

SURVEY

Open Access



# Low-shot learning and class imbalance: a survey

Preston Billion Polak<sup>1\*</sup>, Joseph D. Prusa<sup>1</sup> and Taghi M. Khoshgoftaar<sup>1</sup>

\*Correspondence:  
pbillionpola2019@fau.edu

<sup>1</sup> Florida Atlantic University, 777  
Glades Road, Boca Raton, FL  
33431, USA

## Abstract

The tasks of few-shot, one-shot, and zero-shot learning—or collectively “low-shot learning” (LSL)—at first glance are quite similar to the long-standing task of class imbalanced learning; specifically, they aim to learn classes for which there is little labeled data available. Motivated by this similarity, we conduct a survey to review the recent literature for works which combine these fields in one of two ways, either addressing the obstacle of class imbalance within a LSL setting, or utilizing LSL techniques or frameworks in order to combat class imbalance within other settings. In our survey of over 60 papers in a wide range of applications from January 2020 to July 2023 (inclusive), we examine and report methodologies and experimental results, find that most works report performance at or above their respective state-of-the-art, and highlight current research gaps which hold potential for future work, especially those involving the use of LSL techniques in imbalanced tasks. To this end, we emphasize the lack of works utilizing LSL approaches based on large language models or semantic data, and works using LSL for big-data imbalanced tasks.

**Keywords:** Low-shot learning, Few-shot learning, Zero-shot learning, Class imbalance, Class-imbalanced learning, Machine learning

## Introduction

Class Imbalance (CI)—the scenario when there are many less instances of one type than others—is a challenging problem often faced in machine learning. It is very common in binary tasks such as automated diagnosis, where there are many more negative patients than positive ones; but it is also prevalent in multi-class tasks such as defect classification, where some defects can be much more common—and therefore easier for a classifier to learn—than others. When not accounted for, class imbalance in a dataset can cause a model to become biased towards the more common majority class, incorrectly predicting that most instances will fall into this class. This can lead to potentially disastrous mistakes when the uncommon minority class is also the important one, such as in the case of medical diagnosis tasks, where a biased model would classify diseased patients as healthy. Thus, the question of how to deal with severe CI, especially in binary classification, is a well-studied one [1, 2].

Related but distinct is the task of low-shot learning (LSL), where a model must learn to classify a previously unseen class, given very few training examples [3, 4] (or in the

case of zero-shot learning, no examples [5]). LSL is a newer, more challenging field of machine learning, and LSL techniques have garnered much attention in recent years, as they excel in tasks such as image recognition where there are many common classes with thousands or millions of instances (e.g., “dog” or “car”), but also many rare classes with tens or less (e.g., “abstract portrait” or “pallet”). LSL techniques are also well-suited for applications where all potential rare classes cannot be accounted for in the model’s training dataset, such as in the object detection systems of autonomous driving software [6].

Low-shot learning can be framed as a difficult variant of class-imbalanced learning, as both fields aim to learn the rare classes in a dataset, often with plentiful data provided for other classes. However, there are some notable differences; for example, traditional class-imbalanced settings (such as fraud detection) normally have hundreds of rare-class instances even in severe cases, while many LSL scenarios allow only one to five rare-class instances. This means that common CI countermeasures, such as random under/oversampling, prove insufficient if naïvely applied in a LSL setting. Additionally, while class-imbalanced settings often allow the model to learn all classes simultaneously, LSL settings only provide rare-class examples to the model at evaluation time, necessitating a model that can rapidly learn new classes after it has been trained.

Despite these differences, we believe there is untapped potential in exploring the intersections of these two fields, either in combatting class imbalance within LSL settings, or in harnessing LSL techniques against class imbalance—we believe this latter topic to be especially promising, given the success of LSL models in spite of the setting’s difficulty. Finding no existing surveys on either of these intersections, this paper reviews the recent literature for both, across a wide range of applications. We find that these topics have not been fully explored, with potential for future research to perform more conclusive comparisons of existing methods, or to propose entirely new methods by leveraging existing techniques.

The remainder of this paper is organized as follows: “[Background](#)” section provides background information on both class imbalance and few-shot learning individually, with explanations, definitions of terms, and overviews of approaches to each problem. We define what is meant by “low-shot learning” in this paper (as definitions of few-shot learning are often inconsistent or vague), and we detail the inclusion criteria for our literature review. The remainder of the paper discusses the findings of the review, with in-depth coverage of each work’s methods and experimental findings: “[Solving imbalance within LSL](#)” section covers works which have addressed class imbalance as an obstacle within low-shot learning, and “[Using LSL to solve existing imbalance](#)” section covers works which solve pre-existing class imbalance by borrowing models and techniques from the field of LSL. “[Shortcomings and future research](#)” Section highlights strengths and weaknesses of the current literature and identifies areas with potential for future research, and finally, “[Conclusion](#)” section concludes and summarizes the paper.

## **Background**

### **Low-shot learning**

#### ***Definition***

Broadly, few-shot learning (FSL) refers to a task where a model must learn classes or categories with very few examples; similarly, one-shot learning (OSL) refers to a task

where only one example per rare class is available, and zero-shot learning (ZSL) refers to a task where no examples are available for these classes. In this paper, we collectively refer to FSL, OSL, and ZSL as *low-shot learning* (LSL). While FSL and its variants are not new fields—the term “one-shot learning” has been used in a machine learning context as early as 1997 [7]—it did not receive much research until the mid-2010’s, when it was re-examined with a different approach and a more modern terminology [8]. Low-shot learning tasks have since received greater attention, though specific definitions often vary between works in their scope and terminology.

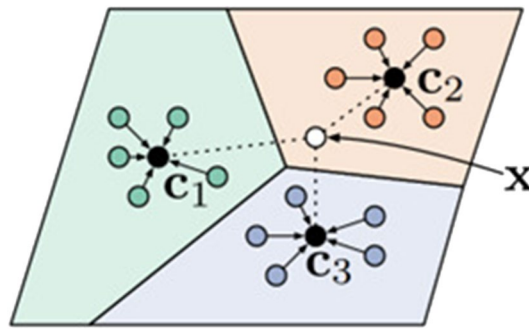
In our survey, we elect to use more restrictive definitions, to prevent confusion and to narrow our search: LSL refers to a task where a model, after training on some “base” classes (usually with many examples), must identify instances of “unseen” or “novel” classes which were not present at all during the model’s training phase; however, a small selection of labeled novel-class examples may be given at test time. This selection, if it exists, is commonly referred to as the support set within LSL, while the unlabeled instance(s) which the model must classify make up the query set. This definition of LSL is primarily meant to exclude works which use the term “few-shot learning” to refer to ordinary supervised-learning settings with rare classes, which we believe are better described as class-imbalanced learning.

Also relevant to our survey are two notable variants of LSL present within the literature. Generalized low-shot learning (GLSL) refers to the scenario where the query set may include instances from the base classes, rather than exclusively containing unseen classes. This is a more challenging setting, as the model must not forget base-class information learned during training, and cannot assume that the query instances belong to a class within the support set. Transductive low-shot learning (TLSL) refers to the scenario where the query set contains multiple instances—usually from different classes—and these instances are classified simultaneously by the model, rather than one at a time. This is a less challenging setting, as the model can use information on the so-far labeled query instances in order to classify the others, in a manner similar to semi-supervised learning. These two variants are not exclusive, and could be applied simultaneously; however, little literature currently exists on this overlap.

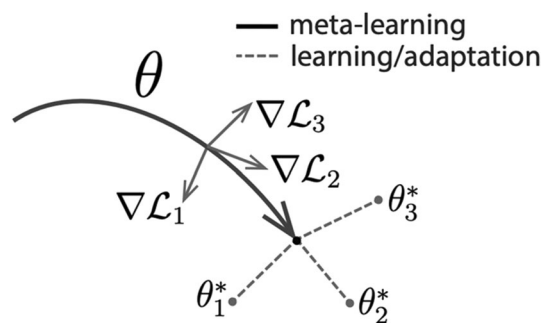
### ***Common approaches to FSL and OSL***

In this section, we highlight the most common approaches to FSL and OSL as found in the results of our literature search, borrowing terminology from a 2022 preprint survey paper by Parnami and Lee [9] which separates FSL methods into metric-based, optimization-based, and model-based techniques. In our review, we found no model-based techniques which qualify under our definition of LSL.

The most common method found in our review was prototypical learning (also known as “contrastive learning,” “pairwise learning,” or “similarity-based learning”) [3], a metric-based framework which changes the model’s objective: instead of learning to classify one input instance into one of many classes, the model takes as input a pair of instances and learns to determine whether they belong to the same class. Test instances are then classified by comparing them to “prototypes” of each class and selecting the class with the most similar prototype (Fig. 1). This approach allows the model to incorporate new classes without requiring significant training data, assuming that the input embedding



**Fig. 1** Depiction of prototypical evaluation in the FSL setting [3].  $c_i$  represents the prototype of the  $i$ th support-set class, and  $x$  represents the query instance



**Fig. 2** Depiction of MAML’s learning process [4]. MAML “optimizes for a representation  $\theta$  that can quickly adapt to new tasks” ( $\theta^*$ )

space can effectively represent a wide range of input data; these embedding spaces are most often found using Siamese neural networks [10] (SNNs) or similar.

The other approach often found is optimization-based meta-learning, where the model is trained in such a way that it can easily adapt to new classes given only a few training examples. Model-agnostic meta-learning (MAML) [4] is a basic implementation of this approach; in this scheme, a “meta-learner” is given many individual “tasks” with their own support and query sets, and it is “meta-trained” to initialize a learner for each task such that it can quickly converge to a new point in the parameter space when given a new task (Fig. 2).

Both prototypical and optimization-based approaches are often trained and validated “episodically,” as in MAML, where instances are given to the model in the form of sets of support and query data; this approach is sometimes used for training even when the model is not evaluated in this manner. Further, episodic training and evaluation is usually framed as an “ $N$ -way- $K$ -shot” setting, with  $N$  referring to the number of unseen classes presented in an episode’s support set and  $K$  referring to the number of samples/shots provided per class, with  $K = 1$  defining OSL. Episodes are generally most challenging with low values for  $K$  and high values for  $N = 2$ .

Finally, a significantly less common approach to LSL is to use large language models (LLMs) such as GPT-3 [11] as an interface for language-based tasks, by simply asking the models to perform tasks while providing examples in full-sentence prompts; the

reasoning behind this is that LLMs often excel at pattern detection and extension, and so are naturally suited to learning new tasks with few examples. Additionally, this method has the potential advantage of not requiring abundant baseclass data, as the LLMs used are most often already trained on vast language data. We do include works in this survey which take this “prompt-based” approach to LSL, though we find that they are comparatively rare in the literature.

### ***Common approaches to ZSL***

For ZSL, most FSL and OSL techniques cannot be applied due to the lack of a support set; thus, one of two strategies are usually employed. For binary classification problems, one-class classification (OCC) [12] is common, where a classifier is trained on data from only one class, and learns to detect outliers during testing; while this concept pre-dates most LSL work, it qualifies as a ZSL technique, as it can effectively classify classes unseen during training. Meanwhile, for multi-class classification problems, most works utilize the semantic information present within the labels themselves, using this information to bootstrap classification. For example, a model may infer that an instance of a “truck” (novel class) will resemble that of a “car” (base class) due to the semantic similarity between the class names; this can be inferred without any examples of the novel class.

There also exists the concept of “zero-shot transfer,” using transfer learning techniques without any supervised fine-tuning on the target domain; however, this setting is distinct from ZSL in that “relevant supervised information” is available during the pre-training process (but not during fine-tuning) [13]. Thus, despite the nominal similarity, we do not include works which work in zero-shot transfer unless they are sufficiently relevant for some other reason (e.g., employing techniques from proper ZSL).

### **Class imbalance**

#### ***Definition***

*Class Imbalance* (CI) is a long-standing obstacle in machine learning [14, 15], referring to when a dataset has many instances of some class or classes (the majority classes) and very few instances of another class or classes (the minority classes). CI is commonly the result of a natural lack of one type of data in the real world, and can appear in binary tasks such as insurance fraud detection, where fraudulent claims are rare; or in multi-class tasks such as image classification, where photographs of objects such as “accordion” are rare. CI is most often quantified as the ratio  $\rho$  between the size of the largest class and the size of the smallest class; in this survey, we will use this metric to specify the severity of imbalance in each work covered, if this information is made available.

A closely related concept to CI is that of foreground-background imbalance in pixel-level image-data tasks (such as object detection or segmentation), where most pixels in each target image are part of the background, rather than any object of interest. Another is that of contrastive/pairwise imbalance which appears when using similarity-based techniques common in LSL (see “[Common approaches to FSL and OSL](#)” sect); when pairing samples from different classes, matching pairs—which are often the pairs of interest—will be much less common than mismatching pairs, especially as the number of classes increases. These two imbalances pose similar problems as traditional CI

despite the differences in their sources, and so we include works in this review which deal with foreground-background imbalance and contrastive imbalance.

Finally, long-tailed learning is a term used in many-class settings where the number of instances per class slowly tapers off, such as in image recognition; the “long tail” refers to the shape of the distribution of class sizes within the dataset. This term is synonymous with CI, and so we include works which address it. Notably, the  $\rho$  metric is often less informative here, due to the much wider range of class sizes common in these tasks.

### ***Intersection with LSL***

Various imbalances can appear within LSL frameworks; the most commonly addressed of these is the natural “data imbalance” between the seen classes with plentiful data and the unseen classes with scarce data. However, in this survey we elect not to include works which focus on this data imbalance, as this is a challenge inherent to all LSL works, rather than an overlap between LSL and CI. Other imbalances are more controllable and therefore more relevant to our review; for instance, a 2023 survey by Ochal et al. [16] distinguishes between “dataset imbalance,” where the data from the entire training set is imbalanced (as in traditional CI) and “task imbalance,” where the data in each episode is imbalanced, either within the support set or query set—though the latter case is only relevant in the context of TLSSL.

### ***Common non-LSL approaches to CI***

Extensive research has been conducted on CI and its countermeasures within more traditional machine learning tasks [1, 2]; these measures include but are not limited to data sampling methods such as random under/oversampling, weighted loss functions such as focal loss, ensemble methods such as bagging and boosting, data augmentation methods such as generative adversarial networks (GANs), and hybrids of these techniques. However, we elect not to go into further detail on these approaches, as they are rarely utilized in the works we cover.

### **Survey methodology**

We restricted our literature search to articles published between January 2020 and July 2023, with exceptions made for sufficiently relevant works we found predating this period (through the references of other papers or through other means).

All collected articles were screened for relevancy, ensuring that they met a minimum quality standard, and that they either addressed CI present within the LSL setting, or used LSL methods or techniques to solve existing CI (“[Solving imbalance within LSL](#)” and “[Using LSL to solve existing imbalance](#)” sections, respectively). Works which use traditional LSL frameworks described in “[Common approaches to FSL and OSL](#)” section (such as episodic training) were more likely to be included than works which do not, though works which deal with tasks similar to LSL (such as incremental learning and one-class classification) were included if sufficiently relevant. As discussed in “[Intersection with LSL](#)” sect, we do not include works which only address the data imbalance between the seen and unseen classes in the LSL setting. Finally, preference was given to works which offer novel solutions or analysis on the overlap between LSL and CI; thus, we often excluded works that simply apply trivial techniques (e.g., a modified loss



function) without deeper exploration, or those that use CI or LSL settings only as ablation studies or footnotes.

To the best of our knowledge, there is only one other survey/review paper which focuses on both LSL and CI, which is a 2023 survey by Ochal et al. [16]. This survey has a much more limited scope than our own, but provides a novel, detailed performance comparison between the works which it does cover. We examine Ochal et al.'s work in more detail in “[Image classification](#)” section.

### **Solving imbalance within LSL**

We first discuss works which address class imbalance as an obstacle present in LSL settings, categorized by the specific task for which LSL is applied. Most works aim to strengthen LSL methods against CI that occurs or could occur in their dataset; however, we also cover works which deal with imbalance which is inherent to the task or method used (such as foreground-background imbalance for object detection). For works which propose models with multiple components, we focus only on those components explicitly stated to be related to or designed for imbalance.

