Information Fusion 106 (2024) 102263

Contents lists available at ScienceDirect

Information Fusion

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

Semi-supervised information fusion for medical image analysis: Recent progress and future perspectives

Ying Weng [a,*,1], Yiming Zhang [a,1], Wenxin Wang [a], Tom Dening [b]

[a] School of Computer Science, University of Nottingham Ningbo China, Ningbo, 315100, China
[b] School of Medicine, University of Nottingham, Nottingham, NG7 2RD, UK

ARTICLE INFO

Keywords:
Semi-supervised learning
Information fusion
Medical imaging
Medical image analysis
Survey

ABSTRACT

Supervised machine learning requires training on the dataset with annotation. However, fine-grained annotation is very expensive to acquire. In the medical image analysis domain, the sheer volume of data and lack of annotation limit the performance of the model. To address these limitations, semi-supervised information fusion has recently emerged as an important and promising paradigm owing to its ability to exploit labelled and unlabelled data and combine information from multiple sources to obtain a more robust and accurate performance. In this survey, we review the recent progress of semi-supervised information fusion for medical image analysis. Moreover, we categorize the state-of-the-art information fusion applications of semi-supervised learning with in-depth analysis. Finally, we discuss the challenges and outline the future perspective.

1. Introduction

Recent advances in machine learning and deep learning have shown promising results in the medicine and healthcare domain [1–4]. Medical imaging plays a crucial role in clinical decision-making, the information extracted from medical images is clinically valuable in many areas such as computer-aided detection, diagnosis, treatment planning, intervention, and therapy. With the increasing demand for medical imaging and the shortage of radiologists, automated methods have become crucial in helping healthcare practitioners. The advancement of deep learning models in medical image analysis has inspired innovations in the field [5]. These techniques enable more accurate and efficient image acquisition, analysis, and interpretation, leading to new insights and improved outcomes in areas such as registration, reconstruction, tracking, segmentation, and image quality assessment.

However, traditional supervised machine learning training requires labels and annotation of medical images that need manual manipulation by experienced radiologists. Moreover, it is difficult to directly transfer trained models to other medical imaging datasets for application due to the differences between various modalities and objective organs. This can be challenging and it is often costly to obtain large annotated datasets. In the medical domain, there exist multiple data modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Ultrasound Imaging (US), X-ray Radiography (X-ray), endoscopy and histology. Only using one modality of data may not achieve accurate performance. Hence, many researchers explored integrating different data modalities or different views of one data modality to improve the overall performance of the model.

To address these challenges, semi-supervised data fusion, which applies both semi-supervised and data fusion techniques, has emerged as a promising paradigm. Integrating these novel techniques with clinical practice has the potential to revolutionize the way that medical imaging is used in healthcare. This leads to better diagnosis and treatment planning and a better understanding of underlying biological processes.

Semi-supervised learning is halfway between supervised and unsupervised learning that has emerged as a solution to the problem of limited labelled data, which is a common issue in many real-world applications of machine learning. It originated as a heuristic self-training method in pattern recognition tasks in the 1960s [6,7]. With the rise of deep learning in recent years, semi-supervised learning has emerged, combining the strengths of deep neural networks in capturing complex image features with the benefits of large amounts of unlabelled data to improve model performance. The availability of large amounts of unlabelled medical images presents an opportunity to leverage semi-supervised learning techniques that use a small amount of labelled data to improve the performance of a model on large amounts of unlabelled data and reduce the need for extensive manual annotation. In the

* Corresponding author.
E-mail address: ying.weng@nottingham.edu.cn (Y. Weng).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.inffus.2024.102263
Received 30 August 2023; Received in revised form 17 January 2024; Accepted 18 January 2024
Available online 22 January 2024

medical field, single data modality may not achieve accurate results compared to multimodal data mining with information fusion techniques. Information fusion integrates different types of data together and can be categorized into three types: data fusion, feature fusion, and decision fusion [8,9].

1.1. Objectives and main contributions of research

This survey provides a comprehensive review on the recent progress of semi-supervised information fusion in medical image analysis. We summarize the background of the semi-supervised information fusion techniques, including the taxonomy of semi-supervised learning method and information fusion. We then address the recent works on medical image analysis applications using semi-supervised information fusion. We also discuss the challenges and provide future perspectives.

During the literature search process, we have identified several published semi-supervised learning survey papers on deep learning and semi-supervised learning [10–17]. In [15–17], the authors explained the principles of traditional semi-supervised learning. In [14], Tajbakhsh et al. specifically analysed the segmentation solutions when faced with imperfect datasets, e.g., scarce or weak annotations. In [11], Cheplygina et al. focused on medical imaging applications with semi-supervised, multiple instance, and transfer learning techniques. In [10,13], a systematic analysis of graph-based semi-supervised learning was presented. Yang et al. [12] described the development of deep semi-supervised learning with representative models. Compared to the existing surveys, we explore the recent progress on semi-supervised information fusion for medical image analysis.

In this survey, we have articulated the following research questions (RQs): **RQ1:** What is the current research progress for semi-supervised information fusion techniques in medical image analysis? **RQ2:** What insights emerge from comparing semi-supervised information fusion methodologies when applied to empirical medical image analysis case studies? **RQ3:** What challenges are encountered while implementing semi-supervised information fusion techniques in medical image analysis? **RQ4:** What are the potential future perspectives for semi-supervised information fusion techniques in medical image analysis?

The contributions of this survey are threefold:

- To the best of our knowledge, this is the first survey on semi-supervised information fusion.
- We provide a comprehensive and in-depth analysis of medical image analysis applications using semi-supervised information fusion.
- We present two case studies to demonstrate practical applications and discuss challenges and future perspectives in this rapidly evolving field.

1.2. Search strategy and organization

A literature search is conducted using the keywords semi-supervised learning, deep learning, information fusion and medical images. In addition, research papers cited in this review are found on four electronic databases - PubMed, IEEE Xplore, Science Direct and Springer Link - for relevant publications between 2018 and 2023 inclusive. The search strings we use in this survey include: (ALL (“Semi-supervised Learning”) OR ALL (“Medical Image”) OR ALL (“Medical Imaging”) OR ALL (“Deep Learning”)). We exclude review articles, non-medical image analysis and non-deep semi-supervised learning articles being the main reasons for exclusion from the title and abstract screening. After full-text reviews, 50 studies have met the inclusion criteria.

The rest of the paper is organized as follows: Section 2 presents the background of semi-supervised and information fusion. Section 3 describes the medical image analysis applications using semi-supervised information fusion. Section 4 analysis two case studies on different datasets. Section 5 discusses the challenges and future directions. Section 6 concludes this survey.

2. Preliminaries

In this section, we present the background of semi-supervised learning and information fusion respectively.

2.1. Overview of semi-supervised learning

There are four basic groups of semi-supervised learning, including pseudo labelling, consistency regularization, graph-based method and generative model. Each method and its subcategories are briefly described below.

2.1.1. Pseudo labelling

Pseudo labelling is a direct and intuitive method which is similar to a wrapper function [18,19]. It first trains the network with originally labelled data and then uses the resulting network to infer labels for the previously unlabelled data, and these predictions are used as pseudo labels in conjunction with labelled data to update the network model. Pseudo labelling is highly versatile, as it can be applied to almost any supervised learning algorithm and used as a subtask for other supervised, semi-supervised or even unsupervised learning methods, although the loss function may vary. This section will focus on purely pseudo-labelling methods, which are simply delineated — according to whether the model processes data in a single view or in multiple views, with the former referred to here as self-training and the latter as co-training.

Self-training The self-training method upgrades the training model itself with both labelled data and pseudo-labels from earlier iterations. This concept of self-training was arguably first attempted in the 1960s to incorporate unlabelled data training, and according to Scudder [6], an untaught adaptive pattern recognition uses its output rather than that of a teacher.

Over time, various improvements have been made to the basic pseudo labelling technique. Nowadays, the self-training paradigm acknowledges a number of design decisions, including techniques for filtering reliable pseudo labels and techniques for weighing the impact of pseudo labels on the final model.

Algorithm 1 An Example of Self-Training Procedure

Input $\{x_l, y_l\}$ from labelled dataset $\mathcal{D}_l$, x_u from unlabelled dataset $\mathcal{D}_u$, iteration numbers T
Output trained model $\mathcal{M}_t$

Train initial $\mathcal{M}_t$ with $\mathcal{D}_l$
while iteration < T **do**
  Generate Pseudo labels $\{P_u\}$ for $\mathcal{D}_u$ with $\mathcal{M}_t$
  Generate Pseudo labels $\{\hat{y}_u\}$ by selected predictions $\{p \in P_u | p$ meets some criteria$\}$
  Update the new Dataset $\mathcal{D}_t$ with $\mathcal{D}_l = \{x_l, y_l\}$ and $\{x_u.\hat{y}_u\}$, new unlabelled $\mathcal{D}_u$
  Fine-tune the training model $\mathcal{M}_t$ with $\mathcal{D}_t$
end while

Co-training In recent years pseudo labelling has been extended to semi-supervised learning with multiple views, where different views can be positioned for different processing methods, and objectives, or aimed at analysing different features. This is referred to as co-training methods, first proposed in [20]. For each data x in dataset $\mathcal{D}_l$, two or more conditionally independent views $v_1, v_2, \ldots, v_n$ are partitioned into multiple different predicting functions $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_n$. For the data in $\mathcal{D}_u$, the prediction generated by each view will supervise the other view in a pseudo labelled manner. The possible views can be arranged in parallel or recursively, and in general, they complement each other, or one functions as a domain and the others as subtasks to complement it.

Due to the different structures, the manner of cross-supervision of multiple views as pseudo labels may also differ. Meanwhile, the quality

of the pseudo label still plays an important role similar to that in self-training, so the techniques to ensure a high confidence level are applicable here as well. It is worth noting that a key difference between self-training and co-training is that the co-training approach optimizes the model in a way that considers multiple perspectives jointly.

Algorithm 2 An Example of Co-Training Procedure

Input $\{x_l, y_l\}$ from labelled dataset $\mathcal{D}_l$, x_u from unlabelled dataset $\mathcal{D}_u$, iteration numbers T
Output trained model $\mathcal{M}_t$, $\mathcal{M}'_t$

```
Train initial dominant model M_t with D_l in view v_1
Train initial sub-model M'_t with D_l in view v_2
while iteration < T do
    Generate Pseudo labels {ŷ_u} by selected predictions {p ∈ P_u|p meets some criteria} for D_u with M_t
    Generate Pseudo labels {ŷ'_u} by selected predictions {p' ∈ P'_u|p' meets some criteria} for D_u with M'_t
    Fine-tune the model M_t with {x'_u, ŷ'_u}
    Fine-tune the model M'_t with {x_u, ŷ_u}
end while
```

2.1.2. Consistency regularization

Unlike pseudo-labelling, this type of approach does not optimize the model directly on the prediction error but focuses on the consistency of predictions [21,22]. The perturbation-based method trains both labelled and unlabelled data simultaneously by adding unsupervised regularization to the unlabelled data. In deep semi-supervised learning, a common skill is to apply an additional penalty component to inconsistent predictions from slightly perturbed inputs. Integrating the basic assumptions of semi-supervised learning, if a realistic perturbation is applied to an unlabelled example, the prediction should be similar to the clean one according to the smoothing assumption, while the probability of switching categories is relatively small due to the clustering assumption, based on the manifold assumption which also acts on the topological space. Concretely, given an unlabelled data x belongs to $\mathcal{D}_u$ and its perturbed version $\tilde{x}$, the objective is to minimize the distance between the two outputs $d(f(x), f(\tilde{x}))$.

Temporal Ensemble Temporal ensemble [23] is a random perturbation method that accelerates the computational time of the Π-model [24]. It uses an ensembled prediction Y_t from previous iterations and a real-time perturbated prediction $\tilde{y}$ to penalize the small changes in outputs, therefore it only requires one propagation for each epoch. In addition, the target of the temporal ensemble aggregates the previously weighted average predictions rather than a single randomly augmented value, increasing the stability of the training process. Y_t of the ensemble output is obtained by updating $Y_t \leftarrow \alpha Y_t + (1-\alpha)\tilde{y}$, where α is a momentum term that controls the range of the ensemble over the training history. More interestingly, the hyperparameters can be further applied in conjunction with uncertainty information, for example, by placing more weight on high-confidence predictions.

Mean Teacher Tarvainen and Valpola [25] proposed Mean Teacher, which separated the teacher model apart, similar to the temporal ensemble where the teacher network was updated from the student network, and the consistency cost was estimated between predictions of the teacher and the stochastic augmentation and dropout predictions of the student (see Tables 1 and 2). The authors named the ensembled prediction method used in the temporal ensemble as Exponential Moving Average (EMA), which was formed by an ensemble of the model's current version and those earlier versions that evaluated the same example, and they modified the EMA method to update the teacher model weights: $\theta'_i = \alpha\theta'_{i-1} + (1-\alpha)\theta_i$.

Mixup Zhang et al. presented Mixup that aimed to alleviate the issues of deep neural networks exhibiting undesirable behaviours such as memorization and sensitivity to adversarial examples [26]. By enforcing linear variation in predictions between samples, the decision boundary was pushed far away from the class boundaries. The proposed augmentation method constructs virtual training examples by

$$\begin{aligned}\tilde{x} &= \lambda x_i + (1-\lambda)x_j \\ \tilde{y} &= \lambda y_i + (1-\lambda)y_j\end{aligned} \tag{1}$$

where (x_i, y_i) and (x_j, y_j) are random examples, and $\lambda \in [0,1]$. By enforcing linear variation in predictions between samples, the decision boundary is pushed far away from the class boundaries.

Interpolation Consistency Training (ICT) Verma et al. [27] implemented this skill in semi-supervised learning, as authors argued that when applying on unlabelled samples the interpolations between random data were still likely to fall in low-density regions and such interpolations were exact ideal locations for consistency-based regularization. The basic idea of the algorithm was to encourage consistent prediction $f(\alpha u_1 + (1-\alpha)u_2) = \alpha f(u_1) + (1-\alpha) f(u_2)$ at interpolations of unlabelled points u_1 and u_2.

