each focusing on medical segmentation as its primary task. All pertinent information from the selected studies is provided in Table 2. In addition, we present ROB and the applicability analyses of each study in Table 3.

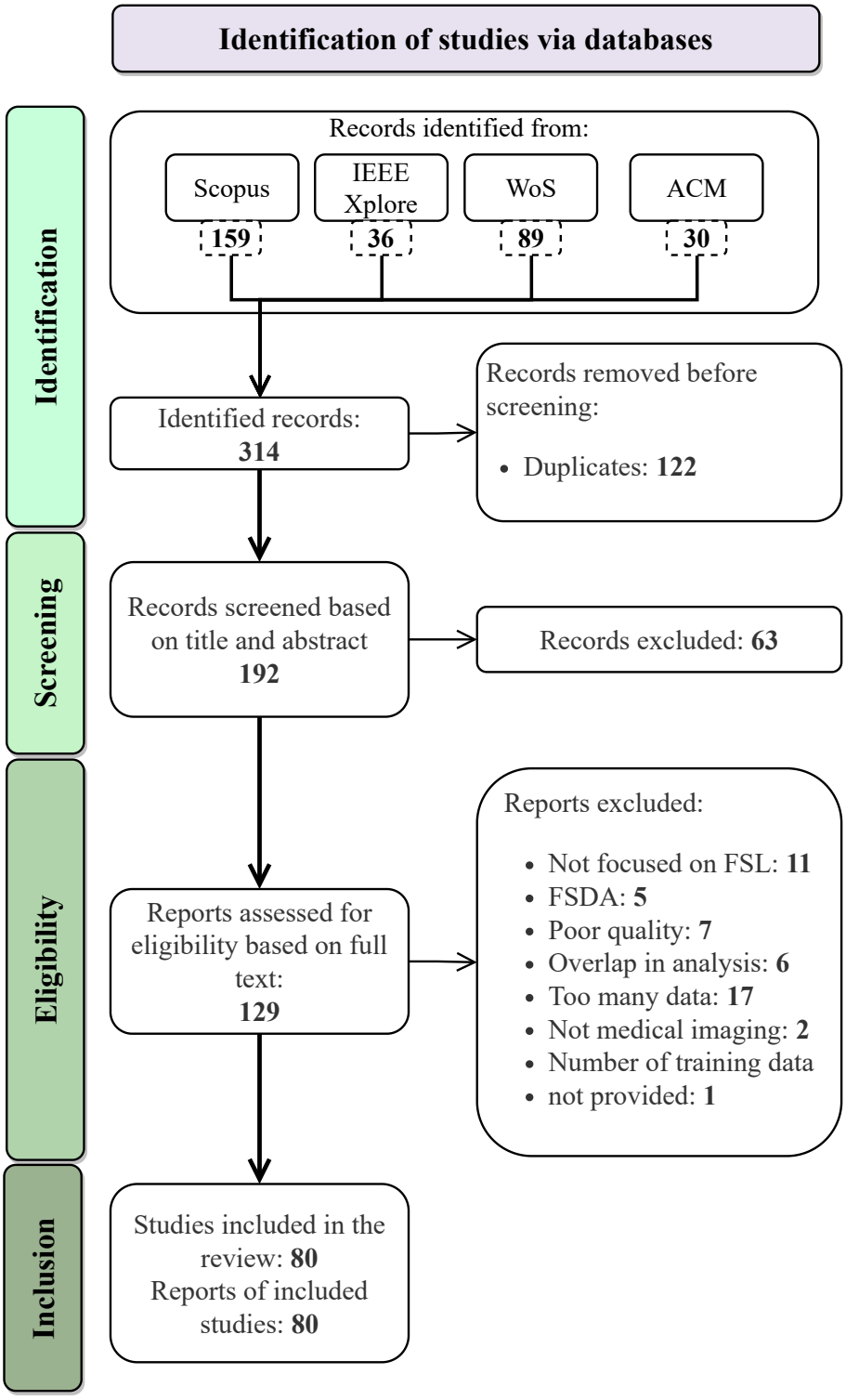

Figure 3: PRISMA flow diagram.

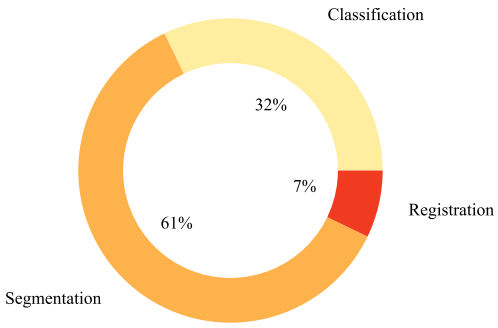

Figure 4: Studies distribution by outcome.

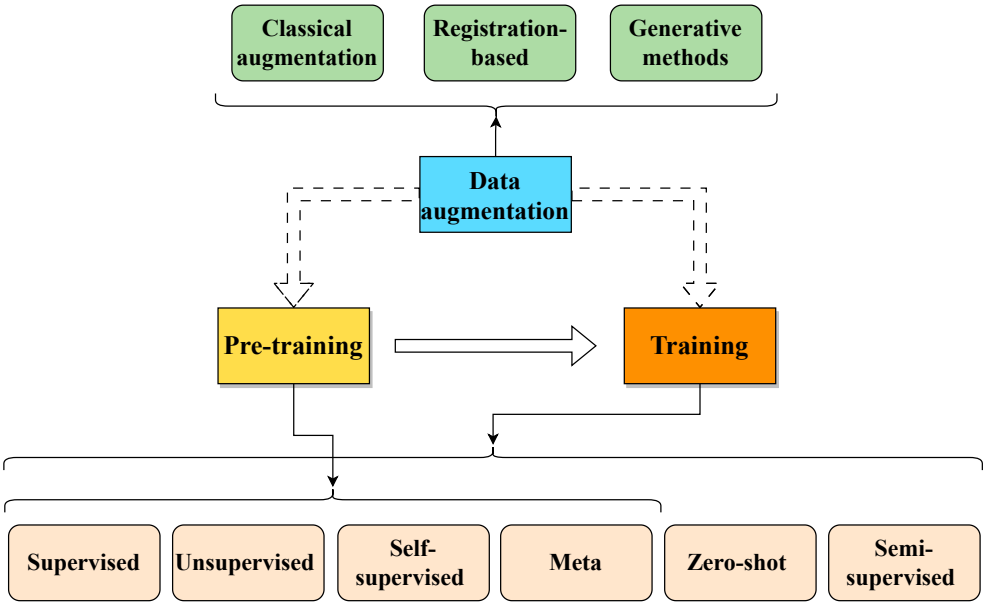

Figure 5: Summary diagram.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
1	Blendowski, Nickisch, and Heinrich [39]	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. [40]	Res2-UNeXT + Data augmentation with Daemons registration algorithm	8-shot	IoU: 0.943 (Cells)	None
3	Chen et al. [10]	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. [41]	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. [42]	Registration model + Similarity model + Patch label fusion	20-shot	Dice: 0.817 (MYO)	Metric learning
6	Ding, Yu and Yang [43]	Registration network + Generative network + Segmentation network	1-shot	Dice: 0.851 (Brain)	Hallucination
7	Farshad et al. [44]	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. [45]	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 [46]	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. [47]	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 [48]	Cascaded U-net + 3D augmentation	From 1-shot to 6-shot	Dice: 0.910 (Kidney)	None

Table 2: FSL studies for medical image segmentation.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
12	Hansen et al. [49]	Anomaly detection-inspired model + SSL	1-shot 2-shot 3-shot	Dice: 0.875 (LV-BP) 0.773 (RV) 0.624 (MYO) 0.833 (Kidney) 0.759 (Spleen) 0.808 (Liver)	Metric learning
13	He et al. [34]	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
14	He et al. [35]	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. [50]	Self-guided Anomaly detection-inspired model	1-shot	Dice: 0.840 (LV) 0.585 (MYO) 0.697 (RV)	Metric learning
16	Joyce and Kozerke [51]	Anatomical model + SSL	1-shot 3-shot 10-shot	Dice: 0.630 (Heart)	None
17	Khadka et al. [52]	Implicit MAML + Attention U-Net	5-shot 10-shot 20-shot	Dice: 0.833 (Skin, nevus)	Initialization
18	Khaled, Han and Ghaleb [53]	Multi-stage GAN	5-shot 10-shot	Dice: 0.940 (Brain, MAS)	None
19	Khandelwal and Yushkevich [54]	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. [55]	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. [56]	3D U-Net) + Prototypical learning + Image alignment module	1-shot	Dice: 0.417 (Prostate, MAS)	Metric learning
22	Lu et al. [57]	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

Table 2: *(continued).*

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
23	Lu and Ye [58]	TractSeg + Knowledge transfer with warmup	1-shot 5-shot	Dice: 0.812 (Brain, WM)	None
24	Ma et al. [59]	Segmentation network + Zero-shot segmentation network + Spatial Context Attention module	0-shot	Dice: 0.882 (Brain, tumour)	None
25	Niu et al. [60]	Conditioner + Segmenter + Symmetrical Supervision Mechanism + Transformer-based Global Feature Alignment module	1-shot	Dice: 0.870 (LV-BP) 0.815 (Kidney) 0.738 (Spleen) 0.729 (Liver)	Metric learning
26	Ouyang et al. [7]	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. [61]	Few-Sample-Fitting	1-shot to 20-shot	Dice: 0.990 (Femur)	None
28	Pham, Dovletov and Pauli [62]	3D U-Net + Imitating encoder + Prior encoder + Joint decoder	1-shot	Dice: 0.776 (Liver)	None
29	Roy et al. [63]	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. [36]	Echo state network + augmented U-Net	5-shot	Dice: 0.640 (Eye, IC)	None
31	Rutter, Lagergren and Flores [64]	CNN for Boundary Optimization	1-shot 3-shot 5-shot	Dice: 0.931 (Cells)	None
32	Shen et al. [65]	Large Deformation Diffeomorphic Metric Mapping model + Sample transformations + Interpolation	1-shot	Dice: 0.883 (Knee)	None
33	Shen et al. [66]	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. [37]	Joint Registration and Segmentation Self-training Framework (JRSS)	5-shot	Dice: 0.795 (Brain, MAS) 0.753 (Abdomen, MAS)	None

Table 2: *(continued)*.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
35	Sun et al. [67]	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. [68]	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. [69]	Generative Style Transfer (Appearance model + Style encoder + Flow model + Flow Adversarial Autoencoder)	1-shot	Dice: 0.835 (Brain, MAS)	None
38	Wang et al. [70]	Label Transfer Network (Atlas-based segmentation + Forward-backward correspondance)	1-shot	Dice: 0.823 (Brain, MAS)	None
39	Wang et al. [71]	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. [8]	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 (AO) 0.798 (PA)	None
41	Wang, Zhou and Zheng [72]	Prototype learning + Self-reference + Contrastive learning	1-shot	Dice: 0.756 (Liver) 0.737 (Spleen) 0.842 (Kidney)	Metric learning
42	Wang et al. [73]	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 [74]	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. [75]	Self-Learning + One-Shot Learning	1-shot	Dice: 0.850 (Spleen) 0.930 (Liver)	None

Table 2: *(continued).*

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
45	Xu and Niethammer [38]	DeepAtlas (Semi-Supervised Learning + Segmentation network + Registration network)	1-shot 5-shot 10-shot	Dice: 0.892 (Knee, MAS) 0.612 (Brain)	None
46	Yu et al. [76]	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 [77]	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. [78]	Spatial and appearance transform models + Semi-supervised learning + Supervised learning	1-shot	Dice: 0.815 (Brain, MAS)	None
49	Zhao et al. [79]	Meta-hallucinator	1-shot 4-shot	Dice: 0.756 (AO) 0.751 (LA) 0.823 (LV) 0.696 (MYO)	Initialization and Hallucination-based
50	Zhou et al. [80]	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