We categorize the surveyed works by the application in which LSL is applied. Additionally, an overview of the surveyed works in this section is shown in Table 1 below, with the name of the paper's proposed model (with “N/A” used when the authors provide no name), the application in which the paper works, the type of imbalance present, its severity  $\rho$  (with “UNSP” used when this is unspecified and “N/A” when not applicable), the amount and type of baseline methods compared to the proposed model, and the proposed model's notable improvements over these baselines. For papers which only compare to variations of their proposed model, we attempt to highlight the performance effects of any CI-related components. For papers which experiment on multiple datasets with varying levels of imbalance, we provide  $\rho$  for the most imbalanced dataset, if this information is available. Double lines between table rows indicate the subsection divisions between the surveyed papers.

We note that there are many works which apply “trivial” CI methods (see “[Common non-LSL approaches to CI](#)” section) independent of the rest of the model, and thus are not included in the body of this survey nor in Table 1. These measures include loss functions such as focal loss [17–21], class-weighted loss [22–25], difficulty-weighted loss [19, 26], and others [27–32]; as well as resampling methods such as random undersampling [33], difficulty-based sampling [34], and other forms of balancing sampling [22, 35, 36] (citations above are not exhaustive). In addition to the works above, we note but do not include a paper by Li et al. which proposes a novel few-shot intent-detection benchmark which contains various forms of class imbalance, but does not propose measures to explicitly counter this imbalance [37].

### **Image classification**

A 2023 paper by Ochal et al. [16] carefully examines the effect of different distributions of minor class imbalance ( $\rho \leq 20$ )—as both task imbalance and dataset imbalance, defined in “[Low-shot learning](#)” section—on ten SOTA few-shot image classification (FSIC) techniques, while also evaluating the impact of simple CI countermeasures such as random sampling. They first find that metric-based models are much more robust

**Table 1** Works covered in “Solving imbalance within LSL” section

Proposed model	Application	Imbalance type	$\rho$	Improvement vs. best baseline	Baseline(s)
No proposed model [16]	LSL image classification	Task, traditional	$\rho \leq 20$	N/A (survey paper)	Many LSL SOTAs
$\alpha$ -TIM [38]	TLSSL image classification	Query set	UNSP	+0–3% Acc.	Many LSL SOTAs
ProtoNet combined [39] <sup>a</sup>	TLSSL image classification	Query set	UNSP	+2% Acc. 1-shot –1.5% Acc. 5-shot	Many LSL SOTAs + [38]
TF-MC [40] <sup>a</sup>	TLSSL image classification	Prediction	UNSP	+2% Acc. 10-shot	Many LSL SOTAs
N/A [41] <sup>a</sup>	LSL image classification	Traditional	UNSP	+2% Acc.	Other proposed variants
N/A [42] <sup>a</sup>	LSL image classification	Traditional	$\rho \leq 1000$	+2–14% Acc. few-shot +10–12% F1 Score	Base model + other proposals
PcGAN [43]	OSL road object classification	Traditional	UNSP	+5% Acc.	Some OSL SOTAs
Post-Scaling [44]	Class-incremental image classification	Traditional	$\rho > 100$	$\pm 1\%$ Acc.	Some SOTAs + CI measures
N/A [45]	LSL semi-supervised image classification	Other	$\rho = 1$	+10–20% Acc.	One non-SOTA model
MAMC-Net [46]	ZSL domain generalization	Traditional	UNSP	+0.8% per-class Acc.	Some ZSDG/ ZSL SOTAs
SCILM [47]	GZSL image classification	Traditional	$\rho \approx 15$	+10% Harm. Mean Acc.	Many ZSL SOTAs
$\mathcal{L}_B T + GP$ [48]	GZSL image classification	Traditional	$\rho \approx 15$	+1.5% Harm. Mean Acc. +8.5% w.r.t. [47]	Many ZSL SOTAs
DUET [6]	GZSL image classification	Traditional	$\rho \approx 15$	+1.5% Harm. Mean Acc. +4% w.r.t. [48]	Many ZSL SOTAs
No Proposed Model [49]	LSL object detection	Traditional	$\rho > 200$	N/A (survey paper)	Some LSL SOTAs
AGCM [50] <sup>a</sup>	LSL object detection	Traditional	$\rho > 200$	+0–5 <i>mAP</i> on novel classes	Many LSL SOTAs
BFS [51]	LSL object detection	Foreground–background	N/A	+15 <i>mAP</i> +5 <i>mAP</i> 10-shot w.r.t. [52]	Some LSL SOTAs
CIR-FSD [52]	LSL object detection	Foreground–background	N/A	+4–17 <i>mAP</i> +7 <i>mAP</i> 3-shot w.r.t. [51]	Some LSL SOTAs
SSL-ALPNet [53]	LSL image segmentation	Foreground–background	N/A	+25–50% Dice	Two LSL SOTAs
AMD-Reg [54]	ZSL sketch-based image retrieval	Traditional	$\rho = 10; 100$	+3–5 <i>mAP</i>	Many SBIR and ZSL SOTAs
GwFrEID [55]	Re-identification	Traditional	UNSP	+6 <i>mAP</i> –9 forgetting ratio	Many Non-LSL SOTAs
SiameseCCR [56]	LSL character recognition	Contrastive	N/A	+13% Top-1 Acc.	Some Non-LSL SOTAs
N/A [57]	Automated CAPTCHA completion	Support set	UNSP	+0–2% Acc. 10-shot	Base model w/o improvements
PRNet [58]	OSL 3D image segmentation	Foreground–background	N/A	+35% Dice	Three OSL SOTAs



**Table 1** (continued)

Proposed model	Application	Imbalance type	$\rho$	Improvement vs. best baseline	Baseline(s)
MRE-Net [59]	LSL 3D image segmentation	Foreground–Background	N/A	+0–6% Dice	One LSL SOTA
UniFewMeta [60]	LSL Nat. language processing	Traditional	UNSP	+0–25% Acc.	Three LSL SOTAs
MELO [61] <sup>a</sup>	Cold-start product recommendation	Task	$5 \leq \rho \leq 25$	–0–0.07 RMSE –0.006 MAE	Base model w/o improvement
N/A [62]	LSL industrial fault classification	Support set	$\rho = 10$	+15–25% Acc.	Other proposed variants
RRPN [63]	LSL industrial fault classification	Traditional	UNSP	+1–1.5% Acc.	Five LSL SOTAs

<sup>a</sup>This work is not published as of September 2023

to task imbalance than optimization-based models on the miniImageNet [64] and CUB [65] datasets, with Chen et al.’s “Baseline++” model being the overall top performer [66]. Ochal et al. theorize that this is because optimization-based methods often use more conventional loss functions, which are vulnerable to CI. Second, they find that applying random oversampling with augmentation (ROS+) during evaluation mostly smooths the difference between these models, bringing the lower performers much closer to the level of the higher-performing models.

Ochal et al. also evaluate the “random-shot” meta-training framework used in previous works, which exposes the model to random task imbalance during training. However, they found that random-shot training actually negatively impacted the performance of almost all models, despite the fact that it more accurately represented the imbalanced evaluation setting; various combinations of ROS and ROS+ during both training and testing proved to be more effective in most cases. Balanced loss functions such as focal loss were also tested and shown to be less effective than ROS. Finally, the authors showed, over many different dataset sizes and distributions, that the models were generally more robust to dataset imbalance than to task imbalance; however, when together, the two imbalances compound to yield a larger drop in accuracy than the simple sum of the two effects. While the authors do not make solid conclusions regarding effective techniques against either type of imbalance, we find this survey to be exceptionally comprehensive in analyzing the effects of these imbalances throughout different scenarios and severities.

A 2021 paper by Veilleux et al. [38] evaluates high-performing TFSL techniques, such as PT-MAP [67] and TIM [68], in a “realistic” setting by adding query-set imbalance. It shows both theoretically and experimentally that many transductive methods rely on the unrealistic assumption of perfect class balance in the query data; in experiments on miniImageNet [64], CUB [65], and tieredImageNet [69], this caused some methods to actually perform worse than inductive baselines (which have less information) when testing on randomly unbalanced query sets ( $\rho$  not specified). A notable exception was the LaplacianShot model [70], which responded well to the query imbalance, presumably because it does not implicitly assume a balanced query set. In addition, Veilleux et al. propose “a generalization of the mutual-information loss based on  $\alpha$ -divergences,” meant to improve robustness to query set distributions. The proposed model using this

loss function, named  $\alpha$ -TIM, outperformed almost all tested models in query-imbalanced TFSL, across multiple datasets and few-shot settings. However, it is not clear how exactly the model is constructed, besides this loss function.

A later paper by Hess and Ditzler [39] builds on [38] by proposing a “maximum log-likelihood” (MLL) method to deal with query imbalance within TLSL; this method is prototypical, and more realistically estimates the distribution of the query set using an exponential probability density function—other details of the model are beyond the scope of this survey. Two versions of the model, using either the MLL metric (ProtoNet MLL) or a combination of MLL and two more basic metrics (ProtoNet Combined), are proposed and tested in LSL and TLSL settings against many SOTA methods from the literature, on the same datasets as in [38]. ProtoNet Combined outperformed all tested models in the transductive 1-shot and inductive settings, and outperformed all except  $\alpha$ -TIM [38] in the less challenging transductive 5-shot setting.

An unpublished 2022 paper by Tao et al. [40] asserts and shows empirically that many TFSIC methods suffer from “imbalanced predictions” at test time, though they are not clear on where this prediction imbalance might come from, nor the motivation behind their method of quantifying this imbalance—the difference between the model’s minimum and maximum number of predictions per-class. However, they do propose a method named “Transductive Fine-tuning with Margin-based uncertainty weighting and Class-balanced normalization” (TF-MC) which aims to give more balanced predictions. Margin-based uncertainty weighting refines and balances the model’s uncertainty measurement on a sample-level basis; meanwhile, class-balanced normalization counteracts the natural tendency for the most-predicted class to have the largest gradient, making it even more likely to be predicted in the future. In experiments on the multiple-domain FSIC dataset Meta-Dataset [71] ( $\rho$  not specified), Tao et al. find that TF-MC yields a small but significant increase to per-class accuracy when compared to other transductive methods applied over various classifier backbones. Ablation studies also confirm both techniques to contribute to the final model performance.

Wertheimer et al. [41]—in a follow-up to a published 2019 paper by Wertheimer and Hariharan [72]—focus on challenges presented by FSIC, including image clutter, inconsistent “granularity” of class labels, and uneven or imbalanced classes. To address the latter problem, they incorporate an approach called “batch folding” into their model, which is based on leave-one-out cross-validation; more specifically, each instance is used as both a support instance and query instances in different “folds” of training, rather than instances only being used as one or the other. While it is unclear how exactly this method is intended to address CI, the authors find that it grants a 4-point gain in model accuracy when added to a basic prototypical network, on a novel meta-learning benchmark task adapted from the iNat2017 dataset [73] ( $\rho$  not specified). Similar results were found in additional experiments using different datasets or settings; however, the batch-folding component was often the smallest contributor to the final performance compared to the two other components (not aimed at CI) of the final model.

A recent paper by Wang et al. [42] claims that the contrastively-trained vision-language models (VLMs) used for LSIC, such as CLIP [74], show promise but perform poorly on imbalanced datasets. They propose a novel “lightweight decoder” to avoid memory issues and better represent tail classes; however, the focus of the paper is on

augmenting VLMs by feeding their output into various CI algorithms, such as focal loss or more recent methods such as Disalign [75]. They test these methods on CLIP on three long-tailed image classification datasets including Places-LT and ImageNet-LT [76] ( $\rho \leq 1000$ ), and compare them to CLIP on its own and with prompt tuning or full fine-tuning (as suggested in earlier literature). While specific results and best-performers varied greatly by dataset, the class-imbalance methods consistently outperformed the baselines in terms of F1 score and accuracy.

A 2022 paper by Deng et al. [43] deals with “intra-class and inter-class” data imbalance within the task of one-shot image classification for road objects, proposing a novel GAN model named PcGAN. Compared to traditional GAN models, PcGAN places more focus on learning a robust embedding space which can function well with only one support instance per class. The model contains a “data reconstruction” generator and a “[image] degradation generator”; the former module separates the important information from the image degradation in real-world test instances, while the latter simulates the effect of a given degradation type on a clean-photo support instance (which the authors refer to as a “prototype”). Test instances can then be classified based on their similarity to each degraded prototype. PcGAN was evaluated quantitatively on two traffic sign/signal datasets ( $\rho$  not specified) against three OSL road-object-classification SOTAs, outperforming all on both seen and unseen classes; it was also evaluated qualitatively through data retrieval and reconstruction experiments, where its output was much clearer on average than the other models tested.

He et al. [44] deal with the task of class-incremental learning (shortened in their paper to CInCL), which is similar to LSL in that new classes can be learned after the initial training phase; more specifically, the paper explores connections between techniques used to combat traditional CI and those used to combat the phenomenon of “catastrophic forgetting” in CInCL. Distinct from the data imbalance problem in LSL, this forgetting is due to limited memory space preventing many seen-class instances from being stored when learning new classes; thus, the seen classes are the minority rather than the unseen ones. They propose a theory-driven CInCL approach known as “post-scaling,” which involves adding a simple fixed layer to the end of a network to compensate for prior shift in test data. Despite the simplicity of this approach, experiments on two datasets (CIFAR-100 [77],  $\rho > 1$ , and a variant of ImageNet [78],  $\rho > 100$ ) showed that it slightly outperformed SOTA CInCL methods on the imbalanced dataset, and outperformed simple CI techniques such as random oversampling in all cases.

Finally, Smith and Conovaloff [45] deal with class performance imbalance in the field of “one-shot semi-supervised learning,” though it is not clear what exactly in their setting is “one-shot”: there is no mention of new or unseen classes, nor the N-way-K-shot paradigm, but they do claim the field “superficially bears similarity to few-shot meta-learning.” Additionally, while they compare their situation to the “traditional” class imbalance problem, they stress that the imbalance here is different, due to the fact that most of the training data is unlabeled.

To deal with this, they first propose using the model-generated pseudo-labels to estimate the class count and establish majority and minority classes in the task. They then propose four separate class balancing methods (to be applied and tested separately rather than in series): the first lowers the pseudo-labeling threshold for minority classes

in order to generate more minority instances, simulating oversampling techniques; the second and third use weighted loss functions to respectively emphasize either all of the minority instances or the high-confidence minority instances; and the fourth simply combines the first and the third methods. The authors only compare their proposed methods to one (non-SOTA) approach from the literature, which all four methods outperform easily. In cross-comparisons, however, they find that the fourth method performs best ( $\sim 91\%$  test accuracy) on the CIFAR-10 dataset [77], while the first performs best ( $\sim 97\%$  test accuracy) for the SVHN [79] dataset—note that these datasets are for digit-recognition and exhibit no natural class imbalance.

### ***Zero-shot image classification***

Due to the difficulty and rarity of the ZSL setting as compared to the FSL or OSL settings, as well as the general difference in approaches used, below we separately cover works which tackle zero-shot image classification (ZSIC).

A 2022 paper by Arfeen et al. [46] handles class imbalance in the less explored, more difficult field of zero-shot domain generalization (ZSDG), which combines the label-space-mismatch challenge of ZSL—in this case with semantic information—with the data-distribution-mismatch challenge of domain generalization. In this setting, CI in the training data may cause a model to learn inaccurate relationships between the seen and unseen classes, thus causing the model to poorly transfer its knowledge between domains. The proposed model, MAMC-Net, adds an “adaptive margin” to each of the semantic classifiers in the model, which is larger for minority classes and smaller for majority classes. MAMC-Net outperformed all tested baselines on the DomainNet benchmark [80] ( $\rho$  not specified), and an ablation study showed the adaptive margin to have a significant positive effect.

The following three papers evaluate their semantic GZSL image classification models on the same three datasets: the balanced CUB [65] and SUN [81] datasets, and the imbalanced AWA2 [82] ( $\rho \approx 15$ ). While none of these papers reference each other, their identical evaluation settings allow us to show direct comparisons of their models’ performances, in which we focus on results for the AWA2 dataset, due to its more realistic class imbalance.