MixMatch In [28] Berthelot et al. further created a holistic semi-supervised approach, MixMatch, which leveraged both labelled and unlabelled samples via shuffling and Mixup to achieve satisfactory performance.

Algorithm 3 Pseudocode for MixMatch algorithm

Input Batch of labelled examples $\mathcal{B}_l = ((x_b, p_b))$, batch of unlabelled examples $\mathcal{B}_u = (u_b)$, sharpening temperature T, number of augmentation N_a, number of iteration N_i
Output trained model $\mathcal{M}$

```
while iteration < N_i do
    x̂_b ← Augment(x_b)
    û_{b,i} ← Augment(u_b)
    q_b = Sharpen ((1/N_a) Σ M(u_{b,a}, θ)); a ∈ {1, ..., N_a}
    B̂_l = (x̂_b, p_b)
    B̂_u = (û_{b,a}, q_b); a ∈ {1, ..., N_a}
    W = Shuffle(Concat(B̂_l, B̂_u))
    B'_l = MixUp(B̂_{l,i}, W_i); i ∈ {1, ..., |B̂_l|}
    B'_u = MixUp(B̂_{u,i}, W_{i+|B̂_l|}); i ∈ {1, ..., |B̂_u|}
    L_mix = (1/(N_a|B'_u|)) Σ L2norm (q, M(u, θ))
    L = L_CE + λ L_mix
    Update M with L
end while
```

2.1.3. Graph-based methods

Graph-based semi-supervised learning (GSSL) has a rich history of development, with roots in both graph theory and machine learning. These techniques involve constructing a graph to represent the relationships between data points, with edges between nodes indicating similarities or proximities between the data points.

Label Propagation In the early days, approaches with graph structure have been popular in semi-supervised learning because the graphs could naturally be used in clustering assumption [29] to identify clusters of similar data points (i.e. nearby vertices) and then assign labels to the clusters; or in the manifold assumption where nodes connected by edges associated with large weights tend to reflect nearby samples on a low-dimensional manifold and should have the same labels [30]. Label propagation and its variants make use of propagation matrices to smooth the constructed graphs in a way that propagates information layer by layer. However, they are limited to propagating the labelled information within the graph structure and normally feed the extracted scores to another downstream learning mechanism for special tasks.

Deep Embedding One solution is to find feasible node embeddings that simultaneously contribute to the properties of the nodes and exploit the graph structure. The main purpose of node embeddings is to

Table 1
Theoretical analysis of typical random perturbation methods.

Methods	Forward propagation for each input per iteration	Consistency on	Loss function	Code source
Temporal Ensemble	Once	Prediction with stochastic augmentations and dropout vs. ensemble prediction	$\mathcal{L}_u = \omega \frac{1}{\|D\|} \sum_{x \in D} \|\|f(x,\eta) - Y_{ema}\|\|^2$	https://github.com/smlaine2/tempens
Mean Teacher	Twice	Predictions from teacher's and student's model	$\mathcal{L}_u = \omega \frac{1}{\|D\|} \sum_{x \in D} \|\|f(x,\eta,\theta) - f(x,\eta',\theta')\|\|^2$	https://github.com/CuriousAI/mean-teacher

Table 2
Theoretical analysis of three typical interpolation perturbation methods.

Methods	Data mixed	Label guess by	Code source
MixUp [26]	Labelled data only	–	https://github.com/facebookresearch/mixup-cifar10
ICT [27]	Unlabelled data only	EMA	https://github.com/vikasverma1077/ICT
MixMatch [28]	Both labelled and unlabelled data	Average of K weakly-augmented prediction + temperature sharpening	https://github.com/google-research/mixmatch

Algorithm 4 Pseudocode for label propagation algorithm

while not converge **do**
  Propagate $\widetilde{Y} \leftarrow T\widetilde{Y}$
  Row-normalize $\widetilde{Y}$
  Preserve labelled data $\widetilde{Y}_l \leftarrow Y_l$
end while

encode nodes as lower dimensional vectors, which can compressively reflect node features and neighbourhood structure.

Deep embedding method, such as Graph Convolution Network (GCN), is one of such techniques for embedding convolutional neural networks into graph-structured data modelling, extending the label smoothing procedure in label propagation to a feature smoothing algorithm. At each convolutional layer, representations of each node's direct neighbours are first aggregated and then propagated to the graph convolution. The general idea of graph convolutions is shown in Eq. (2), where the representations X_l are updated by the original information and the aggregated feature vectors of neighbouring nodes.

$$X_{l+1} = \text{Update}(X_l, \text{Aggregate}(\forall n_l \in \mathcal{N}(X_l))) \tag{2}$$

2.1.4. Generative models

Generative models differ from discriminative models where they assume that the data is generated from a probability distribution and the primary objective is to estimate the data distribution and generate a similar one. Specifically, in machine learning classification tasks, the final step is the same as in discriminative classifiers, i.e., calculating the conditional probability of the target variable [31].

Generative adversarial networks To discover the underlying distribution in the actual data samples, Generative Adversarial Networks (GANs) [32] set up a min–max two-player game between two deep neural network models - a generator and a discriminator. The generator G aims to generate plausible samples by capturing the real data distribution, while the discriminator D is a binary classifier designed to determine whether the samples are real (from the domain) or false (from the generator). The two models are trained adversarially, similar to two opponents who must continually improve their abilities to win the game.

Variational AutoEncoder Another well-known generative model is called Variational Autoencoder (VAE), which was proposed by Kingma and Welling [33]. Similar to traditional autoencoder (AE), it consists of an encoder and a decoder, both of which can be arbitrary models, usually neural networks. The input data x is down-dimensioned by one neural network into a vector of latent variables, z, which is then decoded by another neural network to give generated data identical to the original input data. VAE is trained to balance the model accuracy and the distribution of latent vectors, which is represented by MSE between synthetic and original images and KL divergence between posterior and prior distributions respectively.

Algorithm 5 Pseudocode for GANs

Input data distribution $p_{data}(x)$, noisy prior $p_g(z)$, iteration numbers T
Output trained model Discriminator D, Generator G
while iteration < T **do**
  For k steps **do**
    sample minibatch$\{x_1, x_2, \dots, x_m\} \leftarrow p_{data}(x)$
    sample minibatch$\{z_1, z_2, \dots, z_m\} \leftarrow p_g(z)$
    $G_d \leftarrow \nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} [\log D(x_i) + \log(1 - D(G(z_i)))]$
    update D by ascending its stochastic gradient G_d
  end for
  sample minibatch $\{z_1, z_2, \dots, z_m\} \leftarrow p_g(z)$
  $G_g \leftarrow \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} [\log(1 - D(G(z_i)))]$
  update G by descending its stochastic gradient G_g
end while

Algorithm 6 Pseudocode for VAEs

Input data distribution $p_{data}(x)$, noise distribution $p_\theta(\epsilon)$, iteration numbers T
Output trained model $\mathcal{M}$
while iteration < T **do**
  sample minibatch $\{x_1, x_2, \dots, x_m\} \leftarrow p_{data}(x)$
  samples $\epsilon \leftarrow p(\epsilon)$
  z *gets* Encoder (ϕ, ϵ, x)
  $\mathcal{L}_{KL} = \frac{1}{m} \sum_{i=1}^{m} [-D_{KL}(q_\phi(z|x_i)||p_\theta(z))]$
  $\mathcal{L}_{rec} = \frac{1}{m} \sum_{i=1}^{m} [\log p_\theta(x_i|z_i)]$
  $g \leftarrow \nabla_{\phi,\theta}(\mathcal{L}_{KL} + \mathcal{L}_{rec})$
  update $\mathcal{M}$ with g
end while

2.2. Overview of information fusion

Information fusion is defined as a process that combines single or multiple sources to increase the quality of the information [34]. By applying information fusion, it can process the imperfect raw data to obtain more consistent and reliable information. In principle, information fusion has great advantages over single-source data [35].

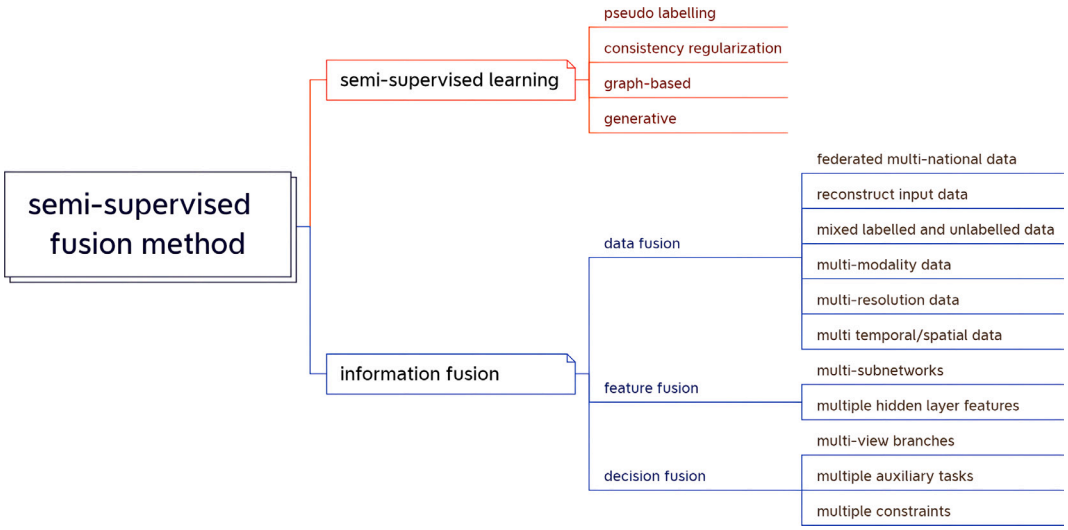

Fig. 1. An overview of semi-supervised information fusion methods.

In [36], Meng et al. summarized three key components of data fusion: data sources, operation, and purpose. In data fusion, single or multiple data sources are gathered from different positions and at different periods. Operation means combining data and refining information. The purpose of data fusion is to improve detection or prediction while reducing error possibilities and enhancing reliability. In terms of the level of information fusion, it can be categorized into three groups: data fusion, feature fusion, and decision fusion.

Data fusion refers to integrating different data modalities or multiple data sources at the beginning before the analysis process. For example, two different data modalities, 3D MRI and 3D CT, are fused for a segmentation task with the consistency regularization semi-supervised method [37].

Feature fusion means several sub-models, sub-networks, or layers in the model learned from sub-data and then aggregated. For instance, in [38], Wang et al. utilized sub-networks for feature fusion with 3D Optical Coherence Tomography (OCT) data.

Decision fusion, also known as late fusion, means that each medical data modality or view of data is trained and makes a prediction, then integrates the decisions of each model to make a final decision. For example, the authors designed multi-view branches and utilized decision fusion for 3D CT segmentation [39].

For designing medical image analysis applications, single-modality medical data may not provide sufficient information. From the perspective of medical data fusion, transforming information from multiple data modalities into a single medical application can improve the efficiency of the application. Moreover, it can also assist medical experts in making decision-making and provide more insights than a single modality. Hence, it is necessary to conduct information fusion in medical image analysis applications to enrich the model with more information.

3. Medical image analysis applications with semi-supervised information fusion

The aim of this survey is to provide a comprehensive and in-depth analysis of medical imaging applications with semi-supervised information fusion methods. As shown in Fig. 1, instead of organizing the structure according to the four classical semi-supervised learning classifications, this survey is designed to classify them according to three different types of information fusion, i.e., data fusion, feature fusion and decision fusion. Following each subtitle, the corresponding medical image applications employing the specified techniques are delineated, for a total of 50 research papers.

3.1. Data fusion applications

This section summarizes semi-supervised information fusion applications in medical image analysis, utilizing various data fusion techniques. Detailed descriptions of these applications can be found in Table 3.

Federate multi-national data Yang et al. combined a pseudo labelling-based deep semi-supervised learning with federated learning to solve the Covid-19 region segmentation problem between multi-national data [40]. The server-client communication scheme allowed the server to aggregate the model weights of all clients, and clients updated the server-assigned weights to fine-tune their local models, sending the new gradients to the server again. Each client with multi-national source data might have completely different annotation availability, and generally be trained in a self-training way. The pseudo labels were generated by the current local model and filtered using a hard threshold to increase the generalizability, with slight augmentation applied during training.

Reconstruct input data In [41], Guo and Yuan presented a pseudo labelling-based semi-supervised algorithm for wireless capsule endoscopy image classification using an adaptive aggregated attention technique with two branches trained jointly, one of which is an abnormal region classifier that outputs texture and structure attention maps. The maps were further fused in another information distiller, in order to emphasize the abnormal regions.

Another pseudo labelling-based semi-supervised study presented by Wang et al. focused on the multi-class infection segmentation task, in which the proposed spatial self-attention network was combined with few-shot techniques [42]. To achieve better segmentation performance, each CT slice was concatenated with its lung mask as the network's input. They demonstrated a re-weighting module for the class imbalance problem, which assigned weights based on the pixel ratio of classes and respective estimated probabilities and obtained reliable pseudo-labels from the trust module by selecting high confidence values.

Huo et al. proposed a consistency regularization method, which added a novel constraint on attention mask consistency in the proposed Dual Consistency-Mean Teacher (DC-MT) framework for grading assessments for knee cartilage defects [43]. Concretely, concentrated attention masks were generated using log-sum exp (LSE) pooling, and the mean squared error (MSE) between attention masks derived from the student and teacher models were calculated to ensure consistency and to make the networks focus on the cartilage regions which can

Table 3
Summary of medical image analysis applications leveraging data fusion.

