Personal identification no. (PIN)	Family name of applicant
Valid 322838	Ben Ayed

SUMMARY OF PROPOSAL FOR PUBLIC RELEASE (Use plain language.)

This plain language summary will be available to the public if your proposal is funded. Although it is not mandatory, you may choose to include your business telephone number and/or your e-mail address to facilitate contact with the public and the media about your research.

Business telephone no. (optional):

E-mail address (optional):

Recently, deep learning methods have shown promising progress in retinal image analysis. However, the deployment of these approaches in real-life clinical and industrial settings, where the requirements in accuracy, interpretability and speed are particularly demanding, remains below expectations due to their current limitations. Indeed, most of the existing models requires access to labeled images annotated by experts and, therefore, trains models on very limited data and narrowly defined tasks. Hence, generalization across the spectrum of pathologies, imaging techniques and patient demographics remains a challenge, making the methods highly sensitive to the variability encountered in real settings. The goal of this partnership with Diagnos Inc. is to develop novel deep learning tools that enable the detection and diagnosis of over 100 ocular and systemic conditions from cost-effective retinal images, as well as the comprehensive analysis of the retinal vasculature, lesions and other subtle biomarkers, enabling interpretable tools. Leveraging expert's knowledge and large-scale amounts of unlabeled images from heterogeneous sources, we aim at building internationally competitive foundation models and adaptation algorithms that could be rapidly deployed for new tasks, are robust to the high variability in real scenarios, and provide reliable measures of the uncertainty of their predictions. Beyond substantially improving the AI-driven products of the industrial partner (in terms of scope, robustness and interpretability), we will address research challenges of keen interest to the wider scientific community in medical image computing. This R&D program could benefit both the industry and society. The ensuing algorithms could be widely deployed for population-based screenings, improving the early detection (hence, timely treatment) of a breadth of vision- and life-threatening conditions from cost-effective retinal images, while reducing health care costs to taxpayers. It will attract and train high-caliber HQPs (5 graduate students and 3 postdocs) in high-demand sectors (AI and medical imaging), yield excellent academic publications, and enhance the competitiveness of the industrial partner.

Other Language Version of Summary (optional).

1. BACKGROUND, OBJECTIVES & ANTICIPATED IMPACT

- Explain the challenge to be addressed, the importance of the topic and the need for new concepts or directions.
- Outline the objectives of the project and briefly explain its anticipated outcomes and impact.
- Position the proposed research relative to other efforts and to the state-of-the-art.

1.a) Application domain & program motivation: According to the WHO report in 2019 [1], more than 418 million people worldwide suffer from eye diseases that could cause blindness, e.g., diabetic retinopathy (DR), glaucoma, age-related macular degeneration (AMD), diabetic macular edema (DME), among dozens of other diseases [2]. Furthermore, at least 1 billion people have a vision impairment that could have been prevented (via early detection) or is yet to be addressed [1]. Deep Learning (DL) algorithms could identify subtle changes in the retinal morphology from cost-effective fundus and optical coherence tomography (OCT) images, enabling the early detection of various eye diseases. Beyond eye diseases, these images provide valuable insights into *systemic* diseases. For instance, the retinal microvascular morphology provides insights into other vascular systems and is clinically established as a cardiovascular risk predictor [3, 4]. Also, the optic nerve and inner retinal layers are clinically linked to neurodegeneration [3]. Given the increasing prevalence of ocular conditions and the limited number of ophthalmologists [5], DL algorithms offer a promising, automated solution for population-wide screenings. They could enable early detections (hence, timely interventions) of a breadth of conditions in cost-effective images, thereby preventing vision loss and predicting life-threatening cardiovascular risks.

1.b) Automated retinal imaging tasks: The strong need outlined above motivated a recent literature on developing specialized DL models for different retinal image classification and segmentation problems [6]. In general, each of these models focuses on a specific task, i.e., a single disease or condition, e.g., the grading of DR [7, 8] or glaucoma [9]. Classification assigns images to labels (or categories). It could be binary, e.g., the presence or absence of a disease, or multi-class such as the different grades of disease severity ("mild", "moderate", "severe", etc.), as in DR grading. Segmentation assigns a label to every pixel, localizing target regions in retinal images, such as vessels, lesions, or anatomical structures [10, 11-13]. It yields comprehensive geometric measures for disease diagnosis. For instance, retinal vessel tortuosity and diameters are clinically established as biomarkers for cardiovascular diseases [4], and the cup-to-disc ratio is a biomarker for Glaucoma [9]. Thus, segmentation yields a clinically interpretable support to the algorithms, which is crucial for wide adoption. More generally, medical image segmentation has attracted wide interest within the research community, for a breadth of pathologies and organs [14].

1.c) Limitations of current models: In retinal imaging, and in the much broader field of medical imaging, most of the existing models rely on *supervised, task-focused learning*, a paradigm that i) requires access to labeled images annotated by experts, and ii) trains models on a specific task, e.g., grading the DR disease [7]. Therefore, generalization across the spectrum of pathologies, imaging techniques and patient demographics remains a challenge [6, 10, 12, 16]. As they focus on a specific condition, and due to the costs of labeling images by clinical experts, such models typically learn from limited amounts of data. This makes the algorithms susceptible to failure whenever the data characteristics change, a difficulty called *domain shift* [15]. In the literature on retinal image classification, there are dozens of conditions whose detection or grading need to be addressed [6], *each typically addressed individually as a supervised classification*. For a few diseases, such as DR or glaucoma, promising performances were achieved thanks to the availability of large-scale labeled datasets (~100K images). However, for a breadth of other conditions, labeled datasets are orders-of-magnitude smaller [17], with a large room to improve the performances. Similarly, due to the prohibitive cost of pixel-level annotations, segmentation datasets are limited to a few dozen subjects or, at best, a few hundred [10]. Besides being trained on limited data, which impedes generalization in real, highly varying settings, there are other limitations of current models. They cannot integrate expert's knowledge beyond the labels, such as the relationships between the classes. For instance, the DR disease is associated with the presence of microaneurysms [17], a type of lesion. Such

domain-specific knowledge could not be integrated in classification loss functions dedicated to DR grading. Also, there is growing evidence in the literature pointing to the poor calibration of the uncertainty estimates of supervised models [18-23]: they assign high confidence to incorrect predictions, an issue related to the inherent properties of the loss functions [20-21]. This issue is significant in safety-critical scenarios, such as medical diagnosis, in which reliable uncertainty estimates are of paramount importance.