Ji et al. [47] propose a simple balanced sampling approach on top of a prototypical model, then add a more complex “feature fusion” technique to account for the fact that certain instances may be more or less representative of their class prototype. This is done by aligning the semantic features of the class with the visual features of instances, creating semantic-guided prototypes for each class. It is admittedly unclear how this component combats imbalance; however, the authors report good performance, with their “SCILM” model outperforming most competitors (19 tested) in both ZSL and GZSL settings on the three datasets listed above plus two others. On AWA2 [82], they report 48.9% accuracy on unseen classes, 77.8% on seen classes, and 60.1% on the harmonic mean of these two.

Ye et al. [48] propose a prototypical model with a neural network to generate the embeddings used for the class prototypes, as well as a “class-balanced” variant of triplet loss which uses balanced sampling for mini-batches and a slightly extended calculation to reduce variance within the batch while maintaining low training times. In

experiments on the three datasets listed above plus two others, the proposed model outperformed many SOTA GZSL models for the balanced datasets, and marginally outperformed all for the imbalanced datasets. On AWA2 [82], they report 62.2% accuracy on unseen classes, 76.7% on seen classes, and 68.7% on the harmonic mean of these two, significantly outperforming [47]. Additionally, on the imbalanced data, they report better training times than all but one compared model, CNZSL [83].

Most recently, Chen et al. [6] propose an “attribute-level contrastive learning” scheme, which contains a novel attribute-based sampling strategy: given an instance with a class and a chosen target attribute, it will select instances of a similar class but without the attribute, or instances of a dissimilar class with the same attribute; this helps the model separate independent classes and attributes, and aims to alleviate the attribute imbalance (the mechanics of this are less clear). Experiments found the proposed DUET model to perform close to or above the level of all tested SOTA models on the three datasets listed above in both ZSL and GZSL settings. On AWA2 [82], they report 63.7% accuracy on unseen classes, 84.7% on seen classes, and 72.7% on the harmonic mean of these two, reasonably outperforming both [47] and [48]. Given that Chen et al. do not make reference to these previous two works, DUET’s higher performance can likely be attributed to its unique use of image attribute information, as well as possibly the classification power of its other components.

### Object detection

A 2021 survey by Majee et al. [49] aims to evaluate existing few-shot object detection (FSOD) techniques under real-world class imbalance. They test four techniques from recent literature on the India Driving Dataset (IDD) [84], deliberately chosen for its unfiltered class imbalance ( $\rho > 200$ ) between common road objects such as “motorcycle” and rare ones such as “trailer”. Tests are performed in two distinct settings: the “same-domain” setting, where 7 of the dataset’s 15 classes are used for training and 3 are used for testing (the remaining 5 are discarded); and the “open-set” setting, where one of the discarded classes (“vehicle-fallback”) is manually broken down into 4 significantly smaller classes and used for testing. The authors found that the method they call “FsDet” (which is simply called “TFA” in its original paper [85]) significantly outperformed the other three methods in both settings, and in some cases reaching similar performances as non-few-shot baselines. Unfortunately, the authors provide no analysis on any of the techniques used, and neither Majee nor [85] theorize on how (or if) TFA/FsDet counters class imbalance.

Agarwal et al. [50] propose a FSOD model named AGCM, which aims to address two weaknesses of meta-learning-based LSL models: catastrophic forgetting (where a model unlearns base classes while learning novel ones) and class confusion (where a model consistently misclassifies one particular class as another). These issues are implied by the authors to be caused or exacerbated by CI, and are solved by an “Attentive Proposal Fusion” module and a “Cosine Margin Cross-Entropy Loss,” respectively. Using these methods, the authors report small but consistent improvements in performance over the other models tested on the IDD [84] ( $\rho > 200$ ) and PASCAL-VOC [86] ( $\rho \approx 50$ ) datasets in a N-way-K-shot setting. Ablation studies are also performed to demonstrate the significance of the chosen methods and hyperparameters.

Two related papers [51, 52] deal with the foreground-background imbalance problem in FSOD for remote sensing images. Models in this field often have a “region proposal system” (RPS) which proposes regions of the image to be classified by other parts of the model. However, these papers found that the foreground-background imbalance of the images would lead to poor performance, due to the RPS not proposing enough “positive anchors” for the unseen classes.

To counteract this, the 2021 paper by Huang et al. [51] used a “balanced fine-tuning strategy” (BFS), which allowed the RPS to create double the amount of region proposals and incorporated a balanced loss function; these measures are to allow the model to find more foreground/positive region proposals for novel classes, as well as more effectively classify these proposals. In experiments on two remote-sensing datasets, NWPU VHR-10 [87] and DIOR [88] ( $\rho$  not applicable), their proposed model outperformed all five tested baselines, and an ablation study showed that all model components contributed to the final performance, though the BFS component only contributed 1–3 *mAP* to the performance.

The more recent and in-depth of the two works, by Wang et al. [52], instead propose fine-tuning the RPS on both the seen and unseen classes, in contrast to the usual method where the RPS is “frozen” after pre-training on the seen classes while the other components are fine-tuned. This measure, along with relaxing the minimum “confidence level” needed for region proposal, increases the amount of “positive anchors” for unseen classes, supposedly solving the imbalance. In experiments on the same two datasets, the proposed CIR-FSD model outperformed all tested baselines – 7 models of varying recency and performance, which did not include the model from [51].

While the model proposed in [51] reports better peak results (+5 *mAP* on 10-shot for NWPU VHR-10’s novel classes), [52] showed better performance with low shots (+7 *mAP* on 3-shot for the same classes) and could attribute more of its performance to its class-imbalance measures. Further research is necessary to directly compare these CI measures with different model components.

### Other image-data tasks

This subsection covers papers which work with image data but do not perform image classification or object detection; this includes papers which deal with image segmentation or retrieval.

Ouyang et al. [53] tackle the foreground-background imbalance problem, among others, for few-shot segmentation of medical images. Their proposed model, named “SSL-ALPNet,” modifies an existing prototypical model named PANet [89] by incorporating “superpixel-based semi-supervised learning” and “adaptive local prototype pooling.” The latter component combats imbalance by creating several prototypes for various sections of the background region, rather than having one larger prototype for the entire category of “background.” When evaluated on three medical scan datasets ( $\rho$  not applicable), SSL-ALPNet significantly outperformed two prior baselines as well as three proposed variants, and achieved respectable performance as compared to two fully-supervised roofline models with access to manually-annotated training data. Overall, experiments are poor, as the authors evaluate relatively few baselines, provide little detail on them, and do not focus on imbalance.



A 2020 paper by Dutta et al. [54] studies “for the first time in literature” the effect of imbalance in the field of zero-shot sketch-based image retrieval (ZS-SBIR), which involves selecting a query image from a database which matches a user-input “sketch” image; the zero-shot component refers to working with classes of images unseen during training. In preliminary experiments, they find that class imbalance negatively affects the embeddings learned by the model, leading to poor generalization to unseen classes, and so they propose an Adaptive Margin Diversity Regularizer (AMD-Reg) module to combat these effects. The key idea borrowed from the Diversity Regularizer (used previously in non-LSL image tasks) is to ensure that semantic class centers are well-separated and spread out in the feature space used by the model; however, Dutta et al. modify this approach by adding an “adaptive margin” to account for class imbalance by giving larger, safer margins to smaller classes. Experiments found AMD-Reg to outperform various sampling and loss-modification techniques when applied to a simple model on artificially imbalanced datasets ( $\rho = 10$  and  $100$ ). AMD-Reg also notably improved the performance of all tested state-of-the-art models on larger datasets, in both ZS-SBIR and its generalized zero-shot variant.

Wu and Gong [55] work in the task of re-identification (Re-ID), where data of a person—usually in the form of video footage—under noisy conditions (camera angle, noise, low resolution, etc.) must be identified as either a previously seen person or as an unseen person. Due to its continual, “active-learning” nature, the authors compare this task to class-incremental learning, with the challenges of ZSL (without semantic information) and CI. The proposed countermeasure is a novel and complex loss function with three components, including one to counteract CI known as “classification coherence loss.” This component of the loss function has two phases: for the first part of training, it is a cross-entropy loss which balances learning new information with retaining old information (a common approach in class-incremental learning); for the second part, a rebalancing component is added. The structure is motivated by the fact that using this rebalancing component for all training would apparently overfit the model on difficult new samples, causing forgetting of old classes.

In experiments on four popular Re-ID benchmarks ( $\rho$  not specified), Wu and Gong found that their method, named GwFReID, outperformed all tested baselines (most built from SOTA incremental and Re-ID approaches) on average and for early sections of training, though it underperforms slightly for later sections, as more classes are introduced; GwFReID also outperformed all baselines when evaluating on new datasets after training. Finally, ablation studies showed that the two-phase structure of the classification coherence loss was optimal, as the final model outperformed versions where the rebalancing component was kept on or off throughout.

A 2020 paper by Chen et al. [56] deals with contrastive imbalance (see “[Intersection with LSL](#)” section) while constructing a LSL Chinese character recognition model, SiameseCCR; while this imbalance is not the focus of this work, the authors outline their novel re-balancing method in detail. In this, an equal number of positive (matching) and negative (mismatching) pairs are first selected and used for training; the  $n$  most difficult instances are chosen based on misclassification rates and similarity scores; on the next iteration, a new set of positive pairs is constructed as before, while a new set of negative pairs is constructed so that all pairs contain at least one difficult instance.



On novel character-recognition datasets ( $\rho$  not specified) SiameseCCR performed exceptionally on the few-shot and one-shot scenarios, significantly outperforming three existing character recognition models (which suffered from overfitting), and achieving  $\geq 98\%$  accuracy even in the top-1 GOSL setting. However, their experiments have some shortcomings—namely, the compared models were neural networks not designed with the LSL setting in mind, and classification accuracy was the only performance metric used. Further, it is unclear how much of SiameseCCR's performance can be attributed to its unique sampling strategy rather than other components of the model.

A recent paper by Wang et al. [57] takes a LSL approach to the task of automatic CAPTCHA completion, but claims that many LSL techniques will perform poorly in realistic settings, due to support-set class imbalance and the multi-domain nature of the task. Their model follows the usual  $N$ -way- $K$ -shot episodic framework—using either ProtoNet [3] or MAML [4] as a base—but they propose two improvements: a basic data augmentation strategy to deal with the cross-domain problem, and “intra-class variance distance weighting,” which adjusts the prototype decision boundaries to counteract the variance caused by the support-set imbalance. While the proposed model was not compared to any others from the literature, experiments on five synthesized CAPTCHA datasets and real-world tests ( $\rho$  not specified) showed both components to have a positive effect on model accuracy in 5-shot and 10-shot scenarios, though distance weighting was not applicable in the 1-shot scenario nor when using MAML as a base.

A 2023 paper by Lei et al. [58] deals with the foreground-background imbalance present in one-shot 3D medical image segmentation, utilizing weak supervision in the form of hand-drawn scribble annotations from the labelled data. Their model, PRNet, uses a “propagation-reconstruction network” to realistically transfer these scribbles to the unlabeled data, thus generating reliable foreground or background markers in the unlabeled 3D images, alleviating the imbalance. These points are then further processed into “pseudo masks” used for class-specific training. The proposed model was evaluated on two medical scan datasets—the  $\rho$  metric is not applicable, but in one dataset the target class makes up only 0.1% of voxels—and drastically outperformed all three OSL methods used as baselines, achieving a 35-point gain in Dice score over from the second-best model, and only a 10-point drop from a gold-standard fully-supervised model.

Cui et al. [59] also deal with (but do not focus on) foreground-background imbalance within FSL image segmentation for 3D medical scan images. Their model, MRE-Net, uses an existing technique known as Online Hard Example Mining (OHEM) [90], an iterative sampling method shown previously to be capable of handling foreground-background imbalance in non-LSL settings. In short, the modified OHEM determines which category is the majority or minority on a pixel-level, collects all minority-class pixels and the  $N$  most difficult majority-class pixels (determined by a loss function), then uses these collected pixels to create a new loss function for the next iteration. In experiments on three segmentation datasets ( $\rho$  not applicable), the proposed MRE-Net outperformed its main competitor, U-Net [91], under different augmentations and numbers of shots, and for some classes approaches the performance of tested fully-supervised methods. Ablation studies showed the OHEM component to contribute a small but significant amount (+4 Dice coefficient) to MRE-Net's final performance.

### Non-image-data tasks

This subsection covers papers which do not work with image data. We find that these works are significantly less common in the LSL literature, likely because most image-based tasks contain many classes over a wide distribution of sizes, and thus lend themselves more easily to the LSL setting.

A 2021 paper by Bragg et al. [60] highlights poor benchmark quality and quantity in the task of few-shot natural language processing, including poor evaluation of class imbalance within this task. They propose and release a benchmark task named “FLEX,” which includes class imbalance (along with many other challenges and features), though the specific types and degrees of imbalance are not specified. Bragg et al. also propose a prompt-based model for this task named UniFew (as well as a version with meta-training, named UniFewMeta), which utilizes an existing question-answering framework known as UnifiedQA [92] in order to streamline the model’s interface and enable more consistent results. UniFew and UniFewMeta achieve good performance on the FLEX benchmark, outperforming the SOTA model on two of three tasks; however, due to the models’ prompt-based nature, as well as a lack of model documentation from the authors, it is difficult to analyze which components, if any, allow UniFew to combat the class imbalance within FLEX.

In a 2023 paper, Kim et al. [61] frame the task of “cold-start” product recommendation (that is, with few prior user interactions) as an LSL problem, by considering each user as a separate task with limited instances (user reviews). They propose a MAML-based [4] approach, and to deal with the imbalance between positive and negative reviews—the severity and direction of which varies per user—they incorporate an adaptive weighted loss, which uses a recurrent encoder to encapsulate the rating distribution in each sequence of user ratings. In experiments on four user recommendation datasets ( $5 \leq \rho \leq 25$ ) and four model backbones, the proposed “MELO” model significantly outperformed MAML alone in terms of both RMSE and MAE.

Yu et al. [62] incorporate a “center loss” combined with softmax loss into a prototypical network for few-shot industrial fault classification (a notably imbalanced task), though they go into less detail on the mechanics of how this counters CI, and also do not directly compare their method to any others from the literature. However, they do compare their model to variants with other bases (MAML [4]) and loss functions (focal loss) on fault diagnosis datasets with artificial support-set imbalance ( $\rho = 10$ ) and show that the prototypical network with center loss yields the highest accuracy and F1 score.

Li et al. [63] also develop a metric-based approach for few-shot industrial fault diagnosis, proposing a “reweighted regularized prototypical network” (RRPN) which uses inter- and intra-class information to produce more discriminative class prototypes. To deal with class imbalance, the authors also incorporate an “intra-class reweighted strategy” (ICRS) which weights individual instances based on their class representativeness, as well as a “balance-enforcing regularization” (BER) term to encourage more balanced class representations and combat overfitting. In experiments on two common industrial fault datasets ( $\rho$  not specified) in 3-shot and 5-shot settings, the proposed RRPN outperformed five SOTA models on average and for rare fault types; an ablation study also showed ICRS and BER to have small but positive effects on model performance.

### Using LSL to solve existing imbalance

Low-shot learning and class imbalanced learning both deal with the same fundamental issue of having limited data of one or multiple classes; thus, many works have proposed anti-CI approaches inspired by LSL techniques.

This section covers any papers that face the class imbalance obstacle (or similar) and use low-shot learning paradigms or techniques, such as prototypical networks or contrastive learning, to solve this obstacle. For works which propose models with multiple components, we generally focus only on those components explicitly stated to be related to or inspired by the LSL setting or its techniques. Due to the relative similarity between the works presented here (as compared to those presented in “[Solving imbalance within LSL](#)” section), we place greater focus on the proposed model’s training and evaluation process when possible.

As in “[Solving imbalance within LSL](#)” section, we categorize the surveyed works by the application in which LSL is applied, and provide an overview of these works in Table 2 below. Unlike Table 1, we do not specify the type of imbalance addressed by each work, as almost all works in this section deal with traditional CI.

