End-to-end multi-task learning approaches for the joint epiretinal membrane segmentation and screening in OCT images

Mateo Gende [a,b], Joaquim de Moura [a,b,*], Jorge Novo [a,b], Marcos Ortega [a,b]

[a] *Grupo VARPA, Instituto de Investigación Biomédica de A Coruña (INIBIC), Universidade da Coruña, Xubias de Arriba, 84, 15006A Coruña, Spain*
[b] *Centro de investigación CITIC, Universidade da Coruña, Campus de Elviña, s/n, 15071A Coruña, Spain*

ABSTRACT

Background and objectives: The Epiretinal Membrane (ERM) is an ocular disease that can cause visual distortions and irreversible vision loss. Patient sight preservation relies on an early diagnosis and on determining the location of the ERM in order to be treated and potentially removed. In this context, the visual inspection of the images in order to screen for ERM signs is a costly and subjective process.
Methods: In this work, we propose and study three end-to-end fully-automatic approaches for the simultaneous segmentation and screening of ERM signs in Optical Coherence Tomography images. These convolutional approaches exploit a multi-task learning context to leverage inter-task complementarity in order to guide the training process. The proposed architectures are combined with three different state of the art encoder architectures of reference in order to provide an exhaustive study of the suitability of each of the approaches for these tasks. Furthermore, these architectures work in an end-to-end manner, entailing a significant simplification of the development process since they are able to be trained directly from annotated images without the need for a series of purpose-specific steps.
Results: In terms of segmentation, the proposed models obtained a precision of 0.760 ± 0.050, a sensitivity of 0.768 ± 0.210 and a specificity of 0.945 ± 0.011. For the screening task, these models achieved a precision of 0.963 ± 0.068, a sensitivity of 0.816 ± 0.162 and a specificity of 0.983 ± 0.068. The obtained results show that these multi-task approaches are able to perform competitively with or even outperform single-task methods tailored for either the segmentation or the screening of the ERM.
Conclusions: These results highlight the advantages of using complementary knowledge related to the segmentation and screening tasks in the diagnosis of this relevant pathology, constituting the first proposal to address the diagnosis of the ERM from a multi-task perspective.

1. Introduction

Recent years have seen significant advances in the development of new and more efficient Computer-Aided Diagnosis (CAD) systems. The advent of modern artificial intelligence algorithms based on deep learning, coupled with the growth in computing capacity, is enabling enormous progress in the field of medical image analysis (Kim et al., 2018; Shen et al., 2017; Litjens et al., 2017). Current deep learning architectures allow for an efficient extraction of visual features from medical images. With these features, deep learning-based CAD systems are able to aid clinicians by performing tasks such as the automatic screening of patients (Cheung et al., 2019; Nielsen et al., 2019) or the

segmentation of pathological or anatomical structures that are relevant for the diagnosis and treatment of several diseases (Fu et al., 2018; Liu et al., 2021; Morano et al., 2021). These systems have demonstrated that their performance can be on par with or even better than that of medical experts (Litjens et al., 2017; Fauw et al., 2018; Gulshan et al., 2016; Ting et al., 2017; Lee et al., 2020).

In the context of medical imaging, Optical Coherence Tomography (OCT) is a non-invasive imaging technique that allows the visualisation of cross-sectional images of tissue at high resolution (Huang et al., 1991). To generate these images, an OCT scanning platform sweeps the tissue with low coherence light, obtaining a depth-wise reading or A-Scan at each point. These readings or samples can be averaged and

combined to form the B-Scans, tomograms or cross-sectional images of the tissue, showing its histological structure. These can be further composed laterally to form three-dimensional C-Scans or OCT volumes that contain a complete visualisation of the analysed tissue (Figure 1). Due to its non-invasiveness and the fact that it uses light to capture images, OCT is an especially valuable technique to study macular structures and diseases (Puliafito et al., 1995). Among its applications, OCT imaging can be used to diagnose different relevant ocular pathologies such as age-related macular degeneration (Kermany et al., 2018; Lee et al., 2017), the leading cause of permanent vision loss in the elderly; diabetic macular oedema (de Moura et al., 2020; Dysli et al., 2019), the leading cause of blindness in the diabetic population; glaucoma (Triolo and Rabiolo, 2020; Mohammadzadeh et al., 2020), the global leading cause of irreversible blindness; or the Epiretinal Membrane (ERM), among others.

The ERM, alternatively known as cellophane maculopathy or macular pucker, is an ocular condition in which a thin, fibrocellular layer develops over the eye retina. It appears over the boundary region between the vitreous humour and the retina. This region is known as the Inner Limiting Membrane (ILM). When the ERM layer starts to harden and contract, it exerts traction over the macula, producing wrinkles or puckers which may translate into a blurring or deformation of the vision, also known as metamorphopsia (Matsumoto et al., 2003). These deformations may also cause a thickening of the macular tissue or even the disappearance of the foveal pit. The key cause for its appearance is the retraction of the vitreous gel from the macula, a process known as posterior vitreous detachment (Johnson, 2010; Snead et al., 2008). Alternatively, the ERM can also appear as a secondary factor to other related ocular diseases, such as diabetic macular oedema (Diabetic Retinopathy Clinical Research Network, 2010; Ophir et al., 2010; Flaxel et al., 2010; Yamamoto et al., 2001; Ghazi et al., 2007), which has a high prevalence of related ERM appearance; or macular hole (Lee et al., 2015; Messmer et al., 1998; Akiba et al., 1996). Treatment for symptomatic ERM presence involves its extraction via *pars plana* vitrectomy. This surgery can help to preserve the sight of the patients. However, many of the deformations caused by ERM contraction may be irreversible, causing permanent vision loss in patients who receive a late diagnosis. Thus, an early diagnosis is paramount in preserving patient vision (Massin et al., 2000; Suh et al., 2009; Rahman and Stephenson, 2014).

The ERM is mostly transparent. Because of this, it is difficult to spot in eye fundus images. In many cases, it cannot be detected until its effects start to develop. In OCT images, however, the ERM appears as a bright layer over the ILM. The appearance of the ERM in OCT images is illustrated in Figure 2. This makes it possible to detect its presence before it starts to deform the macula of the patient and cause irreversible
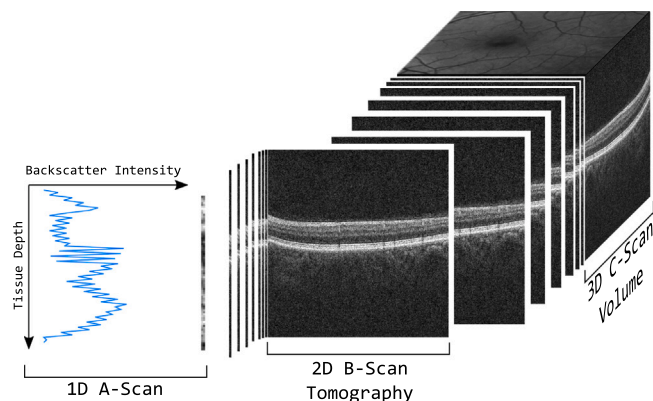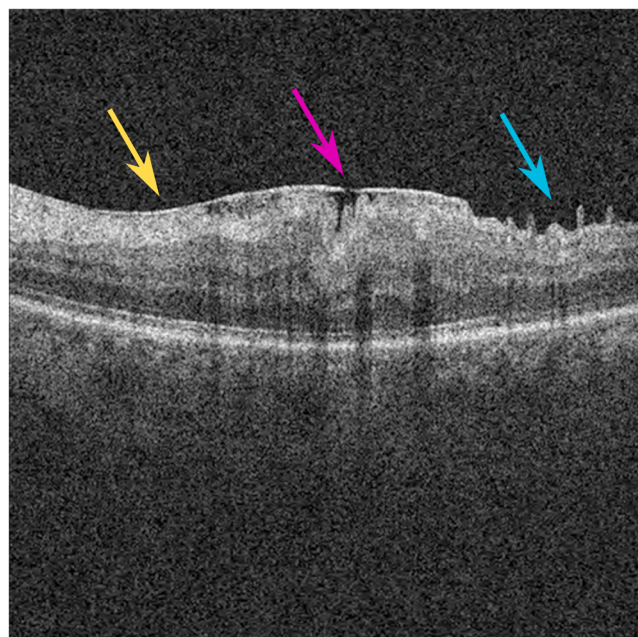

Fig. 2. OCT slice showing ERM presence. *Yellow*: attached ERM. *Magenta*: detached ERM. *Cyan*: deformation caused by ERM contraction.

damage. Thanks to this, OCT has become a standard imaging technique used for the diagnosis and detection of the ERM, allowing an early detection and assessment of its presence, as can be seen in several clinical studies, (for reference (Zhang et al., 2021; Poornachandra et al., 2020; Menteş and Nalçací, 2020). OCT is also used for the pre-operative and post-operative management of patients, helping to predict and explain the visual outcomes from surgery (Goldberg et al., 2014). Furthermore, several studies highlight the relevance of this imaging technique for surgical decision making, planning and assessment in relation to this pathology (Do et al., 2007; Hirano et al., 2010; Falkner-Radler et al., 2010). These studies outline the relevance of detecting the ERM in OCT images in order not only to provide an early diagnosis, but also to study its progression, to estimate the visual outcomes from surgery and, overall, to determine the way it may have an impact on patient care (Chua et al., 2022).

The process of detecting the ERM in OCT images is typically performed by an ophthalmology expert visually inspecting each image. This tedious and time-consuming task can also lead to subjectivity in the diagnosis. Additionally, the existence of a bright spot over the ILM in an OCT image does not univocally mean the presence of ERM, since the effects of the backwards and forwards scattering of light in this imaging technique have a tendency to cause visual artefacts. These commonly take the appearance of speckle noise and bright spots in the images. Because of this, the correct identification of the ERM is a non-trivial matter that requires a careful analysis of the signs of the disease at different scales.

The challenge presented by the objective detection of the ERM in OCT images can be taken as an opportunity for the application of CAD systems that can provide support for the clinicians in performing an early and accurate diagnosis, implying the appropriate treatment. Because of the relevance of a disease such as the ERM, different works have addressed its analysis in the literature. The earlier works, such as the proposal by Wilkins et al. (Wilkins et al., 1996) were semi-automatic and based around the refinement of a manual annotation made by the expert who had to visually inspect the images.