Table 2: *(continued).*

Study ID	Pub. ref.	Risk of Bias					Applicability			
		Part.	Pred	Out.	Analysis	Overall	Part.	Pred.	Out.	Overall
1	Blendowski, Nickisch, and Heinrich [39]	✓	✓	✓	✓	✓	✓	✓	✓	✓
2	Chan et al. [40]	✓	✓	✓	✗	✗	✓	✓	✓	✓
3	Chen et al. [10]	✓	✓	✓	✓	✓	✓	✓	✓	✓
4	Cui et al. [41]	✓	✓	✓	✓	✓	✓	✓	✓	✓
5	Ding, Wangbin et al. [42]	✓	✓	✓	✗	✗	✓	✓	✓	✓
6	Ding, Yu and Yang [43]	✓	✓	✓	✓	✓	✓	✓	✓	✓
7	Farshad et al. [44]	✓	✓	✓	✗	✗	✓	✓	✓	✓
8	Feng et al. [45]	✓	✓	✓	✓	✓	✓	✓	✓	✓
9	Gama, Oliveira and dos Santos [46]	✓	✓	✓	✓	✓	✓	✓	✓	✓
10	Gama et al. [47]	✓	✓	✓	✓	✓	✓	✓	✓	✓
11	Guo, Odu and Pedrosa [48]	✓	✓	✓	✗	✗	✓	✓	✓	✓
12	Hansen et al. [49]	✓	✓	✓	✓	✓	✓	✓	✓	✓
13	He et al. [34]	✓	✓	✓	✗	✗	✓	✓	✓	✓
14	He et al. [35]	✓	✓	✓	✓	✓	✓	✓	✓	✓
15	Jenssen et al. [50]	✓	✓	✓	✗	✗	✓	✓	✓	✓
16	Joyce and Kozerke [51]	✓	✓	✓	✗	✗	✓	✓	✓	✓
17	Khadka et al. [52]	✓	✓	✓	✓	✓	✓	✓	✓	✓
18	Khaled, Han and Ghaleb [53]	✓	✓	✓	✓	✓	✓	✓	✓	✓

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
19	Khandelwal and Yushkevich [54]	✓	✓	✓	✗	✗	✓	✓	✓	✓
20	Kim et al. [55]	✓	✓	✓	✓	✓	✓	✓	✓	✓
21	Li et al. [56]	✓	✓	✓	✓	✓	✓	✓	✓	✓
22	Lu et al. [57]	✓	✓	✓	✓	✓	✓	✓	✓	✓
23	Lu and Ye [58]	✓	✓	✓	✓	✓	✓	✓	✓	✓
24	Ma et al. [59]	✓	✓	✓	✓	✓	✓	✓	✓	✓
25	Niu et al. [60]	✓	✓	✓	✓	✓	✓	✓	✓	✓
26	Ouyang et al. [7]	✓	✓	✓	✓	✓	✓	✓	✓	✓
27	Pham et al. [61]	✓	✓	✓	✗	✗	✓	✓	✓	✓
28	Pham, Dovletov and Pauli [62]	✓	✓	✓	✓	✓	✓	✓	✓	✓
29	Roy et al. [63]	✓	✓	✓	✓	✓	✓	✓	✓	✓
30	Roychowdhury et al. [36]	✓	✓	✓	✗	✗	✓	✓	✓	✓
31	Rutter, Lagergren and Flores [64]	✓	✓	✓	✓	✓	✓	✓	✓	✓
32	Shen et al. [65]	✓	✓	✓	✓	✓	✓	✓	✓	✓
33	Shen et al. [66]	✓	✓	✓	✓	✓	✓	✓	✓	✓
34	Shi et al. [37]	✓	✓	✓	✓	✓	✓	✓	✓	✓
35	Sun et al. [67]	✓	✓	✓	✓	✓	✓	✓	✓	✓
36	Tang et al. [68]	✓	✓	✓	✓	✓	✓	✓	✓	✓
37	Tomar et al. [69]	✓	✓	✓	✓	✓	✓	✓	✓	✓
38	Wang et al. [70]	✓	✓	✓	✓	✓	✓	✓	✓	✓
39	Wang et al. [71]	✓	✓	✓	✓	✓	✓	✓	✓	✓
40	Wang et al. [8]	✓	✓	✓	✓	✓	✓	✓	✓	✓
41	Wang, Zhou and Zheng [72]	✓	✓	✓	✓	✓	✓	✓	✓	✓
42	Wang et al. [73]	✓	✓	✓	✓	✓	✓	✓	✓	✓
43	Wu, Xiao and Liang [74]	✓	✓	✓	✓	✓	✓	✓	✓	✓
44	Wu et al. [75]	✓	✓	✓	✓	✓	✓	✓	✓	✓
45	Xu and Niethammer [38]	✓	✓	✓	✓	✓	✓	✓	✓	✓
46	Yu et al. [76]	✓	✓	✓	✓	✓	✓	✓	✓	✓
47	Yuan, Esteva and Xu [77].	✓	✓	✓	✓	✓	✓	✓	✓	✓
48	Zhao et al. [78]	✓	✓	✓	✓	✓	✓	✓	✓	✓
49	Zhao et al. [79]	✓	✓	✓	✓	✓	✓	✓	✓	✓
50	Zhou et al. [80].	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 3: *(continued)*.

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

Medical application. The segmentation papers within the field of FSL address various anatomical structures and regions, as well as specific lesions such as polyps or tumours. Here's a breakdown of the papers categorized by the anatomical structure(s) investigated. Eighteen papers (36%) focus on liver segmentation; 18 studies (36%) concentrate on kidney segmentation; 17 papers (34%) centre around spleen segmentation; three papers (6%) pertain to psoas segmentation; four (8%) are related to prostate segmentation; three works (6%) involve bladder segmentation; four papers (8%) deal with breast segmentation; one paper (2%) addresses colon segmentation; six (12%) are concerned with stomach segmentation; 12 (24%) are dedicated to brain segmentation; 14 papers (28%) revolve around heart segmentation; three (6%) involve pancreas segmentation; three (6%) pertain to cell segmentation; two papers (4%) are related to lung segmentation; one (2%) focuses on eye segmentation; two papers (4%) involve mandible segmentation; one (2%) addresses duodenum segmentation; two papers (4%) deal with skin segmentation; three papers (6%) are related to knee segmentatio; one (2%) concerns phalanx segmentation; one (2%) deals with hip segmentation; one paper (2%) is dedicated to spine segmentation. For a visual representation of the distribution, please refer to Figure 6.

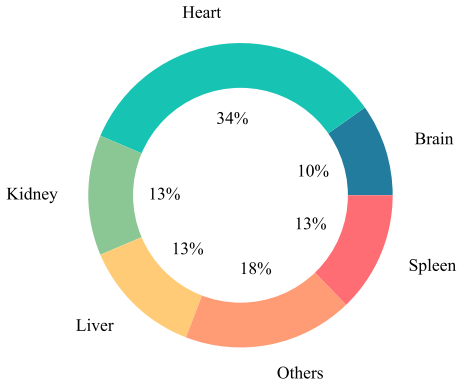

Figure 6: Segmentation studies grouped by the anatomical structure investigated.

Meta-learning methods. 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%) leverage initialization-based methods; 14 studies (28%) utilize metric learning-based techniques; three studies (6%) employ hallucination-based methods; one study (2%) combines both initialization-based and hallucination-based methods. The remaining 28 studies (56%) do not incorporate meta-learning techniques. For a visual representation of the distribution, refer to Figure 7.

K-shot. Among the 50 selected papers, here is the distribution of training shot configurations: 15 studies (30%) utilize k-shot training with k ranging from 2 to 20; 14 studies (28%) perform both OSL and FSL learning; 20 works (40%) exclusively use 1-shot training; 1 paper (2%) employs 0-shot training.

Image modalities. In terms of the imaging modalities utilized in the selected papers, here is the distribution: 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. To examine the behaviour and robustness of the models, the selected studies used different evaluation techniques as follows: 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.

Model performance grouped by organ and meta-learning method. In Figure 8 and Figure 9, we present a summary of the model performance in forest plots, categorized by anatomical structure, w.r.t. Dice score and IoU, respectively. Conversely, in Figure 10 and Figure 11, we depict the performance in terms of Dice and IoU, respectively, by grouping the studies according to the employed meta-learning methods.

Overall pipeline. In Table 4, we outline which steps of the main pipeline are adopted by each segmentation study. Here are the distributions of studies based on their utilization of pre-training, training, and data augmentation techniques:

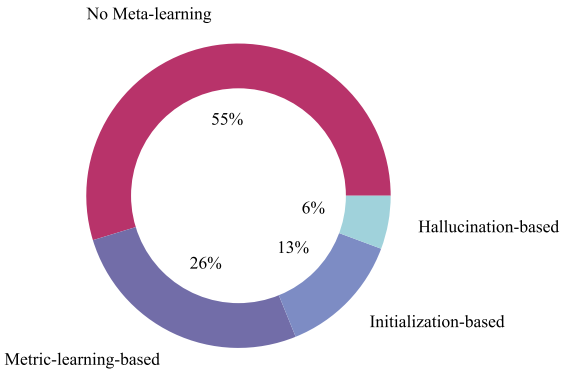

Figure 7: Segmentation studies grouped by meta-learning method employed.

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; six studies (12%) relied on registration-based augmentation. The remaining 24 out of 50 studies (48%) did not employ data augmentation.

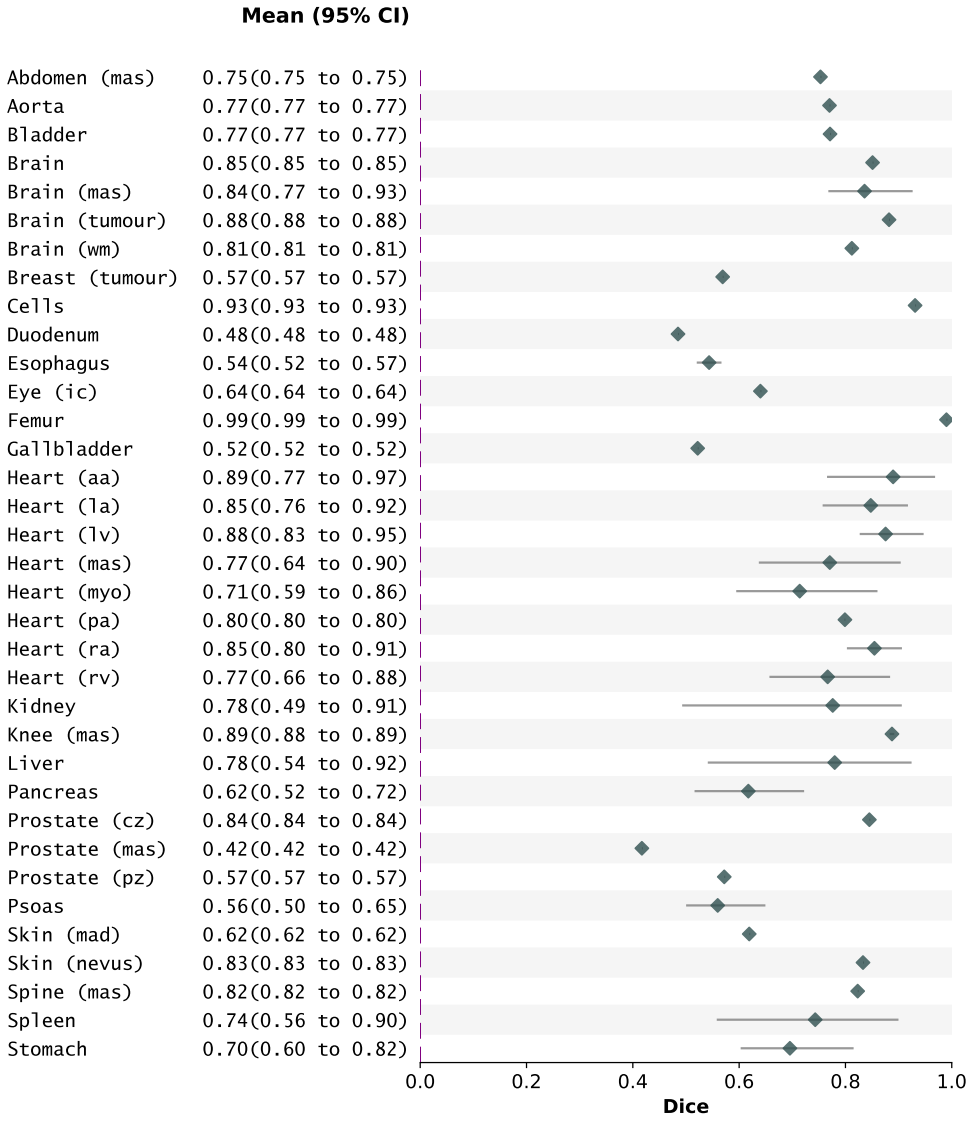

Figure 8: Forest plot of segmentation studies performance based on Dice metric. Studies are grouped by the anatomical structure investigated. AA = Ascending Aorta; IC = Intraretinal Cyst; LA = Left Atrium; LV = Left Ventricle; MAS = Mean Across Structures; MYO = Myocardium; PA = Pulmonary Artery; PZ = Peripheral Zone; RA = Right Atrium; RV = Right Ventricle; TZ = Transitional Zone; WM = White Matter.