1.d) The recent rise of foundation models and the yet-to-explore potential in medical imaging: There is currently a paradigm shift in computer vision and Natural Language Processing (NLP) driven by the growing prevalence of the so-called *foundation models* [24-28], which exploit large-scale amounts of *unlabeled data.* For instance, self-supervised learning [27-28] leverages unlabeled images via some "pretext task", and Vision-Language Models (VLMs) use text captions associated with the images [25-26]. Currently, VLMs are transforming computer vision, emerging as a promising solution towards true generalization. Learning from large amounts of image-text pairs collected over the internet, VLMs yield robust features, providing a powerful alternative to supervised learning. They have shown impressive *few-shot adaptation* [29-32], i.e., when fine-tuned on new tasks using a few labeled samples. This paradigm change is very promising in medical imaging, but at its very beginning. Recent works, including ours [17, 36], explored the potential of foundational image-text [17, 33-34] or segmentation [35-36] models tailored to specific medical domains, and pointed to the poor performances of generalist computer vision models in medical tasks. Generalist models may not capture fine-grained and specialized medical concepts, such as grading disease severity or segmenting subtle regions like lesions. *Therefore, there is a strong need to build specialized methodologies leveraging unlabeled/weakly labeled data and expert's knowledge, which might be extracted from various sources associated with medical images, such as the diagnostic text reports, textual descriptions within the clinical literature, or anatomical constraints*. Also, there are abundant amounts of unlabeled images collected every day. For instance, to date, *the industrial partner's data is approaching one million retinal images*. So far, such abundant data has not been fully exploited by our community, as the focus has been mainly on supervised learning and much smaller labeled data sets.

1.e) Objectives: The overall objective is to build novel DL tools that enable the detection and diagnosis of over 100 ocular and systemic conditions from retinal images, and that could be efficiently adapted to new tasks, while being interpretable, well-calibrated and robust to domain shifts. Beyond substantially improving the AI-driven products of the industrial partner (in terms of scope, robustness and interpretability), we will address research challenges of keen interest to the wider scientific community in medical image computing. Leveraging expert's knowledge and large-scale amounts of unlabeled/weakly labeled images from heterogeneous sources, we aim at specialized and internationally competitive methodologies that could benefit a breadth of medical image classification and segmentation problems. The specific technical objectives are: **(O1) Developing semi-supervised VLMs for retinal image classification:** The goal is to detect and grade over 100 conditions, while improving robustness to domain shifts; **(O2) Developing weakly-supervised foundation models for retinal image segmentation:** The goal is to segment the complex vasculature as well as various lesions and other structures in retinal images. This provides comprehensive estimations of various geometric measures (such as vessel tortuosity), which are clinically established as disease biomarkers, yielding an interpretable support to the developed diagnostic tools; **(O3) Developing specialized few-shot adaptation algorithms**, which efficiently finetune the pre-trained models to new datasets, assuming access to a few labeled samples in the target conditions and limited computation/memory resources: The goal is to enable the models to adapt quickly in real industrial conditions with constant changes in data characteristics; **(O4) Equipping the models with uncertainty estimation mechanisms**: This will provide to practitioners not only disease predictions, but also the confidence that should be placed on such predictions, enabling widespread clinical adoption.

1.f) Anticipated outcomes and impact: This R&D program will expand substantially the commercial product portfolio of the partner, currently limited to DR grading and a handful of other tasks. *Considering the recent retinal-imaging literature, it will mark a transition from standard, narrowly supervised algorithms,*

with limited generalization capabilities, to the pre-training and adaptation of large-scale models leveraging expert's knowledge and unlabeled images (including the partner's data approaching 1M images). Thus, it promises to deliver internationally competitive algorithms, which could be widely deployed for population-based screenings. This could substantially improve the early detection (hence, timely treatment) of a breadth of vision- and life-threatening conditions from cost-effective retinal images, while reducing health care costs to taxpayers. In the broader scope of medical imaging with DL, this program is among the initial but fast-growing community efforts focusing on foundation models and few-shot adaptation. Therefore, following on from the relevant expertise of the academic team and its previous successful collaborations with the industrial partner, we expect this program to: (i) yield excellent academic publications; (ii) attract and train high-caliber HQPs in high-demand sectors (AI and medical imaging); and (iii) yield effective translation into competitive products, expanding the partner's revenue and its current basis of customers.

2. PARTNERSHIP

- List all partner organizations participating in the project. For each, describe their core activities and how they align with the project, their need for the proposed project, and their experience related to it, such as efforts to date to address the challenge.
- Describe each partner organization's active role in the project, including defining the research questions, designing the research plan, collaborating or contributing to the research activities, co-supervising trainees and monitoring progress.
- Describe how the partner organizations will translate, mobilize and/or apply the research results to achieve the intended outcomes.
- Explain the value and importance of each partner organization's involvement and other in-kind contributions to achieving the project's intended outcomes. If applicable, discuss how the combination of partner organizations is beneficial to the project.

2.a) The partner's core activities: Diagnos Inc. is a Montreal-based medical imaging and AI software developer. Founded in the year 2000, the company is a publicly traded corporation specialized in the detection and evaluation of critical health issues from cost-effective retinal images, using computer vision. It has more than 20 years of experience in image and data analysis algorithms, and an excellent history of collaboration with academia. Automated screening of diabetes and its complications has been one of the main focuses of the company. Diagnos tele-ophthalmology platform, called CARA (Computer Assisted Retina Analysis), is a cost-effective tool for screening large numbers of patients in real-time. CARA has helped detect anomalies in retinal images from more than 235K patients, across 16 countries and 131 screening sites, and has been cleared for commercialization by regulatory authorities such as Health Canada, the FDA and the EU. Also, the number of users (i.e., optometrist or ophthalmologist clinics) of Diagnos commercial tools is steadily increasing, and is expected to reach 400 clinics in 2025. The company has partnerships with several clinical institutions. For example, in an ongoing project with the CHUM (Université de Montréal's hospital centre), Diagnos algorithms are used for screening over 3K diabetes patients. To date, Diagnos data totalizes over 800K images, collected through various projects.

2.b) Past efforts and collaborations: Initial developments of CARA focused on DR detection and grading with traditional computer vision algorithms. Moving towards DL, Diagnos has had intensive collaborations with Prof. Ben Ayed, funded by Quebec's PROMPT AI program (2019-23). These have involved the co-supervision of 4 HQPs, translations of prototypes into commercial implementations, several publications [8, 12-13, 17, 37-41], and top-rank performances during international challenges [9, 42]. The partnership has had an impact on the company and yielded internationally competitive methods for DR grading [8, 38], which were integrated to Diagnos' platform. Furthermore, it yielded competitive algorithms for a few other classification and segmentation tasks, related to conditions other than DR, such as the detection of DME [39], Glaucoma [9], AMD [41] and hyper-intensive retinopathy [37]. Currently, the company is working on obtaining Health Canada and FDA approvals for these 4 additional AI modules.