Publication	Citations	Venue	Task	Modality	Data size	Data fusion methods	Semi-supervised method
[40]	182	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	2D CT	1704 slices	Multi-source Data	Pseudo Labelling
[41]	47	Medical Image Analysis, IF:10.9, JCR Q1	Classification	Endoscopy	1800L+1807U	Reconstruct Input Data	Pseudo Labelling
[42]	38	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	2D CT	2414 slices	Reconstruct Input Data	Pseudo Labelling
[43]	7	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	3D CT	140 volumes	Reconstruct Input Data	Consistency Regularization
[44]	37	MICCAI	Segmentation	2D CT	463 slices	Reconstruct Input Data	Consistency Regularization
[45]	8	IEEE Transactions on Multimedia, IF:7.3, JCR Q1	Segmentation	2D MRI, 3D MRI	200 slices, 100 volumes	Reconstruct Input Data	Consistency Regularization
[46]	19	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Segmentation	3D MRI	2956 volumes	Reconstruct Input Data	Consistency Regularization + Generative Model
[37]	43	MICCAI	Segmentation	3D MRI, 3D CT	20 volumes, 20 volumes	Multi-Modality	Consistency Regularization
[47]	445	Medical Image Analysis, IF:10.9, JCR Q1	Classification	Multi-modality image and non-imaging measures	–	Multi-Modality	Graph-based Method
[48]	17	Medical Image Analysis, IF:10.9, JCR Q1	Classification	Multi-modality image and non-imaging measures	–	Multi-Modality	Graph-based Method
[49]	6	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Segmentation	Histology, 2D CT, US, Colonscopy	500 images, 3779 slices, 780 images, 1450 images	Multi-Modality	Graph-based Method
[50]	27	Medical Image Analysis, IF:10.9, JCR Q1	Diagnosis	3D MRI	2525 volumes	Multi-Modality	Generative Model
[51]	11	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Detection	Microscopy	334 images	Multi-Modality	Pseudo Labelling + Generative Model
[52]	33	MICCAI	Segmentation	3D MRI, 3D CT	20 volumes, 20 volumes	Multi-Modality	Consistency Regularization + Generative Model
[53]	32	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Classification	2D US	7000 images	Mixing Labelled and Unlabelled Data	Consistency Regularization
[54]	38	MICCAI	Classification	X-ray, Dermoscopy	224316 images, 10015 images	Mixing Labelled and Unlabelled Data	Consistency Regularization
[55]	27	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Classification	X-ray	43347 images	Mixing Labelled and Unlabelled Data	Consistency Regularization
[56]	69	Medical Image Analysis, IF:10.9, JCR Q1	Regression, Classification	Histology	2978 images, 100k patches	Multi-resolution Patches	Consistency Regularization
[57]	19	Medical Image Analysis, IF:10.9, JCR Q1	Generation	Histology	678 images	Multi-resolution Patches	Generative Model
[58]	8	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Segmentation	Microscopy	134 images	Multi-resolution Patches	Pseudo Labelling + Generative Model
[59]	18	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	2D US video	10530 sequences	Multi-temporal frames	Consistency Regularization
[60]	143	Medical Image Analysis, IF:10.9, JCR Q1	Classification	2D CT	2018 slices	Multi-spatial images	Generative Model
[61]	9	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Contrast Translation	3D MRI	104 subjects	Multi-spatial images	Generative Model

Note: Citations are based on Google Scholar (assessed date: 18 Oct 2023), IF: impact factor, JCR: web of science journal citation report, MICCAI: Medical Image Computing and Computer Assisted Intervention.

both achieve accurate attention masks and boost classification performance simultaneously. Moreover, the constructed slice-level results were further ensembled as fused inputs for subject-level diagnosis via the proposed aggregation network.

In [45], Shu et al. also utilized a consistency regularization method Cross-Mix Monitoring that incorporates Interpolation Consistency Training (ICT) into mean teacher structure and was complemented by a transductive monitor that acted as a bridge between student and teacher models to facilitate knowledge transfer through a projector network. A mix of unlabelled images independent of the teacher's input was fed into the student model, while the outputs of the teacher were also mixed to provide consistency feature guidance for the student. The proposed framework is shown in Fig. 2.

Xu et al. segmented hepatic vessels for CT slices from datasets with different annotation qualities [44]. Auxiliary input information was introduced from the vascular probability maps to avoid the model being overly sensitive to high-intensity regions. The proposed consistency regularization-based self-denoising process used class probability as the confidence to smoothly refine the noisy label and update the low-quality annotation in mixed learning.

In [46], Huang et al. proposed a hybrid method with consistency regularization and generative model for reinforcing consistency learning by implementing a mean teacher basis scheme where one branch is based on various perturbed images and another agent branch is used for reconstructing the image from the perturbed counterpart with an autoencoder. The extracted neuron structure features were again fed into the segmentation network to reinforce the capabilities of consistency learning.

Mix labelled and unlabelled data Nguyen et al. utilized proposed a consistency regularization method ICT in a Siamese network-based model for grading the severity of osteoarthritis of the knee from radiographs [55]. The labelled and unlabelled data were first processed linearly along the Mixup rays by Interpolation Consistency Training (ICT), and then the fused data were randomly perturbed by multiple transformations to form the in- and out-of manifold regularization.

The Mixup procedure could also be applied to latent space, Gyawali et al. presented a consistency regularization method named latent mixing technique for medical image classification [54]. For each layer, the mapping function could be decomposed as $f_l(x) = d_l(e_l(x))$, where e_l encodes the input into latent representation and d_l acts as the decoder

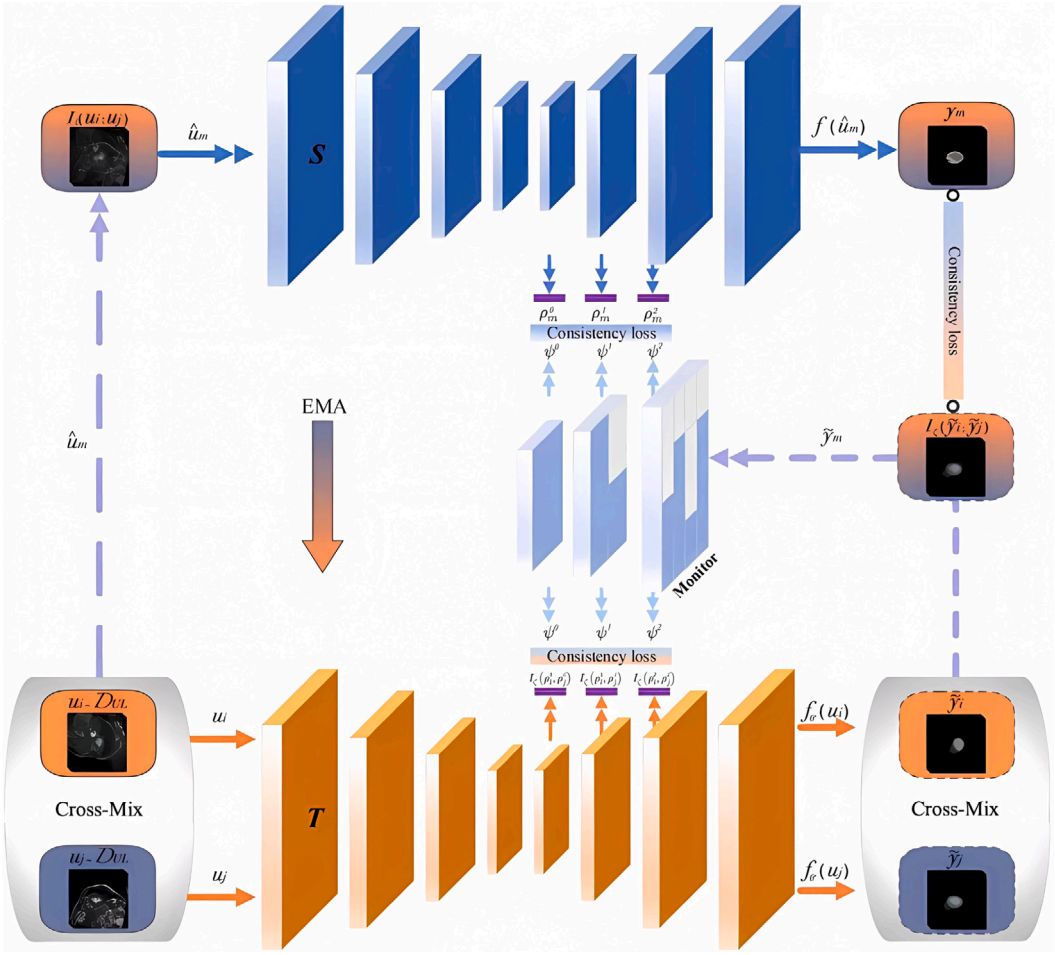

Fig. 2. Cross-Mix Monitoring framework with mixed labelled and unlabelled data for medical image segmentation.

part. The authors implemented MixMatch by packing each input with the encoding function of the guessed labels on the latent space and comparing it with the decoded predictions of the interpolated data samples.

While Meng et al. attempted to classify unseen categories in different domains by applying consistency regularization method through Mixmatch, in which labelled and unlabelled data are mixed in a linear matrix [53]. By extracting features that do not intersect in the latent space through two independent encoders, the mutual information between the two features was minimized to encourage domain-invariant representation learning. Feature clustering constraints on the categorical information were used in order to keep the categorical features consistent between the source and target domain.

Multi-modality data Li et al. leveraged a consistency regularization method called Dual-Teacher to mitigate the need for tedious medical annotations by simultaneously using abundant unlabelled data and widely available cross-modality data [37]. Besides on traditional mean teacher for the intra-domain knowledge transfer, the additional inter-domain teacher instructed with knowledge beneath labelled sources to narrow the modality gap through knowledge distillation. Principally, the inputs for inter-domain teachers were synthesized by a generation model from source samples, and the student model was encouraged to produce consistent outputs as inter-domain teachers, which obtained reliable source domain knowledge.

Wang et al. [50] synthesized high-quality multi-parameter magnetic resonance imaging (mp-MRI) of Clinically Significantly (CS) PCa via generative model GANs with a Stitch layer, without requiring paired Apparent Diffusion Coefficient T2-weighted (ADC-T2 W) images. The authors also combined the prior knowledge of disease features to maximize an auxiliary distance between CS and NonCS images. The synthesizer learns both the marginal distributions of real multi-modal image pairs and the distinguishable visual features of multi-class images.

Graphs are widely used as a natural framework for capturing the interactions between elements represented by the nodes in the graph. This representation allows for the simultaneous inclusion of a large amount of imaging and non-imaging information as well as individual subject characteristics in the disease classification task [47]. The proposed graph convolutional networks integrated imaging and non-imaging data by associating nodes with imaging-based feature vectors while using phenotypic information as edge weights. Concretely, for population graph construction, the authors made careful decisions on both nodes, feature vectors, and edges. Non-imaging phenotypic measures integrated with similarities between imaging data compose the edge weights. A recursive feature elimination strategy using a ridge classifier, a simple autoencoder, a multilayer perceptron and principal component analysis was implemented for imaging feature selection.

In [48], Huang and Chung also took both imaging and non-imaging data into account, a Pairwise Association Encoder was proposed to normalize and rescale the non-imaging features, which consisted of two parallel projection networks with sharing weights and a cosine similarity function to compute hidden features as edge weights.

Zou et al. designed a Graph Flow framework, especially for resource-limited and annotation-limited medical application scenes that utilized knowledge distillation [49]. To exploit the semantic knowledge in deep neural networks, salience graphs were created by extracting and encoding the maximum activation in salience regions. In addition, variation graphs were used to measure the flow of salience graphs

between layers, and knowledge was distilled from the teacher model to the student model.

Xing et al. leveraged the pseudo labelling and generative methods to address cross-modal domain adaptation of microscopy images [51]. The authors first applied the source domain detector to the transformed target training image to implement an effective self-labelling algorithm. A bidirectional mapping between two domain data was used to pair the real label with synthesized source data, and also to pair the real target data with the pseudo labels selected using the self-training algorithm.

Zhao et al. proposed a hybrid method with consistency regularization and generative model that embedded cycle GANs into a mean teacher architecture to apply cross-modality image segmentation tasks [52]. Two teacher models with intra-domain semantic and inter-domain anatomical structural information simultaneously transferred their respective knowledge to the student model. Generating reconstructed images is also a common skill used in conjunction with consistency constraints. The authors successfully constructed two newly augmented data: source-like domain and target-like domain.

Multi-resolution data A self-supervised driven consistency regularization method was proposed by Srinidhi et al. [56], which consisted of a pre-trained self-supervised network for task-agnostic feature representation learning and was finetuned with limited labelled data for task-specific downstream features. The authors exploited the multi-resolution contextual information present in the pyramidal nature of histological whole-slide images to provide alternative supervised signals for representation learning. For unlabelled data input, consistency loss was calculated by predictions from the teacher model with a slight augmentation and the student model with a strong augmentation. Specifically, the teacher model had all layers frozen, while the student updated its last layers and replaced the original teacher at the end of each epoch.

In [57], Li et al. generated high-resolution histopathology images from multi-resolution conditional patch with generative model GANs and selected sufficiently realistic image-label pairs for training. The proposed pyramid of GAN structures shows how each is each responsible for generating and segmenting images at different scales.

Lou et al. performed a hybrid method with pseudo labelling and generative model that self-training and conditional GAN combined paradigm in a label-efficient cell nucleus segmentation framework, allowing the generation of synthesized images from different augmented masks and simultaneously constructing a label-efficient that only selected image patch labels are needed [58]. Multi-scale conditional generators and component-wise discriminators were used to process real and synthetic masks at randomly different scales.

Multi-temporal/spatial data In [59], Wu et al. processed a consistency regularization approach to segment the left ventricle of echocardiography videos, which particularly suffers from irregular and anisotropic cardiac motions and the existence of speckle noise. The proposed method exploited temporal and spatial information by extracting temporal context-aware features from k different encoders, where the relative temporal positions of neighbouring frames could lead to unique weights for each encoder on each layer, thus constituting the feature maps. Besides, the convolutions-based semantic calibration method would formulate a wider bandwidth signal for cardiac motion, significantly reducing the effect of speckle noise by adaptively fusing the bi-directional spatiotemporal semantics of adjacent frames.

In [61], Yurt et al. proposed a generative model named ssGAN to improve learning-based multi-contrast MRI synthesis feasibility, which can synthesize target images directly from under-sampled multi-coil source acquisitions without intermediate reconstruction.