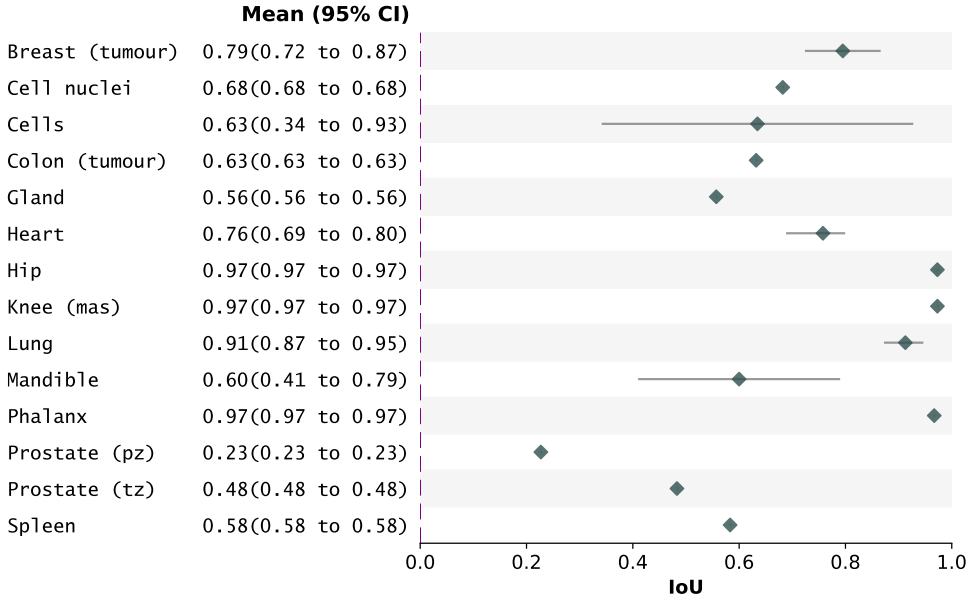

Figure 9: Forest plot of segmentation studies performance based on IoU metric. Studies are grouped by the anatomical structure investigated. IE = Intracranial Ematoma; MAS = Mean Across Structures; PZ = Peripheral Zone; TZ = Transitional Zone.

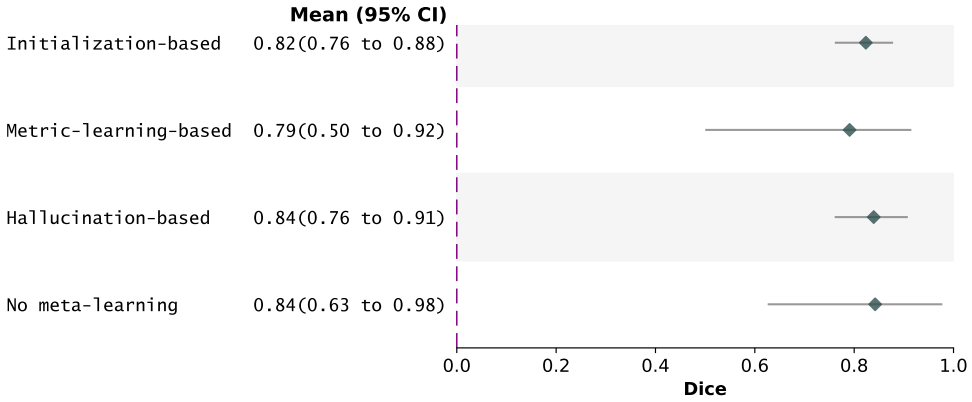

Figure 10: Forest plot of segmentation studies performance based on Dice metric. Studies are grouped by the meta-learning method employed.

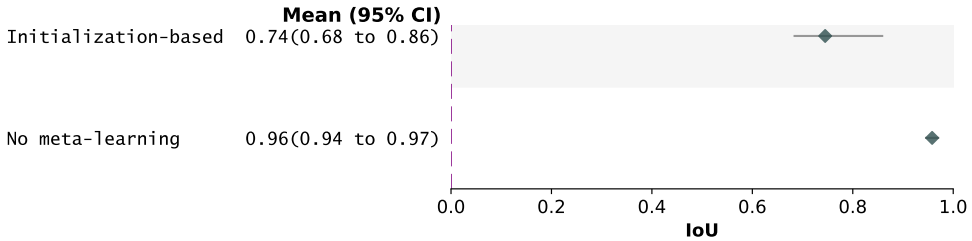

Figure 11: Forest plot of segmentation studies performance based on IoU metric. Studies are grouped by the meta-learning method employed.

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

Table 4: Main pipeline steps adopted by segmentation studies.

Study ID	Pub. ref.	Pre-training	Training	Data augmentation
49	Zhao et al. [79]	None	Meta	Classical and Generative
50	Zhou et al. [80]	Supervised	Supervised	None

Table 4: *(continued)*.

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 relevant information from these selected studies in Table 5. In addition, we provide information concerning ROB and the applicability of each study in Table 6.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
51	Ali et al. [81]	Prototypical network	5-shot	Accuracy: 0.906 (Endoscopic images, MAO)	Metric learning
52	Cai, Hu, and Zheng [82]	Prototypical network + Attention module (CBAM)	20-shot	Accuracy: 0.924 (Brain, MAT)	Metric learning
53	Cai et al. [83]	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 [84]	Siamese Neural Network	1-shot	Accuracy: 0.908 (Breast, MAT)	Metric learning
55	Chen et al. [85]	2D CNN ranking + 2D CNN classification + Heatmap for segmentation	2-shot	AUROC: 0.883 (Breast, LN metastases)	None
56	Chou et al. [86]	Siamese Neural Network (Triple encoder + Triple loss)	1-shot	Accuracy: 0.986 (Brain, classification into contrast type)	None
57	Dai et al. [87]	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 [88]	One-shot Anomaly Detection Framework	1-shot	AUROC: 0.961 (Eye, MAD) 0.955 (Lung ,COVID)	None
59	Jiang et al. [89]	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. [90]	ViT-L/16 + ResNet50 + Metric-learning	1-shot 5-shot 8-shot	Accuracy: 0.346 (Lungs, MAD)	Metric learning

Table 5: FSL studies for medical image classification.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
61	Mahapatra, Ge and Reyes [91]	Self-Supervised Clustering Based Generalized Zero-shot Learning	0-shot	Accuracy: 0.921 (Breast, LN metastases) 0.909 (Lungs, MAD) 0.942 (Eye, DE) 0.911 (Prostate, tumour)	None
62	Maicas et al. [92]	Pre and post-hoc diagnosis and interpretation + 3D DenseNet	4-shot	AUROC: 0,910 (Breast, tumour)	Initialization
63	Mohan et al. [93]	Siamese Network + Classifier	1-shot	Accuracy: 0.930 (Lung, COVID and Pneumonia)	None
64	Moukheiber et al. [94]	DeepVoro Multi-label ensemble	5-shot 10-shot	AUROC: 0.679 (Lung, MAD)	Initialization-based
65	Naren, Zhu and Wang [95]	8 block VGG + MAML++	1-shot to 5-shot	Accuracy: 0.857 (Lung, COVID)	Initialization
66	Ouahab, Ben-Ahmed and Fernandez-Maloigne [96]	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 [97]	DenseNet-121 (feature extractor) + Autoencoder ensemble (classificator)	5-shot	F1-score: 0.440 (Lung, MAD) Recall: 0.490 (Lung,MAD)	None
68	Paul et al. [98]	DenseNet + Vanilla autoencoder)	5-shot	F1-score: 0.470 (Lungs, MAD) AUROC: 0.647 (Lungs, MAD)	None
69	Paul et al. [99]	DenseNet + MVSE network + Self-training	0-shot	Recall: 0.454 (Lungs, MAD)	None
30	Roychowdhury et al. [36]	Echo state network (ParESN) + Target label selection algorithm (TLSA)	5-shot	Accuracy: 0.970 (Eye, IE)	None
70	Singh et al. [100]	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. [101]	VAE + Distribution learning	0-shot 15-shot	AUROC: 0.789 (Pancreas)	None

Table 5: *(continued).*

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
72	Xiao et al. [102]	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. [103]	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. [104]	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 [105]	MAML	1-shot 3-shot 5-shot	Accuracy: 0.788 (VQA-RAD, MAS) 0.614 (PathVQA, MAS)	Initialization
76	Zhu et al. [106]	Query-Relative Loss + Adaptive Hard Margin + Prototypical Network/ Matching Network	1-shot 5-shot	Accuracy: 0.719 (Skin, MAD)	Metric learning

Table 5: *(continued).*

		Risk of Bias					Applicability			
Study ID	Pub. ref.	Part.	Pred	Out.	Analysis	Overall	Part.	Pred.	Out.	Overall
51	Ali et al. [81]	✓	✓	✓	✗	✗	✓	✓	✓	✓
52	Cai, Hu, and Zheng [82]	✓	✓	✓	✓	✓	✓	✓	✓	✓
53	Cai et al. [83]	✓	✓	✓	✓	✓	✓	✓	✓	✓
54	Cano and Cruz-Roa [84]	✓	✓	✓	✗	✗	✓	✓	✓	✓
55	Chen et al. [2]	✓	✓	✓	✗	✗	✓	✓	✓	✓
56	Chou et al. [86]	✓	✓	✓	✗	✗	✓	✓	✓	✓
57	Dai et al. [87]	✓	✓	✓	✓	✓	✓	✓	✓	✓

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
58	Huang, Huang and Tang [88]	✓	✓	✓	✓	✓	✓	✓	✓	✓
59	Jiang et al. [89]	✓	✓	✓	✓	✓	✓	✓	✓	✓
60	Jin et al. [90]	✓	✓	✓	✓	✓	✓	✓	✓	✓
61	Mahapatra, Ge and Reyes [91]	✓	✓	✓	✓	✓	✓	✓	✓	✓
62	Maicas et al. [92]	✓	✓	✓	✓	✓	✓	✓	✓	✓
63	Mohan et al. [93]	✓	✓	✓	✗	✗	✓	✓	✓	✓
64	Moukheiber et al. [94]	✓	✓	✓	✓	✓	✓	✓	✓	✓
65	Naren, Zhu and Wang [95]	✓	✓	✓	✓	✓	✓	✓	✓	✓
66	Ouahab, Ben-Ahmed and Fernandez-Maloigne [96]	✓	✓	✓	✓	✓	✓	✓	✓	✓
67	Paul, Tang and Summers [97]	✓	✓	✓	✓	✓	✓	✓	✓	✓
68	Paul et al. [98]	✓	✓	✓	✓	✓	✓	✓	✓	✓
69	Paul et al. [99]	✓	✓	✓	✓	✓	✓	✓	✓	✓
30	Roychowdhury et al. [36]	✓	✓	✓	✗	✗	✓	✓	✓	✓
70	Singh et al. [100]	✓	✓	✓	✓	✓	✓	✓	✓	✓
71	Vetil et al. [101]	✓	✓	✓	✓	✓	✓	✓	✓	✓
72	Xiao et al. [102]	✓	✓	✓	✓	✓	✓	✓	✓	✓
73	Yan et al. [103]	✓	✓	✓	✓	✓	✓	✓	✓	✓
74	Yarlagadda et al. [104]	✓	✓	✓	✗	✗	✓	✓	✓	✓
75	Zhang, Cui and Ren [105]	✓	✓	✓	✗	✗	✓	✓	✓	✓
76	Zhu et al. [106]	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 6: *(continued)*.

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

Medical application. The classification papers within the FSL domain cover a wide range of anatomical structures and regions, as well as specific lesions. Here's a breakdown of the number of papers categorized by the anatomical structure(s) investigated: two out of 27 studies (7%) perform brain image classification, focusing on different types of tumours and MRI contrast types; six studies (22%) address breast image classification, with four concentrating on breast tumours and two on breast metastases involving nearby lymph nodes; two studies (7%) investigate cell image classification; two studies (7%) focus on cervix image classification; three studies (11%) pertain to colon image classification; four studies (15%) are dedicated to fundus eye image classification, with 2 investigating different diseases; one study (4%) deals with liver disease classification; 11 studies (41%) involve lung image classification; one study (4%) is concerned with pancreas image classification; one study (4%) classifies prostate tumour images; 7 studies (26%) address skin image classification, covering different diseases; one study (4%) investigates esophagus image classification; one study (4%) focuses on stomach image classification. Note that [105] is not included in this analysis as it did not specify which anatomical structures were part of their study. For a visual representation of the distribution, refer to Figure 12.