More recent studies have introduced automatic methods, overcoming the reliance on an initial expert annotation. For reference, with regards to the screening of the ERM, Fang et al. (2017) proposed the classification of multiple macular lesions including the ERM in OCT

Fig. 1. Composition of OCT scans. One dimensional A-Scans containing information relative to the light backscattering measurements can be represented as picture elements and laterally combined into an array forming two dimensional B-Scans or tomograms. B-Scans can be composed by stacking them to form OCT volumes or C-Scans.

images. For their detection, they performed a conventional feature extraction in segmented regions of interest. These features were then used to classify these regions by a machine learning-based system. In the work by Lu et al. (2018), they used deep learning techniques to detect four different pathologies, also including ERM. The authors trained a Convolutional Neural Network (CNN) to classify images into a single pathology and compared their results with those obtained by two experts. The authors reported that the system achieved results comparable to or better than those of the human experts. Kuwayama et al. (2019) studied the application of CNN models for the classification of OCT images. These images showed signs of different retinal diseases, with the ERM being one of them. The convolutional models were able to classify the images into the different pathologies that were considered. The study by Lo et al. (2020) proposed the use of a Residual Neural Network architecture (He et al) for the screening of the ERM in cross-sectional OCT images. The authors reported that the model performed slightly better than non-retinal specialised ophthalmologists. In their work, Sonobe et al. (2018) compared the performance of Support Vector Machines (SVMs) and deep learning methods for the detection of ERM cases in OCT volumes. Their results showed that the deep learning models performed much better than the SVMs. Parra-Mora et al. (2021) used four classifying convolutional neural network architectures for the detection of the ERM in OCT slices. Their results showed high performance in discriminating positive and negative ERM cases. Finally, the works by Baamonde et al. (2017a,b), Baamonde et al. (2019a,b) propose a series of methods for the characterisation of the ERM in OCT images by means of the classification of a series of conventional features.

Conversely, fewer works have addressed the segmentation of the ERM. In Baamonde et al. (2019b), the authors were able to automatically segment the ERM in OCT images by extracting a set of 452 conventional visual features (texture-based, intensity-based and domain-related) from a pre-segmented region of interest and then using the Spatial Uniform ReliefF algorithm to select the ones that contributed the most to the classification process. In their work, they compare the results of using random forest, k-nearest neighbours and SVMs for the classification of different sets of selected features.

In Gende et al. (2021), the segmentation of the ERM by means of the classification of a series of windows extracted around the ILM is proposed. This process involves a set of three steps starting with the segmentation of the region of interest and the extraction of the windows, their classification and the reconstruction of the segmentation. A Densely Connected Convolutional Network (Huang et al., 2017) is used to classify each of these windows. These classifications are then used to reconstruct a segmentation map of the whole eye fundus. The deep learning model showed a greater ability to integrate the surrounding visual information when classifying each window than the classical machine learning methods explored by Baamonde et al. (2019b).

Overall, the works in the literature have shown promising results for the application of deep learning models in the detection of the ERM. Most of these works, however, are limited in their scope to the classification of whole OCT slices into whether the ERM is present or not, with only (Baamonde et al. 2019b) and (Gende et al., 2021) providing a segmentation of the pathological region in the OCT volumes. These two works depend upon a sequence of steps in order to restrict the region to be analysed, effectively obtaining the segmentation by means of a series of patch classifications and a reconstruction of each of the slices. These steps need to be adjusted and may be unreliable when presented with various artefacts or anatomical structures present in macular OCT images. Moreover, due to their reliance on a classification of a series of sliding windows extracted from the images, these methods limit the area to be scanned overall, since the edges of the images cannot be analysed properly. To date, no end-to-end methods for the ERM segmentation have been proposed. Furthermore, there have been no proposals for the diagnosis of the ERM in OCT images that employ multi-task learning.

In this work, we present a study in the application of multi-task deep learning techniques for the joint segmentation and screening of the ERM

in OCT images. In a set of experiments, we compare a series of end-to-end approaches that take advantage of inter-task complementarity in different ways in order to provide an objective and accurate segmentation and screening of the ERM. The first of these approaches leverages the potential of the innermost, highest-level features of the segmentation encoder for the screening of the ERM. This allows the classification head to exploit the most refined visual feature maps that are used in the segmentation. The second approach is based on the use of the complete set of visual feature maps extracted from the images for the classification. This way, taking an extensive approach to the multi-task context by making all of the visual information that is propagated for the segmentation at different levels available to the screening task. Finally, the third approach restricts the ERM screening task to only employ the output maps produced by the segmentation task. This proposal takes the opposite approach to the second one, taking on a restrictive focus and only allowing the visual information that is common to both tasks to be propagated to the classification head. Each of these approaches is tested with three different encoder configurations, representative of the state of the art in medical image classification and segmentation. These architectures make use of deep learning and the ability of convolutional neural networks to be trained directly from the annotated data. This way, the models are able to implicitly learn to extract and select the visual features that are relevant to the task. This eliminates the necessity of performing this process manually through conventional hand-crafted features or the use of feature selection strategies. Since these models function in an end-to-end manner, they are able to produce the segmentation and the classification of the ERM from the images without the need for a delicate series of fine-tuned intermediate steps. Not only does this provide a more reliable and robust system, but it also entails a considerable speedup with regards to the existing methods for the segmentation of the ERM. This also greatly simplifies the development of CAD systems that make use of these approaches, without relying on the tuning and adjustment of each of the steps, conferring them significant advantages during training and inference over the OCT slices. To the extent of our knowledge, this work is the first in the literature to tackle ERM detection using end-to-end multi-task approaches.

This manuscript is structured as follows: In Section 2 we describe the materials and the convolutional architectures employed in this work. Section 3 presents the experimental results and their discussion. Finally, in Section 4 we present the conclusions that were reached and describe possible future lines of work.

2. Materials and methods

In this section, we provide a description of the resources that were used in this study. In particular, the reader will find information about the dataset that was used (Subsection 2.1), as well as the base segmentation architecture that was selected (Subsection 2.2), a detailed description of each of the proposed approaches (Subsection 2.3), the different encoder configurations that were combined with the proposed approaches (Subsection 2.4), the software and hardware resources employed (Subsection 2.5) and finally a detailed explanation of the training and evaluation strategy that was followed (Subsection 2.6).

2.1. Dataset

For the development of this study, an anonymized set of 20 OCT volumes obtained from 20 different patients was used. All the OCT volumes were acquired using a CIRRUS™ HD-OCT Carl Zeiss Meditec confocal scanning laser ophthalmoscope. These images were acquired in accordance with the Declaration of Helsinki, as approved by the local Ethics Committee of Investigation from A Coruña/Ferrol (2014/437) the 24th of November 2014.

Out of the 20 OCT volumes, 8 of them present ERM, while the remaining 12 belong to healthy eyes. These volumes were formed from a total of 2.427 2D B-Scans. 809 of these B-Scans displayed symptoms of

ERM, while the remaining 1.618 showed no signs of ERM. All of the OCT tomograms were manually labelled by an expert, indicating the position of the ILM and the presence of the ERM. A total of 1.308.160 image columns were annotated with ILM position, 251.994 of them showed ERM-related lesions and 1.056.166 contained healthy tissue. These annotations were used as the ground truth for the segmentation and classification process. These B-Scans originally ranged between 479 × 501–693 × 676 pixels of resolution, before being resized to a standard 512 × 512 pixels. Figure 3 displays some examples of OCT slices and their corresponding annotations.

2.2. Network architecture

In this work, we study the application of different end-to-end multi-task deep neural network architectural approaches for the detection of the ERM. These approaches are based on the Feature Pyramid Network (FPN) model (Lin et al., 2017), adapted for semantic segmentation. The FPN models allow for the extraction of visual features at different scales, building high-level semantic feature maps and making predictions at each of them. Feature maps at different scales have proven to be useful for detection and segmentation tasks in medical imaging, and several variations of the original FPN can be found in the literature (for reference (Dimitrakopoulos et al., 2019; Shao et al., 2019; Geng et al., 2020; Cao et al; Chen and Liu). These can be used for the characterisation of the different signs of ERM presence in OCT images at different scale levels, such as its appearance as a hyper-reflective layer, the thickening of the macular tissue or the deformations it causes around the fovea. The basic structure of an FPN adapted for segmentation can be found in Figure 4.

2.3. End-to-end multi-task approaches

The base architecture was extended into three different approaches specifically designed in order to leverage the advantages that inter-task commonality may provide via multi-task learning for the screening and segmentation of the ERM. These architectures approach the multi-task learning in different ways, with the aim of exploring in which way the visual features that are relevant for each of the individual tasks can be of use for the other. These approaches are described as follows:

- **Multi-task via the inner features:** The first approach consists of a classification head that accepts the innermost features of the encoder of the FPN and returns a classification into whether the image displays ERM symptoms or not. This way, the innermost, highest level features can be employed for the classification task, taking advantage of the progressive feature refining of the encoder. This classification

head consists of an adaptive average pooling layer followed by a fully connected layer with two outputs. This structure can be found in other multi-task segmentation and classification neural network architectures (for reference (Amyar et al., 2020; Zhou et al., 2021; Playout et al., 2019). Figure 5 shows a diagram illustrating this variant.

- **Multi-task via the decoder features:** The second approach uses all of the decoder features for the classification. In this structure, all the features employed for the ERM segmentation map are used as inputs in the screening task. The complete set of features is fed to a classification head analogous to the one used in the first approach by employing an additional encoder. In this way, features extracted at different scale levels can be combined with each other in order to obtain an accurate ERM screening. By providing the complete set of feature maps used for segmentation to the screening task, this additional encoder can take advantage of all the visual information that may be relevant to this task, contrasting with the limitation that the first base approach imposes on using only the highest level features. This architecture takes an extensive approach to multi-task learning. Figure 6 summarises this approach.

- **Multi-task via the segmentation maps:** The final approach consists of using the output segmentation maps as an input for the screening task. Since the two outputs of the multi-task model are closely related, we can leverage the segmentation maps as refined features for the ERM classification. As with the second approach, an additional encoder head is used to analyse these maps and produce a classification output. This approach takes a different focus for exploiting multi-task learning. Instead of expanding the feature set to the entirety of the feature maps used for the segmentation, this third approach limits the information that can be employed for the screening to only the output of the segmentation. This way, the screening task is closely guided by the segmentation task, approaching multi-task learning by focusing on the restriction to inter-task commonality. A summary of the structure of this approach can be found in Figure 7.