Xie et al. implemented a generative model-based semi-supervised adversarial autoencoder model for lung nodule classification by extracting multi-view features and constructing knowledge-based collaborative learning models [60]. Multiple views of the overall appearance (OA), voxel value heterogeneity (HVV), and shape heterogeneity (HS) image patches of each nodule were extracted as inputs to the proposed model. Specifically, a learnable layer transferred the representation ability of the reconstruction network to the classification network for better decision-making.

3.2. Feature fusion applications

In this section, we provide an overview of semi-supervised applications in medical image analysis that employ feature fusion techniques. For comprehensive details on these applications, please refer to Table 4.

Multi-subnetworks In [62], Liu et al. assessed paediatric diffusion MRI quality by going from fine to coarse, applying a pseudo labelling method self-training based on hierarchical non-local residual networks, separately from slice to volume to subject, and applying high-confidence pseudo labels layer by layer with a threshold. Another study by the same team described image quality assessment on T1- and T2- paediatric MRI, with self-training techniques applied to both residual networks and random forests [63]. The method consists of two subnetworks: a non-local network for weighted summation of features at local and global locations; and a residual network to avoid the gradient explosion problem. Therefore, the framework is holistic and can effectively combine local and global features.

Zhang et al. took discriminative errors into consideration for the colon gland segmentation task with a pseudo labelling-based dual error-correlation framework [65]. Since they attributed segmentation errors to inter-class similarity and intra-class inconsistency, in addition to having a main segmentation branch, the proposed Discriminative Error Prediction network (DEP-network) has two other auxiliary branches, Bmul-Net and Eintra-Net, which provide inter-class and intra-class error information respectively. These three networks share features in the encoder phase, and in the decoder phase, the global segmentation branch received features from the subbranches in a cross-model concatenating manner. Each network was self-trained by softmax probabilistic refinement of the pseudo labels.

In [67], Zhou et al. proposed a weakly and strongly augmented alignment process guided by the Instance to Prototype Earth Mover Distance (I2PEMD) model in a classification task using a joint prototype tilt with consistent regularization and an unpaired multi-modal segmentation task. By introducing momentum updates, EMD was computed for estimating cross-class prototype relationships. The model was able to use prototype learning and EMD estimation to better align feature distributions, which can mitigate distribution bias between multi-source inputs (e.g., CT and MRI) to help guide prediction selection and maintain label quality.

A generative model named the SD-LayerNet proposed by Fazekas et al. also included VAE components in retinal layer segmentation with disentangled representation [68]. A fully differentiable topological engine was additionally used to convert the surface probability maps to anatomical factors, and the VAE aimed to infer style factors.

In [69], Chartsias et al. utilized a generative model and demonstrated that disentangled representation had considerable potential for medical images, as they proposed the SDNet, which decomposed images into spatial anatomical factors and non-spatial modality factors, via an anatomical encoder and a VAE respectively. These high-level representations were then validated on several tasks such as multi-modality segmentation, multi-task segmentation and regression, and image-to-image synthesis.

Wang et al. [38,66] used pseudo labelling-based method, achieved glaucoma classification in OCT images, a subtask investigating visual archival measures was performed in the regression module. They argued that the relationship between structural and functional changes could positively influence the main classification task and the weighted combination of losses was used for joint training. Pseudo labels were assigned by the similarity between homogeneous images (i.e., Euclidean, cosine and Manhattan distance).

Multiple hidden layer features The method DSAL proposed by Zhao et al. also used a dense condition random field (CRF) to generate pseudo labels and additionally applies active learning to select samples

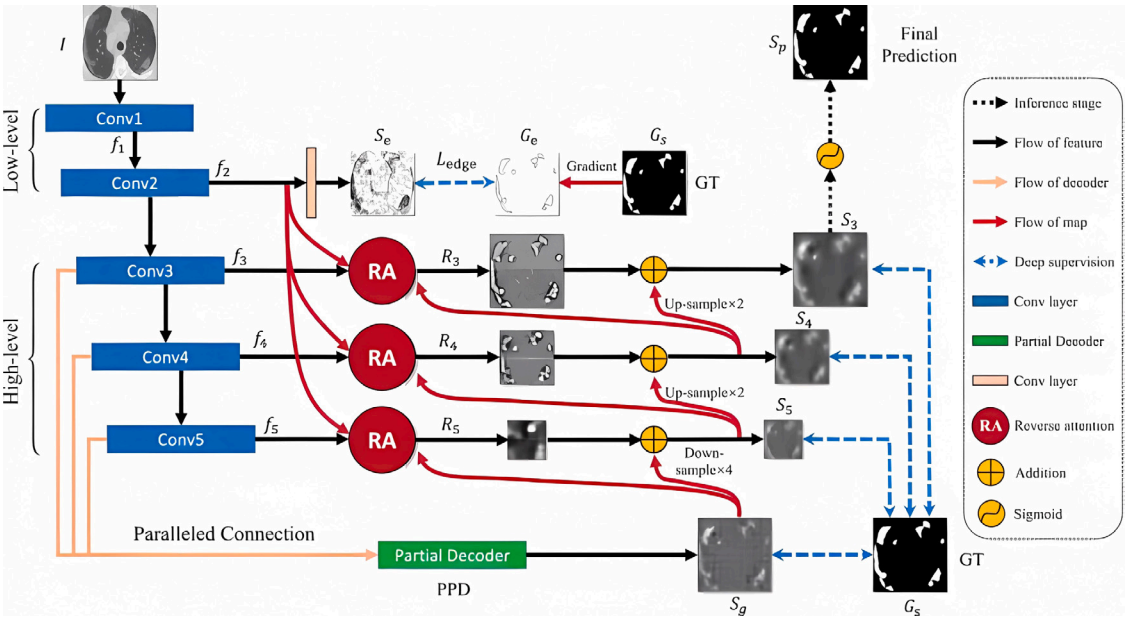

Fig. 3. Lung Infection Segmentation Network (Inf-Net) model with multiple feature fusion modules [64].

Table 4
Summary of medical image analysis applications leveraging feature fusion method.

Publication	Citations	Venue	Task	Modality	Data size	Data Fusion Methods	Semi-supervised Method
[63]	10	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Image quality assessment	3D dMRI	3775L + 11778U	Multiple sub-networks	Pseudo Labelling
[62]	14	IEEE Transactions on Image Processing, IF:10.6, JCR Q1	Image quality assessment	3D MRI	200L+814U	Multiple sub-networks	Pseudo Labelling
[65]	12	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	Histology	378 images	Multiple sub-networks	Pseudo Labelling
[38]	8	MICCAI	Classification	3D OCT	4877 volumes	Multiple sub-networks	Pseudo Labelling
[66]	51	Medical Image Analysis, IF:10.9, JCR Q1	Classification	3D OCT	6108 volumes	Multiple sub-networks	Pseudo Labelling
[67]	1	Medical Image Analysis, IF:10.9, JCR Q1	Classification	Dermoscopy	10015 images	Multiple sub-networks	Consistency Regularization
[68]	2	MICCAI	Segmentation	3D OCT	68L + 391 U	Multiple sub-networks	Generative Model
[69]	142	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation, Generation, Regression	2D MRI, CT	approx 38000, 2580 images	Multiple sub-networks	Generative Model
[70]	41	IEEE Journal of Biomedical and Health Informatics, IF:7.7, JCR Q1	Segmentation	Dermoscopy, X-ray	2000 images, 12611 images	Multiple layer features	Pseudo Labelling
[64]	956	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Segmentation	2D CT	100L+1600U	Multiple layer features	Pseudo Labelling
[71]	17	NeuroImage, IF:5.7, JCR Q1	Segmentation	3D MRI	416 volumes	Multiple layer features	Consistency Regularization
[72]	4	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Segmentation	NIR, Fundus	3600 images, 133 images	Multiple layer features	Generative Model
[73]	40	IEEE Transactions on Medical Imaging, IF:10.6, JCR Q1	Quality improving	OCT	512L + 1408U	Multiple layer features	Generative Model
[74]	50	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	3D CT	196 volumes	Multiple layer features	Consistency Regularization + Generative Model
[75]	59	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	3D MRI, 3D CT	593 volumes, 82 volumes	Multiple layer features	Pseudo Labelling + Consistency Regularization

Note: Citations are based on Google Scholar (assessed date: 18 Oct 2023), IF: impact factor, JCR: web of science journal citation report, MICCAI: Medical Image Computing and Computer Assisted Intervention.

for labelling [70]. The uncertainty score was calculated from the dice coefficients of the prediction masks of the various hidden layers, which were considered to reflect the accuracy and consistency of the prediction. In query selection, the unlabelled data were ranked according to their respective uncertainty, with the least reliable samples receiving strong annotations from experts, and those that were reliable receiving weak annotations via the denseCRF algorithm.

In [64], Fan et al. trained a pseudo labelling-based model named Inf-Net to segment multi-class infection regions from Covid-19 CT scans, which included a parallel partial decoder connected to a reverse attention module as well as an edge attention module for enhanced boundary and feature representation. Each low-level feature obtained was fed to a further convolutional layer to extract high-level features, while multiple inverse attention modules were organized in a cascade, relying on higher-level outputs. The applied semi-supervised approach was based on randomly selected propagation and the generated pseudo labels were used to guide a multi-class labelling framework, which was able to reflect quantitative information of different types of infections, e.g., ground-glass opacity (GGO) and consolidation. The architecture of the proposed model is shown in Fig. 3.

Table 5
Summary of medical image analysis applications leveraging decision fusion.

Publication	Citations	Venue	Task	Modality	Data size	Data Fusion Methods	Semi-supervised Method
[76]	38	MICCAI	Segmentation	3D CT, 3D MRI	210 volumes, 385 volumes	Multi-view branches	Pseudo Labelling
[77]	4	IEEE Transactions on Pattern Analysis and Machine Intelligence, IF:23.6, JCR Q1	Segmentation	2D MRI, 3D MRI, Fundus	200 slices, 60 volumes, 800 images	Multi-view branches	Pseudo Labelling
[39]	161	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	3D CT	82 volumes	Multi-view branches	Pseudo Labelling
[78]	16	Computer Methods and Programs in Biomedicine, IF:6.1, JCR Q1	Segmentation	3D MRI	200 volumes	Multi-view branches	Consistency Regularization
[79]	30	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	3D MRI, CT	53L+460U, 20L+55U	Multi-view branches	Graph-based Method
[80]	19	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	2D CT-PET, 2D MRI	5788 slices, 13850 slices	Multi-view branches	Pseudo Labelling +Consistency Regularization
[81]	3	Artificial Intelligence in Medicine, IF:7.5, JCR Q1	Segmentation	2D CT	91L+913U	Multi-tasks	Pseudo Labelling
[82]	56	Medical Image Analysis, IF:10.9, JCR Q1	Segmentation	3D MRI	150 volumes	Multi-tasks	Consistency Regularization
[83]	209	MICCAI	Segmentation	3D MRI	100 volumes	Multi-tasks	Generative Model
[84]	9	MICCAI	Segmentation	3D MRI	120 volumes	Multiple constraints	Consistency Regularization
[85]	16	Artificial Intelligence in Medicine, IF:7.5, JCR Q1	Segmentation	3D MRI	435 volumes	Multiple constraints	Consistency Regularization
[86]	12	MICCAI	Segmentation	2D CT, 3D MRI	82 slices, 100 volumes	Multiple constraints	Pseudo Labelling +Consistency Regularization

Note: Citations are based on Google Scholar (assessed date: 18 Oct 2023), IF: impact factor, JCR: web of science journal citation report, MICCAI: Medical Image Computing and Computer Assisted Intervention.

In [71], Chen et al. utilized consistency regularization and proposed a novel framework for brain lesion segmentation, which considers the multi-scale consistency based on hierarchical features that are extracted and concatenated across multiple hidden layers. A shape-aware discriminator was embedded using the predicted Signed Distance Maps to force improving segmentation quality from both labelled and unlabelled images.

Shen et al. proposed a generative model and demonstrated SCANet, a tri-branch semi-supervised semantic segmentation network that relied on a recurrent neural network, consistency decoder, and adversarial learning [72]. The multi-scale recurrent networks embedded a pyramid structure from coarse to fine, gradually incorporating semantic information into feature fusion. The generator aimed to provide segmentation predictions, the encoder encouraged synthetic images from the segment's predictions to be consistent with the original input image, and the discriminator was used to identify predictions and ground truth (GT). Generative adversarial models have their potential advantages in disease detection and diagnosis. Multiparametric MRI (mp-MRI) based on combining anatomical T2 W with diffusion-weighted imaging (DWI) and its derived apparent diffusion coefficient (ADC) maps can substantially improve the diagnostic power of prostate cancer (PCa) detection and aggressiveness assessment.

A combination of feature extraction and image reconstruction was applied by He et al. [74], which aimed to utilize the deep priori anatomy and hard region adaptation information for fine renal artery segmentation. The authors utilized a hybrid method with consistency regularization and generative model. The applied dense connection fused multi-receptive field and multi-scale features to adapt the network to intra-scale changes.

For example, Luo et al. presented a hybrid method with consistency regularization and pseudo labelling using uncertainty rectified pyramid consistency, which encouraged latent predictions at various scales to be similar to their average prediction [75]. Only latent predictions that had passed the uncertainty map filter were used as pseudo labels.

A generative model named capsule conditional generative adversarial network (Caps-cGAN) was proposed by Wang et al. for speckle noise denoising of retinal OCT images [73]. A capsule deconvolution module was introduced to effectively fuse the high-resolution and low-semantic features in the lower layers of the network with the high-semantic and low-resolution features in the upper layers.

3.3. Decision fusion applications

This section presents a concise summary of semi-supervised applications in medical image analysis that incorporate decision fusion techniques. Extensive information about these applications is detailed in Table 5.

Multi-view branches Fang and Li utilized pseudo labelling-based method and proposed a DMNet for semantic segmentation, which employs two different decoders that cross-supervised each other [76]. The two decoders used different architectures to introduce diversity and the loss function aimed at minimizing the difference between the two generated masks. Under the assumption of semi-supervised learning, low-entropy masks were also generated to force decision boundaries located in low-density regions.