Meta-learning methods. In the context of classification studies employing FSL, the distribution of meta-learning methods is as follows: six out of 27 studies (22%) utilize initialization-based methods; 10 studies (37%) opt for metric-learning-based algorithms; the remaining 11 studies (41%) do not incorporate any meta-learning techniques. For a visual representation of the distribution, refer to Figure 13.

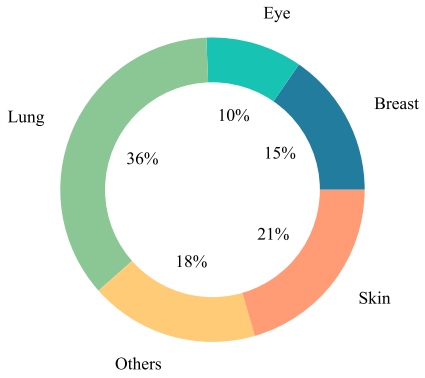

Figure 12: Classification studies grouped by the anatomical structure investigated.

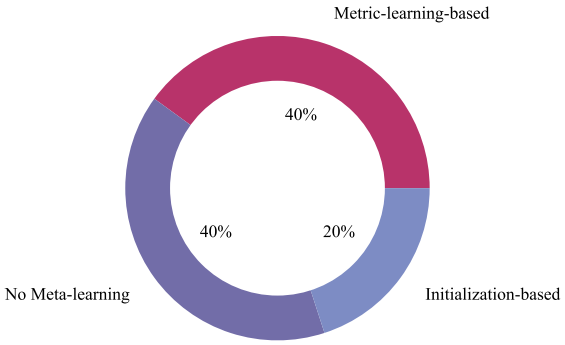

Figure 13: Classification studies grouped by meta-learning method employed.

K-shot. Among the 27 selected studies in the classification domain using FSL, the training configurations are distributed as follows: 13 studies (48%) employ k-shot training with k ranging from 2 to 20; six studies (22%) utilize both OSL and FSL; one study (4%) uses both FSL and ZSL; five studies (19%) exclusively perform 1-shot training; two studies (7%) solely employ 0-shot training.

Image modalities. In the context of classification studies within the FSL domain, the distribution of imaging modalities is as follows: three studies out of 27 (11%) use CT images; three studies (11%) employ MRI images; seven studies (26%) utilize dermoscopic images; 11 studies (41%) rely on X-ray images; three studies (11%) involve fundus images; two studies (7%) make use of microscopic images; nine studies (33%) employ histopathological images; one study (4%) utilizes endoscopy images; one study (4%) involves cytological images; one study (4%) uses OCT images.

Model evaluation. To assess the behaviour and robustness of the models in the selected studies, various evaluation techniques were employed as follows: 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.

Model performance grouped by organ and meta-learning method. In Figure 14, Figure 15, Figure 16, and Figure 17, we present a summary of the model performance in forest plots, categorized by anatomical structure, in terms of Accuracy, AUROC, F1-score, and Recall, respectively. Conversely, in Figure 18, Figure 19, Figure 20, and Figure

21, we depict the performance in terms of Accuracy, AUROC, F1-score, and Recall, respectively, by grouping the studies according to the employed meta-learning methods. For each study, we have considered the highest performance achieved (across different experiments and image modalities). In each forest plot, we provide the mean and the 95% CI across all the studies within the corresponding group. Note that the results of [81] and [105] are not included in the forest plot since they provide average results across different anatomical structures.

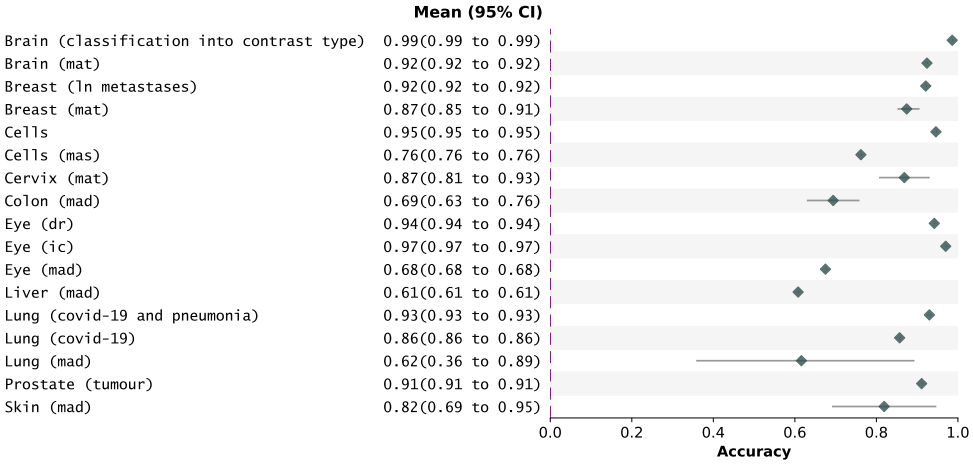

Figure 14: Forest plot of classification studies performance based on Accuracy metric. Studies are grouped by the anatomical structure investigated. DR = Diabetic Retinopathy; IC = Intraretinal Cyst; LN = Limph Nodes; MAD = Mead Across Diseases; MAS = Mean Across Structures; MAT = Mean Across Tumours.

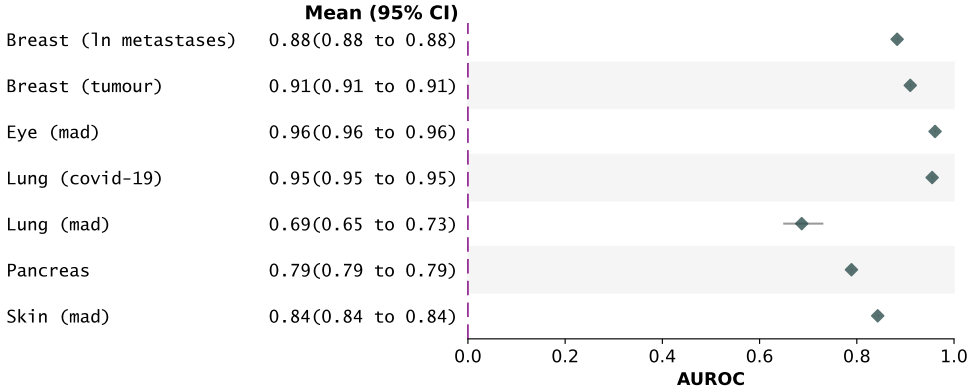

Figure 15: Forest plot of classification studies performance based on AUROC metric. Studies are grouped by the anatomical structure investigated. MAD = Mean Across Diseases.

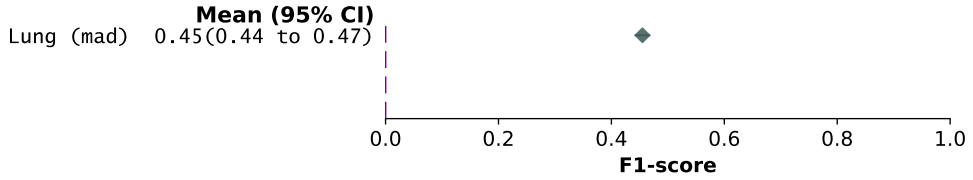

Figure 16: Forest plot of classification studies performance based on F1-score metric. Studies are grouped by the anatomical structure investigated. MAD = Mean Across Diseases.

Overall pipeline. In Table 7, we delineate which stages of the defined pipeline are employed by each study. Here are the distributions of studies based on their utilization of pre-training, training, and data augmentation techniques: 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

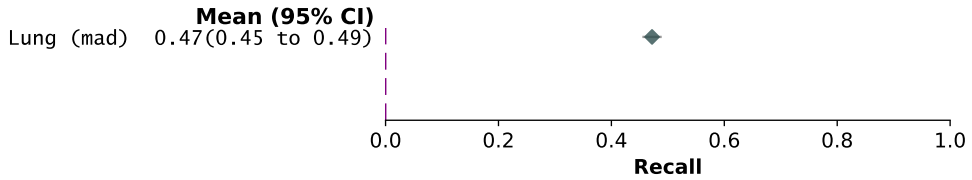

Figure 17: Forest plot of classification studies performance based on Recall metric. Studies are grouped by the anatomical structure investigated. MAD = Mean Across Diseases.

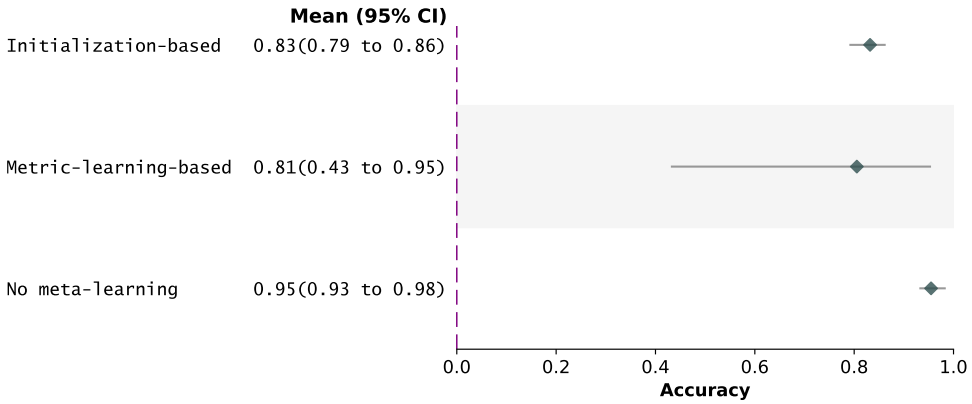

Figure 18: Forest plot of classification studies performance based on Accuracy metric. Studies are grouped by the meta-learning method employed.

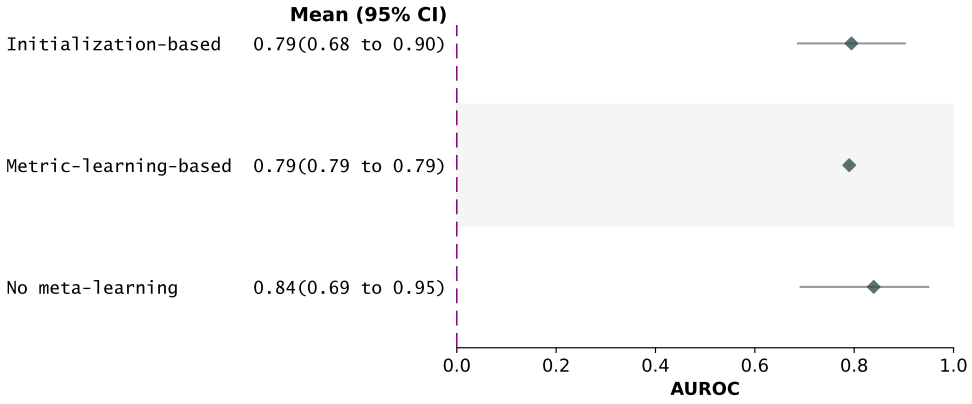

Figure 19: Forest plot of classification studies performance based on AUROC metric. Studies are grouped by the meta-learning method employed.

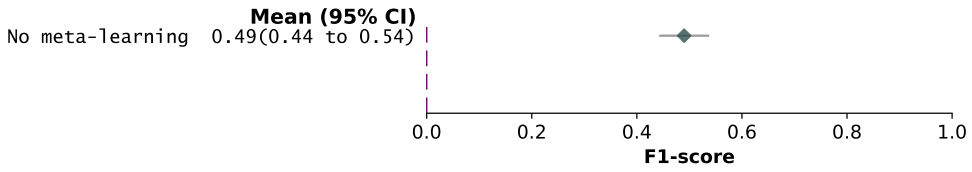

Figure 20: Forest plot of classification studies performance based on F1-score metric. Studies are grouped by the meta-learning method employed.

their main 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%)

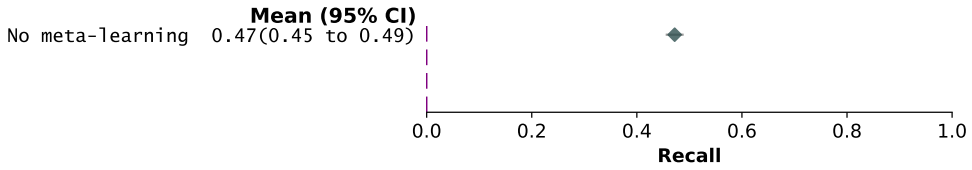

Figure 21: Forest plot of classification studies performance based on Recall metric. Studies are grouped by the meta-learning method employed.

realied on classical data augmentation techniques; two studies (7%) utilized generative methods for data augmentation. The remaining 15 studies did not employ data augmentation.