Regarding the outputs of these approaches, the classification head of each model was tasked with classifying the input B-Scans into whether they display ERM-related tissue or not. Meanwhile, the segmentation heads return two masks with an identical resolution to the original image. These masks display the ILM position and ERM presence, akin to what is shown in Figure 3. The reason for the addition of this ILM-related output mask to the models is twofold. First, due to the overwhelming imbalance that exists in the images between diseased tissue and background. This imbalance is caused by multiple factors, among them, the relative thinness of the ERM when compared to the image height, the
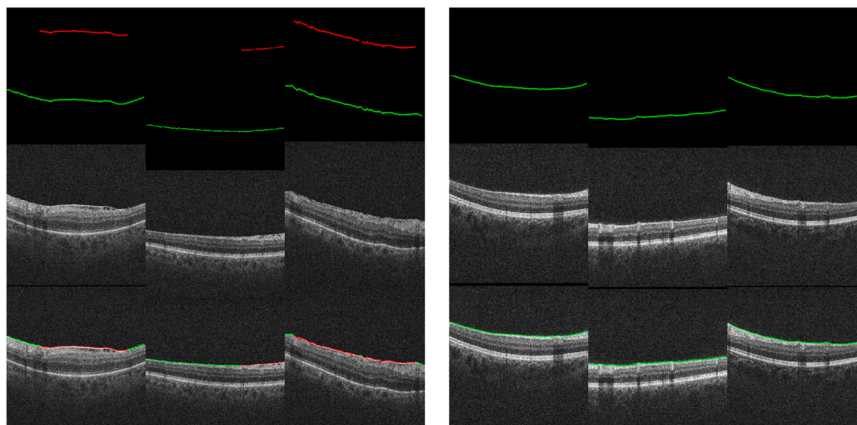

Fig. 3. Images belonging to the dataset, with corresponding annotations. 1st row: ERM annotation. 2nd row: ILM annotation. 3rd row: original OCT B-Scan. 4th row: Composition of the original image with its annotations.
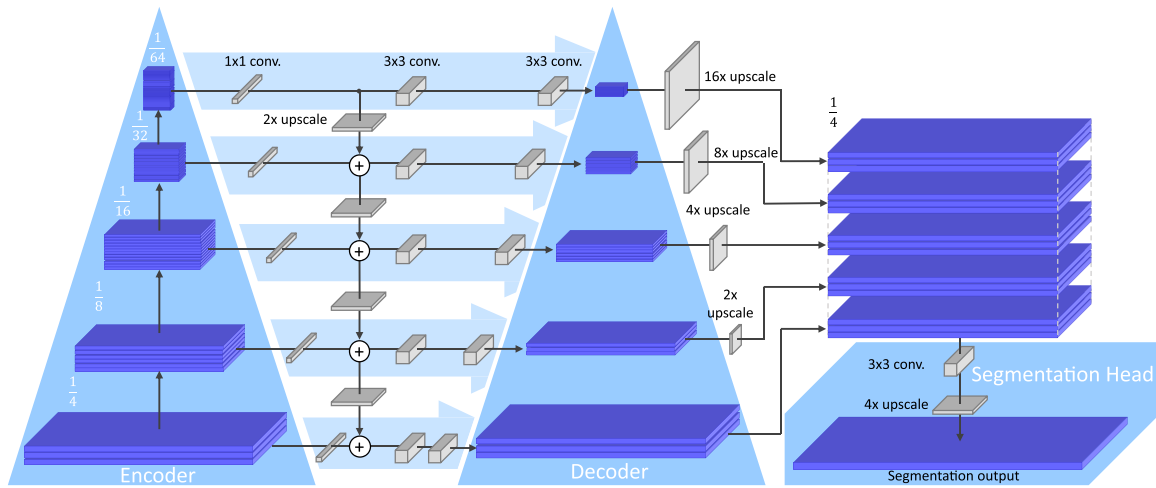
Fig. 4. Basic structure of a FPN architecture adapted for segmentation. At each scale level, a 1 × 1 convolution operation is applied to the features before they are upscaled and concatenated with those of higher levels. These features are then upscaled and fed to a segmentation head which produces the outputs.
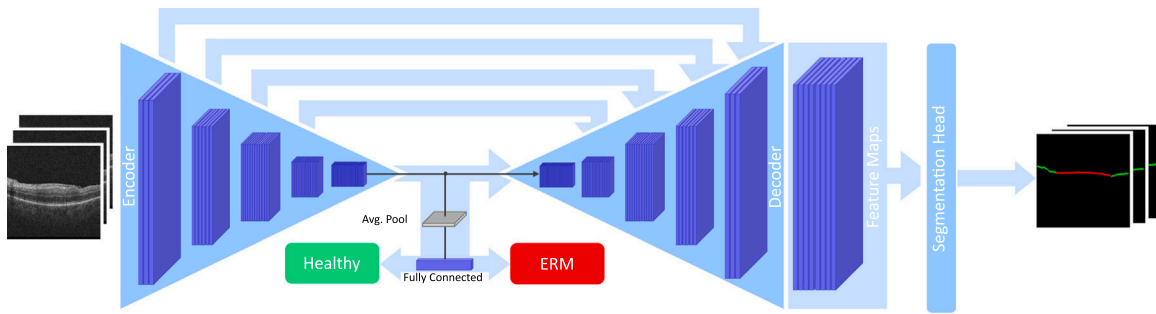

Fig. 5. First Approach: Multi-task via the inner features. The innermost features of the encoder network are used for the classification. This way, the combination of the segmentation encoder and the classification head form a classifier model.
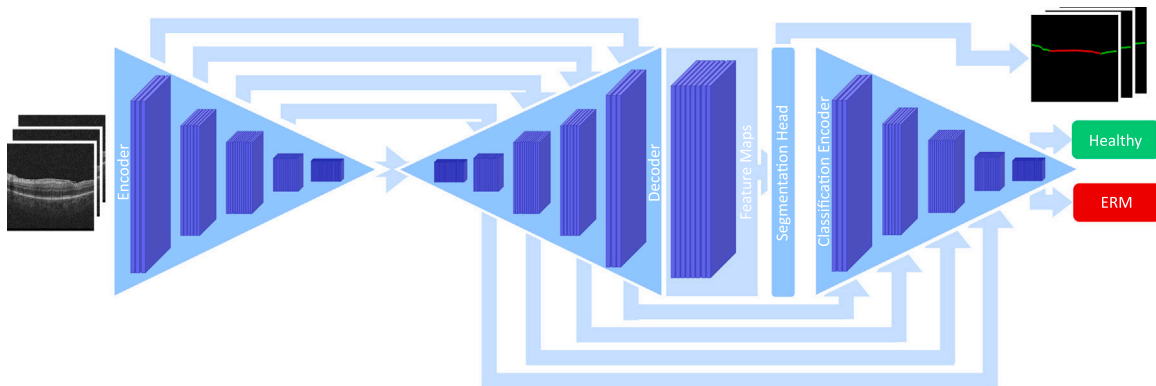

Fig. 6. Second Approach: Multi-task via the decoder features. An additional encoder is used before the classification head. All the feature maps of the network decoder are used for the screening task.

fact that much of the surface of an eye presenting ERM is otherwise healthy and the existing imbalance between healthy and ERM eyes in the dataset. ERM-related positive cases represented around 0.02% of the total pixels in the dataset. Second, by instructing the network to return domain-relevant features, such as the ILM is for ERM segmentation, these can be propagated forward through the network, instead of being regarded in the same way as actual image background such as speckle noise in the vitreous humour.

2.4. Encoder configurations

Each of these approaches weas tested with three different encoders, conforming nine different configurations. Due to the need for an additional encoder for the 2nd and 3rd approaches, the same architecture used to encode the segmentation features was used for screening. In order to perform an exhaustive analysis, these encoders conform a representative set of neural network architectures commonly applied in similar domains in the existing literature (Gende et al., 2021; Khan et al., 2019; Nibali et al., 2017; Apostolopoulos et al. 2020; Samala et al. 2019): .
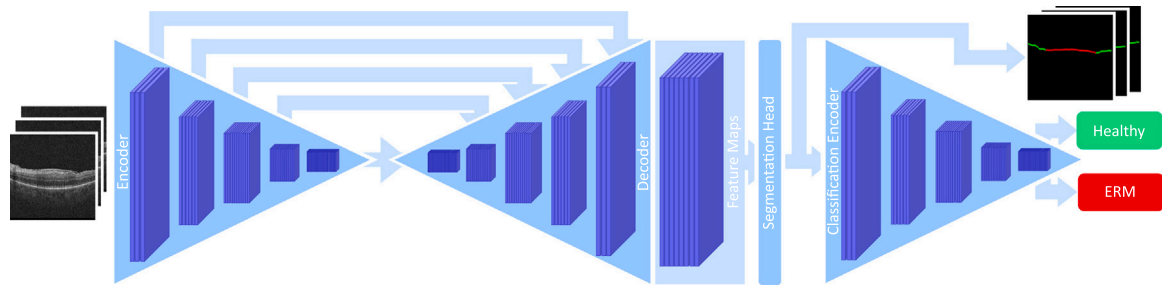
Fig. 7. Third Approach: Multi-task via the segmentation maps. The output segmentation maps of the network are used as inputs for the screening task.

- Densely Connected Convolutional Network (Huang et al., 2017): These networks are based on dense blocks, and the progressive concatenation of the features of earlier layers in order to propagate them ahead. For this work, a DenseNet-121 model was used. DenseNet models consist of four dense blocks, with a transition layer between each of them consisting of a 1×1 convolutional layer and 2×2 max-pooling. For a DenseNet-121 model, these dense blocks are formed by 6, 12, 24 and 16 sets of convolutions each of them formed by a 1×1 followed by a 3×3 convolution. Rectified Linear Unit is used for activation at each layer.
- Residual Neural Network (He et al., 2016): ResNet networks formulate layers as residual functions, similarly propagating features forward by means of identity skip connections. A ResNet-34 model was selected. This model starts with a 7×7 convolutional layer, followed by a 3×3 max-pooling. After this initial layer, this architecture is comprised of four convolutional blocks of 3, 4, 6 and 3 3×3 convolutional layers. Rectified Linear Unit is used for activation after every skip connection.
- Google Inception Network (Szegedy et al., 2015): Inception networks make use of the *network in network* concept (Lin et al., 2014), replacing deeper convolutional layers with inception blocks that perform convolutions using different filter sizes. For this study, the Inception-v4 architecture was selected. This architecture uses a stem block followed by four Inception-A blocks, seven Inception-B blocks and three Inception-C blocks. Each of these Inception blocks is formed by a series of parallel convolutions which are concatenated at the end before using Rectified Linear Unit for activation. For more details of the structure of each Inception block, please refer to the original paper Szegedy et al. (Szegedy et al., 2017).