A multi-organ segmentation framework uncertainty-aware multi-view co-training was developed by Xia et al. which produced data-level view differences by transforming 3D data [39]. Co-trained pseudo labels were computed by fusing the outputs of different views and their respective weights, which were related to the epistemic uncertainty measured by the random dropout configurations in the Bayesian network.

In [78], Xiao et al. utilized consistency regularization and built a dual-teacher module by adding a Transformer in segmentation tasks. Teacher A worked in a UA-MT paradigm while Teacher B and the student models collaborated in the co-training scheme, they created pseudo labels for each other and performed cross-supervision. Another layer of consistency regularization constraint between dual teachers was used to enhance guidance consistency and information interaction. Teacher B's system consisted of a combined CNN and Transformer structure, where the CNN-based student excelled at convolutional computations on local regions, while the Transformer-based teacher had access to remote self-attention on global information.

Ganaye et al. proposed a graph-based semi-supervised method that leveraged adjacent graphs to mitigate segmentation inconsistency under the assumption that all subjects would have the same anatomical adjacencies and inter-region connectivity [79]. The adjacency relationship was represented as a differentiable metric, with an auxiliary loss

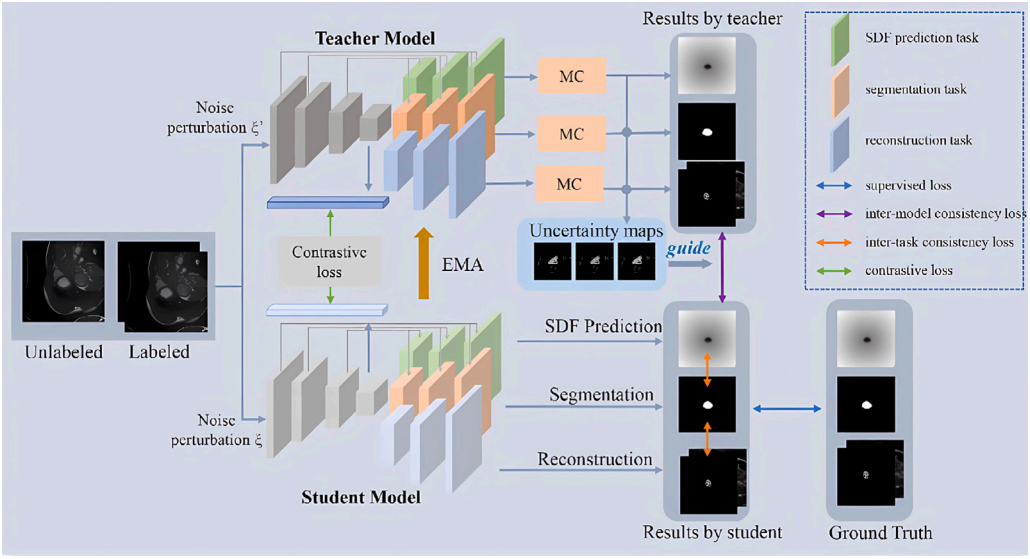

Fig. 4. A triple-uncertainty guided mean teacher model with fused decision of multiple branches [82].

penalty for outputs from regions containing anatomically incorrect adjacencies. The proposed 2.5D network with 6 directional constraints incorporates inference from multiple slicing architectures and generated correspondence maps.

Wu and Zhuang presented a risk estimator for unlabelled data and built a risk-based framework MERU for pseudo labelling-based semi-supervised medical image segmentation, where the prediction risks relied on binary classification [77]. The modified U-net outputs multi-scale predictions, which were integrated to generate the final segmentation with several convolution layers. Pseudo labels generated from the model were used to calculate the respective class portions, which were further used to form unbiased and consistent empirical risk estimates, by assuming all images are from the same distribution. The estimated risk on unlabelled images was partially incorporated into the regularization terms, i.e., positive risks on positive pixels of each class.

Multiple auxiliary tasks In [81], Hao et al. embedded feature importance weighting within the feature maps along with the pseudo label quality into a self-attention module, and iteratively updated the teacher model, which took the initial label data and the pseudo labels that had passed the label grader as training data. An automatic label grader was proposed for the quality control of pseudo labels.

In [82], Wang et al. proposed a consistency regularization-based method with a triple-uncertainty guided network for MRI segmentations. The proposed multi-task framework combined three separate tasks (segmentation, reconstruction, and Signed Distance Field prediction), for each of which the authors performed MC-dropout uncertainty estimation. Auxiliary tasks' results help the segmentation network capture more semantic information and constrain the global geometric shape of the outputs. A contrastive learning-based constraint was also equipped to help the encoder extract more diverse representations to facilitate the performance of the dominant task segmentation. An overview of the proposed model is shown in Fig. 4.

Li et al. utilized generative model and built a multi-task deep network for 3D semantic segmentation with a subtask of predicting signed distance maps using a shared backbone module [83]. The adversarial loss was added for learning shape-aware representations and naturally emphasizing the internal regions of each category, which were further seen as proxies for confidence measurement.

Zhang et al. used hybrid method with consistency regularization and pseudo labelling and exploited the intrinsic correlation of multimodal data, combined with comparative mutual learning for semi-supervised learning, to segment multiple modalities simultaneously [80]. Specifically, two modalities were fed into two independent models in a co-training manner and regularized by the correlations and differences between the modalities. In addition, a novel soft-pseudo label re-learning scheme was implemented via a mean-teacher-like module, which narrowed the gap between the multimodal performances.

Multiple constraints on outputs The work by Xiang et al. fused two sources of uncertainty — aleatoric uncertainty to guide supervised learning and epistemic uncertainty to determine certain predictions as pseudo labels and uncertain data under consistency constraints [86].

In [84], Basak et al. utilized consistency regularization and incorporated fuzzy adaption fusion using the Gompertz function to combine the three proposed class-wise confidence (entropy, variance, averaged probability) into individual confidence scores for each sample. The robust class-wise sampling was embedded into the Mean Teacher structure, with a dynamic modulation of weights for better training stabilization.

Zhang et al. [85] not only incorporated uncertainty estimation into a Monte-Carlo dropout consistency regularization framework but also constructed a two-branch network for multiple tasks performed simultaneously. The final loss function was fused by both within-task consistency and cross-task consistency regularization strategies.

4. Case study

To further illustrate the practical applications of the various deep semi-supervised fusion methodologies explored in this survey, this section presents two comprehensive case studies. Each case study is conducted on a different public dataset with separate statistical analysis and experimental exploration. These case studies provide detailed examinations of these approaches in real-world scenarios.

As depicted in Fig. 8, the majority, exceeding 60%, of the surveyed papers concentrate on segmentation tasks. Consequently, this section delves into two specific medical image segmentation tasks, encompassing the two most frequently utilized modalities: MRI and CT. The first statistical case study provides a thorough analysis of the results obtained by the various methods applied to a specific MRI dataset ACDC [87], for a comprehensive overview of the advantages and disadvantages of each method. On the contrary, the second experimental case study delves into a different classical experimental approach on another dataset 'COVID-19 CT Segmentation Dataset' [88], examining the performance of each method by controlling the labelled ratio and providing an in-depth analysis of the results obtained.

Table 6
Dice scores comparison for different semi-supervised information fusion methods on the ACDC dataset.

Studies	Semi-supervised information fusion method	Labelled data proportion	Dice score	Code available
[45]	Consistency Regularization	10%	0.872	–
	Data Fusion	5%	0.867	
[69]	Generative Models	13%	0.808	here
	Feature Fusion	6%	0.789	
		3%	0.778	
		2%	0.731	
[77]	Pseudo Labelling	6%	0.839	here
	Decision Fusion			
[82]	Consistency Regularization	27%	0.914	–
	Decision Fusion	13%	0.875	
		6%	0.864	
[84]	Consistency Regularization	10%	0.889	–
	Decision Fusion	3%	0.842	
		1%	0.746	

Table 7
Comparison of Dice scores for various semi-supervised information fusion methods applied to the ACDC dataset, categorized according to the proportion of labelled data.

Labelled data proportion	Top 3 results and labelled ratio	Respective study
0%–5%	0.867 (5%)	[45]
	0.842 (3%)	[84]
	0.778 (3%)	[69]
6%–10%	0.889 (10%)	[84]
	0.872 (10%)	[45]
	0.864 (6%)	[82]
>10%	0.914 (27%)	[82]
	0.875 (13%)	[82]
	0.808 (13%)	[69]
100%	0.9302	[89]
	0.9232	[90]
	0.9206	[91]

4.1. Case I: Segmentation for cardiac diagnosis

Dataset One of the extensively used MRI datasets for segmentation tasks within the medical domain is the ACDC (Automated Cardiac Diagnosis Challenge) dataset [87], which consists of short-axis MR-cine T1 3D volumes of cardiac anatomy acquired using 1.5T and 3T scanners of 100 subjects. In this survey, a total of 5 papers use the ACDC dataset for training and testing their semi-supervised learning methods. In this context, a specific case study centred around the ACDC dataset is presented to elucidate the solutions and experimental results covered in this systematic review on semi-supervised learning in medical image analysis.

Quantitative results The presented Table 6 illustrates the findings from five different studies employing various deep semi-supervised fusion methods, each using different proportions of labelled data for training and assessing performance primarily through Dice Scores. Notably, Consistency Regularization consistently demonstrates efficacy across three studies [45,82,84], where Wang et al. [82] showcased a remarkable Dice Score, particularly at a substantial 27% labelled data proportion. Generative models, as investigated in Chartsias's work [69], and Pseudo Labelling, employed by Wu and Xia [77], also exhibit competitive performance. From the part of the fusion method, Decision Fusion is applied in three papers and emerges as a robust strategy with exceptional performance. Within the body of literature, certain studies documented experimental results for multiple labelled datasets, while a few listed only the appropriate proportion of labelled and unlabelled data. Noteworthy, when evaluating results based on the Dice score, the most remarkable performances arise from Wang's [82] and Basak's [84] models with scores of 91.4 and 88.9, respectively, both of which incorporate decision fusion with consistency regularization, hinting at the potential of amalgamating diverse deep learning techniques with semi-supervised learning paradigms to amplify network segmentation proficiency. Wang [82] fused the final decision of multiple branches to balance the specific knowledge captured by different subnetworks, while Basak et al. [84] employed a fuzzy confidence fusion combining three class-wise performance indicators to generate the final score. Additionally, Wang et al. [82] introduced contrastive learning, supported by an ablation study revealing a 1.8%, 0.4%, and 0.1% enhancement in Dice performance when integrating contrast loss with 5, 10, and 20 labelled data instances out of a total of 75 images.

This section categorizes all experimental data in proportion to the labelled data, compares and lists the top three models in each group using dice scores, and additionally describes the control group of the state-of-the-art (SOTA) methods using the fully labelled dataset, which is listed in the last three rows of the Table 7. From the aspect of labelled data proportion, the four categories here are assigned to cover the entire distribution:0%–5%, 6%–10%,> 10% and 100% fully labelled. It is evident that the segmentation results are directly related to the proportion of labelled data. Thus a clear trend emerges: an increase in the amount of labelled data is synonymous with higher performance metrics. For cohorts comprising no more than 5% labelled data, the dice scores consistently fall below the threshold of 0.87. However, as the labelled data percentage ascends to 6% and beyond, the dice scores demonstrate an upward trajectory, consistently surpassing the 0.86 benchmarks. Notably, in scenarios with over 10% labelled data, Wang et al. [82] outperforms with a Dice score of 0.914 at a 27% labelled ratio, which is significantly closer to the results of the SOTA method derived using fully labelled training data, which achieves a score of 0.9302. Interestingly, Wang et al. [82]) also performs well in the 6%–

Table 8
Segmentation results with different labelled proportion by four typical semi-supervised information fusion methods on COVID-19 CT dataset [88].

Semi-supervised method	Detailed fusion application	Labelled data proportion	Dice score	Sensitivity	Specificity
1 Pseudo labelling	Self-training + feature fusion	10%	0.6598	0.8119	0.9312
		20%	0.6744	0.6324	0.9726
		100%	0.7476	0.7526	0.9723
2 Consistency regularization	Mean teacher + data fusion	10%	0.6645	0.6881	0.9021
		20%	0.7328	0.6933	0.9601
		100%	0.7412	0.8212	0.8919
3 Graph-based method	GCN + decision fusion	10%	0.6796	0.6664	0.9789
		20%	0.6822	0.6635	0.9117
		100%	0.7340	0.7481	0.9348
4 Generative model	GANs + feature fusion	10%	0.6913	0.6725	0.9355
		20%	0.7312	0.8124	0.9237
		100%	0.7814	0.7676	0.8979

10% range, emphasizing the adaptability of the method across different labelled data scenarios. While Shu et al. [45] and Chartsias et al. [69] exhibit competitive performance, particularly within the lower labelled proportion groups, showcasing Dice scores of 0.867 and 0.889, in scenarios with 0%–5% labelled and 6%–10% labelled data respectively. This finding highlights the transformative potential of semi-supervised learning, where judiciously combining labelled and unlabelled data is an effective strategy for circumventing data scarcity and improving segmentation accuracy. By demonstrating that semi-supervised learning can produce results comparable to fully supervised learning under resource-constrained conditions, this analysis emphasizes the utility of semi-supervised learning and illuminates its value as an indispensable tool for advancing medical image analysis research.

4.2. Case II: Segmentation for COVID-19

Dataset The COVID-19 CT segmentation dataset [88] is a public lung CT dataset, which comprises 100 axial CT images sourced from over 40 patients diagnosed with COVID-19. These radiological images were gathered and processed by the Italian Society of Medical and Interventional Radiology. A radiologist manually segmented these images, establishing the ground truth used in the training process.

Implementation details The above case study has entailed a comparative analysis of five different methodologies applied to the same MRI dataset. However, these methodologies were tested with varying proportions of labelled data and within diverse experimental settings. In pursuit of a more equitable and comparable case study, this section implement four classical semi-supervised fusion methods that ensure identical labelled data distribution and precisely replicated experimental conditions. Specifically, the experiments were implemented on an RTX 3080 GPU with PyTorch.