Study ID	Pub. ref.	Pre-training	Training	Data augmentation
51	Ali et al. [81]	Supervised	Meta	None
52	Cai, Hu, and Zheng [82]	None	Meta	Classical
53	Cai et al. [83]	Supervised	Meta	Classical
54	Cano and Cruz-Roa [84]	None	Meta	None
55	Chen et al. [85]	Unsupervised	Supervised	None
56	Chou et al. [86]	None	Supervised	None
57	Dai et al. [87]	Supervised	Meta	None
58	Huang, Huang and Tang [88]	None	Supervised	Generative
59	Jiang et al. [89]	Supervised	Meta	Classical
60	Jin et al. [90]	None	Meta	Classical
61	Mahapatra, Ge and Reyes [91]	Supervised	Self-supervised and Supervised	Generative
62	Maicas et al. [92]	Meta	Supervised	None
63	Mohan et al. [93]	Supervised	Supervised	Classical
64	Moukheiber et al. [94]	Supervised	Meta	None
65	Naren, Zhu and Wang [95]	None	Meta	None
66	Ouahab, Ben-Ahmed and Fernandez-Maloigne [96]	Supervised	Meta	Classical
67	Paul, Tang and Summers [97]	Supervised	Supervised	None
68	Paul et al. [99]	Supervised	Zero-shot and Semi-supervised	None
69	Paul et al. [98]	Supervised	Supervised	None
30	Roychowdhury et al. [36]	None	Supervised	Classical
70	Singh et al. [100]	None	Meta	Classical
71	Vétil et al. [101]	None	Zero-shot and Supervised	Classical
72	Xiao et al. [102]	None	Meta	None
73	Yan et al. [103]	Supervised	Meta	Classical
74	Yarlagadda et al. [104]	Supervised	Supervised	None
75	Zhang, Cui and Ren [105]	None	Meta	None
76	Zhu et al. [106]	None	Meta	None

Table 7: Main pipeline steps adopted by classification studies.

4.2.3 Registration

We included six relevant studies, each focusing on medical registration as its primary task. Table 8 summarizes of all the essential information from these selected studies. In addition, we provide information concerning ROB and the applicability of each study in Table 9.

Study ID	Pub. ref.	Algorithm/Pipeline	K-shot	Best performance	Meta-learning type
77	Fechter and Baltas [107]	U-net + Differential spatial transformer module	1-shot	Landmark distance: 1.49 (Lungs) Dice: 0.860 (Heart)	None
78	Ferrante et al. [108]	U-net + Unsupervised learning	1-shot	Dice: 0.920 (Heart) 0.890 (Lungs)	None
79	He et al. [109]	Perception-Correspondence Registration	5-shot	Dice: 0.857 (Heart, MAS) 0.867 (Cervical vertebra, MAS) 0.800 (Brain, MAS)	None
34	Shi et al. [37]	Joint Registration and Segmentation Self-training Framework	5-shot	Dice: 0.759 (Brain, MAS) 0.539 (Abdomen, MAS)	None
45	Xu and Niethammer [38]	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. [110]	CNN + Spatial transformer + similarity loss + smooth loss + cyclic loss	1-shot	TRE: 1.03 (Lung)	None

Table 8: FSL studies for medical image registration.

		Risk of Bias					Applicability			
Study ID	Pub. ref.	Part.	Pred	Out.	Analysis	Overall	Part.	Pred.	Out.	Overall
77	Fechter, Baltas [107]	✓	✓	✓	✓	✓	✓	✓	✓	✓
78	Ferrante et al. [108].	✓	✓	✓	✓	✓	✓	✓	✓	✓
79	He et al. [109]	✓	✓	✓	✓	✓	✓	✓	✓	✓
34	Shi et al. [37]	✓	✓	✓	✓	✓	✓	✓	✓	✓
45	Xu and Niethammer [38]	✓	✓	✓	✓	✓	✓	✓	✓	✓
80	Zhang et al. [110]	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 9: ROB of FSL studies for medical image registration.

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

Medical application. The selected registration papers address a range of anatomical regions. Here's the breakdown of the number of studiescategorized by the anatomical structure investigated: three out of 6 studies (50%) explore brain registration; one study (17%) focuses on the registration of knee bones and cartilages; three studies (50%) delve into heart registration; three studies (50%) concentrate on lung registration; one study (17%) pertains to abdominal registration; one study (17%) deals with cervical vertebra registration. For a visual representation of the distribution, refer to Figure 22.

Meta-learning methods. In the domain of registration studies, all of the selected papers (100 %) do not employ the meta-learning paradigm. For a visual representation of the distribution of these studies according to the anatomical structure investigated, refer to Figure 23.

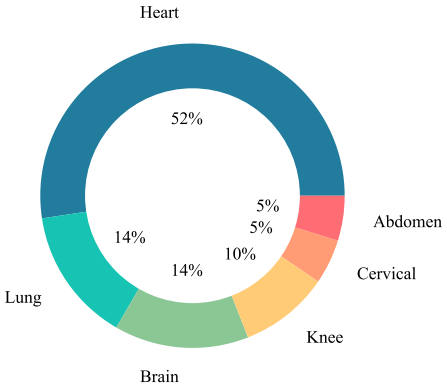

Figure 22: Registration studies grouped by the anatomical structure investigated.

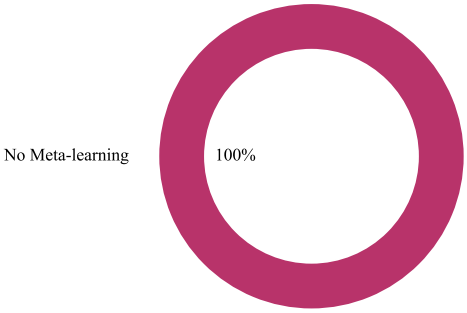

Figure 23: Registration studies grouped by the meta-learning method employed.

K-shot. Among the six selected studies, the distribution of training strategies is as follows: two studies (33%) solely employ FSL; three studies (50%) exclusively investigate OSL; one study (17%) performs both FSL and OSL.

Image modalities. In the context of registration studies, the distribution of imaging modalities used among the selected papers is as follows: four 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 examine the behaviour and robustness of the models in the selected registration studies various evaluation techniques were employed as follows: two studies (33%) utilized only ablation studies; one study (17%) used cross-validation. The remaining studies (50%) did not employ any specific model evaluation technique.

Model performance grouped by organ and meta-learning method. In Figure 24, Figure 25, and Figure 26, we provide a summary of the model performance in forest plots, categorized by anatomical structure, in terms of Dice score, Average Landmark Distance (ALD), and Target Registration Error (TRE), respectively. Conversely, in Figure 27, Figure 28, and Figure 29, we depict the performance in terms of Dice score, ALD, and TRE, respectively, by grouping the studies according to the employed meta-learning methods. For each study, we considered the highest performance achieved (across different experiments and image modalities). In each forest plot, we provided the mean and the 95% CI across all the studies within the corresponding group.

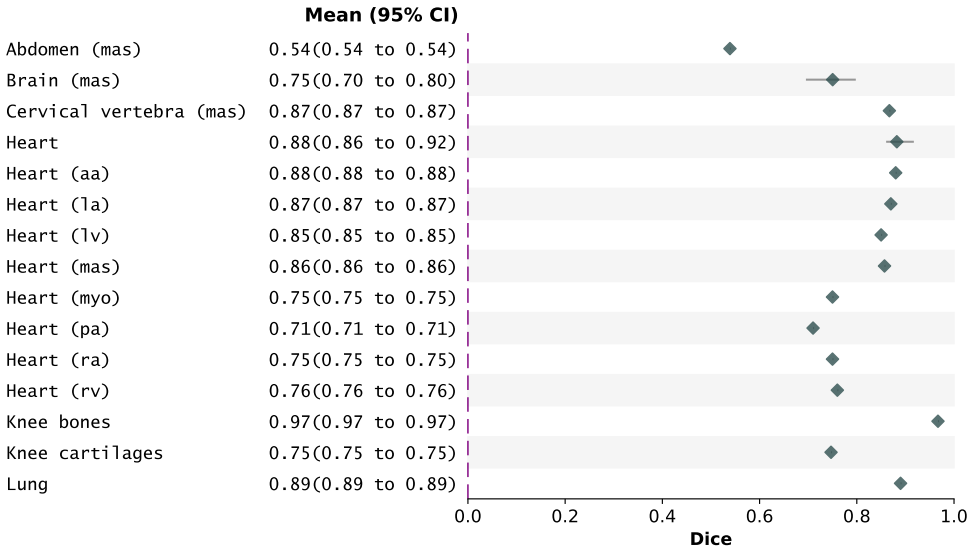

Figure 24: Forest plot of registration studies performance based on Dice metric. Studies are grouped by the anatomical structure investigated.

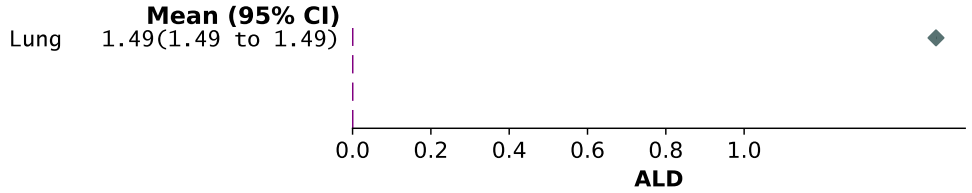

Figure 25: Forest plot of registration studies performance based on ALD metric. Studies are grouped by the anatomical structure investigated.

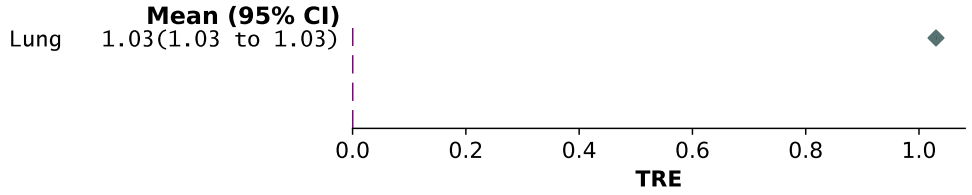

Figure 26: Forest plot of registration studies performance based on TRE metric. Studies are grouped by the anatomical structure investigated.

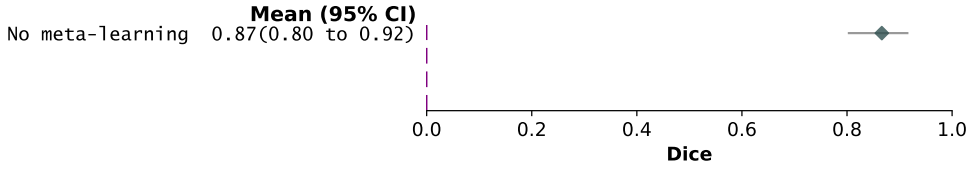

Figure 27: Forest plot of registration studies performance based on Dice metric. Studies are grouped by the meta-learning method employed.

Overall pipeline. In Table 10, we delineate which stages of the defined pipeline are employed by each study. Here are the distributions of studies based on their utilization of pre-training, training, and data augmentation techniques: two out of six studies (33%) employed classical supervised pre-training. The remaining four studies did not employ any pre-training at all. Five out of six studies (83%) utilized supervised training as their primary training approach. One

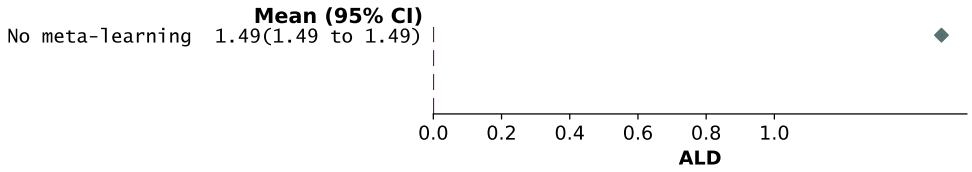

Figure 28: Forest plot of registration studies performance based on ALD metric. Studies are grouped by the meta-learning method employed.

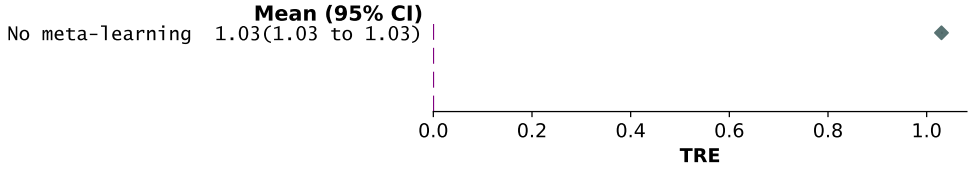

Figure 29: Forest plot of registration studies performance based on TRE metric. Studies are grouped by the meta-learning method employed.

study (17%) employed unsupervised learning for the main training stage. Two out of six studies (33%) used classical data augmentation techniques. The other four studies (67%) did not exploit data augmentation.