A summary of the layer configurations, the memory footprint and the number of trainable parameters of each of the proposed approaches combined with each of the aforementioned encoder configurations can be found in Table 1. For the specific configuration of each encoder architecture, please refer to the original papers.

2.5. Software and hardware resources

This work was developed using Python (version 3.7.7) due to the flexibility that it allows. Moreover, it allows the use of the PyTorch (Paszke et al., 2019) machine learning library (version 1.7.1), which was used for the training and validation of the models. The implementations for the encoders were obtained from the Segmentation Models PyTorch library (Yakubovskiy, 2020). For image processing and data augmentation needs, OpenCV (version 3.4.8) and Numpy (version 1.15.0) were used. With regards to hardware, the training and evaluation process was carried out using an Intel Xeon E5–2640 CPU, an NVIDIA GeForce TITAN X GPU and 64 GB of RAM.

2.6. Training details

To study and compare the different approaches and their applicability for the ERM detection, we partitioned the original dataset following 4-fold cross-validation at the eye level, ensuring that each eye appears in the test set one time. This way, a model was trained for each of the three approaches, using each of the three considered encoder configurations, on each of the four data partitions, for a total of 36 trained models.

With regards to the training process, the encoder models were first initialised with a pre-training on the ImageNet dataset (Deng et al., 2009). This way, making the most of the available data and allowing the convolutional models to be initialised with common visual feature patterns that may be useful for the tasks. Then, each complete model was trained in our dataset for a maximum of 75 epochs, choosing for testing the model at the epoch which produced the smallest validation loss. To balance the more numerous healthy samples, these were undersampled during the training. Binary cross-entropy loss, commonly used for segmentation tasks in the literature (for reference, (Liu et al., 2021; Morano et al., 2021)), was employed as segmentation loss 1.

Table 1

Summary of the activation functions, the estimated memory footprint and the amount of trainable parameters for the proposed approaches with each of the encoder configurations.

	Encoder activation function	Classification activation function	Encoder architecture	Estimated memory footprint (MB)	Total parameters
First Approach Multi-task via Inner Features	Rectified Linear Unit	Sigmoid	DenseNet-121	1091.81	9301,892
			ResNet-18	371.22	13,048,388
			Inception-v4	1091.80	43,573,796
Second Approach Multi-task via Decoder features	Rectified Linear Unit	Sigmoid	DenseNet-121	1177.89	16,647,748
			ResNet-18	396.24	16,222,660
			Inception-v4	1195.36	59,536,164
Third Approach Multi-task via Segmentation Maps	Rectified Linear Unit	Sigmoid	DenseNet-121	2026.63	16,252,612
			ResNet-18	579.47	15,827,524
			Inception-v4	1790.94	59,499,876

$$\ell_{seg,c}(out^s, tar^s) = mean(\{l_{1,c}, ..., l_{N,c}\}^\top)$$
$$l_{n,c} = -w_{n,c}\left[p_c tar^s_{n,c} \cdot log\ \sigma(out^s_{n,c}) + \left(1 - tar^s_{n,c}\right) \cdot log\left(1 - \sigma(out^s_{n,c})\right)\right] \quad (1)$$

where out^s is the segmentation output and tar^s the segmentation ground truth, with p_c, the weight for the positive cases of each class, being 1 for the ILM and 2 for the ERM pixels. Preliminary trials showed that weighing the ERM segmentation loss twice over the ILM provided a good balance between over- and under-segmentation of the ERM. For the screening task, cross-entropy loss was used (Eq. 2).

$$\ell_{cla}(out^c, tar^c) = -out^c[tar^c] + log\left(\sum_j \exp(out^c[j])\right) \quad (2)$$

where out^c is the output class vector and tar^c the ground truth class. These two losses contributed equally in training (Eq. 3).

$$\ell_{total} = \ell_{seg} + \ell_{cla} \quad (3)$$

For all of the models, optimisation was done using Adaptive Moment Estimation (Adam) (Kingma and Ba, 2015), with a learning rate of 1×10^{-5}, $\beta_1 = 0.9$ and $\beta_2 = 0.999$. During the training process, online augmentation was used for the training images. This augmentation consisted of random vertical and horizontal shear operations, horizontal flips, rotations and changes to overall image brightness. A summary of the training conditions that were employed can be found in Table 2.

After training, the models were tested on their corresponding test set according to different statistical metrics commonly reported for similar medical imaging tasks in the literature. The Accuracy (Eq. 4); Sensitivity, or Recall (Eq. 5); Specificity (Eq. 6); Precision (Eq. 7); Dice coefficient, or F-Score (Eq. 8) and Jaccard Index (Eq. 9) of each approach and encoder combination was computed and compared. For the segmentation task, these metrics were evaluated with respect to the annotated columns of the dataset. For the classification task, these metrics were evaluated at the image level, where a positive is an image showing ERM signs.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} \quad (5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Dice = F\text{-}Score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (8)$$

$$Jaccard = \frac{TP}{TP + FP + FN} \quad (9)$$

Finally, these proposed approaches were compared with single-task segmentation and screening methods.

3. Results and discussion

In this section, we cover all the results that were acquired throughout the evaluation of the different multi-task approaches. Additionally, we provide a discussion of said results and a comparison with other single-task methods for ERM segmentation and screening.

3.1. Segmentation results

Regarding the segmentation task, Figure 8 displays the Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves for all the approaches proposed in this work including the areas under each of the curves. Values for each of the considered metrics can be found in Table 3.

The performed evaluation shows that the proposed approaches achieve satisfactory results in terms of segmentation. The first approach, using the encoder inner features obtained positive results, with the best overall segmentation Dice coefficient of 0.757 ± 0.1 being achieved by the ResNet-18 encoder using this approach. The second and third approaches fared comparatively worse, while still achieving competitive results. Since these approaches use, respectively, all of the feature maps and the output segmentation maps as inputs for the classification task, the segmentation task seems to be more affected during the multi-task training process. In this sense, the first approach which acquires the features used for classification from an earlier layer obtained the highest results in terms of Dice score. Figure 9 displays some examples of the segmentation produced by the best performing model. Overall, however, while differences exist between the results of the proposed approaches, most of the configurations using the DenseNet and ResNet encoders describe similar Precision-Recall curves, which are less biased towards the existing imbalances in the dataset. Those configurations that use the Inception-v4 encoder, however, produced comparatively worse results.

3.2. Screening results

With regards to the ERM screening task, ROC and PR curves can be found in Figure 10. Table 4 shows the test results for each of the approaches in terms of the metrics considered.

The results that were obtained for the ERM screening were also satisfactory. Contrasting with the segmentation task, the best overall results were achieved by the second and third approaches using the DenseNet-121 encoder, with F-Scores of 0.881 ± 0.108 and 0.881 ± 0.110, respectively. Between these two approaches, the third approach which restricted the classification inputs to only the segmentation output maps seems to provide more consistent results. This could be indicative that guiding the two tasks closely by restricting the shared features only to that which is common to the two can provide more benefits than sharing all of the inner features as is done in the second proposed approach. The first approach, which only shared the inner, most refined features of the segmentation encoder for the classification task, also achieved close results, with the ResNet-18 configuration achieving an F-Score of 0.847 ± 0.106. Overall, all three approaches achieved satisfactory results when using the DenseNet-121 encoder, while results for the Inception-v4 encoder were generally lower.

3.3. Discussion

These results show that the proposed end-to-end multi-task approaches can provide significant advantages for the diagnosis of the ERM in OCT images. While some of the configurations fared comparatively better than others, all of the proposed approaches were able to perform on par with or even better than the single-task methods in comprehensive metrics.

For the segmentation task, a comparison can be made with the existing state of the art (Gende et al., 2021). This method uses a series of three steps and obtains the segmented area via the classification of a set of windows extracted from a pre-segmented ILM. In order to provide a fair comparison, a window was extracted for each of the ILM pixels and

Table 2
Training and augmentation parameters.

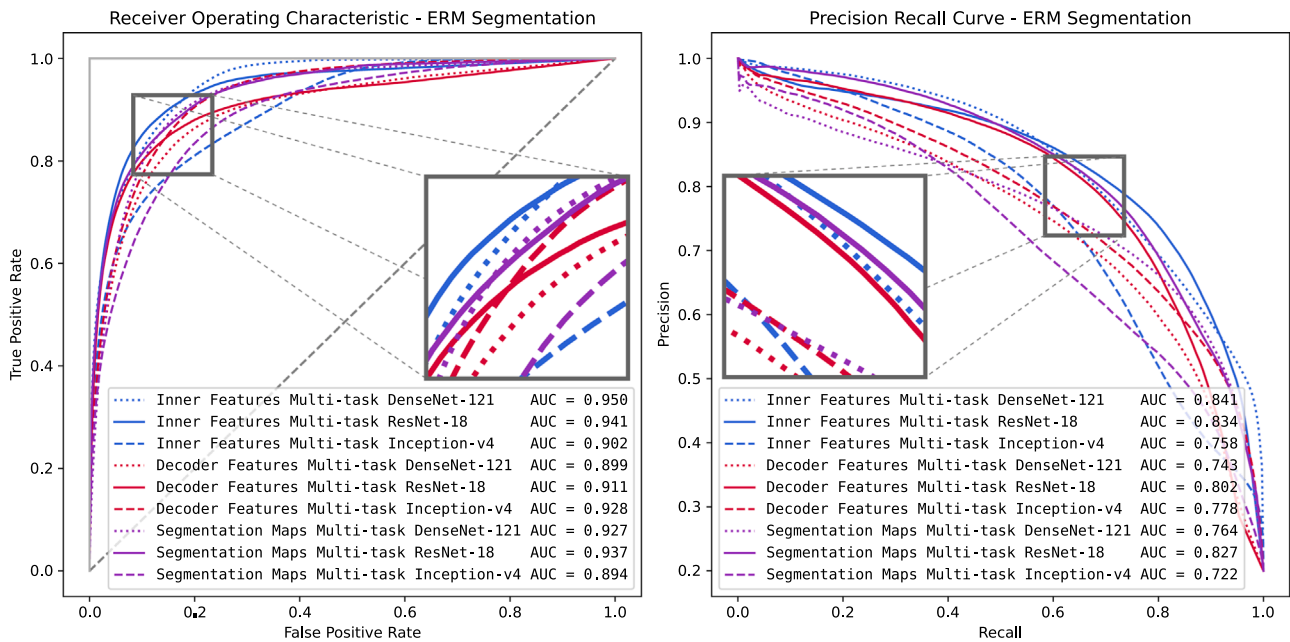| Image size	512×512	Loss	Cross-entropy
Shear aug.	$[-15^o, 15^o]$	Max. epochs	75
Rotation aug.	$[-15^o, 15^o]$	Optimiser	Adam
Brightness aug.	$[\times 0.75, \times 1.25]$	Learning rate	1×10^{-5}
Horizontal rotation	$p = 0.5$	$\beta 1$, $\beta 2$	0.9, 0.999

Fig. 8. Receiver operating characteristic and precision recall curves for the ERM segmentation task.