2.c) Need for the project and how the partner will translate the results: CARA is currently limited to DR grading and a few other tasks. Therefore, Diagnos will continue to work with the applicant's team towards a breadth of other problems, including the detection and grading of dozens of ocular and systemic conditions, as well as the comprehensive analysis of the retinal vasculature, lesions and other subtle biomarkers, enabling interpretable tools. Diagnos has been exposed to the challenges of deploying current supervised models in real conditions, including the lack of labels, dataset shifts and class imbalance. Therefore, the transition to large foundation models and few-shot adaptation promises to make a significant impact on the company, as it could (i) enhance the scope, robustness and interpretability of its AI products, facilitating widespread adoption; (ii) grow its revenue; and (iii) expand its collaborations with clinical and industrial partners. *The ETS-Diagnos team has made progress towards this move and has recently curated large-scale vision-language retinal data [17]*. Also, this partnership has already shown effective knowledge translation over the last years. Following on from this success, the algorithms resulting from this project will be integrated in Diagnos' system. Also, the team will continue publishing in the top venues of the field (such as MICCAI, CVPR, MedIA), which will further increase the company's visibility.

2.d) Cash/in-kind contributions and active role of the partner in the project: Diagnos is investing significantly in R&D and in its strategic alliance with the ÉTS, through the establishment of the *research Chair on Artificial Intelligence in Medical Imaging (AIMI)* for a period of 5 years (2024-29). This will give Diagnos access to state-of-the-art knowledge and expertise in computer vision research, engaging excellent HQPs and maintaining its competitiveness. *Diagnos' financial contribution to the chair held by Prof. Ben Ayed amounts to $500K ($100K per year)*. Considering the indirect costs, this gives a yearly amount of $81K on average to be matched through Alliance and Mitacs grants such as this one. Diagnos' in-kind contribution will be in the form of the time provided by the company's staff, which includes computer scientists and clinical experts (see Team section for details). Diagnos' staff will play an active role in problem definition and in curating large-scale data for prototyping, as well as in algorithm design and updates, validation of experimental methods, programming, optimization and integration of software prototypes in realistic settings, result generation and documentation, and HQP co-supervision. An important part of the in-kind contribution will be the time of an ophthalmologist, Dr Hadi Chakor, Diagnos' Chief Medical Officer, who will bring valuable medical expertise, providing image labels and refining textual descriptions of domain knowledge. The overall value of Diagnos' in-kind contribution is $568K.

3. RESEARCH PLAN

- Specify the research objectives and expected results. Describe the planned research activities, methodology and experimental design.
- Provide approximate timelines for the activities, milestones and deliverables. You may use a Gantt chart, table or diagram.
- Describe how equity, diversity and inclusion are considered in the research process (e.g., research questions, design, methodology, analysis, interpretation and dissemination of results) and how these considerations are integrated where relevant.

For each objective listed in section 1.e), we detail below the planned activities and algorithmic methods. We also discuss the empirical evaluation as well as the EDI considerations and provide an approximate timeline. Along with the emergence of foundational models, and following on from our recent work on the subject, we will address research challenges of keen interest to both Diagnos and the wider scientific community: How to model expert's knowledge and integrate it in foundation models? How to effectively adapt foundation models? How to effectively leverage unlabeled data during pre-training and adaptation?

3.a) O1: Semi-supervised VLMs for medical image classification (Postdoc 1): The developments of VLMs tailored to specific medical domains is an emerging subject [17, 33-34]. We aim at building large-scale, internationally competitive VLMs for retinal imaging, *integrating expert's knowledge in the form of text data as well as large-scale amounts of unlabeled and labeled images from heterogeneous sources*,

including Diagnos data (~800K images) and an assembly of open-access data. The goal is to detect and grade over 100 conditions, including ocular and systemic diseases, while enhancing robustness to domain shifts. We will build on the very recent, potentially impactful work by our team in the context of medial VLMs: *Specifically, we have curated a large-scale vision-language retinal dataset consisting of image-text pairs [17], a pillar step towards achieving (O1)*. More precisely, we enhanced an assembly of 37 open-access, class-labeled datasets (a total of 284K images and 97 classes representing ocular conditions) with text data. We compiled text data from the clinical literature, encoding expert's knowledge implicitly in the data. For each of the 97 classes, we built several text descriptions informing on the relationships with other classes and/or specific biomarkers. This mimics the diagnosis by medical experts, who search for biomarkers (e.g., a lesion) indicating a specific disease. For instance, the disease class "severe DR" is associated with the following text prompts describing related biomarkers: "severe haemorrhages in all quadrants", "venous beading", and "intraretinal microvascular abnormalities". Preliminarily results based on our vision-language data have shown promising performances [17], pointing to the importance of encoding expert's knowledge and to the limitation of generalist computer vision models in medical tasks.

We will investigate composite losses, which integrate complementary terms for training VLMs, leveraging various forms of pairwise data: (i) *image-text pairs*, each including a labeled image and text-encoded expert knowledge associated with the label, as in our vision-language retinal data [17]; (ii) *image-image pairs*, each including an unlabeled image and an augmented view of it (e.g., using random resizing or adding noise and color distortions), as in self-supervised learning methods [27-28]; and (iii) *text-text pairs*, encoding explicitly expert knowledge on the relationships between different concepts, e.g., a text description of a disease and the one describing an associated biomarker. We will expand the text descriptions beyond the class labels in the data, covering a wider range of conditions and leveraging recently released corpuses of textual knowledge derived from ophthalmology textbooks [80]. We will customize contrastive-learning strategies [27]: For each batch of pairs, this amounts to minimizing some distance between the deep representations of associated samples, while maximizing the distances between those that are not associated. We will explore recent transformers architectures, which accommodate image-language representations, as well as autoregressive generative methods to deal with text-text pairs.

3.b) O2: Weakly-supervised models for segmentation: We will investigate large foundation models for three problems, i.e., segmenting (i) lesions, (ii) anatomical structures and (iii) vessels. Leveraging domain knowledge and substantial amounts of unlabeled and/or weakly-labeled images, we will explore specialized unsupervised loss functions and methods. In the literature, retinal image segmentation is often tackled with standard supervised learning [10, 12-13]: The models are trained on very limited data (a few hundred images, at best [10]), due to the prohibitive cost of pixel-level labeling. Therefore, the ensuing methods are seriously affected by domain shifts, and their performances remain below expectations [10, 12]. We will build on the recent literature in computer vision on weakly-supervised segmentation [59, 60-63], including developments by our team [61-63]. The main idea is to leverage unlabeled or weakly-labeled images with some priors, which are encoded in the form of unsupervised losses, thereby learning from substantially larger amounts of data and potentially improving generalisation. In segmentation, a *weak label* takes the form of a global (not pixel-wise) information, e.g., the presence or absence of a region in the image. Thus, large-scale natural-image classification datasets (such as ImageNet) could be used as a form of weakly-labeled data for segmentation. This has motivated an abundant weakly-supervised learning literature in computer vision. However, the subject remains relatively unexplored in retinal imaging.