Study ID	Pub. ref.	Pre-training	Training	Data augmentation
77	Fechter, Baltas [107]	None	Supervised	None
78	Ferrante et al. [108]	None	Unsupervised	None
79	He et al. [109]	None	Supervised	Classical
34	Shi et al. [37]	Supervised	Supervised	None
45	Xu and Niethammer [38]	Supervised	Supervised	Classical
80	Zhang et al. [110]	None	Supervised	None

Table 10: Main pipeline steps adopted by registration studies.

5 Discussion

This review assessed 80 FSL studies applied to the field of medical imaging. We organized these studies into three distinct categories: segmentation, classification, and registration, according to their main outcome. For each category, we collected essential information such as the algorithm or pipeline used, the employed meta-learning methods, the quantity of labelled data utilized during training, and the highest achieved performance. Furthermore, we summarized the outcomes of each category, classifying them based on both the specific anatomical structures investigated and the presence or absence of meta-learning techniques. In addition, we applied the PROBAST method to evaluate both ROB and the applicability of each study in the context of each outcome. Finally, we defined a generic pipeline enclosing all the techniques shared among the selected papers.

Below, we delve into the results derived from our analysis according to the objectives outlined in Sec 1.2.

Studies distribution per outcome. Figure 4 clearly illustrates the predominant focus of FSL studies in medical imaging. Segmentation tasks are the most prominent, constituting the majority at 61%, followed by classification tasks at 32% and registration tasks at 7%. In the following paragraphs, we provide a more detailed exploration of how these studies are further distributed, considering both the specific anatomical structures investigated and the meta-learning methods employed.

Studies distribution and results per anatomical structure investigated. Figure 6 provides an insightful overview of the distribution of anatomical structures studied in segmentation tasks in medical imaging. Notably, the heart emerges as the most extensively investigated anatomical structure, comprising 34% of the studies. Following closely are the kidney, spleen, and liver, each accounting for 13% of the research. The brain also features significantly, representing 10% of the studies. In Figure 12, we shift our focus to the distribution of anatomical structures in classification studies. Here, 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, Figure 22 highlights the distribution of anatomical structures studied in registration tasks. The heart emerges as the most commonly examined organ in this category, representing 52% of the investigations, followed by lungs and brain, accounting for 14% of the studies and the knee for 10%. Finally, the cervical vertebra and abdomen are the main application in 5% of the studies.

Figure 8 provides valuable insights into the performance of segmentation tasks in terms of the Dice score across various anatomical structures. Notably, femur segmentation demonstrates the highest Dice score, although it's worth mentioning that only one study addresses this task, making the result 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. Shifting our focus to IoU, Figure 9 highlights that hip, knee, and phalanx segmentation provide the best IoU results. However, it's important to note that these results are based on a limited number of studies, which may affect their reliability. 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.

Moving on to classification tasks, Figure 14 shows that, in terms of accuracy, the classification of brain images into different contrast type provides the best performance, even tough must be noted that it represents a quite easy task. On the other hand, the classification of skin lesions demonstrates providing a high mean classification accuracy (0.82) across several studies. The worst performing case, instead, is provided by the segmentation of the liver, which yields 0.61 accuracy as a mean across different diseases. Concerning the AUROC metric, in Figure 15 the classification of eye images achieves the highest performance, with a score of 0.96, while the worst performing, even tough computed on a higher number of studies, is the segmentation of the lungs, which provides a 0.69 AUROC as mean across different diseases. Finally, it is worth mentioning that a few studies investigate the classification of different lung diseases,

utilizing the F1-score and Recall metrics. However, the results in both cases are poor, scoring below 0.5, as shown in Figure 16 and Figure 17.

Lastly, turning our attention to the registration task in the following, we examine the key findings. In terms of the Dice score (Figure 24), our analysis suggests that FSL registration achieves the highest result in the registration of knee bones. However, it's important to note that this result should be interpreted cautiously due to its reliance on a single study. On the other hand, when it comes to the registration of whole heart images, there is a consistently high mean Dice score of 0.88, with a small CI derived from several studies. Regarding the ALD and TRE metrics (Figure 25 and Figure 26), it's noteworthy that these metrics are employed only in the context of lung image registration.

Studies distribution and results per meta-learning method employed. Concerning segmentation papers, according to Figure 7, most studies do not employ meta-learning methods for segmentation tasks, accounting for 55% of the investigations. Among the studies that do use meta-learning methods, metric-learning-based approaches are the most commonly utilized, constituting 26% of the selected studies, followed by initialization-based studies (13%) and hallucination-based (6%). Shifting the focus to the classification tasks, Figure 13 reveals that 40% of the studies do not use any meta-learning algorithm. Among the meta-learning algorithms instead, 40 % of all studies employ metric learning-based methods, and 20% initialization-based methods. No classification study instead employ hallucination-based methods. Finally, as for registration purposes, no study employ meta-learning methods.

Regarding the models' performance, Figure 10 reveals that both no-meta-learning methods and hallucination-based methods yield the highest mean Dice scores (0.84). Notably, no-meta-learning methods exhibit a wider CI, which is expected given their application across a greater number of studies. In contrast, metric-learning-based methods, while being the most commonly employed among all the meta-learning methods in segmentation studies, yield slightly lower results, with a mean Dice score of 0.79 and a larger CI. For the IoU metric, as demonstrated in Figure 11, only initialization-based methods and no-meta-learning methods utilize this metric. Remarkably, no-meta-learning methods outperform the others significantly, delivering a notably better performance with a smaller CI. In the context of classification tasks, as illustrated in Figure 18, it becomes evident that studies opting not to utilize meta-learning methods consistently yield the most impressive accuracy results, coupled with a notably narrow CI. Conversely, metric-learning-based methods exhibit notably poorer performance, with a mean accuracy of just 0.81 and a larger CI. Regarding AUROC, as emphasized in Figure 19, the absence of meta-learning methods consistently delivers the most remarkable performance, boasting a mean AUROC of 0.84. Interestingly, initialization-based and metric-learning-based methods both yield a mean AUROC of 0.79 despite metric-learning techniques exhibiting a wider CI. In the context of registration tasks, Figures 27, 28, and 29 illustrate the mean performance metrics for Dice, ALD, and TRE, respectively. In this specific case, all the studies opted for utilising non-meta-learning methods.

Training data, imaging modalities and robustness evaluation distributions. When analyzing the training set size in studies focused on the segmentation task, it's worth highlighting that most of these studies, comprising 58%, incorporate one or more labelled data samples into their training phase. A significant proportion, 40%, also utilize OSL, while only 2% employ ZSL. A similar trend is observed in classification studies, where a substantial 70% of the investigations involve at least one labelled example during their training process, while 19% exclusively rely on OSL. In this context, just 7% of the examined studies make use of ZSL. However, when we turn our attention to registration studies, unlike those in segmentation and classification, mainly rely on OSL (50%). Among the remaining studies, 40% opt for FSL with more than one labelled image in the training set, and only 17% perform ZSL.

In terms of imaging modalities, MRI data are the most commonly utilized in both segmentation and registration studies, constituting 60% and 83% of the cases, respectively. On the other hand, when it comes to classification studies, X-ray imaging takes the lead, being employed in 41% of all studies.

In terms of evaluating the robustness of models, it's worth noting that 74% of the segmentation studies incorporate some form of model robustness evaluation. This typically involves conducting ablation studies and or employing cross-validation techniques. However, considering classification and registration studies, only half of them incorporate robustness evaluation for the models, with ablation studies being the most common approach in these cases.

Identification of a standard pipeline. We present our comprehensive analysis of the examined studies in Figure 5, illustrating the components of the pipelines common to each study. We summarized all the steps into three main clusters: pre-training, training, and data augmentation. We found that pre-training is performed using four paradigms: supervised, unsupervised, self-supervised, or meta-learning. On the other hand, the final training phase encompasses these four methods, including semi-supervised learning and zero-shot learning approaches as well. Regarding data augmentation techniques, we emphasized three methods that are mainly employed: classical data augmentation, which includes geometric or image channel transformations, generative-based augmentation and registration-based augmentation. The latter approach, notably, found extensive application in segmentation studies, where segmentation and registration models are trained jointly.

As anticipated, meta-learning emerged as the prevailing approach for addressing FSL tasks, particularly within segmentation and classification studies. As well, in the realm of segmentation and classification, supervised learning is the most commonly employed method immediately following meta-learning. It's worth noting that the interpretation of classic supervised learning varies based on the specific modules introduced in each study. In segmentation studies, semi-supervised learning is also a commonly employed method involving the joint utilization of labelled and unlabeled data. In contrast, when examining registration studies, almost all employed traditional supervised training, with only one study opting for unsupervised training.

Concerning the application of data augmentation techniques, the predominant practice involves augmenting datasets through standard geometric or colour channel transformations. In segmentation studies, a common approach is to integrate a registration network, trained in conjunction with the segmentation network, to provide additional training data—a technique not observed in classification studies, where classical data augmentation is prevalent, and only two studies explored generative techniques. Finally, among registration studies, those incorporating data augmentation exclusively rely on classical augmentation techniques.

Concluding statements. Our comprehensive analysis offers several valuable insights into the methods employed for addressing FSL tasks. To begin with, although meta-learning methods are widely adopted and successful, they are implemented in a variety of ways. Metric learning-based methods have garnered substantial attention, whereas hallucination-based techniques have not been as extensively explored. Secondly, our analysis highlights that the heart, abdomen, and lungs have been the primary areas of focus in FSL studies. This is likely due to the availability of well-established benchmark datasets such as CHAOS [111], MS-CMRSeg [112], and NIH Chest X-ray [113]. However, there exists untapped potential for researchers to delve into relatively less-explored medical applications, including the prostate, digestive organs, and various bones. Furthermore, we observed that some studies, particularly those in classification and registration tasks, may not conduct comprehensive model investigation analyses. This gap can potentially lead to incomplete and unreliable performance assessments. Lastly, we noted issues related to ROB in some studies. Many studies lack clarity in explaining how they address the FSL task, even when claiming to use reduced amounts of labelled data. In light of these findings, we encourage future researchers in the field to consider the following actions:

- Explore and invest in hallucination-based methods, given their promising performance potential.
- Expand the scope of medical applications investigated, especially in less-explored areas.
- Prioritize thorough model validation and comprehensive analyses to facilitate fair comparisons and the practical implementation of FSL models in clinical settings.

In addition, our analysis underscores the prevalence of meta-learning as a commonly used approach for FSL tasks. However, it also highlights the versatility of alternative methods, including supervised learning with innovative modules and semi-supervised learning, which have proven effective, particularly in segmentation tasks and registration studies. These diverse strategies, coupled with appropriate data augmentation techniques, demonstrate the adaptability of FSL methodologies to address the challenges posed by limited data.

6 Conclusions

In our extensive systematic review, we conducted a thorough examination of the application of FSL in medical image analysis. We categorized the selected studies based on their intended outcome domains, i.e. segmentation, classification, and registration. Our analysis entailed a detailed investigation of these studies, focusing on the specific anatomical structures targeted and the meta-learning methods employed. Moreover, we provided a comprehensive performance summary by grouping the studies according to the anatomical structures studied and the chosen meta-learning techniques. This summary included mean performance values along with a 95% CI. Additionally, we explored supplementary aspects, such as the quantity of training data used, the imaging modalities employed, and the methods used to assess the robustness of the models. We meticulously evaluated each study for its ROB and applicability to ensure the credibility of the findings presented. Finally, we introduced a general pipeline that all the studies in our analysis either partially or fully adopted.

The key findings from our systematic review are as follows. Concerning the outcome domain, segmentation tasks are the most prominently addressed outcome in FSL applied to medical image analysis, and among the various anatomical structures investigated, the abdomen and heart receive the most attention. In terms of the training data, most of the studies demonstrate the effectiveness of FSL by utilizing more than one example during the training phase. Notably, CT and X-ray imaging modalities are the most frequently employed. Regarding the robustness evaluation, our review unveil a significant gap: indeed, a considerable number of studies, particularly those in the classification and registration domains, lack proper robustness assessment, underscoring the need for improved evaluation practices in these areas.