We note but do not include a survey paper by Duan et al. [93] which covers LSL for anomaly detection and mentions the imbalance present within the task, but does not offer any original contribution or analysis against this imbalance.

### Image classification

Bansal et al. [94] address general CI in an unpublished 2021 article, claiming that many traditional methods such as re-balancing or re-weighting lead to overfitting on minority data. They instead propose “MetaBalance,” a meta-learning model based heavily on MAML [4]; the primary value of this structure, according to the authors, is that it allows two CI strategies to be independently applied in the inner and outer training loops. MetaBalance is model-agnostic, and so it can easily be customized for almost any classification setting, with the inner and outer CI measures tunable for each scenario.

MetaBalance is first evaluated on a 10-way image classification task. They compare their model-agnostic method (using a CNN as a base, in this particular experiment) to various traditional CI measures as well as SOTA augmentation techniques, on the CIFAR-100 dataset [77] with different levels of artificial CI ( $\rho = 1000$  or 100). They find that MetaBalance—with ROS in the outer loop and no measure in the inner loop—significantly improves accuracy in both settings, reaching 30% classification accuracy in the severe case and 40% in the moderate case, compared to 23% and 37% from the next best baseline, RUS. For fairness, the authors re-evaluate after re-weighting the base model’s priors to reduce class bias; however, MetaBalance remained the top performer.

Additionally, MetaBalance is evaluated on a facial-recognition task, using a standard facial recognition CNN as a base model. They compare the proposed model to the base CNN model (on its own and with random over-sampling) on a face dataset with an artificial gender imbalance created by removing 90% of female face images. Despite the base CNN’s ability to handle this imbalance, they find that MetaBalance improves

**Table 2** Works covered in “Using LSL to solve existing imbalance” section

Proposed model	Application	$\rho$	Improvement vs. best baseline	Baseline(s)
MetaBalance [94]	Image classification	$\rho = 100$ $\rho = 1000$	+4% Acc. ( $\rho = 100$ ) +6.5% Acc. ( $\rho = 1000$ )	Non-LSL SOTAs + CI Measures
MetaBalance [94]	Facial recognition	$\rho = 10$	+2–3% Acc.	Non-LSL model + CI measure
N/A [95]	Image/video classification	$\rho \leq 1000$	+5–9% Acc. +5% F-Measure	Some LSL SOTAs
DRAGON [96]	Image classification	$\rho \approx 250$	+1–3% Acc. 10-shot	Four LSL SOTAs
N/A [97]	Incremental image classification	UNSP	+25–30% novel-class Acc.	Non-LSL baseline
RF-MML [98]	Aerial photograph classification	$10 \leq \rho \leq 30$	–1–2% Acc. on majority +6–11% Acc. on minority	Non-LSL model + SOTA CI measures
N/A [99]	Dermatology image classification	$\rho > 20$	+1–2% Acc.	Many SOTAs + CI measures
FedFew [100]	Federated medical image classification	$\rho = 10$	+20–25% Acc. +50% precision	Two non-LSL models
fs-WAE [101]	Bearing fault classification	$\rho \leq 20$	+2.5% F-score	Some non-LSL SOTAs
SSN [102]	Character recognition	UNSP	+0.4–0.7% Acc.	Some LSL SOTAs
N/A [103]	Façade defect classification	$\rho \approx 110$	+8–9% overall Acc. +25–30% novel-class Acc.	One LSL model + proposed variants
GMDB-fs [104]	Genetic disorder classification	$\rho \approx 50$	+6–10% Acc.	One non-LSL SOTA
Prototypical Net [105]	Fabric defect classification	$\rho \approx 75$	+4% Acc.	Some non-LSL SOTAs
DualFusion [106]	Incremental road object detection	$\rho \approx 1000$	+8% AP on novel classes –3% AP overall	One LSL SOTA
FSAD-Net [107]	Polyp detection	$\rho > 100$	+7% AUC-ROC	Many LSL SOTAs
SAPNet [108]	Human-object interaction	UNSP	+20% Acc. 5-shot +5% Acc. 1-shot	Some LSL methods
DSCNN [109]	Hardware Trojan detection	UNSP	+5–10% Acc.	Many non-LSL SOTAs
LST [110]	Open-world image segmentation	$\rho \geq 1000$	+10–12% AP on rare classes +2% AP overall	Proposed variants + CI measures
SSF-ViT [111]	Facial expression recognition	$1 \leq \rho \leq 40$	+0–2% Acc. non-LSL +2% Acc. 5-shot	Many non-LSL SOTAs
N/A [112]	Streamer action recognition	UNSP	+20% Acc.	Two SOTA models
N/A [112]	General action recognition	UNSP	+1–3% Acc.	Many SOTA models
N/A [113]	Gestational age estimation	$\rho \approx 16$	–12% Acc.	SOTA non-ML method
SIHTD [114]	Hyperspectral object detection	N/A	+1.3% AUC-ROC +0–30% “Overall” AUC	Some SOTA non-LSL models
MetaBalance [94]	Credit card fraud detection	$\rho \approx 600$	+0.6% AUC-ROC	Many CI measures
MetaBalance [94]	Loan default prediction	$\rho \approx 4$	+1.2% AUC-ROC	Many CI measures
I-SiamIDS [115]	Network intrusion detection	$\rho \approx 650$	+6% F1 Score (majority) +5–25% F1 Score (minority)	Four non-LSL models

**Table 2** (continued)

Proposed model	Application	$\rho$	Improvement vs. best baseline	Baseline(s)
N/A [116]	Network anomaly detection	$\rho \approx 250$	+0.5–1% Acc.	Three non-LSL models
SifterJIT [117]	Software defect prediction	$\rho \approx 10$	+9% AUC-ROC +2.5% F1 Score	One non-LSL SOTA + CI measure
SAT-GAN [118]	Database error detection	$\rho \leq 30$	+0.25% precision $\pm 1\%$ F1 Score	Some non-LSL SOTAs
Meta-IP [119]	Project extension forecasting	$\rho \leq 30$	+0.5–2% ROC-AUC +2–6% BACC	Some CI measures
MVC [120] <sup>a</sup>	3D point cloud segmentation	$\rho \approx 40$	+0.2% Acc. overall +2% Acc. on rare classes	Four non-LSL SOTAs
N/A [121]	Electricity theft detection	$1.5 \leq \rho \leq 9$	+6% AUC-ROC ( $\rho = 9$ ) +10–12% F1 Score ( $\rho = 9$ )	Six non-LSL models
N/A [122]	Electrocardiogram classification	UNSP	+10% Acc. 50-shot +40% Acc. 3-shot	Three non-LSL models
MFCs with FSL [123]	Cough audio diagnosis	$\rho \approx 15$	+1.5% AUC-ROC	Other proposed model + baseline
N/A [124] <sup>a</sup>	Birdsong classification	$\rho \approx 75$	+1.3% Acc. 7-shot +10.5% Acc. 1-shot	Other proposed models
N/A [125]	Utterance classification	UNSP	+4–9% F1 Score	One baseline
SiameseCHEM [126]	Biochemical activity prediction	UNSP	+22% Avg. BACC	Three non-LSL models
Meta-MMFNet [127]	Micro-expression recognition	UNSP	+5% acc.	Many non-LSL SOTAs
CBC [128]	Software defect classification	$5 \leq \rho \leq 50$	–6–12% F1 Score	Two non-LSL rooflines

<sup>a</sup>This work is not published as of September 2023

accuracy further, reaching 90% female accuracy and 92% total accuracy, compared to 87% and 90% from the base model. MetaBalance was finally evaluated in two tabular-data tasks, the results of which are covered in “[Tabular-data tasks](#)” section.

Zhu and Yang [95] propose a novel, technical approach to deal with long-tailed learning in video and image classification. The authors’ approach is inspired by prototypical networks; however, their setting has a higher intra-class variance which they claim cannot be easily encapsulated by one prototype per class. Thus, they instead propose “inflated episodic memory” (IEM), a key-value data storage framework which—in tandem with a novel “region self-attention mechanism” (RSA)—is used to find and store the “most discriminative feature” for each class, which improves generalization to the tail classes. Experiments in both image and video classification tasks (on long-tailed datasets including Places-LT and ImageNet-LT [76],  $\rho \leq 1000$ ) showed the proposed model to marginally but consistently outperform the SOTA, in both “closed-set” and “open-set” settings (with the latter characterized by additional classes added during testing).

A 2021 paper by Samuel et al. [96] deals with general long-tailed learning, proposing a modular (i.e., for any application) model incorporating semantic-based GLSL methods. The proposed model, “DRAGON,” first deals with data scarcity within the tail (few-shot) classes by fusing the predictions of a visual (traditional) sub-model

and a semantic sub-model, with the latter being tuned for tail class prediction. It then deals with the imbalance itself by individually debiasing samples based on the estimated number of samples per class. Both modules are tuned using the confidence values from each sub-model's predictions. Extensive experiments on multiple long-tailed image classification datasets, including ImageNet-LT [76] ( $\rho \leq 250$ ), showed DRAGON to significantly outperform all tested SOTA methods in per-class accuracy, with these results being especially noticeable at higher shots (10 to 20). Remarkably, DRAGON also offered at-or-above SOTA performance when evaluated without any class semantic data—and thus without the model fusion component—presumably by relying on the debiasing component of the model.

A 2020 paper by Patra and Noble [97] works in the task of incremental learning (or “lifelong learning”), where new classes or tasks can be learned after the initial training phase, in the task of medical image object detection. They do this in both balanced and imbalanced settings (defined by the amount of available data from unseen classes), proposing a different model pipeline for each. Both models share a hierarchical classification strategy, with coarse and fine classification stages, and both use the same pre-trained convolutional recurrent “supermodel” for the first stage; however, they differ greatly for the second stage. The model we focus on—for the more imbalanced setting—uses a “similarity-driven few-shot learning regime” where the supermodel outputs are used to direct instances to one of multiple Siamese networks. Experiments conducted on novel medical-image datasets found the proposed LSL model to noticeably outperform the other proposed model on the rare classes, and found both proposed models to significantly outperform a simple transfer-based baseline (used for lack of comparable models from the literature). However, it should be noted that the CI present within their imbalanced setting is unknown but most likely mild, with 500 instances per novel class and an unspecified number of instances per base class.

A short 2020 paper by Guan et al. [98] draws heavily from LSL paradigms to deal with class imbalance in aerial photograph classification. Their model, “Random Fine-Tuning Meta Metric Learning” (RF-MML) first trains  $N$ -way- $K$ -shot on the majority classes only, randomizing  $N$  and  $K$  in each episode to improve robustness and avoid arbitrary hyperparameter selection. The model then similarly fine-tunes on all classes, keeping  $N$  equal to the total number of classes in the dataset. The authors propose two variants of RF-MML, using a prototypical network and a support vector machine (SVM) respectively for the fine-tuning phase; the training phase is always trained as a prototypical network.

Experiments on two artificially imbalanced datasets ( $10 \leq \rho \leq 30$ ) compared the classification accuracy of both variants of RF-MML with a baseline deep neural network model (on its own or with one of two anti-CI measures from the literature). The RF-MML models sacrificed a few points of accuracy on the majority classes for a larger boost on the minority classes, leading to an increase in overall accuracy more noticeable on the more imbalanced datasets. Additionally, the variant with SVM fine-tuning marginally outperformed the variant with prototypical fine-tuning.

A 2020 paper by Weng et al. [99] is dedicated to solving class imbalance within dermatology image classification. They first evaluate many LSL approaches from the literature, both metric- and optimization-based, in both a standard  $N$ -way- $K$ -shot evaluation



setting and a modified “real-world” evaluation setting with a different class split between training, validation, and testing, but identical imbalance ( $\rho > 20$ ). However, they find that even the best-performing of these LSL models, MatchingNet [64], performs only slightly better than a conventional (“CSL”) baseline on rare classes, and performs much worse when the CSL method is augmented with classical CI countermeasures such as sampling or focal loss. In light of this, they propose an ensemble approach combining LSL and CSL methods by simply redirecting majority-class instances to the CSL model (baseline + oversampling) and minority-class instances to the LSL model (MatchingNet), which shows marginally better performance overall than either method alone.

Dong et al. [100] address CI in the field of federated (decentralized, i.e., different data is stored in different clients) partially-labeled classification of medical images. They propose “FedFew,” a three-stage system: the first stage is a self-supervised class-agnostic embedding function, the second stage is a partially-supervised “energy-based” classifier for the common classes, and the third stage (the one drawn from LSL) is a nearest-neighbor model which uses “dual prototypes” to classify the rare classes. The “energy” from the second stage is used to determine whether an instance is likely to be a rare class and should thus go to the third stage. Experiments on an artificially imbalanced ( $\rho = 10$ ) chest X-ray dataset showed FedFew to consistently outperform two baselines from the literature in all metrics (accuracy, recall, F1, and especially precision) on the rare classes.

A 2021 paper by Pei et al. [101] deals with class imbalance ( $\rho \leq 20$ ) in fault diagnosis of rolling bearings. This imbalance is often dealt with using a Wasserstein auto-encoder (WAE) [129] for robust data augmentation; the authors propose an augmented version (fs-WAE) which adds an optimization-based FSL technique named Reptile [130], a variant of MAML [4]. They train fs-WAE using multiple tasks created from their dataset, which differ in the selection of samples and optimization function used. In experiments evaluating augmentation quality on two datasets, fs-WAE was shown to generally outperform other baselines, including WAE and fs-WAE without Reptile. The exception to this was the WGAN-GP model, which outperformed fs-WAE on the more imbalanced dataset in terms of accuracy, but took significantly more computing power to train.

Liu et al. [102] propose a contrastive approach to combat the class imbalance and data scarcity in the task of ancient character recognition. The authors go into detail on the technical aspects of their model; important points include a residual-learning inspired embedding network which reduces overfitting while preserving detail, a novel “soft similarity contrast loss function” which prevents over-optimizing as can occur in normal contrastive loss, and a novel “cumulative class prototype” which is more robust to outliers and deviations in each class. After preliminary experiments to confirm that these measures were effective, experiments were conducted in a 20-way-1-shot setting on a variety of character recognition datasets, with unspecified CI but a wide range of sizes, including a “big data” dataset with 1.1 million instances. The proposed model, SSN, marginally but consistently outperformed all SOTA networks tested (+0.3–0.7% accuracy) on all three datasets whose results were included in the paper. Variants of SSN were also tested which did not train on any of the classes used for evaluation; these variants suffered only very minor drops in accuracy (0.2–0.5%).

A 2022 paper by Cui et al. [103] utilizes LSL to counter the severe class imbalance ( $\rho \approx 110$ ) in building façade defect classification. More specifically, they propose an



“extensible classifier” which uses “weight imprinting” to synthesize embeddings for novel classes from limited data; a contrastive learning module integrated into the loss function to improve the embedding function; and a hard negative mining (HNM) module added to the contrastive learning module to reduce the pair imbalance within the contrastive learning process. Their model is trained on the base classes and allowed to imprint on one to ten support samples of each novel class during evaluation. For a lack of comparable façade defect classification models, the authors compared their model to a simple LSL baseline as well as versions of their model with reduced or removed HNM; they found that the proposed model outperformed under any number of shots, especially for the novel classes, achieving 63.5% accuracy on novel classes with only one shot and 83.0% with ten.

A 2022 paper by Sümer et al. [104] deals with the CI ( $\rho \approx 50$ ) within the task of facial phenotype recognition for genetic disorders using LSL, proposing a standard prototypical network trained  $N$ -way- $K$ -shot, using a deep CNN as a feature encoder. The proposed model, “GMDB-fs”, was compared against one other SOTA (non-LSL) model, and outperformed this model significantly (+6–10% classification accuracy); additionally, the proposed model was more effective when the encoder was pre-trained on the MS1MV2 facial recognition dataset [131], and 10-way or 15-way episodic training was found to be the most effective.

Zhan et al. [105] face CI ( $\rho \approx 75$ ) in fabric defect classification, and propose a prototypical network (using a CNN feature extractor) trained under an  $N$ -way- $K$ -shot paradigm, reasoning that the artificial class balance within the support sets will circumvent the CI within the dataset itself. After establishing that their “Prototypical Net” network achieves the best performance when trained 2-way- $K$ -shot ( $K = 1$  or 4), they directly compare its performance to deep CNNs commonly used in the literature, finetuned on fabric data. While Prototypical Net consistently outperformed all of these models, the authors were unclear on whether any of the models tested, proposed or baseline, were evaluated traditionally or  $N$ -way- $K$ -shot.