(O2.a) Weakly-supervised lesion segmentation leveraging expert knowledge (PhD1): We will develop foundation models for segmenting dozens of lesion types, leveraging weak labels and expert knowledge. We will build a large assembly of retinal image classification datasets as a form of weakly-labeled data (over 600K images in total): This includes open-access data totalizing 284K images and 97 classes [17], as well as Diagnos' class-labeled data totalizing over 350K images, with labels for DR, DME or AMD, among other conditions. The difficulty is that the class information in such data is mostly associated with

diseases (not lesions). We will use the textual data that we compiled recently in [17], which associates diseases to lesions and other biomarkers. For instance, the class "mild DR" is associated with microaneurysms and hard exudates [17]. This transforms retinal image classification data into weakly-labeled data for lesion segmentation, leveraging *class activations maps (CAMs)* methodologies in computer vision [59, 60-61]. CAMs are byproducts of classification tasks, highlighting the image's regions that influence the classification decisions. These regions could be used as pixel-level pseudo labels to train segmentation networks. We will investigate multi-term losses integrating two types of supervised-learning terms, one based on the pseudo-labels and weakly-labeled data and the other on an assembly of open-access fully labeled data for retinal lesion segmentation [64-66], which totalizes over 900 images. To deal with extreme region imbalance in lesion segmentation, we will explore dedicated losses. This includes, for instance, decision boundary losses, as introduced in our award-winning work in [67].

(O2.b) Semi-supervised retinal-structure segmentation with anatomical constraints (PhD2): We will investigate foundation models for segmenting the retinal anatomical structures, such as the optic disc, optic cup, macula and fovea, leveraging large amounts of unlabeled images with shape priors. We will curate nearly 1M unlabeled/weakly-labeled images from heterogeneous sources, integrating open-access classification data (~200K images) and Diagnos data (~800K images). Such unlabeled data will be used in conjunction with an assembly of labeled open-access datasets built for these problems [68-70], which totalizes near 1K images. We will explore formulations integrating: (i) unsupervised constraints encoding knowledge about the shape of the target regions; and (ii) supervised-learning terms. Such shape information could be expressed mathematically as functions of the network's predictions at the unlabeled images, e.g., via invariant shape moments. Our recent, award-winning work in [58] showed promising results along a related line of research, as we were able to supervise organ segmentation with a few inequality constraints on shape moments. To accommodate several constraints, we will explore customized optimization strategies, building on our recent work of augmented Lagrangian principles in DL [71].

(O2.c) Semi-supervised vessel segmentation models with data-driven priors (PhD3): We will investigate large-scale foundation models for segmenting the retinal vessels, using the unlabeled data mentioned above (~1M images) and an assembly of labeled vessel-segmentation data [12], compiled from 10 datasets from heterogeneous sources, and totalizing over 500 images. For instance, we will customize unsupervised losses, such as the Shannon entropy [55] or graph-Laplacian regularization [54], to retinal vessel segmentation. Such data-driven priors encode some desirable assumptions on the classifier's decision boundary, given unlabeled samples, and were mostly deployed in image classification [54, 55]: Entropy minimization prescribes that classifier's boundaries should not occur at dense regions of the feature space, and Laplacian regularization encourages nearby samples (in the input space) to have similar predictions.

3.c) O3: Few-shot adaptation for medical imaging: In the computer vision community, there is currently a staggering interest in adapting foundation models in few-shot regimes (i.e., using a few labeled samples in the target conditions) [29, 30-32, 46], including recent works from our team [31-32]. Yet, these methods are designed for computer vision and NLP tasks and rely on many *ad hoc* choices, which are still not well understood [43], raising interesting questions: *What model parameters to update [44-45]? What loss function to optimize [31, 46] and do the hyper-parameters depend on the target task [32]*? Recent experiments suggest that the best choices depend on the target data characteristics and shifts [32, 36], which might limit the applicability of these methods in medical imaging. We will explore how to fine-tune effectively the models in medical imaging segmentation and classification tasks. Also, how to make the techniques hyper-parameter free and computationally efficient are important questions in practice.

(O3.a) Few-shot adaptation of vision-language classifiers (PhD4): We will investigate optimization-based formulations, which probe the most influential trainable parameters, e.g., via gradient activations [44], low-rank approximations [47] and sparsity constraints [48], thereby finding automatically "what to update". Such Parameter-Efficient Fine-Tuning (PEFT) methods have recently gained popularity in NLP

[47-48], as they tune a fraction of the parameters, easing storage and optimization burden while improving performances. As pointed out recently in [49], PEFT methods are promising yet need to be customized to medical imaging tasks. Also, building on the optimization insights we provided in [31], we will explore Majorize-Minimize optimization strategies, in which the hyper-parameters (e.g., step size) are implicit, and how to learn the hyper-parameters via algorithm unrolling strategies [50]. Another interesting avenue is the investigation of losses designed for fine-tuning VLMs, e.g., via integrating information from both the vision and language encoders [31-32] and/or via encoding expert knowledge in the form of text.

(O3.b) Few-shot adaptation of segmentation models (Postdoc2): We will build on the expertise of our team in few-shot segmentation of natural images [51-52], and design adapters customized for medical imaging. For instance, we will explore transductive inference [53-55], which makes joint predictions for all the pixels of a testing image, leveraging their statistics with priors that are appropriate for segmentation. To do so, we will optimize losses integrating: (i) supervised terms using a few labeled images; and (ii) unsupervised terms embedding domain-knowledge constraints (e.g. anatomical/shape priors for retinal structures), spatial regularization (e.g., graph-Laplacian terms [54]), and/or discriminative clustering objectives (e.g., the Shannon entropy [55]). To manage clustering at the pixel level (i.e., millions of data points per image), we will build on the expertise of our team in optimization for large-scale clustering [56-57] and investigate dedicated approximations/bounds of the objective functions for efficient inference.

3.d) O4: Uncertainty calibration: We will develop uncertainty calibration strategies that are tailored to foundation models, including segmentation and VLMs, and to their adaptation. So far, the literature on calibration has mostly focused on image classification and on standard supervised learning [20-22].

(O4.a): Uncertainty calibration for vision-text classifiers (Postdoc3): Our very recent experimental work in [72] pointed to the critical issue of miscalibration of vision-language adapters, so far overlooked in the literature. We empirically showed that vision-language adapters, such as prompt-learning methods, substantially degrade calibration in the presence of domain shifts. We will extend and customize our margin-based label smoothing methods [21-22], to account for the text knowledge. For instance, we will examine adding margin-based inequality constraints on the vision-language scores to fine-tune losses, enabling to control in a flexible way the trade-offs between the confidence and accuracy of the predictions.

(O4.b): Uncertainty calibration for segmentation (PhD5): The calibration literature is mostly focused on image classification [20-21], except a few recent works that examined segmentation (e.g., [73]). Yet, these tackled the problem as individual-pixel classification, omitting the structural relationships between pixels. We will develop calibration methods tailored to segmentation. For instance, we will explore structural pairwise constraints encouraging nearby pixels (both in terms of spatial coordinates and image/feature space) to have similar confidence, building on concepts from the Conditional Random Fields (CRFs) literature in computer vision [63]. We will explore various measures of pixel-level uncertainty (e.g., the Shannon entropy). To accommodate a large set of pixel-level constraints, we will deploy customized optimizers, following on from our recent work on augmented Lagrangian methods in DL [71].