The preprocessing phase followed the method outlined by Fan et al. [64] and consisted of uniformly resizing the images to 352×352, with on-the-fly randomized scale augmentations in subsequent training. Out of 100 CT images, 80 were used for the training and validation phase and the remaining 20 for testing purposes. Each experiment implements a different deep semi-supervised fusion method at three labelled data scales: 10%, 20% and 100%. This approach was adopted to fully evaluate the segmentation performance and to assess the efficacy of utilizing unlabelled data.

Quantitative results The four methods included in this part are mainly classifies by semi-supervised learning methods: Pseudo labelling, Consistency regularization, Graph-based method and Generative model. Three different fusion techniques were also considered in the implementation. Among them, the self-training scheme referenced the feature fusion approach of Fan et al. [64], which enhanced the boundary and feature representations using additional attention modules. Mean teacher, a typical semi-supervised learning method, has been applied as a benchmark to various consistency regularization models [43–45], and combined with data fusion by reconstructing the input data. For instance, the experiment used probability maps as auxiliary input information [44] which were extended to the normal mean teacher model. While Graph Neural Networks (GNN) could further leverage adjacent graph information for decision fusion [79] using an auxiliary loss penalty for anatomical outputs. In terms of generative models, the experiment reproduced the disentangled feature representation of a multi-subnetwork model structure based on Generative Adversarial Networks (GAN), demonstrated by Chartsias [69]. The Table 8 below lists the segmentation results produced by four different methods with labelled data proportions and compares them by dice score, sensitivity, and specificity.

The Table 8 presents segmentation outcomes derived from four distinct semi-supervised fusion methods, each evaluated across three varying labelled data proportions (10%, 20% and 100%). The pseudo labelling, employing self-training with feature fusion, demonstrated a noticeable improvement in performance with an increase in the percentage of labelled data. While consistency regularization utilizing a mean-teacher approach coupled with data fusion, showcased the highest efficacy at 20% labelled data, achieving a dice score of 0.7328, closely approaching the result produced by the fully labelled model (dice score of 0.7412). The graph fusion decision method, integrating GNN with decision fusion, exhibited relatively stable performance across different labelled data ratios, underscoring the robustness of its approach. The generative model utilizing GAN with feature fusion, displayed a notable enhancement in all metrics as the labelled data proportion increased. Moreover, it showed substantial potential in both lower labelled and fully labelled data scenarios, with dice scores of 0.6913 and 0.7814 respectively. Overall, these results highlight the subtle dynamic relationship between the proportion of labelled data and segmentation performance for each method. It is worth noting that some methods continue to improve as the labelled data increases, while others exhibit relative stability in performance metrics.

The average specificity values for the twelve experiments exceeded 0.9, indicating a reasonably high accuracy for the prediction of negative pixels, whereas the sensitivity values varied considerably depending on the method. This may be attributed to the inherent imbalance between background and foreground pixels because in most cases there are more background pixels than foreground pixels. Models tend to prioritize predicting the majority class (background) at the expense of the minority class (disease mask). Furthermore, the varying disease conditions across different patients, including ground glass, consolidation, and pleural effusion, contribute to the complexity.

Qualitative results Fig. 5 illustrates the four disease levels ((a) to (d)) and the segmentation results of the four semi-supervised information fusion methods with 20% labelled data. It is worth noting that when the infected region covers a large portion of the lung, more ground truth is labelled, making it easier to capture complex disease details and textures. In contrast, smaller infected regions provide less

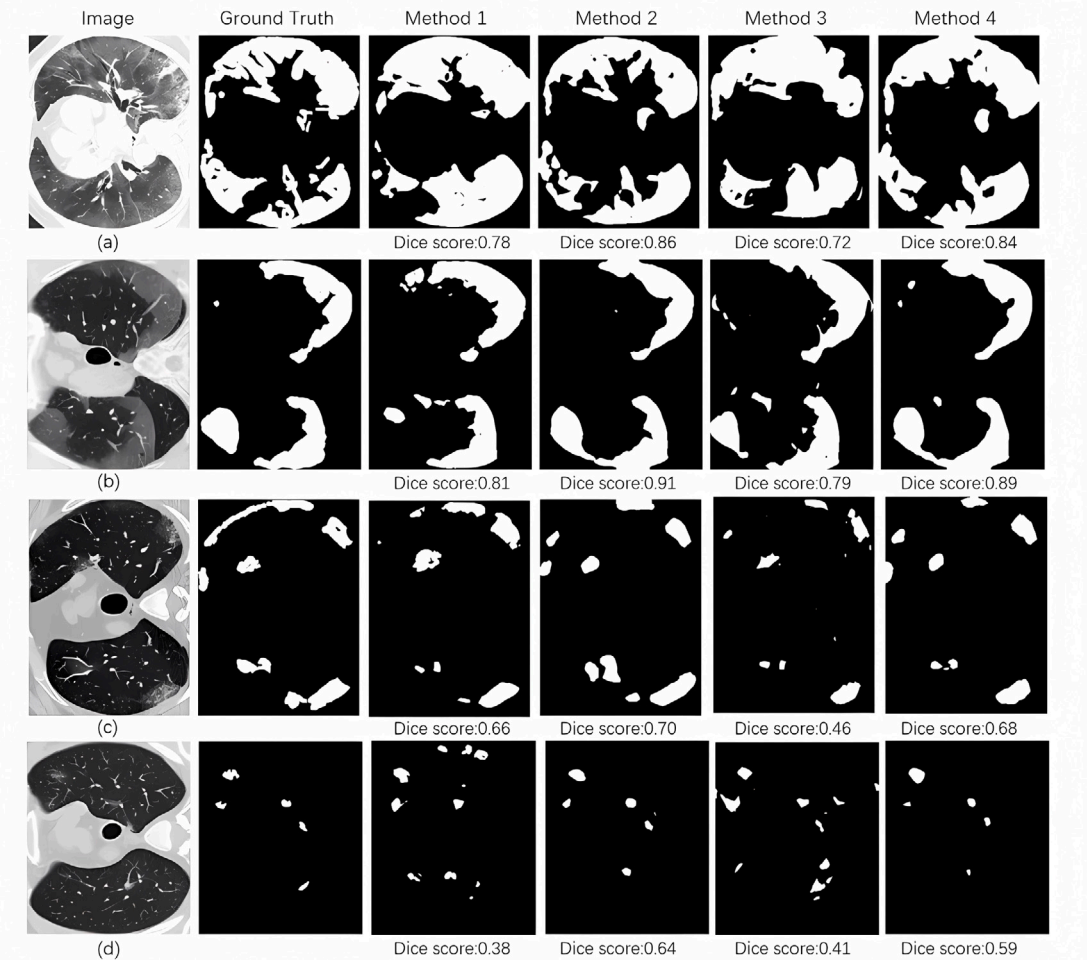

Fig. 5. Visualization of segmentation results from various methods on four COVID-19 chest CT images from a public dataset [88]. Methods include: (1) Pseudo Labelling, (2) Consistency Regularization, (3) Graph-Based Method, and (4) Generative Model. The bottom of each figure displays its Dice score.

detail in the ground truth, making the segmentation task more challenging, especially in semi-supervised training scenarios. The average Dice scores from (a) to (d) show a clear correlation with the disease regions obtained in the ground truth images.

It can be noted that although sometimes the Dice scores are similar, each method performs differently in terms of segmentation results, as shown in Fig. 5. For example, pseudo labelling with feature fusion (Method 1) outperforms at capturing the boundaries and anatomical features. Consistency regularization with data fusion (Method 2) benefits from additional possibility maps, enhancing overall segmentation accuracy. The graph-based model with decision fusion (Method 3) strikes a balance of multiple considerations, yielding smooth results. The generative method with feature fusion (Method 4) integrates auxiliary feature representations resulting in superior performance.

In general, the observed dice scores ranged from 0.6 to 0.7. While an increase in the proportion of labelled data tended to correlate with higher dice scores to some extent, this effect was not consistently significant across all methods, particularly in the case of graph-based methods. As for the consistency regularization and generative models, the fully labelled data did not improve the performance much compared to the pseudo labelling method with 20% labelled data. The relatively small size of the dataset may be one of the reasons, potentially introducing challenges of overfitting and limited representativeness, and increasing complexities in achieving significant performance improvements even with more labelled data.

In conclusion, we have evaluated the efficacy of various semi-supervised information fusion techniques in segmentation tasks by using two real-world datasets. Our findings indicate that these semi-supervised methods, while promising, are currently underperformed compared to their fully supervised counterparts. We observe a positive correlation between the volume of labelled data and the performance of the semi-supervised techniques. This suggests that there is room for improvement in these methods, particularly compared to fully supervised learning techniques. The decision fusion approach with consistency regularization has yielded the best results in the cardiac diagnosis dataset. Conversely, for the COVID-19 diagnosis dataset, the generative model with feature fusion has emerged as the most effective in terms of the Dice score.

5. Discussion on challenges and future perspectives

Semi-supervised information fusion has attracted attention to improve model performance through unlabelled data in medical image analysis. This section discusses current trends of the existing semi-supervised information fusion algorithms and various combinations of other deep learning techniques. This section also investigates the open research questions, challenges, and potential future directions. We address the previously stated research questions in details. To address RQ1, Section 3 thoroughly reviews the current research progress in semi-supervised information fusion techniques applied to medical image analysis. In Section 4, we empirically compare semi-supervised information fusion methodologies to offer insights that respond to RQ2. Section 5.2 explores the challenges that arise during the implementation of these techniques in corresponding to RQ3. In Section 5.3,

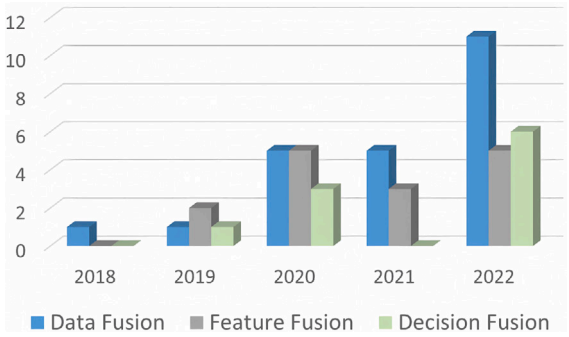

Fig. 6. A temporal overview of the distribution of research papers with different fusion techniques during the years from 2018 to 2022.

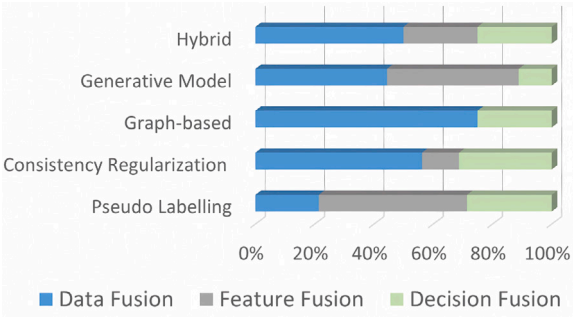

Fig. 7. Analysis of the utilization of fusion techniques in amalgamation with different semi-supervised learning methods.

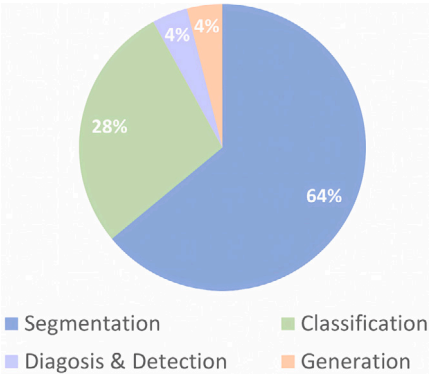

Fig. 8. A categorization of the distribution of research papers on different fusions according to various medical image analysis tasks.

we discuss potential future directions for semi-supervised information fusion techniques in medical image analysis to provide answers to RQ4.

5.1. Research progress

In our survey, we have meticulously selected 50 research articles that align with the thematic focus of our study. Notably, these articles have been published in premier journals classified under JCR Q1 or in leading conferences in the field. Upon conducting an in-depth assessment of the publishers, we have discerned that the predominant publishers are as follows: "Medical Image Analysis" with 22 articles, "IEEE Transactions on Medical Imaging" with 11 articles, and "MICCAI" accounting for 10 articles. Collectively, these eminent publishers constitute 84% of the papers we surveyed, underscoring their significant contribution to the research discourse on this topic.

To examine the overall trends in the usage of different methods, Fig. 6 shows the distribution of different research papers using each information fusion method across different years. The general trend from 2018 to 2022 demonstrates an increase in the total number of papers related to fusion techniques in the domain of semi-supervised medical image analysis and in particular a significant increase in the number of papers on data fusion. This ascending trend illustrated a growing recognition of the involvement of fusion techniques in addressing complex challenges in medical image processing and analysis, thus further highlighting the importance of fusion methods in the medical image research community.

Fig. 7 shows an insightful analysis of the utilization of fusion techniques in amalgamation with five semi-supervised learning methods, i.e., pseudo labelling, consistency regularization, graph-based method, generative model and hybrid method. In summary, pseudo labelling and regularized are particularly well represented in the fusion domain, demonstrating their adaptabilities to the diverse fusion strategies. Although graph-based and generative model-based fusion methods present relatively few examples, both categories do exhibit instances of data fusion, and generative models extensively display notable applications in the feature fusion area. The hybrid categories have a balanced distribution, showing their adaptabilities to the various fusion strategies and semi-supervised learning methods.

To explore the correlation between the different methods and various medical imaging tasks, Fig. 8 provides an overview on the distribution of research papers across different methods and task categories. Among the included 50 publications, the two most prominent tasks are segmentation and classification, with 32 and 14 papers respectively, and attract significant attention in traditional medical imaging analysis. It is also obvious that feature fusion and data fusion are versatile and frequent choices, to offer potential improvements in the various aspects of segmentation, classification and generation tasks. Meanwhile, the presence of data fusion in both diagnosis and generation categories demonstrates its adaptability. And interestingly, all of these instances include semi-supervised generative models, due to their unique creative ability that differs from other semi-supervised methods.