Table 3
ERM segmentation task results. In bold, the best result for Jaccard index and dice coefficient.

|	Encoder	Accuracy	Sensitivity	Specificity	Precision	Jaccard	Dice
First approach Multi-task via inner features	DenseNet-121	0.906 ± 0.023	0.659 ± 0.161	0.965 ± 0.027	0.831 ± 0.079	0.568 ± 0.095	0.721 ± 0.080
	ResNet-18	0.910 ± 0.047	0.768 ± 0.210	0.945 ± 0.011	0.760 ± 0.050	0.622 ± 0.165	**0.757 ± 0.140**
	Inception-v4	0.870 ± 0.054	0.703 ± 0.314	0.910 ± 0.050	0.646 ± 0.080	0.503 ± 0.201	0.649 ± 0.204
Second approach Multi-task via decoder features	DenseNet-121	0.883 ± 0.048	0.643 ± 0.281	0.943 ± 0.045	0.742 ± 0.081	0.506 ± 0.199	0.652 ± 0.202
	ResNet-18	0.904 ± 0.041	0.674 ± 0.213	0.960 ± 0.014	0.796 ± 0.039	0.571 ± 0.167	0.715 ± 0.149
	Inception-v4	0.884 ± 0.063	0.725 ± 0.166	0.920 ± 0.045	0.693 ± 0.134	0.557 ± 0.157	0.705 ± 0.137
Third approach Multi-task via segmentation maps	DenseNet-121	0.890 ± 0.053	0.704 ± 0.141	0.937 ± 0.027	0.725 ± 0.111	0.562 ± 0.132	0.712 ± 0.118
	ResNet-18	0.906 ± 0.042	0.673 ± 0.241	0.965 ± 0.016	0.823 ± 0.063	0.573 ± 0.187	0.713 ± 0.170
	Inception-v4	0.820 ± 0.093	0.454 ± 0.238	0.933 ± 0.051	0.699 ± 0.026	0.382 ± 0.172	0.525 ± 0.178
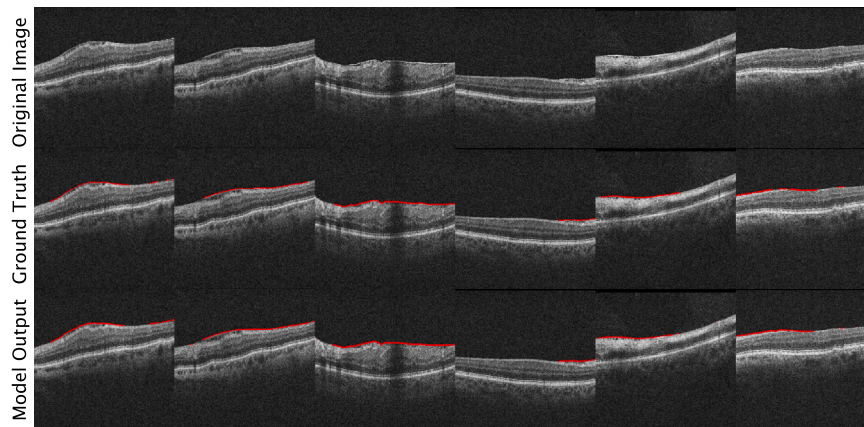
Fig. 9. Examples of the ERM segmentation outputs produced by the first proposed approach with a ResNet-18 encoder overlaid on the original OCT slices.

the segmentation results were obtained by classifying each of these windows. This way, both the proposed approaches and the classification-based approach share a common segmentation ground truth.

All of the proposed approaches were able to achieve very similar results to the single-task, multi-step method of sliding window classification performed by the state of the art (Table 5), with the first approach obtaining better results than the single-task method. These approaches

also entail a considerable simplification of the development process, since they are able to be trained directly from annotated data, while the single-task method relies on a set of ad hoc, fine-tuned steps that involve the detection of the region of interest via active contour models. This reliance on a pre-segmentation step based on active contour models also means that the state of the art method would not be able to correctly analyse the ILM whenever the contour is not adjusted correctly, which can happen due to many of the different visual artefacts that may be
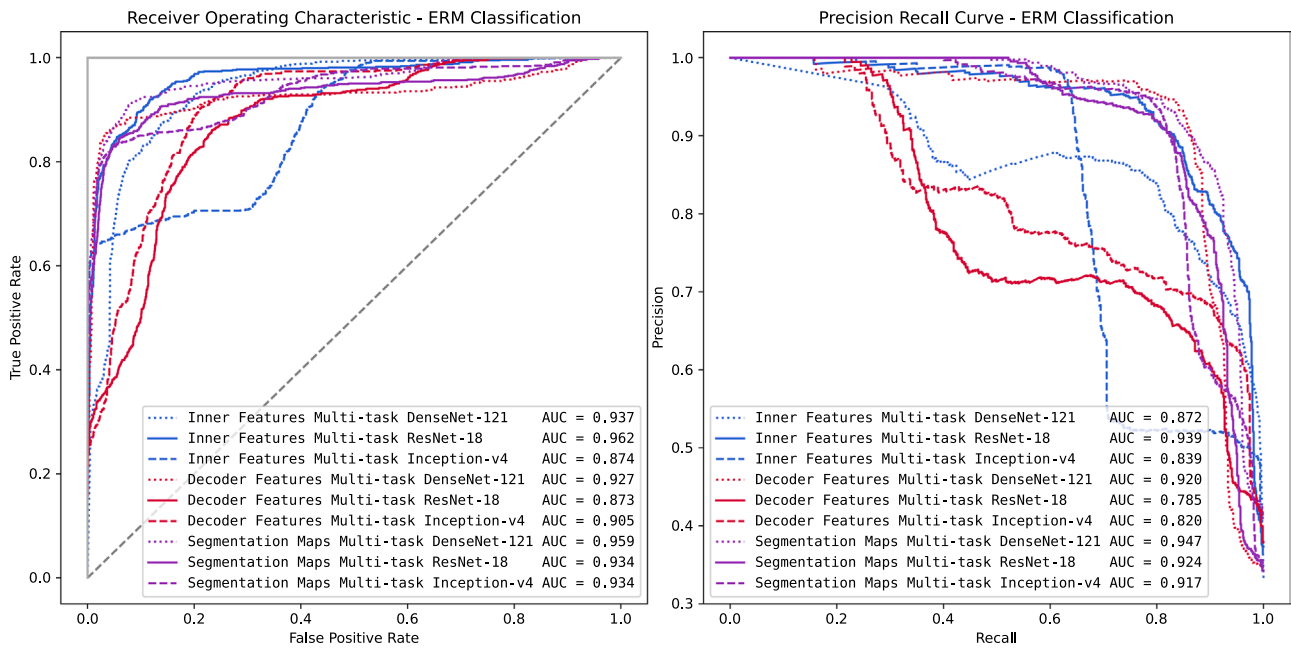
Fig. 10. Receiver operating characteristic and precision recall curves for the classification task.

Table 4

ERM screening task results. In bold, the best results in terms of F-Score.

	Encoder	Accuracy	Sensitivity	Specificity	Precision	F-Score
First approach Multi-Task via inner features	DenseNet-121	0.894 ± 0.066	0.835 ± 0.105	0.929 ± 0.116	0.887 ± 0.164	0.847 ± 0.075
	ResNet-18	0.927 ± 0.053	0.816 ± 0.162	0.983 ± 0.033	0.963 ± 0.068	0.874 ± 0.106
	Inception-v4	0.686 ± 0.332	0.950 ± 0.035	0.561 ± 0.507	0.644 ± 0.362	0.733 ± 0.252
Second approach multi-task via decoder features	DenseNet-121	0.928 ± 0.061	0.852 ± 0.147	0.965 ± 0.042	0.922 ± 0.085	**0.881 ± 0.108**
	ResNet-18	0.809 ± 0.271	0.838 ± 0.220	0.795 ± 0.399	0.825 ± 0.322	0.785 ± 0.243
	Inception-v4	0.826 ± 0.205	0.911 ± 0.098	0.787 ± 0.270	0.753 ± 0.294	0.806 ± 0.210
Third approach multi-task via segmentation maps	DenseNet-121	0.919 ± 0.078	0.888 ± 0.083	0.934 ± 0.085	0.878 ± 0.143	**0.881 ± 0.110**
	ResNet-18	0.897 ± 0.112	0.855 ± 0.223	0.917 ± 0.141	0.870 ± 0.185	0.842 ± 0.172
	Inception-v4	0.899 ± 0.088	0.832 ± 0.180	0.930 ± 0.110	0.880 ± 0.160	0.840 ± 0.134

Table 5

Comparison between the state of the art and the results from the best performing encoder from each of the proposed approaches for the ERM segmentation task.

	Encoder	Accuracy	Sensitivity	Specificity	Precision	Jaccard	Dice
State of the art (Gende et al., 2021)	DenseNet-121	0.883 ± 0.003	0.824 ± 0.072	0.895 ± 0.022	0.656 ± 0.102	0.577 ± 0.098	0.728 ± 0.079
1st Approach: Inner Features	ResNet-18	0.910 ± 0.047	0.768 ± 0.210	0.945 ± 0.011	0.760 ± 0.050	**0.622 ± 0.165**	**0.757 ± 0.140**
2nd Approach: Feature Maps	ResNet-18	0.904 ± 0.041	0.674 ± 0.213	0.960 ± 0.014	0.796 ± 0.039	0.571 ± 0.167	0.715 ± 0.149
3rd Approach: Segmentation Maps	ResNet-18	0.906 ± 0.042	0.673 ± 0.241	0.965 ± 0.016	0.823 ± 0.063	0.573 ± 0.187	0.713 ± 0.170

present in the OCT images. Furthermore, since the proposed approaches can analyse each OCT image as a whole, they do not suffer from the limitations that the sliding windows pose with regard to dealing with image borders.