3.e) Empirical methods and EDI aspects (O1, O2, O3 and O4): The data will be distributed into train/validation/fine-tuning/test subsets that reflect disease prevalence in realistic setting and ensure fair distributions in terms of subject demographics and data sources. We will validate the models in different steps: i) foundational pre-training; ii) fine-tuning and adaptation; and iii) testing. To simulate concepts and domain shifts, we will build splits where certain concepts (e.g., a type of lesion) and datasets are rare or absent during pre-training but seen during adaptation and testing. We will take measures to mitigate biases that may occur in the models, e.g., due to a lack of diversity or an imbalance in the data, which could yield biased decisions towards certain concepts or demographic groups. For instance, we will ensure that training data is balanced across the broadest possible diversity of groups and deploy specialized losses that minimize bias during training [74]. As for evaluating performances, we will use measures that are standard

for each type of tasks: The Kappa and F1 scores for classification [37]; the Dice and Hausdorff measures for segmentation [13] and the Expected Calibration Error [18] for calibration.

3.f) Timeline: This R&D program and will be realized according to the following approximate timeline:

	Year	1	2	3	4	5
O1, O2, O3, O4: Experimental evaluations & documentation & integration in CARA	**O1:** Semi-supervised VLMs for medical image classification **(Postdoc1)**					
	O2.a: Weakly-supervised lesion segmentation leveraging expert knowledge **(PhD1)**					
	O2.b: Semi-supervised structure segmentation with anatomical constraints **(PhD2)**					
	O2.c: Semi-supervised vessel segmentation models with data-driven priors **(PhD3)**					
	O3.a: Few-shot adaptation of vision-language classifiers **(PhD4)**					
	O3.b: Few-shot adaptation of segmentation models **(Postdoc2)**					
	O4.a: Uncertainty calibration for vision-text classifiers **(Postdoc3)**					
	O4.b: Uncertainty calibration for segmentation **(PhD5)**					

4. TEAM

- List the applicant, any co-applicants, key participating staff of the partner organizations and any other key academic team members. For each, explain how their knowledge, expertise, experience and contributions align with the proposed project and describe their role in the project, as well as their roles and capabilities in training and mentoring trainees.
- Briefly describe the plan for managing the project, along with the qualifications, roles and responsibilities of the team members involved in this respect.

The academic applicant, I. Ben Ayed, is a well-recognized expert in the field, with a track-record of over 170 fully peer-reviewed articles, 2 technical books, and 7 approved US patents. Most of his articles are published in the topmost venues in computer vision, machine learning and medical imaging. He has an *h-index* of 54 and over 10K citations (Google Scholar). He gave over 50 invited talks and 7 tutorials on subjects related to this project, at flagship conferences, e.g., [77-79]. He contributed to the training of 37 graduate students and postdocs, many of whom interacted successfully with the industry. His team collected several international distinctions, such as the best-paper awards at the MIDL'21 conference [58] and at the MICCAI'23 Workshop on foundations models [36], as well as top-rank positions in international challenges. The academic co-applicant, J. Dolz, is also a prolific expert in the field, with over 110 publications, over 6K citations and an *h-index* of 39. Recent work from the applicants tackled various aspects pertaining to this project, including VLMs [17, 31-32], few-shot learning [53-55], adaptation [75-76], calibration [72-73], optimization for DL [71], as well as methods customized for medical imaging and segmentation tasks [12-13, 17, 23, 51-52, 67]. The co-applicants will (i) manage the funding associated with the project and monitor progress; (ii) hire and co-supervise HQPs; and (iii) participate in the algorithmic aspects, sharing state-of-the-art knowledge in specialized methodologies and co-authoring publications. The industrial partner's staff participating in the project include 3 data modeling and algorithm development specialists (R. Kobbi, J. Chelbi and M-A Racine) and a clinical expert (Dr. H. Chakor), Diagnos' chief medical officer, who will bring very valuable domain knowledge and extensive dataset-related work (image annotations, textual descriptions, etc.). Each of these staff members has over 10 years of experience in the development of retinal image analysis algorithms and over 10 academic publications on the subject. They will play an active role in (i) problem definition, data curation and model training; (ii) programming, experimental evaluations and integration of software prototypes in CARA system; (iii) publications; and iv) HQP co-supervision. More details on the role of each participating staff member are provided in the in-kind contribution document. The interactions between ÉTS and Diagnos will include bi-weekly meetings, to set expectations, monitor progress, and develop the algorithmic aspects, data sets, and evaluation protocols. The team will continue using effective collaboration tools, such as GitHub, to collaborate on codes, and Overleaf for documentation and publications. This also

enables effective tracking and documentation of progress. The involved HQPs are expected to work in Diagnos' offices 1-3 days per week, which gives them access to the partner's data and to regular technical interactions with its staff. Finally, this R&D program is a part of Prof. Ben Ayed's industrial Chair with Diagnos. The Chair's governance is ensured by a management committee composed of 4 members (2 from ÉTS and 2 from Diagnos), whose roles will be to adopt the program, validate the annual budget and follow up on the resolutions adopted during committee meetings. A minimum of 2 meetings per year is planned.

5. TRAINING PLAN

- Describe the learning experiences the project will provide, including the nature of interactions between trainees (undergraduate and graduate students, postdoctoral fellows) and partner organizations.
- Describe the research and professional skills that trainees will develop through these experiences and through their roles in the project.
- Explain how the research and professional skills gained by the trainees will prepare them for their future careers.
- Describe challenges to equity, diversity and inclusion in the context of your project's training environment and specify concrete practices you will implement to address them. You are encouraged to cite evidence supporting the proposed practices and to describe how you will monitor and adapt your actions based on non-demographic indicators of success.

5.a) Learning experiences, professional skills and career prospects: This program will involve 8 HQPs, and integrate them in a collaborative environment, which includes internationally recognized expertise in computer vision, optimization, deep learning, and medical imaging, as well as established interactions with the industrial partner (since 2019) and clinical experts. Following on from our past contributions to HQP training, we will continue to lead the students so that they perform internationally visible research, publish in the topmost venues, participate in knowledge-translation activities, and secure excellent positions.

Academic environment and training approach: Prof. Ben Ayed holds regular one-on-one and group meetings with HQPs, enabling them to work in collaborative settings. In group meetings (including those involving industrial and clinical collaborators), HQPs present methods, progress and results, thereby improving their communication and knowledge-translation skills. Also, such interactions promote collaborations and new ideas. The students will be part of the ÉTS laboratory of imagery, vision and artificial intelligence (LIVIA), one of the largest computer vision labs in Canada, gathering 13 PIs and over 70 HQPs. Thus, HQPs will benefit from LIVIA's modern infrastructure, which includes state-of-the-art GPU servers. They will also benefit from interactions with researchers at the LIVIA and other AI labs in Montreal, and attend regular seminars in these labs, gaining knowledge in DL, vision, signals, and NLP.