### Other image-data tasks

This subsection covers papers which work with image data but do not perform image classification.

Tambwekar et al. [106] address class imbalance in road object recognition, framing the problem as a few-shot batch-incremental setting and building off of a previous LSL model, FewX [132]. Their proposed model, DualFusion, uses two detector models: a (non-LSL) Faster R-CNN model [133] trained on only the base classes, and a modified FewX model which is fine-tuned on the novel class data. These two models are connected with a fusion network which allows the detectors to pass information between each other in an intermediate stage, and combines the feature outputs of the detectors at the end. An experiment on the IDD benchmark [84] ( $\rho > 200$ ) compared DualFusion to each of its individual submodels for incremental learning; as expected, it significantly outperformed both in the general case, only losing to FewX when evaluating exclusively on novel classes. A second experiment on the COCO benchmark [134] ( $\rho \approx 1000$ ) compared DualFusion with a SOTA technique, ONCE, finding that DualFusion

underperformed somewhat in the general case (10.6 vs 13.7 AP) but significantly outperformed on novel classes (9.9 vs 1.2 AP).

Tian et al. [107] address the imbalance ( $\rho > 100$ ) in the binary task of polyp classification in colonoscopy images, proposing a LSL method inspired by one-class classification (OCC). The proposed FSAD-Net consists of a feature encoder pre-trained on exclusively negative instances (inliers), and a “score inference network” trained on negative instances and few positive instances (outliers), optimized with a contrastive loss. The score generated by the latter module represents how close the input is to the center of the latent feature space, and therefore how likely it is to be an inlier. Experiments showed that FSAD-Net outperformed all tested SOTA ZSL/FSL baselines by a significant margin, achieving +0.07 AUC-ROC over the next-best model. FSAD-Net and all FSL baselines were trained using only 30–40 abnormal images. However, all  $\sim 13,000$  images used in the dataset are sourced from only 18 colonoscopy videos; this raises the possibility of overfitting, which is not addressed by the authors.

A 2021 paper by Ji et al. [108] deals with the extreme class imbalance within human-object interaction (HOI) recognition, where a machine must identify both the object within the input image and the action being performed on it. They propose a prototypical network, “SAPNet,” trained  $N$ -way- $K$ -shot, first using a novel “SIGMA” module utilizing semantic label information to address the two-dimensionality of the task. For the prototypical classification itself, they then propose two competing modules: a “prototypes shift” (PS) method which incorporates the query sample into the estimation of each support class prototype, and a “hallucinatory graph prototypes” (HGP) method which uses a small network to hallucinate novel support samples in order to avoid bias in class prototypes.

Experiments on two datasets ( $\rho$  not specified), each tested with two different class splits, showed both versions of SAPNet to outperform all tested methods, including LSL methods from general (non-HOI) literature. Of the two versions, the model with the HGP module outperformed slightly, reaching 72% 5-way-5-shot accuracy on one dataset, compared to 70.8% from the version with the PS module (and 53.3% from the next-best method, MAML [4]). Notably, its performance was worse in the 5-way-1-shot scenario, only outperforming MAML by 5% accuracy.

Sharma et al. [109] propose a similarity-based model in the field of hardware trojan (HT) detection, where images of integrated circuits (ICs) are checked for malicious trojans installed during production; this field is naturally plagued by data scarcity and CI. To handle these issues, they propose a novel “deep Siamese CNN” (DSCNN) model which uses novel deep CNN sub-models as feature extractors. They train, validate, and test their model on small synthetic datasets created from two popular HT detection benchmarks, with positive instances created by manually augmenting trojan-free image data. Validation and testing are performed using a modified 2-way-5-shot paradigm, where the support set consists only of negative instances, and the query image is compared to these for a similarity score which is thresholded for the final classification. Experiments found that the proposed DSCNN model significantly outperformed (by at least 5–10 points of accuracy) all four tested HT detection models from the literature. However, the level of imbalance in the training data is unclear: the authors first explain that their synthetic datasets have 50 positive and 50 negative instances (implying  $\rho =$

1), but later claim that their model “effectively handles the class imbalance” within this dataset.

A 2020 paper by Hu et al. [110] deals with the extreme class imbalance in “open-world” image segmentation by framing the setting as a few-shot incremental learning problem, which they approach by “segmenting the tail.” This refers to the process of splitting the training data into balanced but progressively smaller (i.e., fewer-shot) subsets of classes, allowing the model to incrementally learn each group after pre-training on the largest classes. Along with a balanced replay system to maintain class balance while preventing catastrophic forgetting, they propose a meta weight generator (MWG) module which initializes new-class classifiers from combinations of semantically similar old-class classifiers, e.g., a “drone” classifier initialized as a combination of “fan” and “airplane” classifiers.

In experiments on the LVIS image segmentation dataset [135], with instances per class ranging from only 1 to over 1000, the proposed “LST” model outperformed the base model (on its own or with various resampling methods from the literature) on all classes except the largest. A version of the proposed model without the MWG module was also evaluated, and outperformed slightly on the “middle” classes but was otherwise inferior to the full model. Unfortunately, no SOTA methods were compared to the proposed model.

In a 2023 paper, Chen et al. [111] address the class imbalance and data scarcity within facial expression recognition from image and video data, by proposing a model which combines elements of LSL and semi-supervised learning. Their model, “SSF-ViT,” first pre-trains a vision transformer on four self-supervised “pretext” tasks with unlabeled face data, then fine-tunes on a fully supervised expression recognition task. Finally, when evaluating on real-world data, a prototypical network is constructed (using the vision transformer as an encoder), and the input data is grouped into support and query sets for episodic evaluation. On six expression image datasets with a wide range of imbalance levels ( $1 \leq \rho \leq 40$ ), SSF-ViT performed at or above the level of thirteen expression-recognition SOTAs, and outperformed two LSL models in additional 5-shot and 1-shot experiments.

A short paper by He et al. [112] deals with CI within the relatively unexplored field of action recognition from streamer video footage (streamer action recognition, or SAD). The authors propose a two-phase model which first pre-trains (non-episodically) to learn video features and then meta-trains a prototypical network 5-way-5-shot. Preliminary experiments using different similarity functions found the proposed model to yield the best performance when trained with an inner-product similarity and evaluated with cosine similarity. The proposed model significantly outperformed two SOTA methods from the literature on a novel few-shot SAD dataset (+20% accuracy under one-shot), and six others on general human-action datasets (+1–3% accuracy). Imbalance levels  $\rho$  were not provided for any datasets used.

A 2021 paper by Romanov et al. [113] addresses class imbalance ( $\rho \approx 16$ ) and data scarcity, in the field of newborn gestational age estimation from image data, and proposes two approaches, tested separately: a prototypical network and MAML [4]. However, because there is too little postnatal image data to train these models “traditionally,” the authors opt to pre-train on balanced non-medical image datasets—specifically,

either miniImageNet [64], CelebA [136], or both—before evaluating 5-way-5-shot on the entire base of relevant data without fine-tuning. Due to the lack of comparable models from the literature, direct comparisons could only be conducted between the two proposed models. In these, MAML (meta-trained on both image datasets) showed the best performance, yielding accuracies of  $\sim 53\%$  for face and ear images and  $\sim 40\%$  for foot images. The authors also compare this model to the “Ballard score,” a manually-evaluated metric which is the SOTA for postnatal gestational age estimation; while the proposed model underperforms the Ballard score by  $\sim 12\%$ , they note that this metric simultaneously has access to foot, face, and ear information (plus extraneous markers), and theorize that an ensemble of the proposed MAML networks could perform on par with the Ballard score.

Shi et al. [114] deal with the foreground-background imbalance within hyperspectral object detection by combining LSL with transfer learning, proposing a novel semi-supervised domain adaptive few-shot learning model. The LSL portion of the proposed model is a simple prototypical network trained 2-way-1-shot to learn appropriate embeddings; other components of the model include a novel  $MDS^2F^2$  convolutional network for pre-processing, an RCA attention mechanism/subnetwork, and a complex semi-supervised domain adaptation framework with a discriminatively boosted loss function. In experiments against five SOTA models from the literature on hyperspectral image data ( $\rho$  not applicable), the proposed “SIHTD” model was the best performer by a significant margin in terms of AUC-ROC, though it performed marginally worse with respect to other metrics tested (such as area under the threshold vs false-alarm-rate curve).

### Tabular-data tasks

Bansal et al. [94] evaluate their MAML-based [4] model, MetaBalance, on two tabular tasks: credit-card fraud detection and loan-default prediction (details on the model and its results on image-based tasks can be found in the “For Image Classification” subsection). Of these two, the credit-card task is much larger ( $\sim 300,000$  instances) and has more severe imbalance ( $\rho \approx 600$ ), but the authors find that it is nonetheless easier to classify than the loan task ( $\rho \approx 4$ ). Evaluating on AUC-ROC and using a traditional neural network as a base, experiments on the credit-card task found MetaBalance to outperform all tested classical CI measures, including RUS, SMOTE, and Edited Nearest Neighbors (ENN); experiments on the loan task found that the best-performing variant of MetaBalance used ENN in the outer loop and no CI measure in the inner loop.

Two papers by Bedi et al. in 2020 [137] and 2021 [115] utilize Siamese networks to combat the class imbalance in network intrusion detection. The first paper proposes SiamIDS, a standalone traditional Siamese network (explicitly borrowed from LSL literature), while the follow-up paper proposes the improved I-SiamIDS, an ensemble model which additionally includes XGBoost and deep neural network classifiers, with each of the three classifiers applied in series and fed into a second XGBoost classifier for the final output. Experiments conducted in the second paper compared SiamIDS and I-SiamIDS to other common classifiers (such as Random Forest and a CNN) on two intrusion detection datasets, including the NSL-KDD benchmark ( $\rho \approx 650$ ) [138]; results showed I-SiamIDS to generally outperform the other classifiers in terms of F1 scores and AUC-ROC, with this behavior being more consistent on minority classes.

Notably, the authors do not compare their models to traditional anti-CI measures such as data sampling or cost-sensitive learning, despite these methods' effectiveness in imbalanced binary classification [1].

A short paper by Huang et al. [116] proposes a "gated few-shot" model to combat the imbalance in network anomaly detection. They utilize a similarity-based approach, using a CNN as an encoder, and a "gate structure" which pre-emptively determines whether the query instance belongs to a seen or unseen class from the support set. The model is trained and evaluated on a reduced version of the NSL-KDD dataset ( $\rho \approx 250$ ) [138], and trained episodically on only some selected base classes. Experimental results were mixed; no SOTA methods from the literature were compared, and the proposed method was roughly on par with a traditional 1-nearest-neighbor approach for unknown classes and a simple SVM approach for known classes, marginally outperforming both for the overall dataset.

A 2021 paper by Gesi et al. [117] works in "just-in-time" (JIT) software defect prediction, and highlights the often-neglected class imbalance ( $\rho \approx 10$ ) in the field; they propose a split approach where majority instances are classified using a pre-existing non-LSL neural network, DeepJIT [139], while minority instances are redirected to a Siamese neural network. Experiments on two defect prediction datasets found that the proposed SifterJIT model outperformed the original DeepJIT (on its own or with oversampling) by a noticeable margin on the rare classes (+0.09 AUC-ROC, +0.025 F1 score).

Wu and Wang [118] deal with database error detection, a naturally imbalanced task which is often addressed using data augmentation methods. However, they propose ZSL in the form of one-class classification (OCC), allowing their model to train using no unclean (minority-class) data whatsoever. More specifically, they train a GAN which uses self-attention-based encoder-decoder modules as a generator, and a CNN as a discriminator; feeding this model clean (majority-class) data allows it to learn the distribution of clean data, allowing it to discriminate unclean data evaluated after training. Experiments were performed on 5 datasets with various types of data errors (though the model was only tasked to distinguish between "clean" and "unclean") and class imbalance ( $\rho \leq 30$ ), comparing the proposed SAT-GAN model to six error-detection baselines, four of which were statistical or rule-based rather than ML-based. SAT-GAN significantly outperformed all non-ML methods, and achieved performance on par with (and occasionally exceeding) that of the best baseline, AUG [140], despite the fact that SAT-GAN had no access to minority-class data.

A 2022 paper by Li et al. [119] proposes a model named "Meta-IP" to deal with CI in the task of project extension forecasting; this model employs a transfer learning strategy framed as a LSL setting: a simulated source dataset is used as the support set for MAML, while the (real-world) target domain dataset is used as the query set. While the authors explain that using MAML allows for the choice of two independent loss functions and sampling strategies (for the inner and outer training loops), they do not explain which losses and strategies were chosen or tested in practice. While poorly explained, experiments on project extension datasets ( $\rho \leq 30$ ) showed that Meta-IP consistently outperformed all "traditional" anti-CI measures tested, including SMOTE and bagging, in terms of AUC and especially BACC.

### Other tasks

This subsection covers works which do not fall into any of the above categories, i.e., work with neither image data nor tabular data.

In a 2020 preprint, Chen et al. [120] deal with “3D point cloud segmentation,” a field whose relative obscurity causes data issues, including scarcity and class imbalance; to deal with this, they borrow LSL techniques from the similar field of 2D image segmentation. They propose a prototypical network, adapted for point segmentation with a novel “Multi-View Comparison Convolution” (MVC) module to generate different “views” (embeddings) of support instances. In experiments on a novel point segmentation benchmark ( $\rho \approx 40$ ), the authors report that their model outperformed all four tested SOTA models on the rarest classes, and yielded slightly better performance overall (+0.2% classification accuracy over next-best).

A recent paper by Gao et al. [121] avoids the imbalance within electricity theft detection (from electricity usage time series data) by reframing the task as a one-class classification problem, using a contrastive learning approach to detect outliers. After multiple layers of pre-processing—using techniques unique to electricity monitoring—a contrastive network is trained on a large sample of negative (non-theft) data, which is used to create the model’s “support set” when evaluating on test data. In experiments on multiple electricity-theft datasets with a range of artificial imbalance and scarcity conditions ( $1.5 \leq \rho \leq 9$ ), the proposed model consistently outperformed six other machine learning models constructed as baselines in terms of F1 score, FPR, and AUC-ROC.

Gupta et al. [122] address class imbalance and data scarcity in electrocardiogram (ECG) classification by using a Siamese network with CNN feature extractors which is evaluated (but not trained) under a LSL paradigm. They use five ECG time series datasets ( $\rho$  not specified): two for training, two for validation, and the largest dataset for testing, which is done under a traditional 5-way- $K$ -shot paradigm. Their proposed SCNN model outperformed all tested competitors—a traditional similarity-based time-series classification method, a simple nearest-neighbor algorithm, and an LSTM convolutional model, all of which were also evaluated  $K$ -shot—and proved very robust to the number of shots given (performance remained the same from  $K = 3$  through 50).

A 2021 paper by Bhosale et al. [123] proposes, among other methods, a contrastive learning approach to deal with the class imbalance present in the task of COVID-19 diagnosis from cough sounds. After pre-processing the audio into “MFCC” sound features, they train their FSL model under a 2-way- $K$ -shot paradigm, aiming to circumvent class imbalance by giving equal shots for each class. They also use a triplet loss in order to better adapt to the variability within negative (COVID-free) instances. While they do not compare this method to others in the literature, they do compare to a baseline as well as an alternative SVM-based (non-LSL) approach using a wider set of features and balanced weighting, on a curated cough sound dataset ( $\rho \approx 15$ ). They found that the FSL approach outperformed these marginally, reaching an AUC of 0.719 on the test dataset (compared to 0.699 and 0.706), but noted that specificity was poor throughout all models ( $\sim 0.50$  on the test dataset).