When compared to other approaches that make use of classical machine learning techniques by classifying a selected set of features (Baamonde et al., July 2019a), we can see that the accuracies of 0.910, 0.904 and 0.906 achieved by each of the proposed end-to-end architectures are higher than those reported by the single task method when using Random Forests (0.894), Naive Bayes (0.855), Multi-layer Perceptrons (0.869) or k-Nearest Neighbours (0.890). Compared to the Support Vector Machines used for the segmentation in Baamonde et al. (2019b), which achieved the best results out of the classical machine learning techniques that were compared for the classification of image windows, the proposed methodologies achieved better results than the classical ones in terms of Dice Score (0.757 vs 0.670) and Jaccard Index (0.622 vs 0.515), metrics that are less biased towards the existing imbalance in this problem. These results highlight the advantages of using deep

learning to obtain the segmentation of the ERM in a straightforward manner, without the need for intermediate steps combined with classical machine learning techniques.

From a technical point of view, the proposed approaches also achieve a considerable speedup when compared with the ones existing in the literature, as shown in Table 6. The proposed approaches are able to take advantage of deep learning and their end-to-end structure to directly analyse the images, resulting in a drastic reduction of inference times by omitting the pre-processing, region-of-interest detection and feature extraction stages. Furthermore, the omission of this delicate series of stages confers these models greater reliability and flexibility. By not needing an initial segmentation of the ILM to restrict the problem to a region of interest and an extraction of windows, these approaches are able to provide an accurate segmentation and classification of the ERM even in cases where these stages would be unsuccessful.

In terms of architecture complexity, the simpler models seem to perform better than the more complex ones, with the configurations that used Inception-v4 as an encoder achieving comparatively worse results

Table 6
Inference times for the analysis of a single image, calculated as the average of 100 images. Note the inference time for (Baamonde et al., 2019b) and (Gende et al., 2021) include the time for each of the steps of the feature extraction methodology.

	Encoder Architecture	Inference Time (GPU)	Inference Time (CPU)
Hand-crafted features (Baamonde et al., 2019b)	–	–	116.651
Deep learning (Gende et al., 2021)	DenseNet-121	22.108	67.493
First approach Multi-task via inner features	DenseNet-121	0.047	0.461
	ResNet-18	0.018	0.211
	Inception-v4	0.098	0.682
Second approach Multi-task via decoder features	DenseNet-121	0.076	0.560
	ResNet-18	0.021	0.243
	Inception-v4	0.126	0.772
Third approach Multi-task via segmentation maps	DenseNet-121	0.086	0.767
	ResNet-18	0.025	0.305
	Inception-v4	0.152	1.090

than the others. For ease of comparison, please refer to Table 1 for the number of parameters in each of the proposed architectures.

Regarding the screening task, a comparison was made between the proposed multi-task approaches and single-task methods. For this comparison, we include the results of two single-task, classification-only approaches (Table 7). The first of these approaches consists of a classifier that receives the original OCT slices as inputs and classifies them into whether they contain diseased tissue or not. The second approach uses the segmentation maps produced by another model as inputs. The models that were employed to obtain these maps were those based on the first proposed approach, multi-task via inner features with a ResNet-18 encoder, which were selected based on the results obtained in the Dice Coefficient and Jaccard Index segmentation metrics. Both of the single-task methods were tested with the DenseNet-121 and the ResNet-18 architectures.

All of the proposed approaches were able to achieve better results than the single-task classification methods, indicating the advantages of multi-task learning. Furthermore, both of the single-task direct classification configurations that relied on the OCT slices failed to produce satisfactory results, proving unable to classify the positive test samples while all of the proposed approaches can provide an accurate end-to-end screening of ERM-positive images. Additionally, although the second single-task approach that used the segmentation maps as inputs were also able to classify the ERM-positive images, all of the proposed multi-task approaches were able to obtain better results. This test can serve as an indication of the potential of multi-task training. Additionally, these single-task models rely on an additional architecture to provide the segmentation maps. This comparison shows that multi-task learning can provide significant advantages over the single-task screening of ERM images, enabling the models to directly classify OCT slices and achieving better results than the two-stage classification of segmentation maps.

Overall, among the proposed configurations, the results for the

different approaches are similar for both tasks. The first proposed approach, using the inner encoder features for the classification task, proved generally more stable, especially with regards to the segmentation task, in which it achieved the best results. The second approach obtained more variable results between the different encoder configurations and tasks. This approach made all of the visual feature maps available for the classification task, which could explain the increased variability since the loss of both tasks affect the feature maps more directly. Finally, the third proposed approach achieved the highest and most stable overall results in the classification task, at the cost of generally lower segmentation results. This approach guides the screening of the ERM more closely than the other two, by only using the segmentation output as information in order to classify the images. While this simplifies the screening task, it increases the impact of the classification loss on the segmentation output. Regarding the best performing configuration, the multi-task via the inner features approach combined with a ResNet-18 encoder provided a satisfactory balance between segmentation and classification results, with a segmentation Dice coefficient of 0.757 ± 0.140, compared to the 0.728 ± 0.079 achieved by the state of the art, as well as a classification F-Score of 0.874 ± 0.106, compared to the best result achieved by the single-task DenseNet-121 classifier of 0.800 ± 0.163, which also depended on another segmentation model to perform its task. While the absence of an ERM-focused public dataset prevents a direct comparison, these proposed approaches can achieve performance results that are competitive with other single-task, state of the art proposals, which were trained and evaluated on different private datasets. Moreover, the approaches proposed in this work also constitute the first multi-task end-to-end proposal able to provide a simultaneous ERM segmentation and screening in the literature.

4. Conclusions

In this work, we propose three different approaches for the end-to-end multi-task simultaneous segmentation and classification of the ERM in OCT images. These approaches can automatically perform a screening over the OCT slices, helping to identify images that may contain signs of the ERM presence as well as to detect and show such signs as a segmentation mask over said images. These can be employed to facilitate the diagnosis and monitoring process of patients with ERM.

The models proposed in this work approach the multi-task segmentation and classification of the ERM in three different ways: first, using the innermost, highest-level encoder features for the classification; second, classifying by using the whole set of feature maps employed in the decoder; and finally, using the output segmentation maps as inputs of the classification head, while propagating the classification loss in a multi-task manner. These three approaches exploit the complementarity of the ERM segmentation and screening tasks in different ways, by employing the highest-order information of the segmentation task, by leveraging all the relevant features that are contained in the feature maps or by limiting the information used in the classification to only that which is relevant to the segmentation, respectively for each of the three approaches. These approaches were also combined with four different state of the art encoder configurations, in order to comprehensively

Table 7
Comparison between the single-task classification methods and the results from the best performing encoder from each of the proposed approaches for the ERM screening task.

	Encoder	Accuracy	Sensitivity	Specificity	Precision	F-Score
OCT slice single-task classification	ResNet-18	0.605 ± 0.139	0.040 ± 0.053	0.886 ± 0.229	0.276 ± 0.485	0.050 ± 0.058
	DenseNet-121	0.697 ± 0.029	0.088 ± 0.176	0.990 ± 0.013	0.222 ± 0.444	0.126 ± 0.252
Output maps single-task classification	ResNet-18	0.767 ± 0.269	0.887 ± 0.086	0.707 ± 0.368	0.710 ± 0.335	0.761 ± 0.240
	DenseNet-121	0.857 ± 0.128	0.850 ± 0.117	0.860 ± 0.134	0.761 ± 0.198	0.800 ± 0.163
1st approach: Inner features	ResNet-18	0.927 ± 0.053	0.816 ± 0.162	0.983 ± 0.033	0.963 ± 0.068	0.874 ± 0.106
2nd approach: Feature maps	DenseNet-121	0.928 ± 0.061	0.852 ± 0.147	0.965 ± 0.042	0.922 ± 0.085	$\mathbf{0.881 \pm 0.108}$
3rd approach: Segmentation maps	DenseNet-121	0.919 ± 0.078	0.888 ± 0.083	0.934 ± 0.085	0.878 ± 0.143	$\mathbf{0.881 \pm 0.110}$

evaluate their behaviour on each of the tasks.

The obtained results show that these multi-task approaches were able to outperform single-task segmentation and classification methods, indicating the potential of combining different related tasks into a single architecture and taking advantage of task complementarity in multi-task training. Additionally, it should be noted that these end-to-end models are able to provide both the segmentation and classification outputs directly from the OCT images, without relying on a set of ad hoc steps tailored to the specific tasks. By bypassing the feature extraction and selection process, these models can be trained directly from the data, simplifying the development of CAD systems and considerably speeding up training and inference times. Furthermore, their independence from intermediate stages such as the extraction of the ILM and the acquisition of the classifiable windows allows these models to work even when these stages would fail to adjust correctly. Overall, the proposed approaches demonstrated the significant advantages of the end-to-end multi-task training for ERM detection and segmentation, helping to provide robustness and objectivity to the diagnosis of such a relevant eye condition.

Regarding future work, current plans include taking advantage of the 3-dimensional relation of OCT slices in order to explore the use of 3D CNN architectures for the volumetric segmentation of the ERM. Furthermore, a separation in the different forms that the ERM may present can be incorporated into the classification and segmentation process, such as its attachment or detachment from the macula or the different stages it may be in. By doing this, more information about the state of the patient can be provided so that it can be used to aid in the treatment of this and other ocular diseases.

CRediT authorship contribution statement

Role Authors Conceptualization **Mateo Gende, Joaquim de Moura, Jorge Novo, Marcos Ortega** Methodology **Mateo Gende, Joaquim de Moura, Jorge Novo, Marcos Ortega** Software **Mateo Gende** Validation **Joaquim de Moura, Jorge Novo, Marcos Ortega** Formal analysis **Mateo Gende, Joaquim de Moura, Jorge Novo, Marcos Ortega** Investigation **Mateo Gende, Joaquim de Moura, Jorge Novo, Marcos Ortega** Data Curation **Mateo Gende, Joaquim de Moura, Jorge Novo, Marcos Ortega** Writing – draft **Mateo Gende** Writing – review & editing **Joaquim de Moura, Jorge Novo, Marcos Ortega** Visualisation **Mateo Gende** Supervision **Joaquim de Moura, Jorge Novo, Marcos Ortega** Project administration **Joaquim de Moura, Jorge Novo, Marcos Ortega** Funding acquisition **Jorge Novo, Marcos Ortega**.