Interactions with the partner: Collaborations between Diagnos and Prof. Ben Ayed's team has started in 2019, with 4 HQPs involved (3 postdocs and 1 PhD student). This yielded top-ranked algorithms during international challenges [9, 37], translations of research prototypes into products, and publications in excellent venues (e.g., [8,13,18]). We will follow on from these successful interactions, which include bi-weekly meetings, to set expectations, discuss progress, elaborate on the technical details of algorithm designs, data sets, evaluation protocols, and performance goals. HQPs are expected to work in Diagnos' offices 1-3 days per week during the project. They will have access to the partner's data and interact with its staff, collaborating on codes and publications. These knowledge-translation activities and industrial experiences in high-demand areas make HQPs uniquely qualified to reach positions in industrial labs.

Tackling timely research problems and developing technical skills: Each student will be involved in a cutting-edge project, which gives rise to algorithmic problems. With the recent rise of foundation models, the questions stated in this R&D program are timely and currently attract wide interest. Technically, these questions involve the design of original objective functions, effective optimization schemes to tackle them,

and understanding the statistical assumptions underlying them. This requires in-depth understanding of advanced concepts in mathematical optimization, statistics, DL, algorithms, image analysis, NLP, and experimental validation, all essential skills for the HQP's future careers. With regular personalized and group meetings, HQPs will directly benefit from the state-of-the-art knowledge of our team. Specifically, our team has recently made several methodological contributions, e.g., in VLMs [31-32, 81], few-shot learning [51-55], calibration [21-23], adaptation [75-76] and optimization for DL [71], which led to several publications at the topmost venues in vision and learning (such as CVPR and NeurIPS). This technical know-how will serve as a solid basis for this training program. Furthermore, new HQPs will promptly learn from the excellent programming skills and up-to-date libraries that already exist in our group.

High-quality publications and other internationally visible activities: Prof. Ben Ayed's team has been very active in publishing in the most competitive and visible venues in vision, learning, and medical imaging (such as CVPR, NeurIPS and MedIA), with HQPs as first authors. We will continue along this path, enabling HQPs to acquire the necessary skills to publish in these prestigious venues and to maximize the visibility of their works. Indeed, HQP involved in past collaborations between Diagnos and Ben Ayed's team published papers in these venues [8, 13, 21-23, 32, 72-73]. Such publications enhance career prospects and are key elements in hiring decisions. Also, pursuing an open-science tradition in our team, HQPs will be encouraged to publish software prototypes on their GitHub pages, which trigger further visibility, feedback and collaboration opportunities. Finally, HQPs will have the opportunity to attend the flagship conferences of the field, and to participate in international challenges and training activities.

5.b) EDI considerations: HQPs in our team are part of a collaborative and highly diversified environment (both culturally and scientifically). We will continue to promote our group's diversity and its engaging atmosphere, accommodating health, family, cultural, religious, and other constraints. Despite the growing EDI efforts in our computer science community, the percentage of female students remains low. Indeed, the representation of women graduating from university-level computer science programs in Canada is 22.8% [82], which is 10 to 20% lower than many other industrialized world countries [83-84]. The recent statistics in [83-84] show substantial differences from one country to another, which suggests that there might be significant biases and cultural differences at play. According to data from our graduate programs at the ÉTS (2024-2025), female students represent 30% of the population, and 62% of graduate students are international. These proportions are also reflected in our LIVIA lab (over 70 graduate students). While this represents a relatively strong diversity, and despite the concrete actions at our institution, female students remain under-represented. This is also the case for other groups including persons with disabilities, indigenous people, visible-minority groups and LGBTQ2+. Therefore, we will pursue a concrete EDI action plan and evaluate/update it with regular inputs from our institution's EDI advisors and from HQPs to i) increase the inclusion of underrepresented groups; and ii) mitigate unconscious biases:

- In HQP hiring, we will advertise positions as widely as possible and reach out to venues focused on under-represented groups, e.g., the WiCV (Women in Computer Vision) Workshop, held regularly during our major conferences. We will also seek help from the ÉTS EDI advisors, to disseminate gender-neutral, diversity-oriented postings to a breadth of venues (societies, institutions, women+ in IT, etc.), with ample time allowed for applying. Team diversity will be privileged when candidates have similar qualifications.
- The researchers have been following regular institutional training sessions on EDI, unconscious biases and inclusive leadership [85]. They involve all group members, independently of their background or seniority, in decisions pertaining to the group's work environment and interactions. Following recent EDI guidelines [86], they will arrange anonymous surveys on a yearly basis, prompting HQPs to: (i) rate various lab function aspects (such as HQP integration in the team, work environment, lab rules, supervision and progress, and time management); (ii) provide feedback on identified issues and the solutions implemented; and (iii) suggest improvements. In addition, regular one-on-one and group meetings will enable to collect further feedback, detect issues, elaborate solutions, and evaluate the effectiveness of the pursued actions. This ensures HQP's well-being and increases awareness of HQP's career goals, while maintaining

confidentiality. Also, a conflict/complaint management protocol will be put in place to resolve issues. Furthermore, the researchers will follow development sessions during the project, and will continue applying the prescribed EDI actions. Also, HQPs will be encouraged to follow such sessions, to ensure they are aware of (i) effective work-environment practices that promote EDI and (ii) the EDI implications of their research (as indicated in the methods above). Finally, we will organize regular social activities to integrate HQPs, with timing and themes that respect physical, cultural, religious and family constraints. At the management level, measures will be taken to accommodate the schedules and location of work for HQPs with disabilities or family responsibilities, and to ensure equitable conference participation.

6. IMPACT AND BENEFITS TO CANADA

- Explain how and the extent to which the proposed research will generate new knowledge in the natural sciences or engineering disciplines and/or develop or advance new technologies.
- Considering the partner organizations' plans to use the research results, discuss how the project will lead to new or improved technologies, products, processes, services, policies, standards or regulations in Canada.
- Describe how and the extent to which the project's intended outcomes will lead to economic, environmental, and/or other societal benefits to Canada and Canadians.