A preprint article by Rentería et al. [124] uses Siamese neural networks to combat class imbalance and data scarcity in birdsong syllable classification. While they propose no changes to the standard Siamese structure, they do test variations with five different



encoder sub-models, including one with no encoder (comparable to a nearest-neighbor approach). In experiments on an existing birdsong dataset ( $\rho \approx 75$ ), they found that the model versions with LSTM encoders and with no encoders were the highest performers, with the former slightly outperforming with a high number of shots (91.3% accuracy when  $K = 7$ ) and the latter outperforming with very a low number (64.6% when  $K = 1$ ). They also claim that these methods yield better accuracy on their dataset than any other method proposed in the literature.

A 2021 paper by Sunder and Fosler-Lussier [125] deals with the long tail in utterance classification (human speech recognition) using a pairwise approach. Specifically, the proposed model uses a “mixup strategy” which creates artificial data as combinations of any two classes. The model is then trained using a novel loss function which combines contrastive loss and a “mixup” loss, and final classifications on an instance are made by balancing this output with the model output on the unmixed input. Though the authors do not compare their model to any from the literature, experiments on two long-tailed datasets ( $\rho$  not specified) with three encoder bases find that their pairwise strategy improves performance over a simpler cross-entropy loss (about +6% F1 score).

Fernández-Llaneza et al. [126] propose an “ $N$ -shot learning” method (misnomer for LSL) to deal with CI within the task of biochemical activity prediction (from textual molecule representations), as both a binary and categorical classification task. The authors propose a Siamese recurrent neural network with a self-attention mechanism, in tandem with heavy data augmentation and random oversampling; the model is evaluated by comparing input samples to a random sample of  $N$  training instances (hence “ $N$ -shot learning”). Experiments compared the proposed “SiameseCHEM” to three common ML classifiers (MLP, RF, and SVM) using one of two molecule representations in the binary classification task; SiameseCHEM was shown to outperform these significantly in terms of MCC on all five datasets tested ( $\rho$  not specified). Unfortunately, no SOTA activity prediction models were tested for comparison, and no comparisons were performed at all for the categorical classification task.

Wenjuan et al. [127] work in the task of micro-expression recognition using facial video data, a task with class imbalance and scarce data. Their proposed “Meta-MMFNet” model uses a non-LSL “feature fusion” module, which combines the “optical flow” and “frame difference” information between consecutive frames (common techniques for video processing), followed by a LSL-inspired meta-learning “model fusion” module. This module appears to be a prototypical network which takes as input the weighted sum (hence “fusion”) of two DNN models fine-tuned on micro-expression and macro-expression data, respectively. Evaluating under a  $N$ -way- $K$ -shot paradigm over three datasets ( $\rho$  not specified, but described as “highly unbalanced”), experiments showed performance on par with or exceeding SOTA techniques (+5% accuracy over next-best), with especially good results when identifying the “surprise” emotion. An ablation study showed that the model fusion method yielded on average a better performance than either DNN model did alone, though the micro-expression sub-model was quite close.

Finally, Patil and Ravindran [128] avoid CI within software defect classification (from full-sentence textual defect descriptions) by proposing a completely unsupervised “concept-based classification” (CBC) approach, which the authors categorize as zero-shot learning—despite it being more difficult than classical ZSL, due to the complete lack of



labeled data for even the majority classes. CBC is completely semantic: the model constructs a pre-processed and indexed corpus of semantic Wikipedia articles relating to each defect type (“concepts”), and represents each input/label description as a combination of these concepts. They then use a similarity-based classifier to determine the label which most closely matches each input. Experiments conducted on two defect datasets ( $5 \leq \rho \leq 50$ ) showed that despite a complete lack of labeled data, CBC yielded slightly inferior but comparable performance to SOTA fully- and semi-supervised defect classification rooflines ( $-0.05$  to  $-0.10$  F1 score from the fully-supervised model).

### Shortcomings and future research

In this section, we highlight shortcomings of the literature covered in this survey, first covering general critiques present throughout many of the works covered, then briefly reviewing notable research gaps within works which deal with imbalance within LSL (“[Solving imbalance within LSL](#)” section) and then the gaps within works which propose LSL techniques against imbalance (“[Using LSL to solve existing imbalance](#)” section).

First, as a more general criticism, a common occurrence in the literature is poor documentation and standardization, especially in regards to LSL. For instance, the phrase “few-shot learning” itself is often confused in the literature, with many papers (not included in this survey) using the term to simply refer to data scarcity. We believe better standardization of LSL definitions and terminology would allow for more streamlined comparison and communication between similar works and approaches. A 2022 survey on LSL by Parnami and Lee [9] offers what we believe to be the most constructive taxonomy so far for LSL, including a sufficiently restrictive definition for the problem setting, and the “metric-based”/“optimization-based” terminology used throughout this paper.

A less important but similar point is the lack of documentation and standardization in regards to CI. Many papers neglect to provide detail on the severity of class imbalance present within their datasets, despite this information being both quite important to the difficulty of the problem, especially in binary or few-class classification, and relatively simple to quantify using  $\rho$  or other metrics. Additionally, many papers do not compare their proposed methods with traditional measures such as data sampling or cost-sensitive learning, despite the simplicity and effectiveness of these approaches in many scenarios [1].

We also note the general lack of cross-comparison and cross-reference between the papers and models covered in this review; however, this is largely due to the wide range of applications and settings throughout LSL literature—ranging from open-set object detection to software defect prediction—making it difficult or impossible for authors (or ourselves) to directly compare the performances of LSL models. However, we do take the opportunity to compare the reported performances of those few works which do evaluate their models on the same datasets.

### Imbalance within LSL

Many papers were found which addressed CI and its variants within LSL settings, over a wide range of applications and techniques. We believe this category to be better-explored as compared to LSL-against-CI, likely because there are simply fewer unique

challenges or variants within the LSL setting as compared to machine learning as a whole, leading to fewer possible combinations and areas of overlap to research.

Nonetheless, we do observe that most of the works covered do not deal with the generalized or transductive variants of LSL (explained in “[Low-shot learning](#)” section), or at least do not identify themselves as such. We found only three papers which explicitly work with the transductive setting [38–40], and only four which explicitly work under or experiment with the generalized setting [6, 47, 48, 54] (though many papers may well deal with GLSL without using this term). While this is less important for TLSSL—which is a less challenging and arguably less realistic setting than standard LSL—we believe that GLSL, being a more realistic setting for multi-class problems, should be further studied with respect to imbalance.

### **LSL against CI**

Many papers were found which applied LSL techniques and ideas to CI learning, over a wide range of settings and approaches. We believe this category to be lesser-explored as compared to CI-within-LSL, as well as the one with more overall potential. Below we highlight some notable gaps.

There was very little overlap between LSL, CI, and big data settings; while this term has no agreed-upon quantitative definition in the literature, we found only a few papers which dealt with datasets of more than 100,000 instances. We believe this is an area with high research potential, not only to further the existing literature on class imbalance in big data [1], but also because we theorize that the episodic training methods common in LSL may prove effective at compartmentalizing the large swaths of available training data in these settings.

Also of note is the lack of works which utilize techniques or methods from ZSL against class imbalance. In this survey, we only found three relevant papers which utilized a one-class classification (OCC) approach [107, 118, 121], and only three papers which utilized semantic information [96, 108, 128] as is common in ZSL techniques; of these six papers, only two [108, 128] cite a relation to ZSL specifically. While OCC for class imbalance (outside of the LSL context) is comparatively well-studied [141], we believe semantic LSL approaches to have high potential for solving CI in multi-class problems, and possibly even binary problems with sufficiently creative approaches.

Finally, we mention the complete lack of works found which utilize LLM-based methods for imbalanced tasks. This approach is admittedly less consistent than more traditional LSL techniques, due to the generative nature of LLMs, the requirement of using natural-language input prompts, and the incompatibility of some models with image or tabular data; regardless, we believe there is potential in experimenting with this approach for imbalanced text-based classification tasks such as word sense disambiguation [142].

### **Conclusion**

This survey paper examined over 60 works in a wide variety of applications from the last 3 years (2020 to mid-2023) which combined the fields of low-shot learning and class imbalanced learning, either by addressing imbalance within low-shot settings, or using low-shot learning techniques and frameworks to combat class imbalance elsewhere. We

aimed to comprehensively find and report all recent literature which falls into these categories, and we found that each area has been explored, with generally successful results and to varying degrees of completeness. In particular, we noted the lack of literature covering LLM-based approaches to imbalanced textual tasks, semantic-based approaches to class imbalance in any application, or LSL methods for imbalanced big data tasks; we believe all of these areas, especially the latter two, to hold great potential for future research.

#### Abbreviations

CI	Class imbalance
FSL	Few-shot learning
GLSL	Generalized low-shot learning
LSL	Low-shot learning
MAML	Model-agnostic meta-learning
OSL	One-shot learning
TLSL	Transductive low-shot learning
ZSL	Zero-shot learning

#### Acknowledgements

We would like to thank the reviewers in the Data Mining and Machine Learning Laboratory at Florida Atlantic University.

#### Author contributions

PBP performed the literature review and wrote the manuscript. JDP provided feedback and advice throughout the manuscript's development. TMK suggested the topic to PBP and guided the research direction. All authors read and approved the final manuscript.

#### Funding

Not applicable.

#### Availability of data and materials

Not applicable.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 23 September 2023 Accepted: 28 October 2023

Published online: 02 January 2024

#### References

1. Leevy JL, Khoshgoftaar TM, Bauder RA, Seliya N. A survey on addressing high-class imbalance in big data. *J Big Data*. 2018;5(1):42. <https://doi.org/10.1186/s40537-018-0151-6>.
2. Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. *J Big Data*. 2019;6(1):27. <https://doi.org/10.1186/s40537-019-0192-5>.
3. Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: *Advances in neural information processing systems*, vol. 30. Curran Associates, Inc.; 2017. [https://proceedings.neurips.cc/paper\\_files/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html](https://proceedings.neurips.cc/paper_files/paper/2017/hash/cb8da6767461f2812ae4290eac7cbc42-Abstract.html). Accessed 30 June 2023.
4. Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: *Proceedings of the 34th international conference on machine learning*. PMLR; 2017. ISSN: 2640-3498. p. 1126–35. <https://proceedings.mlr.press/v70/finn17a.html>. Accessed 11 June 2023.
5. Xian Y, Schiele B, Akata Z. Zero-shot learning—the good, the bad and the ugly; 2017. p. 4582–91. [https://openaccess.thecvf.com/content\\_cvpr\\_2017/html/Xian\\_Zero-Shot\\_Learning\\_-\\_CVPR\\_2017\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2017/html/Xian_Zero-Shot_Learning_-_CVPR_2017_paper.html). Accessed 23 Oct 2023.
6. Chen Z, Huang Y, Chen J, Geng Y, Zhang W, Fang Y, Pan JZ, Chen H. DUET: cross-modal semantic grounding for contrastive zero-shot learning. In: *Proceedings of the AAAI conference on artificial intelligence*. 2023;37(1):405–13. <https://doi.org/10.1609/aaai.v37i1.25114>. Number: 1. Accessed 24 Aug 2023.

7. Sussman GJ, Yip K. Sparse representations for fast, one-shot learning. *AI Memos*; 1997. <https://dspace.mit.edu/handle/1721.1/6673>. Accessed 23 Oct 2023.
8. Triantafillou E, Zemel R, Urtasun R. Few-shot learning through an information retrieval lens. In: *Advances in neural information processing systems*, vol. 30. Curran Associates, Inc.; 2017. <https://proceedings.neurips.cc/paper/2017/hash/01e9565cecc4e989123f9620c1d09c09-Abstract.html>. Accessed 23 Oct 2023.
9. Parnami A, Lee M. Learning from few examples: a summary of approaches to few-shot learning; 2022. <https://arxiv.org/abs/2203.04291v1>. Accessed 30 June 2023.
10. Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature Verification using a “Siamese” time delay neural network. In: *Advances in neural information processing systems*, vol. 6. Morgan-Kaufmann; 1993. <https://doi.org/10.1142/S0218001493000339>. <https://proceedings.neurips.cc/paper/1993/hash/288cc0ff022877bd3df94bc9360b9c5d-Abstract.html>. Accessed 29 June 2023.
11. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;33:1877–901.
12. Tax DMJ. One-class classification: concept learning in the absence of counter-examples. PhD Thesis, Delft University of Technology (TU Delft); 2002. <https://www.elibrary.ru/item.asp?id=5230402>. Accessed 29 June 2023.
13. Zhai X, Wang X, Mustafa B, Steiner A, Keyzers D, Kolesnikov A, Beyer L. LiT: zero-shot transfer with locked-image text tuning; 2022. p. 18123–33. [https://openaccess.thecvf.com/content/CVPR2022/html/Zhai\\_LiT\\_Zero-Shot\\_Transfer\\_With\\_Locked-Image\\_Text\\_Tuning\\_CVPR\\_2022\\_paper.html](https://openaccess.thecvf.com/content/CVPR2022/html/Zhai_LiT_Zero-Shot_Transfer_With_Locked-Image_Text_Tuning_CVPR_2022_paper.html). Accessed 30 June 2023.
14. Van Hulse J, Khoshgoftaar TM, Napolitano A. Experimental perspectives on learning from imbalanced data. In: *Proceedings of the 24th international conference on machine learning. ICML '07*. Association for computing machinery, New York, NY, USA; 2007. p. 935–42. <https://doi.org/10.1145/1273496.1273614>. Accessed 23 Oct 2023.
15. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21(9):1263–84. <https://doi.org/10.1109/TKDE.2008.239>.
16. Ochal M, Patacchiola M, Vazquez J, Storkey A, Wang S. Few-shot learning with class imbalance. *IEEE Trans Artif Intell*. 2023. <https://doi.org/10.1109/TAI.2023.3298303>.
17. Zhou J, Li B, Wang P, Li P, Gan W, Wu W, Yan J, Ouyang W. Real-time visual object tracking via few-shot learning; 2021. arXiv preprint [arXiv:2103.10130](https://arxiv.org/abs/2103.10130). <https://doi.org/10.48550/arxiv.2103.10130>
18. Zhang Z, Zhou J, Liang X. Zero-shot learning based on semantic embedding for ship detection. In: *2020 3rd international conference on unmanned systems (ICUS)*. IEEE; 2020. p. 1152–6. <https://doi.org/10.1109/ICUS50048.2020.9274981>
19. Rahman S, Khan S, Barnes N. Improved visual-semantic alignment for zero-shot object detection. In: *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, p. 11932–9; 2020. Issue: 07
20. Gao Y, Hou R, Gao Q, Hou Y. A fast and accurate few-shot detector for objects with fewer pixels in drone image. *Electronics*. 2021;10(7):783. <https://doi.org/10.3390/electronics10070783>.
21. Wolters P, Daw C, Hutchinson B, Phillips L. Proposal-based few-shot sound event detection for speech and environmental sounds with perceivers. arXiv preprint; 2021. [arXiv:2107.13616](https://arxiv.org/abs/2107.13616)
22. Huang K, Geng J, Jiang W, Deng X, Xu Z. Pseudo-loss confidence metric for semi-supervised few-shot learning. In: *Proceedings of the IEEE/CVF international conference on computer vision*; 2021. p. 8671–80. <https://doi.org/10.1109/ICCV48922.2021.00855>
23. Yang H, Cai S, Sheng H, Deng B, Huang J, Hua X-S, Tang Y, Zhang Y. Balanced and hierarchical relation learning for one-shot object detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*; 2022. p. 7591–600.
24. Liu T, Yang Y, Fan W, Wu C. Few-shot learning for cardiac arrhythmia detection based on electrocardiogram data from wearable devices. *Dig Signal Process*. 2021;116: 103094. <https://doi.org/10.1016/j.dsp.2021.103094>.
25. Olah J, Baruah S, Bose D, Narayanan S. Cross domain emotion recognition using few shot knowledge transfer; 2021. arXiv preprint [arXiv:2110.05021](https://arxiv.org/abs/2110.05021)
26. Zhang L, Zhou S, Guan J, Zhang J. Accurate few-shot object detection with support-query mutual guidance and hybrid loss. In: *2021 IEEE/cvf conference on computer vision and pattern recognition (CVPR)*; 2021. p. 14419–27. <https://doi.org/10.1109/CVPR46437.2021.01419>. ISSN: 2575-7075
27. Ma S, Li X, Tang J, Guo F. A zero-shot method for 3d medical image segmentation. In: *2021 IEEE international conference on multimedia and expo (ICME)*. IEEE; 2021. p. 1–6. <https://doi.org/10.1109/icme51207.2021.9428261>
28. Masihullah S, Garg R, Mukherjee P, Ray A. Attention based coupled framework for road and pothole segmentation. In: *2020 25th international conference on pattern recognition (ICPR)*. IEEE; 2021. p. 5812–9. <https://doi.org/10.1109/ICPR48806.2021.9412368>
29. Hua Y, Yi D. Synthetic to realistic imbalanced domain adaption for urban scene perception. *IEEE Trans Ind Inform*. 2021;18(5):3248–55. <https://doi.org/10.1109/TII.2021.3107785>.
30. Yuan Y, Chen W, Chen T, Yang Y, Ren Z, Wang Z, Hua G. Calibrated domain-invariant learning for highly generalizable large scale re-identification. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*; 2020. p. 3589–98.
31. Zhang L, Zhang C, Quan S, Xiao H, Kuang G, Liu L. A class imbalance loss for imbalanced object recognition. *IEEE J Select Top Appl Earth Obs Remote Sens*. 2020;13:2778–92. <https://doi.org/10.1109/JSTARS.2020.2995703>.
32. Li Q, Zhang Y, Sun S, Zhao X, Li K, Tan M. Rethinking semantic-visual alignment in zero-shot object detection via a softplus margin focal loss. *Neurocomputing*. 2021;449:117–35. <https://doi.org/10.1016/j.neucom.2021.03.073>.
33. Olson AW, Cucu A, Bock T. Multi-class zero-shot learning for artistic material recognition; 2020. arXiv preprint [arXiv:2010.13850](https://arxiv.org/abs/2010.13850)<https://doi.org/10.48550/arxiv.2010.13850>
34. Wu C, Wang B, Liu S, Liu X, Wu P. TD-sampler: learning a training difficulty based sampling strategy for few-shot object detection. In: *2022 7th international conference on cloud computing and big data analytics (ICCCBDA)*. IEEE; 2022. p. 275–9.