5.2. Current challenges

However, there are several challenges that need to be addressed for semi-supervised information fusion to reach its full potential.

Data bias and imbalance A prominent challenge is the problem of data bias and imbalance. Semi-supervised learning models may be biased towards the existing labelled data, resulting in poor generalization to new and unseen data for data fusion. This is a common problem in real-world scenarios, especially in medical images, where the valid sample of each modality is quite limited and there is even less guarantee of annotation by an expert which might cause variations and shifts in data distributions, and differences in data quality, noise and feature representations.

Data bias may arise from the heterogeneous nature of radiographic equipment, such as variations in brand and software versions, which can significantly impact data uniformity. Moreover, diverse patient cohorts from various medical centers and countries can introduce further biases. Addressing these challenges requires amalgamating multi-center data and expanding dataset sizes to meet the demands of deep learning algorithms. However, the ethical considerations, and financial implications associated with data collection and expert annotation may pose significant barriers. Given these constraints, data augmentation emerges as a pragmatic and effective strategy [92]. By synthesizing new data from the existing datasets, data augmentation can effectively alleviate bias and imbalance concerns while enhancing the model's generalizability to diverse and unseen data.

Lack of theoretical analysis The lack of a clear theoretical understanding of how semi-supervised learning and information fusion techniques work and how to optimize them effectively is another challenge. This leads to difficulties in selecting suitable models and designing effective algorithms for various tasks. All semi-supervised approaches are based on well-known assumptions made for semi-supervised scenarios, but solid theoretical support is still lacking. Quantifying the

Table 9
Distribution of other methods applied for medical image application.

Methods	Studies included
Attention mechanism	[41–43,64,81]
Contrastive learning	[80,82,85]
Knowledge distillation	[37,49]
Self-supervised Learning	[56,68]
Active Learning	[58,70]
Transfer Learning	[60]
Transformer model	[78]
Federated Learning	[40]
Few-shot Learning	[42]

uncertainty in fused data and propagating it efficiently throughout the analysis process is a challenging task as theoretical support is similarly lacking. Meanwhile, the characteristics of black-box schemes in deep learning may introduce incomprehension and mistrust into the end result, a more transparent and interpretable model is inevitably needed in medical clinical applications.

Security Medical image analysis is a safety-critical domain, and data security is always a challenge for the AI community. For semi-supervised information fusion, during multiple data modalities of medical imaging are analysed and fusion, the researchers should maintain the safety and the high security of these data and prevent reverse engineering of the sensitive medical data and any ethical issues that may arise. Federated learning stands out as a powerful paradigm, holding potential to uphold the privacy and security integrity of medical imaging data [93]. To fully benefit from this paradigm, it is important to integrate semi-supervised information fusion techniques in future developments. This will help advance the field of medical imaging by maximizing the advantages offered by federated learning.

Absence of fair comparison and external validation While numerous studies have tested their algorithms on public datasets, inconsistencies in experimental setups may compromise fair comparisons. Variations in implementation details, such as software versions and hardware differences, can also influence the fair assessment of these algorithms. Furthermore, the diversity in medical imaging modalities and targeted organs could bias the selection of SOTA semi-supervised information fusion algorithms.

Additionally, the lack of validation on external datasets and multi-center data undermines the reliability of these results in a clinical context. In this field, DL models are trained and tested on datasets from a single hospital or institution, which may inadvertently introduce the risk of information leakage and compromise the generalizability of the findings. Future studies should incorporate external datasets in their validation processes. This approach would strengthen the models and improve their applicability across diverse clinical settings. Developing a comprehensive benchmark dataset, encompassing medical images of various modalities and organs from multiple sources, would be instrumental in overcoming this challenge and facilitating a balanced comparison of different algorithms.

5.3. Future perspectives

The evidence strongly suggests that the right combination of methods can achieve better performance than applying these methods separately. By leveraging the strengths of different approaches, synergistic effects can be achieved due to the potential synergies and complementary nature of different semi-supervised information fusion approaches. Generally speaking, it is crucial to balance the effectiveness and efficiencies of the selected methods, carefully consider the compatibility and potential interactions between them and avoid any conflicting assumptions or limitations.

By observing the included studies, a number of incorporated mechanisms (excluding the mentioned semi-supervised fusion methods) have been applied for better model performance, see Table 9. Among the total of included 50 studies that work on semi-supervised information fusion in medical imaging analysis, 5 studies consider attention mechanism in their models, and 3 directly apply or fuse the contrastive learning techniques. 2 studies each use knowledge distillation, self-supervised learning and active learning. There are a few other publications that may try methods such as the Transformer model, federated learning and few-shot learning in their proposed solutions. From another point of view, it presents the generalization of semi-supervised fusion as its possible application in various perspectives.

Designing robust and scalable methods for semi-supervised fusion remains an ongoing challenge. The listed methods from Table 9 could be referred to some extent as they have shown excellent performance in the various phases of medical imaging tasks. For example, active learning techniques are valuable in semi-supervised learning scenarios, where the model can select the most informative instances for annotation to improve performance [70]. As for the attention mechanism, it can help identify the most informative and relevant parts of the input data, and can also help the model adapt to changes in distribution by dynamically adjusting the attention weights to improve model performance and make effective usage of unlabelled data. Federated learning can resolve the security challenge and make medical image data secure sharing between multi-centers feasible [40].

The interpretability of deep learning models is crucial, particularly in critical fields such as medical image analysis. To achieve this, researchers should explore Explainable Artificial Intelligence (XAI) methodologies, including Class Activation Map (CAM) and Gradient-weighted Class Activation Mapping (Grad-CAM) [94]. These advanced techniques provide a clear understanding of the model's predictions to aid in decision making and increase transparency. Moreover, insights gained from XAI methodologies can provide valuable feedback to improve the design and effectiveness of semi-supervised information fusion methods. Ultimately, this can result in more reliable and interpretable models to foster trust and acceptance of these technologies in essential medical settings.

The advancement in large-scale language models and foundation models such as "Segment Anything" [95] opens up new horizons for future research. It is important for researchers to utilize the capabilities inherent in these foundation models. Alongside the progression in theoretical understanding of semi-supervised information fusion techniques, when integrated with methodologies like foundation model, transfer learning, federated learning, and self-supervised learning, the deployment of semi-supervised information fusion in clinical settings becomes an achievable prospect.

The future perspective of semi-supervised information fusion is promising, as there is a vast amount of unlabelled data available for various applications, but labelled data is often costly to obtain. Semi-supervised learning in conjunction with information fusion has the potential to significantly reduce the need for labelled data and make full use of existing data while still achieving high accuracy.

6. Conclusion

The domain of semi-supervised information fusion represents a rapidly evolving and promising research area with significant practical implications in medical imaging. This comprehensive survey carries out an inclusive examination of semi-supervised learning methods, including pseudo labelling, consistency regularization, graph-based methods, generation models, and information fusion (data fusion, feature fusion, and decision fusion). A total of 50 publications using semi-supervised information fusion for medical image analysis are included and analysed in this survey. Furthermore, the advantages and disadvantages of each classic method and future challenges are discussed. Two case studies on different datasets are conducted to analyse and investigate the nuanced effects of different semi-supervised fusion methods on different medical image tasks by statistically and experimentally

testing. With further research and development, we sincerely believe that semi-supervised information fusion has the potential to unlock new applications and enable more efficient and effective machine learning across a wide range of domains. Particularly, this approach demonstrates potential in improving diagnostic accuracy, representing a significant advancement in the application of machine learning within the realms of medicine and healthcare.

CRediT authorship contribution statement

Ying Weng: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, review & editing, Supervision, Project administration. **Yiming Zhang:** Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft, review & editing. **Wenxin Wang:** Methodology, Data Curation, Software, Visualization, Formal analysis, Writing – original draft. **Tom Dening:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the Ningbo Major Science & Technology Project 2022Z126.

References

[1] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, J. Dean, A guide to deep learning in healthcare, Nature Med. 25 (1) (2019) 24–29.
[2] Z. Chen, Y. Cao, S. He, Y. Qiao, Development of models for classification of action between heat-clearing herbs and blood-activating stasis-resolving herbs based on theory of traditional Chinese medicine, Chinese Med. 13 (2018) 1–11.
[3] Z. Chen, M. Zhao, L. You, R. Zheng, Y. Jiang, X. Zhang, R. Qiu, Y. Sun, H. Pan, T. He, et al., Developing an artificial intelligence method for screening hepatotoxic compounds in traditional Chinese medicine and western medicine combination, Chinese Med. 17 (1) (2022) 58.
[4] Z. Chen, Y. Jiang, X. Zhang, R. Zheng, R. Qiu, Y. Sun, C. Zhao, H. Shang, Resnet18dnn: prediction approach of drug-induced liver injury by deep neural network with ResNet18, Brief. Bioinform. 23 (1) (2022).
[5] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
[6] H. Scudder, Probability of error of some adaptive pattern-recognition machines, IEEE Trans. Inform. Theory 11 (3) (1965) 363–371.
[7] S.C. Fralick, Learning to recognize patterns without a teacher, IEEE Trans. Inform. Theory 13 (1967) 57–64.
[8] G. Muhammad, F. Alshehri, F. Karray, A. El Saddik, M. Alsulaiman, T.H. Falk, A comprehensive survey on multimodal medical signals fusion for smart healthcare systems, Inf. Fusion 76 (2021) 355–375.
[9] F. Behrad, M.S. Abadeh, An overview of deep learning methods for multimodal medical data mining, Expert Syst. Appl. 200 (2022) 117006.
[10] Y. Chong, Y. Ding, Q. Yan, S. Pan, Graph-based semi-supervised learning: A review, Neurocomputing 408 (2020) 216–230.
[11] V. Cheplygina, M. de Bruijne, J.P. Pluim, Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis, Med. Image Anal. 54 (2019) 280–296.
[12] X. Yang, Z. Song, I. King, Z. Xu, A survey on deep semi-supervised learning, IEEE Trans. Knowl. Data Eng. 35 (9) (2022) 8934–8954.
[13] Z. Song, X. Yang, Z. Xu, I. King, Graph-based semi-supervised learning: A comprehensive review, IEEE Trans. Neural Netw. Learn. Syst. 34 (11) (2023) 8174–8194.
[14] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J.N. Chiang, Z. Wu, X. Ding, Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation, Med. Image Anal. 63 (2020) 101693.
[15] J.E. Van Engelen, H.H. Hoos, A survey on semi-supervised learning, Mach. Learn. 109 (2) (2019) 373–440.
[16] G.-J. Qi, J. Luo, Small data challenges in big data era: A survey of recent progress on unsupervised and semi-supervised methods, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2020) 2168–2187.
[17] Y. Chen, M. Mancini, X. Zhu, Z. Akata, Semi-supervised and unsupervised deep visual learning: A survey, IEEE Trans. Pattern Anal. Mach. Intell. (2022).
[18] D.-H. Lee, et al., Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Challenges in Representation Learning, ICML, Vol. 3, No. 2, Atlanta, 2013, p. 896.
[19] Y. Grandvalet, Y. Bengio, Semi-supervised learning by entropy minimization, Adv. Neural Inf. Process. Syst. 17 (2004).
[20] A. Blum, T.M. Mitchell, Combining labeled and unlabeled data with co-training, in: COLT' 98, 1998.
[21] P. Bachman, O. Alsharif, D. Precup, Learning with pseudo-ensembles, Adv. Neural Inf. Process. Syst. 27 (2014).
[22] Y. Duan, Z. Zhao, L. Qi, L. Wang, L. Zhou, Y. Shi, Y. Gao, Mutexmatch: semi-supervised learning with mutex-based consistency regularization, IEEE Trans. Neural Netw. Learn. Syst. (2022).
[23] S. Laine, T. Aila, Temporal ensembling for semi-supervised learning, 2016, arXiv:arXiv:1610.02242.
[24] M. Sajjadi, M. Javanmardi, T. Tasdizen, Regularization with stochastic transformations and perturbations for deep semi-supervised learning, in: Advances in Neural Information Processing Systems, Vol. 29, Curran Associates, Inc., 2016.
[25] A. Tarvainen, H. Valpola, Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results, 2017, https://arxiv.org/abs/1703.01780v6.
[26] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: beyond empirical risk minimization, 2018, arXiv:1710.09412.
[27] V. Verma, K. Kawaguchi, A. Lamb, J. Kannala, A. Solin, Y. Bengio, D. Lopez-Paz, Interpolation consistency training for semi-supervised learning, Neural Netw. 145 (2022) 90–106, arXiv:1903.03825.
[28] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, C. Raffel, MixMatch: a holistic approach to semi-supervised learning, 2019, arXiv:1905.02249.
[29] O. Chapelle, A. Zien, Semi-supervised classification by low density separation, in: R.G. Cowell, Z. Ghahramani (Eds.), Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, in: Proceedings of Machine Learning Research, vol. R5, PMLR, 2005, pp. 57–64, Reissued by PMLR on 30 March 2021..
[30] X. Zhu, Z. Ghahramani, Learning from labeled and unlabeled data with label propagation, Citeseer, 2002.
[31] H. Gm, M.K. Gourisaria, M. Pandey, S.S. Rautaray, A comprehensive survey and analysis of generative models in machine learning, Comp. Sci. Rev. 38 (2020) 100285.
[32] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, 2014, arXiv:arXiv:1406.2661.
[33] D.P. Kingma, M. Welling, Auto-encoding variational Bayes, 2013, arXiv:arXiv:1312.6114.
[34] L. Wald, Some terms of reference in data fusion, IEEE Trans. Geosci. Remote Sens. 37 (3) (1999) 1190–1193.
[35] D.L. Hall, J. Llinas, An introduction to multisensor data fusion, Proc. IEEE 85 (1) (1997) 6–23.
[36] T. Meng, X. Jing, Z. Yan, W. Pedrycz, A survey on machine learning for data fusion, Inf. Fusion 57 (2020) 115–129.
[37] K. Li, S. Wang, L. Yu, P.-A. Heng, Dual-teacher: Integrating intra-domain and inter-domain teachers for annotation-efficient cardiac segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23, Springer, 2020, pp. 418–427.
[38] X. Wang, H. Chen, L. Luo, A.-r. Ran, P.P. Chan, C.C. Tham, C.Y. Cheung, P.-A. Heng, Unifying structure analysis and surrogate-driven function regression for glaucoma OCT image screening, in: D. Shen, T. Liu, T.M. Peters, L.H. Staib, C. Essert, S. Zhou, P.-T. Yap, A. Khan (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2019, pp. 39–47.
[39] Y. Xia, D. Yang, Z. Yu, F. Liu, J. Cai, L. Yu, Z. Zhu, D. Xu, A. Yuille, H. Roth, Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation, Med. Image Anal. 65 (2020) 101766.
[40] D. Yang, Z. Xu, W. Li, A. Myronenko, H.R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang, W. Zhu, G. Carrafiello, F. Patella, M. Cariati, H. Obinata, H. Mori, K. Tamura, P. An, B.J. Wood, D. Xu, Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan, Med. Image Anal. 70 (2021) 101992.
[41] X. Guo, Y. Yuan, Semi-supervised WCE image classification with adaptive aggregated attention, Med. Image Anal. 64 (2020) 101733.
[42] X. Wang, Y. Yuan, D. Guo, X. Huang, Y. Cui, M. Xia, Z. Wang, C. Bai, S. Chen, SSA-net: spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning, Med. Image Anal. 79 (2022) 102459.