In medical imaging, this program is among the initial but fast-growing efforts transitioning from narrowly supervised algorithms, with limited generalization capabilities, to larger-scale foundation models enriched with unlabeled data and expert's knowledge. This paradigm shift calls for new DL methods, modeling domain knowledge and mitigating the difficulties inherent to medical data. The applicants have had an outstanding, internationally visible research record in the field. Therefore, we expect this R&D program to 1) yield excellent academic publications; and 2) attract and train high-caliber HQPs at the intersection of AI, computer vision and medical imaging, both of which are high-demand sectors in Canada. This will contribute to keeping Canada at the cutting edge in these growing areas. It could benefit the private sector, enhancing job creation and economic activity, even beyond medical imaging, in a breadth of areas in industrial computer vision R&D. In retinal imaging, we intend to build internationally competitive algorithms, which enable the evaluation of over 100 conditions, expanding substantially the product portfolio of the partner and increasing its revenue. The number of users (i.e., optometrist and ophthalmologist clinics) of Diagnos commercial tools is steadily increasing, and is expected to reach 400 clinics in 2025 and beyond 4,000 clinics by the end of this R&D program. This growth could translate into a tenfold revenue increase, without considering the impact of addressing a much wider range of conditions, which could substantially increase the revenue generated per clinic. Also, the increased effectiveness of AI-based screening tools makes them appealing to healthcare institutions, potentially yielding wider adoption and implementations in the Canadian health systems. This could enable population-wide screenings and improve the early detection (hence, timely interventions) of a breadth of conditions and biomarkers from cost-effective images, potentially preventing or delaying vision loss and predicting life-threatening cardiovascular risks. Therefore, the outcomes could have substantial economic benefits to Canada, reducing health-care costs to taxpayers. For instance, according to [87], AI algorithms for early DR screening could save up to 85 CAD$ per patient in Quebec, which corresponds to a total saving of 68 million CAD$ for diabetic patients in the province. Furthermore, according to the American Academy of Ophthalmology, every dollar invested in preventing eye diseases generates an estimated saving of seven dollars for the healthcare system [87]. Finally, and beyond retinal imaging, this program could impact the wider area of medical imaging AI: The developed methods could benefit other clinical domains, such as radiology or histology, potentially improving a breadth of tasks pertaining to disease detection, diagnosis, treatment planning, and outcome prediction. The potential of improved health care research and practices from AI developments is huge, and so is its economic potential, with a market size evaluated at USD 19.27B in 2023, estimated at 26.69B in 2024 and anticipated to reach around USD 613.81B by 2034.

REFERENCES

- Use this section to provide a list of relevant literature references. Do not refer readers to websites for additional information on your proposal. Do not introduce hyperlinks.
- These pages are not included in the page count.

In the references' list below, bold font indicates papers authored by the co-applicants. The sign '' indicates papers resulting from previous collaborations between the co-applicants and the proposed industrial partner.*

[1] World report of vision, *World Health Organization*, 2019.

[2] Cen et al., Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks, *Nature Communications*, 2021.

[3] Zhou et al., A foundation model for generalizable disease detection from retinal images, *Nature* 2023.

[4] Tap et al., Associations of Retinal Microvascular Diameters and Tortuosity with Blood Pressure and Arterial Stiffness, *Hypertension*, 2019.

[5] Mookiah et al., Computer-aided diagnosis of diabetic retinopathy: A review, *Computers in Biology and Medicine*, 2013.

[6] Li et al. Applications of deep learning in fundus images: A review, *Medical Image Analysis*, 2021.

[7] Liu et al., Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge, *Patterns*, 2022.

[8] Galdran et al., Cost-Sensitive Regularization for Diabetic Retinopathy Grading from Eye Fundus Images, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2020.

[9] Wu et al., GAMMA challenge: Glaucoma grAding from Multi-Modality imAges, *Medical Image Analysis*, 2023.

[10] Lyu et al., AADG: Automatic Augmentation for Domain Generalization on Retinal Image Segmentation, *IEEE Transactions on Medical Imaging*, 2022.

[11] Yang et al., Anomaly-guided weakly supervised lesion segmentation on retinal OCT Images, *Medical Image Analysis*, 2024.

[12] *Galdran et al., State-of-the-art retinal vessel segmentation with minimalistic models, *Nature Scientific Reports*, 2022.

[13] *Liu et al., Do we really need dice? The hidden region-size biases of segmentation losses, *Medical Image Analysis*, 2024.

[14] Antonelli et al., The medical segmentation decathlon, *Nature Communications*, 2022.

[15] Finlayson et al., The clinician and dataset shift in artificial intelligence, *The New England Journal of Medicine*, 2021.

[16] Bai et al., Unsupervised Lesion-Aware Transfer Learning for Diabetic Retinopathy Grading in Ultra-Wide-Field Fundus Photography, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2022.

[17] *Silva-Rodriguez et al., A Foundation Language-Image Model of the Retina (FLAIR): Encoding Expert Knowledge in Text Supervision, *Medical Image Analysis*, 2025.

[18] Guo et al., On calibration of modern neural networks, *International Conference on Machine Learning (ICML)*, 2017.

[19] Minderer et al.: Revisiting the calibration of modern neural networks, *Neural Information Processing Systems (NeurIPS)*, 2021.

[20] Mukhoti et al., Calibrating deep neural networks using focal loss, *Neural Information Processing Systems (NeurIPS)*, 2020.

[21] Liu et al.: The devil is in the margin: Margin-based label smoothing for network calibration, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[22] Liu et al., Class adaptive network calibration, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[23] Murugesan et al., Trust your neighbours: Penalty-based constraints for model calibration, *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2023.

[24] Brown et al., Language models are few-shot learners, *Neural Information Processing Systems (NeurIPS)*, 2020.

[25] Radford et al., Learning Transferable Visual Models From Natural Language Supervision, *International Conference on Machine Learning (ICML)*, 2021.

[26] Wang et al., Image as a Foreign Language: BEIT Pretraining for All Vision and Vision-Language Tasks, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[27] Chen et al., A simple framework for contrastive learning of visual representations, *International Conference on Machine Learning (ICML)*, 2020.

[28] Caron et al., Emerging properties in self-supervised vision transformers, *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[29] Zhou et al., Learning to Prompt for Vision-Language Models, *International Journal of Computer Vision*, 2022.

[30] Yao et al., Visual-language prompt tuning with knowledge-guided context optimization, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[31] Huang et al., LP++: A Surprisingly Strong Linear Probe for Few-Shot CLIP, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[32] Silva-Rodriguez et al., A Closer Look at the Few-Shot Adaptation of Large Vision-Language Models, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[33] Zhang et al., Knowledge-enhanced visual-language pretraining on chest radiology images, *Nature Communications*, 2023.

[34] Huang et al., A visual–language foundation model for pathology image analysis using medical twitter, *Nature medicine*, 2023.

[35] Mazurowski et al., Segment anything model for medical image analysis: An experimental study, *Medical Image Analysis*, 2023.

[36] Silva-Rodriguez et al., Towards foundation models and few-shot parameter-efficient fine-tuning for volumetric organ segmentation, *MICCAI International Workshop on Foundation Models for General Medical AI (MedAGI)*, 2023.

[37] *Silva-Rodriguez et al., Exploring the Transferability of a Foundation Model for Fundus Images: Application to Hypertensive Retinopathy, *Computer Graphics International Conference (CGI)*, 2023.