35. Chen R, Chen T, Hui X, Wu H, Li G, Lin L. Knowledge graph transfer network for few-shot recognition. In: Proceedings of the AAAI conference on artificial intelligence, vol. 34; 2020. p. 10575–82. <https://doi.org/10.48550/arxiv.1911.09579>. Issue: 07.
36. Malik V, Kumar A, Veppa J. Exploring the limits of natural language inference based setup for few-shot intent detection; 2021. arXiv preprint [arXiv:2112.07434](https://arxiv.org/abs/2112.07434). <https://doi.org/10.48550/arxiv.2112.07434>
37. Li G, Zhai Y, Chen Q, Gao X, Zhang J, Zhang Y. Continual few-shot intent detection. In: Proceedings of the 29th international conference on computational linguistics; 2022. p. 333–43.
38. Veilleux O, Boudiaf M, Piantanida P, Ben Ayed I. Realistic evaluation of transductive few-shot learning. *Adv Neural Inf Process Syst*. 2021;34:9290–302.
39. Hess S, Ditzler G. A maximum log-likelihood method for imbalanced few-shot learning tasks; 2022. <https://arxiv.org/abs/2211.14668>.
40. Tao R, Chen H, Sawides M. Towards class-balanced transductive few-shot learning; 2022. [https://openreview.net/forum?id=ZA\\_F5AU-byh](https://openreview.net/forum?id=ZA_F5AU-byh).
41. Wertheimer D, Tang L, Bajjal D, Mittal P, Talwar A, Hariharan B. Few-shot learning in long-tailed settings. <https://daviswer.github.io/files/tpami2021paper.pdf>.
42. Wang Y, Yu Z, Wang J, Heng Q, Chen H, Ye W, Xie R, Xie X, Zhang S. Exploring vision-language models for imbalanced learning; 2023. <https://link.springer.com/article/10.1007/s11263-023-01868-w>.
43. Deng L, He C, Xu G, Zhu H, Wang H. PcGAN: a noise robust conditional generative adversarial network for one shot learning. *IEEE Trans Intell Transp Syst*. 2022;23(12):25249–58.
44. He C, Wang R, Chen X. A tale of two CILs: the connections between class incremental learning and class imbalanced learning, and beyond. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2021. p. 3559–69.
45. Smith LN, Conovaloff A. Building one-shot semi-supervised (BOSS) learning up to fully supervised performance. *Front Artif Intell*. 2022;5: 880729. <https://doi.org/10.3389/frai.2022.880729>.
46. Arfeen A, Dutta T, Biswas S. Handling class-imbalance for improved zero-shot domain generalization. *BMVC*; 2022.
47. Ji Z, Yu X, Yu Y, Pang Y, Zhang Z. Semantic-guided class-imbalance learning model for zero-shot image classification. *IEEE Trans Cybernetics*. 2021;52(7):6543–54.
48. Ye C, Barnes N, Petersson L, Tsuchida R. Efficient Gaussian process model on class-imbalanced datasets for generalized zero-shot learning. In: 2022 26th international conference on pattern recognition (ICPR). IEEE; 2022. p. 2078–85.
49. Majee A, Agrawal K, Subramanian A. Few-shot learning for road object detection. In: AAAI workshop on meta-learning and MetaDL challenge. PMLR; 2021. p. 115–26.
50. Agarwal A, Majee A, Subramanian A, Arora C. Attention guided cosine margin for overcoming class-imbalance in few-shot road object detection; 2022. <https://ieeexplore.ieee.org/document/9707579>
51. Huang X, He B, Tong M, Wang D, He C. Few-shot object detection on remote sensing images via shared attention module and balanced fine-tuning strategy. *Remote Sens*. 2021;13(19):3816. <https://doi.org/10.3390/rs13193816>.
52. Wang Y, Xu C, Liu C, Li Z. Context information refinement for few-shot object detection in remote sensing images. *Remote Sens*. 2022;14(14):3255.
53. Ouyang C, Biffi C, Chen C, Kart T, Qiu H, Rueckert D. Self-supervision with superpixels: training few-shot medical image segmentation without annotation. In: Computer vision-ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. Springer; 2020. p. 762–80.
54. Dutta T, Singh A, Biswas S. Adaptive margin diversity regularizer for handling data imbalance in zero-shot sbir. In: Proceedings, Part V 16, computer vision-ECCV 2020: 16th European conference, Glasgow, UK, August 23–28. Springer; 2020. p. 349–64.
55. Wu G, Gong S. Generalising without forgetting for lifelong person re-identification. In: Proceedings of the AAAI conference on artificial intelligence, vol. 35; 2021. p. 2889–97. <https://doi.org/10.1609/aaai.v35i4.16395>
56. Chen Z, Ma W, Xu N, Ji C, Zhang Y. SiameseCCR: a novel method for one-shot and few-shot Chinese CAPTCHA recognition using deep Siamese network. *IET Image Process*. 2020;14(12):2855–9. <https://doi.org/10.1049/iet-ipc.2019.0618>.
57. Wang Y, Wei Y, Zhang Y, Jin C, Xin G, Wang B. Few-shot learning in realistic settings for text CAPTCHA recognition. *Neural Comput Appl*. 2023;35(15):10751–64. <https://doi.org/10.1007/s00521-023-08262-0>.
58. Lei W, Su Q, Jiang T, Gu R, Wang N, Liu X, Wang G, Zhang X, Zhang S. One-shot weakly-supervised segmentation in 3D medical images. *IEEE Trans Med Imaging*. 2023. <https://doi.org/10.1109/TMI.2023.3294975>.
59. Cui H, Wei D, Ma K, Gu S, Zheng Y. A unified framework for generalized low-shot medical image segmentation with scarce data. *IEEE Trans Med Imaging*. 2020;40(10):2656–71. <https://doi.org/10.1109/TMI.2020.3045775>.
60. Bragg J, Cohan A, Lo K, Beltagy I. Flex: unifying evaluation for few-shot nlp. *Adv Neural Inf Process Syst*. 2021;34:15787–800.
61. Kim M, Yang Y, Ryu JH, Kim T. Meta-learning with adaptive weighted loss for imbalanced cold-start recommendation; 2023. <https://dl.acm.org/doi/10.1145/3583780.3614965>.
62. Yu T, Guo H, Zhu Y. Center loss guided prototypical networks for unbalance few-shot industrial fault diagnosis. *Mobile information systems*. Hindawi; 2022.
63. Li K, Shang C, Ye H. Reweighted regularized prototypical network for few-shot fault diagnosis. *IEEE Trans Neural Netw Learn Syst*. 2022. <https://doi.org/10.1109/TNNLS.2022.3232394>.
64. Vinyals O, Blundell C, Lillicrap T, kavukcuoglu k, Wierstra D. Matching networks for one shot learning. In: Advances in neural information processing systems, vol. 29. Curran Associates, Inc.; 2016. <https://proceedings.neurips.cc/paper/2016/hash/90e1357833654983612fb05e3ec9148c-Abstract.html>. Accessed 11 June 2023.
65. Wah C, Branson S, Welinder P, Perona P, Belongie S. The Caltech-UCSD Birds-200-2011 Dataset. [https://www.vision.caltech.edu/datasets/cub\\_200\\_2011/](https://www.vision.caltech.edu/datasets/cub_200_2011/).
66. Chen W-Y, Liu Y-C, Kira Z, Wang Y-CF, Huang J-B. A closer look at few-shot classification; 2018. <https://openreview.net/forum?id=HkxLXnAcFQ>. Accessed 03 June 2023.

67. Hu Y, Gripon V, Pateux S. Leveraging the feature distribution in transfer-based few-shot learning; 2021. [arXiv: arXiv:2006.03806](https://arxiv.org/abs/2006.03806) [cs, stat]. <http://arxiv.org/abs/2006.03806>. Accessed 10 May 2023.
68. Boudiaf M, Masud ZI, Rony J, Dolz J, Piantanida P, Ayed IB. Transductive information maximization for few-shot learning; 2020. [arXiv: arXiv:2008.11297](https://arxiv.org/abs/2008.11297) [cs, stat]. <http://arxiv.org/abs/2008.11297>. Accessed 10 May 2023.
69. Ren M, Triantafillou E, Ravi S, Snell J, Swersky K, Tenenbaum JB, Larochelle H, Zemel RS. Meta-learning for semi-supervised few-shot classification; 2018. <https://openreview.net/forum?id=HJcSzz-CZ>. Accessed 18 July 2023.
70. Ziko I, Dolz J, Granger E, Ayed IB. Laplacian regularized few-shot learning. In: International conference on machine learning. PMLR; 2020. p. 11660–70.
71. Triantafillou E, Zhu T, Dumoulin V, Lamblin P, Evci U, Xu K, Goroshin R, Gelada C, Swersky K, Manzagol P-A, Larochelle H. Meta-dataset: a dataset of datasets for learning to learn from few examples; 2019. <https://openreview.net/forum?id=rkgAGAVKPr>. Accessed 21 July 2023.
72. Wertheimer D, Hariharan B. Few-shot learning with localization in realistic settings. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR); 2019. p. 6551–60. <https://doi.org/10.1109/CVPR.2019.00672>. ISSN: 2575-7075.
73. Van Horn G, Mac Aodha O, Song Y, Cui Y, Sun C, Shepard A, Adam H, Perona P, Belongie S. The INaturalist species classification and detection dataset. 2018. p. 8769–78. [https://openaccess.thecvf.com/content\\_cvpr\\_2018/html/Van\\_Horn\\_The\\_INaturalist\\_Species\\_CVPR\\_2018\\_paper.html](https://openaccess.thecvf.com/content_cvpr_2018/html/Van_Horn_The_INaturalist_Species_CVPR_2018_paper.html). Accessed 18 July 2023.
74. Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, Krueger G, Sutskever I. Learning transferable visual models from natural language supervision; 2021. [arXiv: arXiv:2103.00020](https://arxiv.org/abs/2103.00020) [cs]. <https://doi.org/10.48550/arXiv.2103.00020>. Accessed 04 Sept 2023.
75. Zhang S, Li Z, Yan S, He X, Sun J. Distribution alignment: a unified framework for long-tail visual recognition; 2021. p. 2361–70. [https://openaccess.thecvf.com/content\\_CVPR2021/html/Zhang\\_Distribution\\_Alignment\\_A\\_Unified\\_Framework\\_for\\_Long-Tail\\_Visual\\_Recognition\\_CVPR\\_2021\\_paper.html](https://openaccess.thecvf.com/content_CVPR2021/html/Zhang_Distribution_Alignment_A_Unified_Framework_for_Long-Tail_Visual_Recognition_CVPR_2021_paper.html). Accessed 04 Sept 2023.
76. Liu Z, Miao Z, Zhan X, Wang J, Gong B, Yu SX. Large-scale long-tailed recognition in an open world; 2019. p. 2537–46. [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Liu\\_Large-Scale\\_Long-Tailed\\_Recognition\\_in\\_an\\_Open\\_World\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Liu_Large-Scale_Long-Tailed_Recognition_in_an_Open_World_CVPR_2019_paper.html). Accessed 25 July 2023.
77. Krizhevsky A. Learning multiple layers of features from tiny images; 2009. <https://www.semanticscholar.org/paper/Learning-Multiple-Layers-of-Features-from-Tiny-Krizhevsky/5d90f06bb70a0a3dc62413346235c02b1aa086>. Accessed 21 July 2023.
78. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition; 2009. p. 248–55. <https://doi.org/10.1109/CVPR.2009.5206848>. ISSN: 1063-6919.
79. Netzer Y, Wang T, Coates A, Bissacco A, Wu B, Ng AY. Reading digits in natural images with unsupervised feature learning. In: NIPS workshop on deep learning and unsupervised feature learning 2011; 2011. [http://ufldl.stanford.edu/housenumbers/nips2011\\_housenumbers.pdf](http://ufldl.stanford.edu/housenumbers/nips2011_housenumbers.pdf). Accessed 21 July 2023.
80. Peng X, Bai Q, Xia X, Huang Z, Saenko K, Wang B. Moment matching for multi-source domain adaptation; 2019. p. 1406–15. [https://openaccess.thecvf.com/content\\_ICCV\\_2019/html/Peng\\_Moment\\_Matching\\_for\\_Multi-Source\\_Domain\\_Adaptation\\_ICCV\\_2019\\_paper.html](https://openaccess.thecvf.com/content_ICCV_2019/html/Peng_Moment_Matching_for_Multi-Source_Domain_Adaptation_ICCV_2019_paper.html). Accessed 19 July 2023.
81. Xiao J, Hays J, Ehinger KA, Oliva A, Torralba A. SUN database: large-scale scene recognition from abbey to zoo. In: 2010 IEEE computer society conference on computer vision and pattern recognition; 2010. p. 3485–92. <https://doi.org/10.1109/CVPR.2010.5539970>. ISSN: 1063-6919.
82. Xian Y, Lampert CH, Schiele B, Akata Z. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Trans Pattern Anal Mach Intell.* 2019;41(9):2251–65. <https://doi.org/10.1109/TPAMI.2018.2857768>.
83. Skorokhodov I, Elhoseiny M. Class normalization for (continual)? Generalized zero-shot learning; 2021. [arXiv: arXiv:2006.11328](https://arxiv.org/abs/2006.11328) [cs, stat]. <https://doi.org/10.48550/arXiv.2006.11328>. <http://arxiv.org/abs/2006.11328>. Accessed 06 June 2023.
84. Varma G, Subramanian A, Namboodiri A, Chandraker M, Jawahar CV. IDD: a dataset for exploring problems of autonomous navigation in unconstrained environments. In: 2019 IEEE winter conference on applications of computer vision (WACV); 2019. p. 1743–51. <https://doi.org/10.1109/WACV.2019.00190>. ISSN: 1550-5790.
85. Wang X, Huang TE, Darrell T, Gonzalez JE, Yu F. Frustratingly simple few-shot object detection; 2020. [arXiv: arXiv:2003.06957](https://arxiv.org/abs/2003.06957) [cs]. <https://doi.org/10.48550/arXiv.2003.06957>. <http://arxiv.org/abs/2003.06957>. Accessed 02 June 2023.
86. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *Int J Comput Vis.* 2010;88(2):303–38. <https://doi.org/10.1007/s11263-009-0275-4>.
87. Cheng G, Han J, Zhou P, Guo L. Multi-class geospatial object detection and geographic image classification based on collection of part detectors. *ISPRS J Photogrammetry Remote Sens.* 2014;98:119–32. <https://doi.org/10.1016/j.isprsjprs.2014.10.002>.
88. Li K, Wan G, Cheng G, Meng L, Han J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J Photogrammetry Remote Sens.* 2020;159:296–307. <https://doi.org/10.1016/j.isprsjprs.2019.11.023>.
89. Wang K, Liew JH, Zou Y, Zhou D, Feng J. PANet: few-shot image semantic segmentation with prototype alignment. In: 2019 IEEE/CVF international conference on computer vision (ICCV); 2019. p. 9196–205. <https://doi.org/10.1109/ICCV.2019.00929>. ISSN: 2380-7504.
90. Shrivastava A, Gupta A, Girshick R. Training region-based object detectors with online hard example mining; 2016. p. 761–9. [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/html/Shrivastava\\_Training\\_Region-Based\\_Object\\_CVPR\\_2016\\_paper.html](https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Shrivastava_Training_Region-Based_Object_CVPR_2016_paper.html). Accessed 26 May 2023.
91. Ronneberger O, Fischer P, Brox T. 2015; U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention—MICCAI 2015. Lecture notes in computer science.* Cham: Springer. p. 234–41. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)