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.

Acknowledgements

References

Akiba, J., Ishiko, S., Hikichi, T., Ogasawara, H., Yanagiya, N., Yoshida, A., 1996. Imaging of epiretinal membranes in macular holes by scanning laser ophthalmoscopy. Am. J. Ophthalmol. 121 (2), 177–180. https://doi.org/10.1016/s0002-9394(14)70582-3.

Amyar, A., Modzelewski, R., Li, H., Ruan, S., 2020. Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: classification and segmentation. Comput. Biol. Med. 126, 104037 https://doi.org/10.1016/j.compbiomed.2020.104037.

Apostolopoulos, I.D., Aznaouridis, S.I., Tzani, M.A., 2020. Extracting possibly representative COVID-19 biomarkers from x-ray images with deep learning approach and image data related to pulmonary diseases. J. Med. Biol. Eng. 40 (3), 462–469. https://doi.org/10.1007/s40846-020-00529-4.

Baamonde, S., de Moura, J., Novo, J., Ortega, M., 2017a. Automatic detection of epiretinal membrane in oct images by means of local luminosity patterns. In: Rojas, I., Joya, G., Catala, A. (Eds.), Advances in Computational Intelligence. Springer International Publishing, Cham, pp. 222–235. ISBN 978-3-319-59153-7.

Baamonde, S., de Moura, J., Novo, J., Rouco, J., Ortega, M., 2017. Feature definition and selection for epiretinal membrane characterization in optical coherence tomography images. Image Analysis and Processing - ICIAP. Springer International Publishing, pp. 456–466. https://doi.org/10.1007/978-3-319-68548-9_42.

Baamonde, S., de Moura, J., Novo, J., Charlón, P., Ortega, M., 2019a. Automatic identification and characterization of the epiretinal membrane in OCT images. Biomed. Opt. Express 10 (8). https://doi.org/10.1364/boe.10.004018.

Baamonde, S., de Moura, J., Novo, J., Charlón, P., Ortega, M., . 2019. Automatic identification and intuitive map representation of the epiretinal membrane presence in 3d OCT volumes. Sensors 19 (23), 5269. https://doi.org/10.3390/s19235269.

Diabetic Retinopathy Clinical Research Network, 2010. Vitrectomy outcomes in eyes with diabetic macular edema and vitreomacular traction. Ophthalmology 117 (6), 1087–1093. https://doi.org/10.1016/j.ophtha.2009.10.040.

Cao, Z., Zhang, K., Wu, J., 2020. FPB: Improving multi-scale feature representation inside convolutional layer via feature pyramid block. In: Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), pp. 1666–1670.

Chen, Y., Liu, Y., 2021. Automatic segmentation of hippocampal subfields MRI based on FPN-DenseVoxNet. In: Proceedings of the 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), pp. 58–62. IEEE, 2021.10.1109/acctcs52002.2021.00020.

Cheung, C.Y., Tang, F., Ting, D.S.W., Tan, G.S.W., Wong, T.Y., 2019. Artificial intelligence in diabetic eye disease screening. Asia-Pac. J. Ophthalmol. 8 (2), 158–164. https://doi.org/10.22608/apo.201976.

Chua, P.Y., Sandinha, M.T., Steel, D.H., 2022. Idiopathic epiretinal membrane: progression and timing of surgery. Eye 36 (3), 495–503. https://doi.org/10.1038/s41433-021-01681-0.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255.10.1109/CVPR.2009.520 6848.

Dimitrakopoulos, P., Sfikas, G., Nikou, C., 2019. Nuclei detection using residual attention feature pyramid networks. In: Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 109–114. 10.11 09/BIBE.2019.00028.

Do, D.V., Cho, M., Nguyen, Q.D., Shah, S.M., Handa, J.T., Campochiaro, P.A., Zimmer-Galler, I., Sung, J.U., Haller, J.A., 2007. Impact of optical coherence tomography on surgical decision making for epiretinal membranes and vitreomacular traction. Retina 27 (5), 552–556. https://doi.org/10.1097/iae.0b013e31802c518b.

Dysli, M., Rückert, R., Munk, M.R., 2019. Differentiation of underlying pathologies of macular edema using spectral domain optical coherence tomography (sd-oct). Ocul. Immunol. Inflamm. 27 (3), 474–483. https://doi.org/10.1080/09273948.2019.1603313. PMID: 31184556.

Falkner-Radler, C.I., Glittenberg, C., Hagen, S., Benesch, T., Binder, S., 2010. Spectral-domain optical coherence tomography for monitoring epiretinal membrane surgery. Ophthalmology 117 (4), 798–805. https://doi.org/10.1016/j.ophtha.2009.08.034.

Fang, L., Yang, L., Li, S., Rabbani, H., Liu, Z., Peng, Q., Chen, X., 2017. Automatic detection and recognition of multiple macular lesions in retinal optical coherence tomography images with multi-instance multilabel learning. J. Biomed. Opt. 22 (6), 066014 https://doi.org/10.1117/1.jbo.22.6.066014.

Fauw, J.D., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., van den Driessche, G., Lakshminarayanan, B., Meyer, C., Mackinder, F., Bouton, S., Ayoub, K., Chopra, R., King, D., Karthikesalingam, A., Hughes, C.O., Raine, R., Hughes, J., Sim, D.A., Egan, C., Tufail, A., Montgomery, H., Hassabis, D., Rees, G., Back, T., Khaw, P.T., Suleyman, M., Cornebise, J., Keane, P.A., Ronneberger, O., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat. Med. 24 (9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6.

Flaxel, C.J., Edwards, A.R., Aiello, L.P., Arrigg, P.G., Beck, R.W., Bressler, N.M., Bressler, S.B., Ferris, F.L., Gupta, S.K., Haller, J.A., Lazarus, H.S., Qin, H., 2010. Factors associated with visual acuity outcomes after vitrectomy for diabetic macular edema. Retina 30 (9), 1488–1495. https://doi.org/10.1097/iae.0b013e3181e7974f.

Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. IEEE Trans. Med. Imaging 37 (7), 1597–1605. https://doi.org/10.1109/tmi.2018.2791488.

Gende, M., De Moura, J., Novo, J., Charlón, P., Ortega, M., 2021. Automatic segmentation and intuitive visualisation of the epiretinal membrane in 3d oct images using deep convolutional approaches. IEEE Access 9, 75993–76004. https://doi.org/10.1109/ACCESS.2021.3082638.

Geng, L., Li, S., Xiao, Z., Zhang, F., 2020. Multi-channel feature pyramid networks for prostate segmentation, based on transrectal ultrasound imaging. Appl. Sci. 10 (11) https://doi.org/10.3390/app10113834.

Ghazi, N.G., Ciralsky, J.B., Shah, S.M., Campochiaro, P.A., Haller, J.A., 2007. Optical coherence tomography findings in persistent diabetic macular edema: the vitreomacular interface. Am. J. Ophthalmol. 144 (5), 747–754. https://doi.org/10.1016/j.ajo.2007.07.012.

Goldberg, R.A., Waheed, N.K., Duker, J.S., 2014. Optical coherence tomography in the preoperative and postoperative management of macular hole and epiretinal membrane. Br. J. Ophthalmol. 98 (Suppl 2), ii20–ii23. https://doi.org/10.1136/bjophthalmol-2013-304447.

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P. C., Mega, J.L., Webster, D.R., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 316 (22), 2402. https://doi.org/10.1001/jama.2016.17216.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. IEEE, June 2016.10.1109/cvpr.2016.90.

Hirano, Y., Yasukawa, T., Ogura, Y., 2010. Optical coherence tomography guided peeling of macular epiretinal membrane. Clin. Ophthalmol. 5, 27–29.

Huang, D., Swanson, E.A., Lin, C.P., Schuman, J.S., Stinson, W.G., Chang, W., Hee, M.R., Flotte, T., Gregory, K., Puliafito, C.A., Fujimoto, J.G., 1991. Optical coherence tomography. Science 254 (5035), 1178–1181. https://doi.org/10.1126/science.1957169.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2261–2269. 10.1109/CVPR.2017.243.

Johnson, M.W., 2010. Posterior vitreous detachment: evolution and complications of its early stages. Am. J. Ophthalmol. 149 (3), 371–382. https://doi.org/10.1016/j.ajo.2009.11.022.

Kermany, D.S., Goldbaum, M., Cai, W., Valentim, C.C., Liang, H., Baxter, S.L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M.K., Pei, J., Ting, M.Y., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., Shi, A., Zhang, R., Zheng, L., Hou, R., Shi, W., Fu, X., Duan, Y., Huu, V.A., Wen, C., Zhang, E.D., Zhang, C.L., Li, O., Wang, X., Singer, M.A., Sun, X., Xu, J., Tafreshi, A., Lewis, M.A., Xia, H., Zhang, K., 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172 (5), 1122–1131. https://doi.org/10.1016/j.cell.2018.02.010.

Khan, S., Islam, N., Jan, Z., Ud Din, I., Rodrigues, J.J.P.C., 2019. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognit. Lett. 125, 1–6. https://doi.org/10.1016/j.patrec.2019.03.022.

Kim, J., Hong, J., Park, H., 2018. Prospects of deep learning for medical imaging. Precis. Future Med. 2 (2), 37–52. https://doi.org/10.23838/pfm.2018.00030.

Kingma, D.P., Ba, J., 2015. Adam: a method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings, 2015.

Kuwayama, S., Ayatsuka, Y., Yanagisono, D., Uta, T., Usui, H., Kato, A., Takase, N., Ogura, Y., Yasukawa, Y., 2019. Automated detection of macular diseases by optical coherence tomography and artificial intelligence machine learning of optical coherence tomography images. J. Ophthalmol. 1–7 https://doi.org/10.1155/2019/6319581.

Lee, C.H., Kwoen, E.Y., Cho, N.C., Kim, W.J., 2015. Analysis of leading diseases causing epiretinal membrane and comparison of prognosis after epiretinal membrane peeling. J. Korean Ophthalmol. Soc. 56 (10), 1586. https://doi.org/10.3341/jkos.2015.56.10.1586.