[43] J. Huo, X. Ouyang, L. Si, K. Xuan, S. Wang, W. Yao, Y. Liu, J. Xu, D. Qian, Z. Xue, Q. Wang, D. Shen, L. Zhang, Automatic grading assessments for knee MRI cartilage defects via self-ensembling semi-supervised learning with dual-consistency, Med. Image Anal. 80 (2022) 102508.
[44] Z. Xu, D. Lu, Y. Wang, J. Luo, J. Jayender, K. Ma, Y. Zheng, X. Li, Noisy labels are treasure: mean-teacher-assisted confident learning for hepatic vessel segmentation, in: M. de Bruijne, P.C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2021, pp. 3–13.
[45] Y. Shu, H. Li, B. Xiao, X. Bi, W. Li, Cross-mix monitoring for medical image segmentation with limited supervision, IEEE Trans. Multimedia (2022) 1.
[46] W. Huang, C. Chen, Z. Xiong, Y. Zhang, X. Chen, X. Sun, F. Wu, Semi-supervised neuron segmentation via reinforced consistency learning, IEEE Trans. Med. Imaging 41 (11) (2022) 3016–3028.
[47] S. Parisot, S.I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, D. Rueckert, Disease prediction using graph convolutional networks: application to autism spectrum disorder and alzheimer's disease, Med. Image Anal. 48 (2018) 117–130.
[48] Y. Huang, A.C.S. Chung, Disease prediction with edge-variational graph convolutional networks, Med. Image Anal. 77 (2022) 102375.
[49] W. Zou, X. Qi, W. Zhou, M. Sun, Z. Sun, C. Shan, Graph flow: cross-layer graph flow distillation for dual efficient medical image segmentation, IEEE Trans. Med. Imaging (2022) 1.
[50] Z. Wang, Y. Lin, K.-T.T. Cheng, X. Yang, Semi-supervised mp-MRI data synthesis with StitchLayer and auxiliary distance maximization, Med. Image Anal. 59 (2020) 101565.
[51] F. Xing, T. Cornish, T. Bennett, D. Ghosh, Bidirectional mapping-based domain adaptation for nucleus detection in cross-modality microscopy images, IEEE Trans. Med. Imaging 40 (10) (2021) 2880–2896.
[52] Z. Zhao, K. Xu, S. Li, Z. Zeng, C. Guan, MT-UDA: towards unsupervised cross-modality medical image segmentation with limited source labels, in: M. de Bruijne, P.C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, C. Essert (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, vol. 12901, Springer International Publishing, Cham, 2021, pp. 293–303.
[53] Q. Meng, J. Matthew, V. Zimmer, A. Gomez, D.F.A. Lloyd, D. Rueckert, B. Kainz, Mutual information-based disentangled neural networks for classifying unseen categories in different domains: application to fetal ultrasound imaging, IEEE Trans. Med. Imaging 40 (2) (2021) 722–734.
[54] P.K. Gyawali, S. Ghimire, P. Bajracharya, Z. Li, L. Wang, Semi-supervised medical image classification with global latent mixing, in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2020, pp. 604–613.
[55] H. Nguyen, S. Saarakkala, M. Blaschko, A. Tiulpin, Semixup: in- and out-of-manifold regularization for deep semi-supervised knee osteoarthritis severity grading from plain radiographs, IEEE Trans. Med. Imaging 39 (12) (2020) 4346–4356.
[56] C.L. Srinidhi, S.W. Kim, F.-D. Chen, A.L. Martel, Self-supervised driven consistency training for annotation efficient histopathology image analysis, Med. Image Anal. 75 (2022) 102256.
[57] W. Li, J. Li, J. Polson, Z. Wang, W. Speier, C. Arnold, High resolution histopathology image generation and segmentation through adversarial training, Med. Image Anal. 75 (2022) 102251.
[58] W. Lou, H. Li, G. Li, X. Han, X. Wan, Which pixel to annotate: A label-efficient nuclei segmentation framework, IEEE Trans. Med. Imaging (2022) 1.
[59] H. Wu, J. Liu, F. Xiao, Z. Wen, L. Cheng, J. Qin, Semi-supervised segmentation of echocardiography videos via noise-resilient spatiotemporal semantic calibration and fusion., Med. Image Anal. 78 (2022) 102397.
[60] Y. Xie, J. Zhang, Y. Xia, Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT, Med. Image Anal. 57 (2019) 237–248.
[61] M. Yurt, O. Dalmaz, S. Dar, M. Ozbey, B. Tınaz, K. Oguz, T. Çukur, Semi-supervised learning of MRI synthesis without fully-sampled ground truths, IEEE Trans. Med. Imaging (2022) 1.
[62] S. Liu, K.H. Thung, W. Lin, P.T. Yap, D. Shen, Real-time quality assessment of pediatric MRI via semi-supervised deep nonlocal residual neural networks, IEEE Trans. Image Process. 29 (2020) 7697–7706.
[63] S. Liu, K.H. Thung, W. Lin, D. Shen, P.T. Yap, Hierarchical nonlocal residual networks for image quality assessment of pediatric diffusion MRI with limited and noisy annotations, IEEE Trans. Med. Imaging 39 (11) (2020) 3691–3702.
[64] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, L. Shao, Inf-net: automatic COVID-19 lung infection segmentation from CT images, IEEE Trans. Med. Imaging 39 (8) (2020) 2626–2637.
[65] Z. Zhang, C. Tian, H.X. Bai, Z. Jiao, X. Tian, Discriminative error prediction network for semi-supervised colon gland segmentation, Med. Image Anal. 79 (2022) 102458.
[66] X. Wang, H. Chen, A.-R. Ran, L. Luo, P.P. Chan, C.C. Tham, R.T. Chang, S.S. Mannil, C.Y. Cheung, P.-A. Heng, Towards multi-center glaucoma OCT image screening with semi-supervised joint structure and function multi-task learning., Med. Image Anal. 63 (2020) 101695.
[67] Q. Zhou, R. Wang, G. Zeng, H. Fan, G. Zheng, Towards bridging the distribution gap: instance to prototype earth mover's distance for distribution alignment, Med. Image Anal. 82 (2022) 102607.
[68] B. Fazekas, G. Aresta, D. Lachinov, S. Riedl, J. Mai, U. Schmidt-Erfurth, H. Bogunović, SD-LayerNet: semi-supervised retinal layer segmentation in OCT using disentangled representation with anatomical priors, in: L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, S. Li (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, in: Lecture Notes in Computer Science, Springer Nature Switzerland, Cham, 2022, pp. 320–329.
[69] A. Chartsias, T. Joyce, G. Papanastasiou, S. Semple, M. Williams, D.E. Newby, R. Dharmakumar, S.A. Tsaftaris, Disentangled representation learning in cardiac image analysis, Med. Image Anal. 58 (2019) 101535.
[70] Z. Zhao, Z. Zeng, K. Xu, C. Chen, C. Guan, DSAL: deeply supervised active learning from strong and weak labelers for biomedical image segmentation, IEEE J. Biomed. Health Inf. 25 (10) (2021) 3744–3751.
[71] G. Chen, J. Ru, Y. Zhou, I. Rekik, Z. Pan, X. Liu, Y. Lin, B. Lu, J. Shi, MTANS: multi-scale mean teacher combined adversarial network with shape-aware embedding for semi-supervised brain lesion segmentation, NeuroImage 244 (2021) 118568.
[72] N. Shen, T. Xu, Z. Bian, S. Huang, F. Mu, B. Huang, Y. Xiao, J. Li, Scanet: a unified semi-supervised learning framework for vessel segmentation, IEEE Trans. Med. Imaging (2022) 1.
[73] M. Wang, W. Zhu, K. Yu, Z. Chen, F. Shi, Y. Zhou, Y. Ma, Y. Peng, D. Bao, S. Feng, L. Ye, D. Xiang, X. Chen, Semi-supervised capsule cGAN for speckle noise reduction in retinal OCT images, IEEE Trans. Med. Imaging 40 (4) (2021) 1168–1183.
[74] Y. He, G. Yang, J. Yang, Y. Chen, Y. Kong, J. Wu, L. Tang, X. Zhu, J.-L. Dillenseger, P. Shao, S. Zhang, H. Shu, J.-L. Coatrieux, S. Li, Dense biased networks with deep priori anatomy and hard region adaptation: semi-supervised learning for fine renal artery segmentation, Med. Image Anal. 63 (2020) 101722.
[75] X. Luo, G. Wang, W. Liao, J. Chen, T. Song, Y. Chen, S. Zhang, D.N. Metaxas, S. Zhang, Semi-supervised medical image segmentation via uncertainty rectified pyramid consistency, Med. Image Anal. 80 (2022) 102517.
[76] K. Fang, W.-J. Li, Dmnet: difference minimization network for semi-supervised segmentation in medical images, in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, in: Lecture Notes in Computer Science, Springer International Publishing, Cham, 2020, pp. 532–541.
[77] F. Wu, X. Zhuang, Minimizing estimated risks on unlabeled data: a new formulation for semi-supervised medical image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. (2022) 1–17.
[78] Z. Xiao, Y. Su, Z. Deng, W. Zhang, Efficient combination of CNN and transformer for dual-teacher uncertainty-guided semi-supervised medical image segmentation, Comput. Methods Programs Biomed. 226 (2022) 107099.
[79] P.-A. Ganaye, M. Sdika, B. Triggs, H. Benoit-Cattin, Removing segmentation inconsistencies with semi-supervised non-adjacency constraint, Med. Image Anal. 58 (2019) 101551.
[80] S. Zhang, J. Zhang, B. Tian, T. Lukasiewicz, Z. Xu, Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation, Med. Image Anal. 83 (2023) 102656.
[81] D. Hao, M. Ahsan, T. Salim, A. Duarte-Rojo, D. Esmaeel, Y. Zhang, D. Arefan, S. Wu, A self-training teacher-student model with an automatic label grader for abdominal skeletal muscle segmentation, Artif. Intell. Med. 132 (2022) 102366.
[82] K. Wang, B. Zhan, C. Zu, X. Wu, J. Zhou, L. Zhou, Y. Wang, Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with contrastive learning, Med. Image Anal. 79 (2022) 102447.
[83] S. Li, C. Zhang, X. He, Shape-aware semi-supervised 3D semantic segmentation for medical images, in: A.L. Martel, P. Abolmaesumi, D. Stoyanov, D. Mateus, M.A. Zuluaga, S.K. Zhou, D. Racoceanu, L. Joskowicz (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, vol. 12261, Springer International Publishing, Cham, 2020, pp. 552–561.
[84] H. Basak, S. Ghosal, R. Sarkar, Addressing class imbalance in semi-supervised image segmentation: a study on cardiac MRI, in: L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, S. Li (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, in: Lecture Notes in Computer Science, Springer Nature Switzerland, Cham, 2022, pp. 224–233.
[85] Y. Zhang, R. Jiao, Q. Liao, D. Li, J. Zhang, Uncertainty-guided mutual consistency learning for semi-supervised medical image segmentation, Artif. Intell. Med. 138 (2023) 102476.
[86] J. Xiang, P. Qiu, Y. Yang, FUSsnet: fusing two sources of uncertainty for semi-supervised medical image segmentation, in: L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, S. Li (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2022, in: Lecture Notes in Computer Science, Springer Nature Switzerland, Cham, 2022, pp. 481–491.
[87] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M.A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V.A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K.H. Maier-Hein, P.M. Full, I. Wolf, S. Engelhardt, C.F. Baumgartner, L.M. Koch, J.M. Wolterink, I.

Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, P.-M. Jodoin, Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? IEEE Trans. Med. Imaging 37 (11) (2018) 2514–2525.

[88] H.B.J. MedSeg, COVID-19 CT segmentation dataset, 2020, URL https://medicalsegmentation.com/covid19/.

[89] A. Tragakis, C. Kaul, R. Murray-Smith, D. Husmeier, The fully convolutional transformer for medical image segmentation, 2023, arXiv:2206.00566.

[90] M.M. Rahman, R. Marculescu, Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation, 2023, arXiv:2303.16892.

[91] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, Y. Yu, Nnformer: interleaved transformer for volumetric segmentation, 2022, arXiv:2109.03201.

[92] E. Goceri, Medical image data augmentation: techniques, comparisons and interpretations, Artif. Intell. Rev. (2023) 1–45.

[93] N. Rieke, J. Hancox, W. Li, F. Milletari, H.R. Roth, S. Albarqouni, S. Bakas, M.N. Galtier, B.A. Landman, K. Maier-Hein, et al., The future of digital health with federated learning, NPJ Digit. Med. 3 (1) (2020) 119.

[94] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond, Inf. Fusion 77 (2022) 29–52.

[95] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, et al., Segment anything, 2023, arXiv preprint arXiv:2304.02643.