[38] *Liu et al., A domain adaptation method for deep learning based automatic diabetic retinopathy grading, *Investigative Ophthalmology & Visual Science*, 2023.

[39] *A Galdran, H Chakor, W Kabir, R Kobbi, B Liu, J Dolz, I Ben Ayed, An automatic deep learning-based system for screening and management of DME, *Diabetes Research and Clinical Practice*, 2022.

[40] *Murugesan et al., Automatic screening and progress with AI-assisted OCT in retinal macular oedema detection, *Investigative Ophthalmology & Visual Science*, 2023.

[41] *Kabir et al., An AI-Assisted System for Early Screening of Age-Related Macular Degeneration (ARMD) from Real-World Fundus Images, *Diabetes Research and Clinical Practice*, 2024

[42] Qian et al., HRDC challenge: a public benchmark for hypertension and hypertensive retinopathy classification from fundus images, *The Visual Computer*, 2024.

[43] Fu et al., On the effectiveness of parameter-efficient fine-tuning, *AAAI Conference on Artificial Intelligence*, 2023.

[44] He et al., Sensitivity-Aware Visual Parameter-Efficient Fine-Tuning, *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

[45] Zanella and Ben Ayed, Low-Rank Few-Shot Adaptation of Vision-Language Models, *CVPR Workshop on Prompting in Vision*, 2024.

[46] Zhang et al., Training-free Adaption of CLIP for Few-shot Classification, *European Conference on Computer Vision (ECCV)*, 2022.

[47] Hu et al., LoRA: Low-Rank Adaptation of Large Language Models, *International Conference on Learning Representations (ICLR)*, 2022.

[48] Ding et al., Sparse Low-rank Adaptation of Pre-trained Language Models, *Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

[49] Dutt et al., Parameter-Efficient Fine-Tuning for Medical Image Analysis: The Missed Opportunity, *Medical Imaging with Deep Learning (MIDL)*, 2024.

[50] Huang et al., Unrolled variational Bayesian algorithm for image blind deconvolution, *IEEE Transactions on Image Processing*, 2022.

[51] Hajimiri et al., A Strong Baseline for Generalized Few-Shot Semantic Segmentation, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[52] Boudiaf et al., Few-Shot Segmentation Without Meta-Learning: A Good Transductive Inference Is All You Need? *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[53] Martin et al., Transductive Zero-Shot and Few-Shot CLIP, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[54] Ziko et al., Laplacian Regularized Few-Shot Learning, *International Conference on Machine Learning (ICML)*, 2020.

[55] Boudiaf et al., Transductive Information Maximization for Few-Shot Learning, *Neural Information Processing (NeurIPS)*, 2020.

[56] Ziko et al., Scalable Laplacian K-modes, *Neural Information Processing Systems (NeurIPS)*, 2018.

[57] Ziko et al., Variational Fair Clustering, *AAAI Conference on Artificial Intelligence*, 2021.

[58] Kervadec et al., Beyond pixel-wise supervision for segmentation: A few global shape descriptors might be surprisingly good, *Medical Imaging with Deep Learning (MIDL)*, 2021.

[59] Ru et al., Learning affinity from attention: End-to-end weakly-supervised semantic segmentation with transformers, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[60] Chen et al., Class re-activation maps for weakly-supervised semantic segmentation, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[61] Murugesan et al., Prompting classes: Exploring the Power of Prompt Class Learning in Weakly Supervised Semantic Segmentation, *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2024.

[62] Kervadec et al., Constrained-CNN losses for weakly supervised segmentation, *Medical Image Analysis (MedIA)*, 2019.

[63] Tang et al., On Regularized Losses for Weakly-supervised CNN Segmentation, *European Conference on Computer Vision (ECCV)*, 2018.

[64] Ting et al., Diabetic retinopathy: Global prevalence, major risk factors, screening practices and public health challenges: A review, *Clinical and Experimental Ophthalmology*, 2016.

[65] Decenciere et al., TeleOphta: Machine learning and image processing methods for teleophthalmology, *IRBM*, 2013.

[66] Li et al., Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening, *Information Sciences.*, 2019.

[67] Kervadec et al., Boundary loss for highly unbalanced segmentation, *Medical Image Analysis*, 2021.

[68] Sivaswamy et al., A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis, *JSM Biomedical Imaging Data Papers*, 2015.

[69] Orlando et al., REFUGE challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs, *Medical Image Analysis*, 2020.

[70] Fumero et al., RIM-ONE: An open retinal image database for optic nerve evaluation, *International Symposium on Computer-Based Medical Systems (CBMS)*, 2011.

[71] Rony et al., Augmented Lagrangian Adversarial Attacks, *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[72] Murugesan et al., Robust Calibration of Large Vision-Language Adapters, *European Conference on Computer Vision (ECCV)*, 2024.

[73] Murugesan et al., Calibrating Segmentation Networks with Margin-based Label Smoothing, *Medical Image Analysis*, 2023.

[74] Ricci-Lara et al., Addressing fairness in artificial intelligence for medical imaging, *Nature Communications*, 2022.

[75] Boudiaf et al., Parameter-free Online Test-time Adaptation, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[76] Bateson et al., Source-Free Domain Adaptation for Image Segmentation, *Medical Image Analysis*, 2022.

[77] Ben Ayed et al., Foundation Models for Medical Imaging, *Tutorial at the Medical Image Computing and Computer-Assisted Intervention Conference (MICCAI)*, 2024.

[78] Ben Ayed et al., Few-Shot Learning and Adaptation (In the Era of Large Image-Language Models), *Tutorial at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[79] Ben Ayed et al., Learning with Limited Supervision, *Tutorial at the Medical Image Computing and Computer-Assisted Intervention Conference (MICCAI)*, 2022.

[80] Wu et al., MM-Retinal: Knowledge-Enhanced Foundational Pretraining with Fundus Image-Text Expertise, *Medical Image Computing and Computer-Assisted Intervention Conference (MICCAI)*, 2024.

[81] Zanella et al., Boosting Vision-Language Models with Transduction, *Neural Information Processing Systems (NeurIPS)*, 2024.

[82] Computing Research Association, The CRA Taulbee Survey, **2022.**

[83] OECD dataset showing distribution of graduates by gender in fields of education (including the field of Information and Communication Technologies), **2021.**

[84] UNESCO dataset showing female share of graduates in Information and Communication Technologies, **2017-2018**.

[85] Randel et al., Inclusive leadership: Realizing positive outcomes through belongingness and being valued for uniqueness, *Human Resource Management Review*, **2018**

[86] Guide pour repérer les défis reliés à l'équité, la diversité et l'inclusion vécus par les membres d'une équipe de recherche, *CFSG chaire pour les femmes en science et en génie, Université de Sherbrooke*, **2023**

[87] Intelligence artificielle & santé: 13 cas d'entreprises québécoises pour transformer la santé, *Fiches de l'École de l'intelligence artificielle en santé du CHUM (ÉIAS)*, **2021**