Concerning the meta-learning techniques employed, metric-learning-based approaches are the predominant choice among meta-learning methods, despite providing poorer results w.r.t. other meta-learning and non-meta-learning methods. Among non-meta-learning approaches, classical supervised learning with custom modules, as well as semi-supervised learning, are commonly applied. In general, non-meta-learning methods perform better w.r.t meta-learning ones. Finally, in terms of data augmentation, most studies address the challenge of limited data by incorporating data augmentation techniques. Classical augmentation methods are the most widely employed for this purpose. Our systematic review is intended to serve as a valuable resource for future researchers in the field, offering guidance on areas of anatomical interest and methodological exploration that warrant further investigation. Our ultimate goal is to promote the advancement and broader adoption of FSL techniques within the medical imaging domain by addressing identified gaps, emphasizing robustness evaluation and showing an overview of the methods currently used in the SOTA.

References

[1] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, and others. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[2] Wei-Yu Chen, Yen-Cheng Liu, Zsolt Kira, Yu-Chiang Frank Wang, and Jia-Bin Huang. A closer look at few-shot classification. *arXiv preprint arXiv:1904.04232*, 2019.

[3] Mahdi Rezaei and Mahsa Shahidi. Zero-shot learning and its applications from autonomous vehicles to COVID-19 diagnosis: A review. *Intelligence-based medicine*, 3:100005, 2020. Publisher: Elsevier.

[4] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020. Publisher: ACM New York, NY, USA.

[5] Viet-Khoa Vo-Ho, Kashu Yamazaki, Hieu Hoang, Minh-Triet Tran, and Ngan Le. Chapter 19 - Neural architecture search for medical image applications. In Hien Van Nguyen, Ronald Summers, and Rama Chellappa, editors, *Meta Learning With Medical Imaging and Health Informatics Applications*, The MICCAI Society book Series, pages 369–384. Academic Press, 2023. ISBN 978-0-323-99851-2.

[6] Jannatul Nayem, Sayed Sahriar Hasan, Noshin Amina, Bristy Das, Md Shahin Ali, Md Manjurul Ahsan, and Shivakumar Raman. Few Shot Learning for Medical Imaging: A Comparative Analysis of Methodologies and Formal Mathematical Framework. *arXiv preprint arXiv:2305.04401*, 2023.

[7] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. Self-supervised Learning for Few-shot Medical Image Segmentation. *IEEE Transactions on Medical Imaging*, 2022. Publisher: IEEE.

[8] Wenji Wang, Qing Xia, Zhiqiang Hu, Zhennan Yan, Zhuowei Li, Yang Wu, Ning Huang, Yue Gao, Dimitris Metaxas, and Shaoting Zhang. Few-shot learning by a cascaded framework with shape-constrained pseudo label assessment for whole heart segmentation. *IEEE Transactions on Medical Imaging*, 40(10):2629–2641, 2021. Publisher: IEEE.

[9] Xiaoyan Wang, Yiwen Yuan, Dongyan Guo, Xiaojie Huang, Ying Cui, Ming Xia, Zhenhua Wang, Cong Bai, and Shengyong Chen. SSA-Net: Spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning. *Medical Image Analysis*, 79:102459, 2022. Publisher: Elsevier.

[10] Chen Chen, Chen Qin, Cheng Ouyang, Zeju Li, Shuo Wang, Huaqi Qiu, Liang Chen, Giacomo Tarroni, Wenjia Bai, and Daniel Rueckert. Enhancing MR image segmentation with realistic adversarial data augmentation. *Medical Image Analysis*, 82:102597, 2022. Publisher: Elsevier.

[11] Timothy Hospedales, Antreas Antoniou, Paul Micaelli, and Amos Storkey. Meta-learning in neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5149–5169, 2021. Publisher: IEEE.

[12] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR, 2017.

[13] Alex Nichol, Joshua Achiam, and John Schulman. On first-order meta-learning algorithms. *arXiv preprint arXiv:1803.02999*, 2018.

[14] Sachin Ravi and Hugo Larochelle. Optimization as a model for few-shot learning. In *International conference on learning representations*, 2016.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997. Publisher: MIT press.

[16] Ke Li and Jitendra Malik. Learning to optimize. *arXiv preprint arXiv:1606.01885*, 2016.

[17] Richard Bellman. A Markovian decision process. *Journal of mathematics and mechanics*, pages 679–684, 1957. Publisher: JSTOR.

[18] Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy Lillicrap. Meta-learning with memory-augmented neural networks. In *International conference on machine learning*, pages 1842–1850. PMLR, 2016.

[19] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 6, 1993.

[20] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, and others. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, page 0. Lille, 2015.

[21] Elad Hoffer and Nir Ailon. Deep metric learning using triplet network. In *International workshop on similarity-based pattern recognition*, pages 84–92. Springer, 2015.

[22] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, and others. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016.

[23] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[24] Mohamed Elhoseiny, Babak Saleh, and Ahmed Elgammal. Write a classifier: Zero-shot learning using purely textual descriptions. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2584–2591, 2013.

[25] Adam Santoro, David Raposo, David G Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. A simple neural network module for relational reasoning. *Advances in neural information processing systems*, 30, 2017.

[26] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018.

[27] Bharath Hariharan and Ross Girshick. Low-shot visual recognition by shrinking and hallucinating features. In *Proceedings of the IEEE international conference on computer vision*, pages 3018–3027, 2017.

[28] Yu-Xiong Wang, Ross Girshick, Martial Hebert, and Bharath Hariharan. Low-shot learning from imaginary data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7278–7286, 2018.

[29] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, and others. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, 88:105906, 2021. Publisher: Elsevier.

[30] Mingxuan Gu, Sulaiman Vesal, Ronak Kosti, and Andreas Maier. 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*, pages 20–25. Springer, 2022.

[31] Matthew R Keaton, Ram J Zaveri, and Gianfranco Doretto. CellTranspose: Few-shot Domain Adaptation for Cellular Instance Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 455–466, 2023.

[32] Shaohua Li, Xiuchao Sui, Jie Fu, Huazhu Fu, Xiangde Luo, Yangqin Feng, Xinxing Xu, Yong Liu, Daniel SW Ting, and Rick Siow Mong Goh. Few-shot domain adaptation with polymorphic transformers. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 330–340. Springer, 2021.

[33] Robert F Wolff, Karel GM Moons, Richard D Riley, Penny F Whiting, Marie Westwood, Gary S Collins, Johannes B Reitsma, Jos Kleijnen, Sue Mallett, and PROBAST Group†. Probast: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of internal medicine*, 170(1):51–58, 2019.

[34] Yuting He, Tiantian Li, Guanyu Yang, Youyong Kong, Yang Chen, Huazhong Shu, Jean-Louis Coatrieux, Jean-Louis Dillenseger, and Shuo Li. Deep complementary joint model for complex scene registration and few-shot segmentation on medical images. In *European Conference on Computer Vision*, pages 770–786. Springer, 2020.

[35] Yuting He, Rongjun Ge, Xiaoming Qi, Yang Chen, Jiasong Wu, Jean-Louis Coatrieux, Guanyu Yang, and Shuo Li. Learning Better Registration to Learn Better Few-Shot Medical Image Segmentation: Authenticity, Diversity, and Robustness. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. Publisher: IEEE.

[36] Sohini Roychowdhury. Few Shot Learning Framework to Reduce Inter-observer Variability in Medical Images. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 4581–4588. IEEE, 2021.

[37] Huabang Shi, Liyun Lu, Mengxiao Yin, Cheng Zhong, and Feng Yang. Joint few-shot registration and segmentation self-training of 3D medical images. *Biomedical Signal Processing and Control*, 80:104294, 2023. Publisher: Elsevier.

[38] Zhenlin Xu and Marc Niethammer. 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*, pages 420–429. Springer, 2019.

[39] Maximilian Blendowski, Hannes Nickisch, and Mattias P Heinrich. How to learn from unlabeled volume data: Self-supervised 3d context feature learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 649–657. Springer, 2019.

[40] Sixian Chan, Cheng Huang, Cong Bai, Weilong Ding, and Shengyong Chen. Res2-UNeXt: a novel deep learning framework for few-shot cell image segmentation. *Multimedia Tools and Applications*, 81(10):13275–13288, 2022. Publisher: Springer.

[41] Hengji Cui, Dong Wei, Kai Ma, Shi Gu, and Yefeng Zheng. A unified framework for generalized low-shot medical image segmentation with scarce data. *IEEE Transactions on Medical Imaging*, 40(10):2656–2671, 2020. Publisher: IEEE.

[42] Wangbin Ding, Lei Li, Xiahai Zhuang, and Liqin Huang. Cross-modality multi-atlas segmentation using deep neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 233–242. Springer, 2020.

[43] Yuhang Ding, Xin Yu, and Yi Yang. Modeling the probabilistic distribution of unlabeled data for one-shot medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1246–1254, 2021. Issue: 2.

[44] Azade Farshad, Anastasia Makarevich, Vasileios Belagiannis, and Nassir Navab. 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*, pages 45–55. Springer, 2022.

[45] Ruiwei Feng, Xiangshang Zheng, Tianxiang Gao, Jintai Chen, Wenzhe Wang, Danny Z Chen, and Jian Wu. Interactive few-shot learning: Limited supervision, better medical image segmentation. *IEEE Transactions on Medical Imaging*, 40(10):2575–2588, 2021. Publisher: IEEE.

[46] Pedro HT Gama, Hugo Oliveira, and Jefersson A dos Santos. Learning to Segment Medical Images from Few-Shot Sparse Labels. In *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 89–96. IEEE, 2021.

[47] Pedro Henrique Targino Gama, Hugo Neves Oliveira, Jose Marcato, and Jefersson Dos Santos. Weakly Supervised Few-Shot Segmentation Via Meta-Learning. *IEEE Transactions on Multimedia*, 2022. Publisher: IEEE.

[48] Junyu Guo, Ayobami Odu, and Ivan Pedrosa. Deep learning kidney segmentation with very limited training data using a cascaded convolution neural network. *PloS one*, 17(5):e0267753, 2022. Publisher: Public Library of Science San Francisco, CA USA.

[49] Stine Hansen, Srishti Gautam, Robert Jenssen, and Michael Kampffmeyer. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels. *Medical Image Analysis*, 78:102385, 2022. Publisher: Elsevier.

[50] Robert Jenssen, Stine Hansen, Srishti Gautam, and Michael Kampffmeyer. A self-guided anomaly detection-inspired few-shot segmentation network. 2022.

[51] Thomas Joyce and Sebastian Kozerke. 3D medical image synthesis by factorised representation and deformable model learning. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 110–119. Springer, 2019.

[52] Rabindra Khadka, Debesh Jha, Steven Hicks, Vajira Thambawita, Michael A Riegler, Sharib Ali, and Pål Halvorsen. Meta-learning with implicit gradients in a few-shot setting for medical image segmentation. *Computers in Biology and Medicine*, 143:105227, 2022. Publisher: Elsevier.

[53] Afifa Khaled, Jian-Jun Han, and Taher A Ghaleb. Multi-Model Medical Image Segmentation Using Multi-Stage Generative Adversarial Networks. *IEEE Access*, 10:28590–28599, 2022. Publisher: IEEE.

[54] Pulkit Khandelwal and Paul Yushkevich. 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*, pages 73–84. Springer, 2020.

[55] Soopil Kim, Sion An, Philip Chikontwe, and Sang Hyun Park. Bidirectional rnn-based few shot learning for 3d medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 1808–1816, 2021. Issue: 3.

[56] Yiwen Li, Yunguan Fu, Qianye Yang, Zhe Min, Wen Yan, Henkjan Huisman, Dean Barratt, Victor Adrian Prisacariu, and Yipeng Hu. 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)*, pages 1–5. IEEE, 2022.

[57] Yuhang Lu, Kang Zheng, Weijian Li, Yirui Wang, Adam P Harrison, Chihung Lin, Song Wang, Jing Xiao, Le Lu, Chang-Fu Kuo, and others. Contour transformer network for one-shot segmentation of anatomical structures. *IEEE transactions on medical imaging*, 40(10):2672–2684, 2020. Publisher: IEEE.

[58] Qi Lu and Chuyang Ye. Knowledge transfer for few-shot segmentation of novel white matter tracts. In *International Conference on Information Processing in Medical Imaging*, pages 216–227. Springer, 2021.

[59] Shiqiang Ma, Xuejian Li, Jijun Tang, and Fei Guo. A zero-shot method for 3d medical image segmentation. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021.

[60] Yao Niu, Zhiming Luo, Sheng Lian, Lei Li, Shaozi Li, and Haixin Song. Symmetrical Supervision with Transformer for Few-shot Medical Image Segmentation. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1683–1687. IEEE, 2022.