Lee, C.S., Baughman, D.M., Lee, A.Y., 2017. Deep learning is effective for classifying normal versus age-related macular degeneration oct images. Ophthalmol. Retin. 1 (4), 322–327. https://doi.org/10.1016/j.oret.2016.12.009.

Lee, J.-H., Kim, Y.-T., Lee, J.-B., Jeong, S.-N., 2020. A performance comparison between automated deep learning and dental professionals in classification of dental implant systems from dental imaging: a multi-center study. Diagnostics 10 (11), 910. https://doi.org/10.3390/diagnostics10110910.

Lin, M., Chen, Q., Yan, S., 2014. Network in network.In Y. Bengio and Y. LeCun, editors, 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14–16, 2014, Conference Track Proceedings, 2014.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944. 10.1109/CVPR.2017.106.

Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. Med. Image Anal. 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005.

Liu, X., Song, L., Liu, S., Zhang, Y., 2021. A review of deep-learning-based medical image segmentation methods. Sustainability 13 (3). https://doi.org/10.3390/su13031224.

Lo, Y.-C., Lin, K.-H., Bair, H., Sheu, W.H.-H., Chang, C.-S., Shen, Y.-C., Hung, C.-L., 2020. Epiretinal membrane detection at the ophthalmologist level using deep learning of optical coherence tomography. Sci. Rep. 10 (1) https://doi.org/10.1038/s41598-020-65405-2.

Lu, W., Tong, Y., Yu, Y., Xing, Y., Chen, C., Shen, Y., 2018. Deep learning-based automated classification of multi-categorical abnormalities from optical coherence tomography images. Transl. Vis. Sci. Technol. 7 (6) https://doi.org/10.1167/tvst.7.6.41.

Massin, P., Allouch, C., Haouchine, B., Metge, F., Paques, M., Tangui, L., Erginay, A., Gaudric, A., 2000. Optical coherence tomography of idiopathic macular epiretinal membranes before and after surgery. Am. J. Ophthalmol. 130 (6), 732–739. https://doi.org/10.1016/s0002-9394(00)00574-2.

Matsumoto, C., Arimura, E., Okuyama, S., Takada, S., Hashimoto, S., Shimomura, Y., 2003. Quantification of metamorphopsia in patients with epiretinal membranes. Investig. Opthalmol. Vis. Sci. 44 (9), 4012. https://doi.org/10.1167/iovs.03-0117.

Menteş, J., Nalçacı, S., 2020. An unusual case: self-separation of an idiopathic epiretinal membrane. Turk. J. Ophthalmol. 50 (1), 56–58. https://doi.org/10.4274/tjo.galenos.2019.62372.

Messmer, E.M., Heidenkummer, H.-P., Kampik, A., 1998. Ultrastructure of epiretinal membranes associated with macular holes. Graefe's Arch. Clin. Exp. Ophthalmol. 236 (4), 248–254. https://doi.org/10.1007/s004170050072.

Mohammadzadeh, V., Fatehi, N., Yarmohammadi, A., Lee, J.W., Sharifipour, F., Daneshvar, R., Caprioli, J., Nouri-Mahdavi, K., 2020. Macular imaging with optical coherence tomography in glaucoma. Surv. Ophthalmol. 65 (6), 597–638. https://doi.org/10.1016/j.survophthal.2020.03.002.

Morano, J., Hervella, Álvaro S., Novo, J., Rouco, J., 2021. Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images. Artif. Intell. Med. 118, 102116 https://doi.org/10.1016/j.artmed.2021.102116.

de Moura, J., Samagaio, G., Novo, J., AlmuiñaVarela, P., Fernández, M., Ortega, M., 2020. Joint diabetic macular edema segmentation and characterization in oct images. J. Digit. Imaging 33, 06. https://doi.org/10.1007/s10278-020-00360-y.

Nibali, A., He, Z., Wollersheim, D., 2017. Pulmonary nodule classification with deep residual networks. Int. J. Comput. Assist. Radiol. Surg. 12 (10), 1799–1808. https://doi.org/10.1007/s11548-017-1605-6.

Nielsen, K.B., Lautrup, M.L., Andersen, J.K., Savarimuthu, T.R., Grauslund, J., 2019. Deep learning-based algorithms in screening of diabetic retinopathy: a systematic review of diagnostic performance. Ophthalmol. Retin. 3 (4), 294–304. https://doi.org/10.1016/j.oret.2018.10.014.

Ophir, A., Martinez, M.R., Mosqueda, P., Trevino, A., 2010. Vitreous traction and epiretinal membranes in diabetic macular oedema using spectral-domain optical coherence tomography. Eye 24 (10), 1545–1553. https://doi.org/10.1038/eye.2010.80.

Parra-Mora, E., Cazañas-Gordon, A., Proença, R., daSilvaCruz, L.A., 2021. Epiretinal membrane detection in optical coherence tomography retinal images using deep learning. IEEE Access 9, 99201–99219. https://doi.org/10.1109/ACCESS.2021.3095655.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S., 2019. Pytorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, 32, pp. 8026–8037 (Curran Associates, Inc.). ⟨https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf⟩.

Playout, C., Duval, R., Cheriet, F., 2019. A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images. IEEE Trans. Med. Imaging 38 (10), 2434–2444. https://doi.org/10.1109/TMI.2019.2906319.

Poornachandra, B., James, E., Aseem, A., Jayadev, C., Sanjay, S., Ashwini, K.T., Gadde, S. G.K., Yadav, N.K., 2020. Multimodal imaging of an idiopathic florid vascularised epiretinal membrane: course, treatment, and outcome. Indian J. Ophthalmol. 68 (10), 2302–2304. https://doi.org/10.4103/ijo.IJO_2363_19.

Puliafito, C.A., Hee, M.R., Lin, C.P., Reichel, E., Schuman, J.S., Duker, J.S., Izatt, J.A., Swanson, E.A., Fujimoto, J.G., 1995. Imaging of macular diseases with optical coherence tomography. Ophthalmology 102 (2), 217–229. https://doi.org/10.1016/s0161-6420(95)31032-9.

Rahman, R., Stephenson, J., 2014. Early surgery for epiretinal membrane preserves more vision for patients. Eye 28 (4), 410–414. https://doi.org/10.1038/eye.2013.305.

Samala, R.K., Chan, H.-P., Hadjiiski, L.M., Richter, C., Zhou, C., Wei, J., 2019. Analysis of deep convolutional features for detection of lung nodules in computed tomography. In: Hahn, H.K., Mori, K. (Eds.), Medical Imaging 2019: Computer-Aided Diagnosis. SPIE, pp. 185–190. https://doi.org/10.1117/12.2512208.

Shao, Q., Gong, L., Ma, K., Liu, H., Zheng, Y., 2019. Attentive CT lesion detection using deep pyramid inference with multi-scale booster. Lecture Notes in Computer Science. Springer International Publishing,, pp. 301–309. https://doi.org/10.1007/978-3-030-32226-7_34.

Shen, D., Wu, G., Suk, H.-I., 2017. Deep learning in medical image analysis. Annu. Rev. Biomed. Eng. 19 (1), 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442.

Snead, D.R.J., James, S., Snead, M.P., 2008. Pathological changes in the vitreoretinal junction 1: epiretinal membrane formation. Eye 22 (10), 1310–1317. https://doi.org/10.1038/eye.2008.36.

Sonobe, T., Tabuchi, H., Ohsugi, H., Masumoto, H., Ishitobi, N., Morita, S., Enno, H., Nagasato, D., 2018. Comparison between support vector machine and deep learning, machine-learning technologies for detecting epiretinal membrane using 3d-OCT. Int. Ophthalmol. 39 (8), 1871–1877. https://doi.org/10.1007/s10792-018-1016-x.

Suh, M.H., Seo, J.M., Park, K.H., Yu, H.G., 2009. Associations between macular findings by optical coherence tomography and visual outcomes after epiretinal membrane removal. Am. J. Ophthalmol. 147 (3), 473–480. https://doi.org/10.1016/j.ajo.2008.09.020.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. Proceedings of the Thirty-First

AAAI Conference on Artificial Intelligence, AAAI'17. AAAI Press, pp. 4278–4284 https://arxiv.org/abs/1602.07261.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. 2015. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9.10.1109/CVPR.2015.7298594.

Ting, D.S.W., Cheung, C.Y.-L., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., Hamzah, H., Garcia-Franco, R., Yeo, I.Y.S., Lee, S.Y., Wong, E.Y.M., Sabanayagam, C., Baskaran, M., Ibrahim, F., Tan, N.C., Finkelstein, E.A., Lamoureux, E.L., Wong, I.Y., Bressler, N.M., Sivaprasad, S., Varma, R., Jonas, J.B., He, M.G., Cheng, C.-Y., Cheung, G.C.M., Aung, T., Hsu, W., Lee, M.L., Wong, T.Y., 2017. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 318 (22), 2211. https://doi/org/10.1001/jama.2017.18152.

Triolo, G., Rabiolo, A., 2020. Optical coherence tomography and optical coherence tomography angiography in glaucoma: diagnosis, progression, and correlation with functional tests. Ther. Adv. Ophthalmol. 12 https://doi.org/10.1177/2515841419899822.

Wilkins, J.R., Puliafito, C.A., Hee, M.R., Duker, J.S., Reichel, E., Coker, J.G., Schuman, J. S., Swanson, E.A., Fujimoto, J.G., 1996. Characterization of epiretinal membranes using optical coherence tomography. Ophthalmology 103 (12), 2142–2151. https://doi.org/10.1016/s0161-6420(96)30377-1.

Yakubovskiy, P., 2020. Segmentation models pytorch.⟨https://github.com/qubvel/segmentation_models.pytorch⟩, 2020.

Yamamoto, T., Akabane, N., Takeuchi, S., 2001. Vitrectomy for diabetic macular edema: the role of posterior vitreous detachment and epimacular membrane. Am. J. Ophthalmol. 132 (3), 369–377. https://doi.org/10.1016/s0002-9394(01)01050-9.

Zhang, S., Xue, C.-Y., Liu, Y.-J., Zhang, W.-W., Xie, Z.-G., 2021. Macular pucker, an atypical clinical presentation of ocular toxoplasmosis: a case report. BMC Ophthalmol. 21 (1), 222. https://doi.org/10.1186/s12886-021-01983-7.

Zhou, Y., Chen, H., Li, Y., Liu, Q., Xu, X., Wang, S., Yap, P.-T., Shen, D., 2021. Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images. Med. Image Anal. 70, 101918 https://doi.org/10.1016/j.media.2020.101918.