92. Khashabi D, Min S, Khot T, Sabharwal A, Tafjord O, Clark P, Hajjishirzi H. UNIFIEDQA: crossing format boundaries with a single qa system. in: findings of the association for computational linguistics: EMNLP. Association for Computational Linguistics, Online 2020; 2020. p. 1896–907. <https://doi.org/10.18653/v1/2020.findings-emnlp.171>. <https://aclanthology.org/2020.findings-emnlp.171>. Accessed 19 May 2023.
93. Duan R, Li D, Tong Q, Yang T, Liu X, Liu X. A survey of few-shot learning: an effective method for intrusion detection. Security and communication networks. Hindawi Limited; 2021.
94. Bansal A, Goldblum M, Cherepanova V, Schwarzschild A, Bruss CB, Goldstein T. MetaBalance: high-performance neural networks for class-imbalanced data; 2021. arXiv preprint [arXiv:2106.09643](https://arxiv.org/abs/2106.09643). <https://doi.org/10.48550/arxiv.2106.09643>
95. Zhu L, Yang Y. Inflated episodic memory with region self-attention for long-tailed visual recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 4344–53.
96. Samuel D, Atzmon Y, Chechik G. From generalized zero-shot learning to long-tail with class descriptors. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision; 2021. p. 286–95. <https://doi.org/10.1109/wacv48630.2021.00033>.
97. Patra A, Noble JA. Hierarchical class incremental learning of anatomical structures in fetal echocardiography videos. *IEEE J Biomed Health Inform.* 2020;24(4):1046–58. <https://doi.org/10.1109/jbhi.2020.2973372>.
98. Guan J, Liu J, Sun J, Feng P, Shuai T, Wang W. Meta metric learning for highly imbalanced aerial scene classification. In: ICASSP 2020–2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2020. p. 4047–51. <https://doi.org/10.1109/ICASSP40776.2020.9052900>
99. Weng W-H, Deaton J, Natarajan V, Elsayed GF, Liu Y. Addressing the real-world class imbalance problem in dermatology. In: Machine learning for health. PMLR; 2020. p. 415–29.
100. Dong N, Kampffmeyer M, Voiculescu I. learning underrepresented classes from decentralized partially labeled medical images. In: International conference on medical image computing and computer-assisted intervention. Springer; 2022. p. 67–76.
101. Pei Z, Jiang H, Li X, Zhang J, Liu S. Data augmentation for rolling bearing fault diagnosis using an enhanced few-shot Wasserstein auto-encoder with meta-learning. *Meas Sci Technol.* 2021;32(8): 084007. <https://doi.org/10.1088/1361-6501/abe5e3>.
102. Liu X, Gao W, Li R, Xiong Y, Tang X, Chen S. One shot ancient character recognition with siamese similarity network. *Sci Rep.* 2022;12(1):14820. <https://doi.org/10.1038/s41598-022-18986-z>.
103. Cui Z, Wang Q, Guo J, Lu N. Few-shot classification of façade defects based on extensible classifier and contrastive learning. *Autom Constr.* 2022;141: 104381.
104. Sumer O, Hellmann F, Hustinx A, Hsieh T-C, André E, Krawitz P. Few-shot meta-learning for recognizing facial phenotypes of genetic disorders. *Stud Health Technol Inform.* 2023;302:932–6. <https://doi.org/10.3233/SHTI230312>.
105. Zhan Z, Zhou J, Xu B. Fabric defect classification using prototypical network of few-shot learning algorithm. *Comput Ind.* 2022;138: 103628.
106. Tambwekar A, Agrawal K, Majee A, Subramanian A. Few-shot batch incremental road object detection via detector fusion. In: Proceedings of the IEEE/CVF international conference on computer vision; 2021. p. 3070–7. <https://doi.org/10.1109/iccvw54120.2021.00341>
107. Tian Y, Maicas G, Pu LZCT, Singh R, Verjans JW, Carneiro G. Few-shot anomaly detection for polyp frames from colonoscopy. In: Medical image computing and computer assisted intervention-MICCAI 2020: , Lima, Peru, October 4-8, 2020, Proceedings, Part VI 23. Springer; 2020. p. 274–84. <https://doi.org/10.48550/arxiv.2006.14811>
108. Ji Z, Liu X, Pang Y, Ouyang W, Li X. Few-shot human-object interaction recognition with Semantic-guided attentive prototypes network. *IEEE Trans Image Process Publ IEEE Signal Process Soc.* 2021;30:1648–61. <https://doi.org/10.1109/TIP.2020.3046861>.
109. Sharma R, Sharma G, Pattanaik M. A few shot learning based approach for hardware trojan detection using deep siamese cnn. In: 2021 34th international conference on VLSI design and 2021 20th international conference on embedded systems (VLSID). IEEE; 2021. p. 163–8. <https://doi.org/10.1109/VLSID51830.2021.00033>
110. Hu X, Jiang Y, Tang K, Chen J, Miao C, Zhang H. Learning to segment the tail. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2020. p. 14045–54. <https://doi.org/10.1109/cvpr42600.2020.01406>
111. Chen X, Zheng X, Sun K, Liu W, Zhang Y. Self-supervised vision transformer-based few-shot learning for facial expression recognition. *Inf Sci.* 2023;634:206–26. <https://doi.org/10.1016/j.ins.2023.03.105>.
112. He C, Zhang J, Yao J, Zhuo L, Tian Q. Meta-Learning Paradigm and CosAttn for Streamer Action Recognition in Live Video. *IEEE Signal Process Lett.* 2022;29:1097–101.
113. Romanov S, Song H, Valstar M, Sharkey D, Henry C, Triguero I, Torres MT. Few-shot learning for postnatal gestational age estimation. In: 2021 international joint conference on neural networks (IJCNN). IEEE; 2021. p. 1–8. <https://doi.org/10.1109/IJCNN52387.2021.9534239>
114. Shi Y, Li J, Li Y, Du Q. Sensor-independent hyperspectral target detection with semisupervised domain adaptive few-shot learning. *IEEE Trans Geosci Remote Sens.* 2020;59(8):6894–906. <https://doi.org/10.1109/TGRS.2020.3032528>.
115. Bedi P, Gupta N, Jindal V. I-SiamIDS: an improved Siam-IDS for handling class imbalance in network-based intrusion detection systems. *Appl Intell.* 2021;51:1133–51. <https://doi.org/10.1007/s10489-020-01886-y>.
116. Huang S, Liu Y, Fung C, An W, He R, Zhao Y, Yang H, Luan Z. a gated few-shot learning model for anomaly detection. In: 2020 international conference on information networking (ICOIN); 2020. p. 505–9. <https://doi.org/10.1109/ICOIN48656.2020.9016599>. ISSN: 1976-7684
117. Gesi J, Li J, Ahmed I. An empirical examination of the impact of bias on just-in-time defect prediction. In: Proceedings of the 15th ACM/IEEE international symposium on empirical software engineering and measurement (ESEM); 2021. p. 1–12. <https://doi.org/10.1145/3475716.3475791>
118. Wu X, Wang N. Detecting errors with zero-shot learning. *Entropy (Basel, Switzerland).* 2022;24(7):936. <https://doi.org/10.3390/e24070936>.



119. Li M, Zhang Y, Han D, Zhou M. Meta-IP: an imbalanced processing model based on meta-learning for IT project extension forecasts. *Mathematical problems in engineering*; 2022. <https://www.hindawi.com/journals/mpe/2022/3140301/>.
120. Chen X, Zhang C, Lin G, Han J. Compositional prototype network with multi-view comparison for few-shot point cloud semantic segmentation; 2020. arXiv preprint [arXiv:2012.14255](https://arxiv.org/abs/2012.14255)<https://doi.org/10.48550/arxiv.2012.14255>
121. Gao A, Mei F, Zheng J, Sha H, Guo M, Xie Y. Electricity theft detection based on contrastive learning and non-intrusive load monitoring. *IEEE Trans Smart Grid*. 2023. <https://doi.org/10.1109/TSG.2023.3263219>.
122. Gupta P, Bhaskarpandit S, Gupta M. Similarity Learning based few shot learning for ECG time series classification. In: 2021 digital image computing: techniques and applications (DICTA). IEEE; 2021. p. 1–8. <https://doi.org/10.1109/DICTA52665.2021.9647357>
123. Bhosale S, Tiwari U, Chakraborty R, Kopparapu SK. Contrastive learning of cough descriptors for automatic COVID-19 preliminary diagnosis. *Interspeech*; 2021. <https://doi.org/10.21437/Interspeech.2021-1249>
124. Rentería S, Vallejo EE, Taylor CE. Birdsong phrase verification and classification using siamese neural networks. *bioRxiv*. Cold Spring Harbor Laboratory. 2021–03; 2021. <https://doi.org/10.1101/2021.03.16.435625>.
125. Sunder V, Fosler-Lussier E. Handling class imbalance in low-resource dialogue systems by combining few-shot classification and interpolation. In: ICASSP 2021–2021 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE; 2021. p. 7633–7. <https://doi.org/10.1109/ICASSP39728.2021.9413405>
126. Fernández-Llaneza D, Ulander S, Gogishvili D, Nittinger E, Zhao H, Tyrchan C. Siamese Recurrent neural network with a self-attention mechanism for bioactivity prediction. *ACS Omega*. 2021;6(16):11086–94. <https://doi.org/10.1021/acsomega.1c01266>.
127. Wenjuan G, Zhang Y, Wang W, Cheng P, Sabate JG. Meta-MMFNet: meta-learning based multi-model fusion network for micro-expression recognition. *ACM transactions on multimedia computing, communications and applications*. Association for Computing Machinery (ACM); 2022.
128. Patil S, Ravindran B. Predicting software defect type using concept-based classification. *Empir Softw Eng*. 2020;25(2):1341–78. <https://doi.org/10.1007/s10664-019-09779-6>.
129. Tolstikhin I, Bousquet O, Gelly S, Schoelkopf B. Wasserstein auto-encoders; 2019. [arXiv:1711.01558](https://arxiv.org/abs/1711.01558) [cs, stat]. <https://doi.org/10.48550/arXiv.1711.01558>. <http://arxiv.org/abs/1711.01558>. Accessed 16 Aug 2023.
130. Nichol A, Achiam J, Schulman J. On first-order meta-learning algorithms; 2018. [arXiv:1803.02999](https://arxiv.org/abs/1803.02999) [cs]. <https://doi.org/10.48550/arXiv.1803.02999>. <http://arxiv.org/abs/1803.02999>. Accessed 16 June 2023.
131. Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: additive angular margin loss for deep face recognition; 2019. p. 4690–9. [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Deng\\_ArcFace\\_Additive\\_Angular\\_Margin\\_Loss\\_for\\_Deep\\_Face\\_Recognition\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Deng_ArcFace_Additive_Angular_Margin_Loss_for_Deep_Face_Recognition_CVPR_2019_paper.html). Accessed 26 July 2023.
132. Fan Q, Zhuo W, Tang C-K, Tai Y-W. Few-shot object detection with attention-rpn and multi-relation detector; 2020. [arXiv:1908.01998](https://arxiv.org/abs/1908.01998) [cs]. <https://doi.org/10.48550/arXiv.1908.01998>. <http://arxiv.org/abs/1908.01998>. Accessed 18 June 2023.
133. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks; 2016. [arXiv:1506.01497](https://arxiv.org/abs/1506.01497) [cs]. <https://doi.org/10.48550/arXiv.1506.01497>. <http://arxiv.org/abs/1506.01497>. Accessed 18 June 2023.
134. Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: common objects in context. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T, editors. *Computer vision—ECCV 2014 Lecture notes in computer science*. Cham: Springer; 2014. p. 740–55. [https://doi.org/10.1007/978-3-319-10602-1\\_48](https://doi.org/10.1007/978-3-319-10602-1_48).
135. Gupta A, Dollar P, Girshick R. LVIS: a dataset for large vocabulary instance segmentation; 2019. p. 5356–64. [https://openaccess.thecvf.com/content\\_CVPR\\_2019/html/Gupta\\_LVIS\\_A\\_Dataset\\_for\\_Large\\_Vocabulary\\_Instance\\_Segmentation\\_CVPR\\_2019\\_paper.html](https://openaccess.thecvf.com/content_CVPR_2019/html/Gupta_LVIS_A_Dataset_for_Large_Vocabulary_Instance_Segmentation_CVPR_2019_paper.html). Accessed 25 July 2023.
136. Liu Z, Luo P, Wang X, Tang X. Deep learning face attributes in the wild. In: *Proceedings of international conference on computer vision (ICCV)*; 2015. [https://openaccess.thecvf.com/content\\_iccv\\_2015/papers/Liu\\_Deep\\_Learning\\_Face\\_ICCV\\_2015\\_paper.pdf](https://openaccess.thecvf.com/content_iccv_2015/papers/Liu_Deep_Learning_Face_ICCV_2015_paper.pdf).
137. Bedi P, Gupta N, Jindal V. Siam-IDS: handling class imbalance problem in intrusion detection systems using siamese neural network. *Procedia Comput Sci*. 2020;171:780–9. <https://doi.org/10.1016/j.procs.2020.04.085>.
138. Tavallaee M, Bagheri E, Lu W, Ghorbani AA. A detailed analysis of the KDD CUP 99 data set. In: 2009 IEEE symposium on computational intelligence for security and defense applications; 2009. p. 1–6. <https://doi.org/10.1109/CISDA.2009.5356528>. ISSN: 2329-6275
139. Hoang T, Khanh Dam H, Kamei Y, Lo D, Ubayashi N. DeepJIT: an end-to-end deep learning framework for just-in-time defect prediction. In: 2019 IEEE/ACM 16th international conference on mining software repositories (MSR); 2019. p. 34–45. <https://doi.org/10.1109/MSR.2019.00016>. ISSN: 2574-3864.
140. Heidari A, McGrath J, Ilyas IF, Rekatsinas T. HoloDetect: few-shot learning for error detection. In: *Proceedings of the 2019 international conference on management of data*. SIGMOD '19. Association for Computing Machinery, New York, NY, USA 2019. p. 829–846. <https://doi.org/10.1145/3299869.3319888>. <https://dl.acm.org/doi/10.1145/3299869.3319888>. Accessed 12 June 2023.
141. Seliya N, Abdollah Zadeh A, Khoshgoftaar TM. A literature review on one-class classification and its potential applications in big data. *J Big Data*. 2021;8(1):122. <https://doi.org/10.1186/s40537-021-00514-x>.
142. Navigli R. Word sense disambiguation: a survey. *ACM Comput Surv*. 2009;41(2):10–11069. <https://doi.org/10.1145/1459352.1459355>.

## Publisher's Note

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