[61] Duc Duy Pham, Melanie Lausen, Gurbandurdy Dovletov, Sebastian Serong, Stefan Landgraeber, Marcus Jäger, and Josef Pauli. U-Net in Constraint Few-Shot Settings. In *Bildverarbeitung für die Medizin 2020*, pages 280–285. Springer, 2020.

[62] Duc Duy Pham, Gurbandurdy Dovletov, and Josef Pauli. Using Anatomical Priors for Deep 3D One-shot Segmentation. In *BIOIMAGING*, pages 174–181, 2021.

[63] Abhijit Guha Roy, Shayan Siddiqui, Sebastian Pölsterl, Nassir Navab, and Christian Wachinger. 'Squeeze & excite'guided few-shot segmentation of volumetric images. *Medical image analysis*, 59:101587, 2020. Publisher: Elsevier.

[64] Erica M Rutter, John H Lagergren, and Kevin B Flores. 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*, pages 190–198. Springer, 2019.

[65] Zhengyang Shen, Zhenlin Xu, Sahin Olut, and Marc Niethammer. Anatomical data augmentation via fluid-based image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 318–328. Springer, 2020.

[66] Xiaoang Shen, Guokai Zhang, Huilin Lai, Jihao Luo, Jianwei Lu, and Ye Luo. PoissonSeg: Semi-Supervised Few-Shot Medical Image Segmentation via Poisson Learning. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1513–1518. IEEE, 2021.

[67] Liyan Sun, Chenxin Li, Xinghao Ding, Yue Huang, Zhong Chen, Guisheng Wang, Yizhou Yu, and John Paisley. Few-shot medical image segmentation using a global correlation network with discriminative embedding. *Computers in biology and medicine*, 140:105067, 2022. Publisher: Elsevier.

[68] Hao Tang, Xingwei Liu, Shanlin Sun, Xiangyi Yan, and Xiaohui Xie. Recurrent mask refinement for few-shot medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3918–3928, 2021.

[69] Devavrat Tomar, Behzad Bozorgtabar, Manana Lortkipanidze, Guillaume Vray, Mohammad Saeed Rad, and Jean-Philippe Thiran. Self-Supervised Generative Style Transfer for One-Shot Medical Image Segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1998–2008, 2022.

[70] Shuxin Wang, Shilei Cao, Dong Wei, Renzhen Wang, Kai Ma, Liansheng Wang, Deyu Meng, and Yefeng Zheng. 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*, pages 9162–9171, 2020.

[71] Shuxin Wang, Shilei Cao, Dong Wei, Cong Xie, Kai Ma, Liansheng Wang, Deyu Meng, and Yefeng Zheng. Alternative Baselines for Low-Shot 3D Medical Image Segmentation—An Atlas Perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 634–642, 2021. Issue: 1.

[72] Runze Wang, Qin Zhou, and Guoyan Zheng. 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, pages 514–523. Springer, 2022.

[73] Heying Wang, Qince Li, Yongfeng Yuan, Ze Zhang, Kuanquan Wang, and Henggui Zhang. Inter-subject registration-based one-shot segmentation with alternating union network for cardiac MRI images. *Medical Image Analysis*, 79:102455, 2022. Publisher: Elsevier.

[74] Huisi Wu, Fangyan Xiao, and Chongxin Liang. 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*, pages 417–434. Springer, 2022.

[75] Yixuan Wu, Bo Zheng, Jintai Chen, Danny Z Chen, and Jian Wu. 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*, pages 244–254. Springer, 2022.

[76] Qinji Yu, Kang Dang, Nima Tajbakhsh, Demetri Terzopoulos, and Xiaowei Ding. A location-sensitive local prototype network for few-shot medical image segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 262–266. IEEE, 2021.

[77] Zheng Yuan, Andre Esteva, and Ran Xu. MetaHistoSeg: A Python Framework for Meta Learning in Histopathology Image Segmentation. In *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*, pages 268–275. Springer, 2021.

[78] Amy Zhao, Guha Balakrishnan, Fredo Durand, John V Guttag, and Adrian V Dalca. Data augmentation using learned transformations for one-shot medical image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8543–8553, 2019.

[79] Ziyuan Zhao, Fangcheng Zhou, Zeng Zeng, Cuntai Guan, and S Kevin Zhou. 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*, pages 128–139. Springer, 2022.

[80] Hong-Yu Zhou, Hualuo Liu, Shilei Cao, Dong Wei, Chixiang Lu, Yizhou Yu, Kai Ma, and Yefeng Zheng. Generalized Organ Segmentation by Imitating One-shot Reasoning using Anatomical Correlation. In *International Conference on Information Processing in Medical Imaging*, pages 452–464. Springer, 2021.

[81] Sharib Ali, Binod Bhattarai, Tae-Kyun Kim, and Jens Rittscher. Additive angular margin for few shot learning to classify clinical endoscopy images. In *International Workshop on Machine Learning in Medical Imaging*, pages 494–503. Springer, 2020.

[82] Aihua Cai, Wenxin Hu, and Jun Zheng. Few-shot learning for medical image classification. In *International Conference on Artificial Neural Networks*, pages 441–452. Springer, 2020.

[83] Anping Cai, Leiting Chen, Yongqi Chen, Jiahao Fang, Mengqi Sun, and Zhou Chuan. Pre-MocoDiagnosis: Few-Shot Ophthalmic Diseases Recognition using Contrastive Learning. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2059–2066. IEEE, 2022.

[84] Fabian Cano and Angel Cruz-Roa. 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*, volume 11330, pages 66–73. SPIE, 2020.

[85] Jiaojiao Chen, Jianbo Jiao, Shengfeng He, Guoqiang Han, and Jing Qin. Few-shot breast cancer metastases classification via unsupervised cell ranking. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(5):1914–1923, 2019. Publisher: IEEE.

[86] Yiyu Chou, Samuel W Remedios, John A Butman, and Dzung L Pham. Automatic classification of MRI contrasts using a deep Siamese network and one-shot learning. In *Medical Imaging 2022: Image Processing*, volume 12032, pages 110–114. SPIE, 2022.

[87] Zhiyong Dai, Jianjun Yi, Lei Yan, Qingwen Xu, Liang Hu, Qi Zhang, Jiahui Li, and Guoqiang Wang. PFEMed: Few-shot medical image classification using prior guided feature enhancement. *Pattern Recognition*, 134:109108, 2023. Publisher: Elsevier.

[88] Weikai Huang, Yijin Huang, and Xiaoying Tang. 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*, pages 1–11. Springer, 2022.

[89] Hongyang Jiang, Mengdi Gao, Heng Li, Richu Jin, Hanpei Miao, and Jiang Liu. Multi-Learner Based Deep Meta-Learning for Few-Shot Medical Image Classification. *IEEE Journal of Biomedical and Health Informatics*, 27(1):17–28, 2022. Publisher: IEEE.

[90] Yufei Jin, Huijuan Lu, Wenjie Zhu, Ke Yan, Zhigang Gao, and Zhao Li. 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)*, pages 616–622. IEEE, 2021.

[91] Dwarikanath Mahapatra, Zongyuan Ge, and Mauricio Reyes. Self-Supervised Generalized Zero Shot Learning For Medical Image Classification Using Novel Interpretable Saliency Maps. *IEEE Transactions on Medical Imaging*, 2022. Publisher: IEEE.

[92] Gabriel Maicas, Andrew P Bradley, Jacinto C Nascimento, Ian Reid, and Gustavo Carneiro. Pre and post-hoc diagnosis and interpretation of malignancy from breast DCE-MRI. *Medical Image Analysis*, 58:101562, 2019. Publisher: Elsevier.

[93] Vishnu Mohan. Detection of COVID-19 from Chest X-ray Images: A Deep Learning Approach. In *2021 Ethics and Explainability for Responsible Data Science (EE-RDS)*, pages 1–7. IEEE, 2021.

[94] Dana Moukheiber, Saurabh Mahindre, Lama Moukheiber, Mira Moukheiber, Song Wang, Chunwei Ma, George Shih, Yifan Peng, and Mingchen Gao. 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*, pages 112–122. Springer, 2022.

[95] Tarun Naren, Yuanda Zhu, and May Dongmei Wang. 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*, pages 1–9, 2021.

[96] Achraf Ouahab, Olfa Ben-Ahmed, and Christine Fernandez-Maloigne. 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*, pages 115–125. Springer, 2022.

[97] Angshuman Paul, Yu-Xing Tang, and Ronald M Summers. Fast few-shot transfer learning for disease identification from chest x-ray images using autoencoder ensemble. In *Medical Imaging 2020: Computer-Aided Diagnosis*, volume 11314, pages 33–38. SPIE, 2020.

[98] Angshuman Paul, Thomas C Shen, Yifan Peng, Zhiyong Lu, and Ronald M Summers. Learning Few-Shot Chest X-Ray Diagnosis Using Images From The Published Scientific Literature. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 344–348. IEEE, 2021.

[99] Angshuman Paul, Thomas C Shen, Sungwon Lee, Niranjan Balachandar, Yifan Peng, Zhiyong Lu, and Ronald M Summers. Generalized zero-shot chest x-ray diagnosis through trait-guided multi-view semantic embedding with self-training. *IEEE Transactions on Medical Imaging*, 40(10):2642–2655, 2021. Publisher: IEEE.

[100] Rishav Singh, Vandana Bharti, Vishal Purohit, Abhinav Kumar, Amit Kumar Singh, and Sanjay Kumar Singh. MetaMed: Few-shot medical image classification using gradient-based meta-learning. *Pattern Recognition*, 120: 108111, 2021. Publisher: Elsevier.

[101] Rebeca Vétil, Clément Abi-Nader, Alexandre Bône, Marie-Pierre Vullierme, Marc-Michel Rohé, Pietro Gori, and Isabelle Bloch. 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*, pages 464–473. Springer, 2022.

[102] Junsheng Xiao, Huahu Xu, DiKai Fang, Chen Cheng, and HongHao Gao. Boosting and rectifying few-shot learning prototype network for skin lesion classification based on the internet of medical things. *Wireless Networks*, pages 1–15, 2021. Publisher: Springer.

[103] Jin Yan, Kaiyuan Feng, Hongyu Zhao, and Kai Sheng. 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)*, pages 387–391. IEEE, 2022.

[104] Dig Vijay Kumar Yarlagadda, Praveen Rao, Deepthi Rao, and Ossama Tawfik. A system for one-shot learning of cervical cancer cell classification in histopathology images. In *Medical Imaging 2019: Digital Pathology*, volume 10956, pages 216–221. SPIE, 2019.

[105] Ce Zhang, Qingshan Cui, and Shaolong Ren. Few-shot Medical Image Classification with MAML Based on Dice Loss. In *2022 IEEE 2nd International Conference on Data Science and Computer Application (ICDSCA)*, pages 348–351. IEEE, 2022.

[106] Wei Zhu, Haofu Liao, Wenbin Li, Weijian Li, and Jiebo Luo. 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*, pages 330–339. Springer, 2020.

[107] Tobias Fechter and Dimos Baltas. One-shot learning for deformable medical image registration and periodic motion tracking. *IEEE transactions on medical imaging*, 39(7):2506–2517, 2020. Publisher: IEEE.

[108] Enzo Ferrante, Ozan Oktay, Ben Glocker, and Diego H Milone. On the adaptability of unsupervised CNN-based deformable image registration to unseen image domains. In *International Workshop on Machine Learning in Medical Imaging*, pages 294–302. Springer, 2018.

[109] Yuting He, Tiantian Li, Rongjun Ge, Jian Yang, Youyong Kong, Jian Zhu, Huazhong Shu, Guanyu Yang, and Shuo Li. Few-shot learning for deformable medical image registration with perception-correspondence decoupling and reverse teaching. *IEEE Journal of Biomedical and Health Informatics*, 26(3):1177–1187, 2022. Publisher: IEEE.

[110] Yunlu Zhang, Xue Wu, H Michael Gach, Harold Li, and Deshan Yang. GroupRegNet: a groupwise one-shot deep learning-based 4D image registration method. *Physics in Medicine & Biology*, 66(4):045030, 2021. Publisher: IOP Publishing.

[111] A Emre Kavur, N Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, and others. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis*, 69:101950, 2021. Publisher: Elsevier.

[112] Xiahai Zhuang. Multivariate mixture model for myocardial segmentation combining multi-source images. *IEEE transactions on pattern analysis and machine intelligence*, 41(12):2933–2946, 2018. Publisher: IEEE.

[113] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. 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*, pages 2097–2106, 2017.