Master thesis and

internships opportunities

at

INESC TEC- Center of biomedical engineering

Proposal integrated in the project ScreenDR - Image Analysis and Machine Learning Platform for Innovation in Diabetic Retinopathy Screening:

Risk assessment of diabetic macular edema from eye fundus images

The Diabetic Macular Edema (DME) is one the most prevalent causes of visual loss in industrialized countries. However, such visual impairment can be reduced or even prevented by the early diagnosis of this pathology. Currently, the risk of DME is assessed by examining the fundus of the eye and looking for the presence of exudates near the region of the fovea. The risk of DME is evaluated according to the distance of exudates to the center of the fovea.
Despite being a relatively straightforward task for humans, this is a very time which could be performed through the use of automatic image analysis and machine learning methodologies.
The main goal of this dissertation is the development and comparison of two approaches for the assessment of the risk of DME, one using image analysis methods and another using machine learning methods.

Data efficient deep learning. Strategies for training with a small dataset applied to lesion detection in eye fundus images

Deep learning has dramatically improved the state-of-the-art in object detection, speech recognition, robotics, and many other domains. Whether it is superhuman performance in object recognition or beating human players in Go, the astonishing success of deep learning is achieved by deep neural networks trained with huge amounts of training examples and massive computing resources. Other problem domains, such as personalized healthcare, are characterized as small-data problems in which the ability to learn in a sample-efficient manner is a necessity. Thus, data-efficient machine learning approaches, which can be defined as approaches which have the ability to learn in complex domains without requiring large quantities of data, are warranted.
Currently, deep-learning approaches used for the biomedical imaging field are either pre-trained in large image datasets (e.g. ImageNet) or use trivial data-augmentation strategies to generate artificial data.
The main goal of this master thesis is to investigate strategies that harness the power of deep learning while achieving good results with less training data and in particular less human supervision. Strategies investigated can include semi-supervised learning, few-shot learning, and learning models that can be reliably used in simulator-based inference. Developed approaches should be tailored for detection of pathologies from eye fundus images and assessed in publicly available datasets.

Proposal integrated in the project LUCAS - Lung cancer screening - A non-invasive methodology for early diagnosis:

Lung cancer gene mutation prediction using transfer learning

Reduced datasets represent the biggest and most common limitation in biomedical engineering studies. Transfer learning has been successfully employed in data scarce situations, with model knowledge being effectively transferred across tasks/domains. Transfer learning will be applied as fine-tuning a network pre-trained on medical images for a different task. In this work, will be used pre-trained networks for cancer screening and fine-tuned for gene mutations status classification.

Feature engineering for lung cancer gene mutation prediction

Previous results from the project "Lung Cancer Screening - A non-invasive methodology for early diagnosis" and literature suggest that the most relevant information to predict the mutation status in lung cancer might be the combination of features from the nodule and other lung structures. Quantitative features extracted from cancer nodule have been used to create predictive models for gene mutation status and screening. Novel quantitative feature from external structures to the nodule will give relevant information to the machine learning models and it will improve the accuracy of diagnosis.

Synthesizing artificial thoracic CT images using Machine Learning

Abstract:

Privacy issues, and a large amount of time and effort to submit a protocol to the ethical committees and get approval, make the clinical data extremely difficult to obtain. Additionally, there are other indirect barriers to access data such as fees and management requirements. A reproducible and clinically viable predictive model needs a large and heterogeneous cohort of patients and methods capable of coping with the inherent data heterogeneity. This project will be dedicated to implement an efficient method, based on Generative Adversarial Networks (GAN), to synthesize artificial thoracic CT images helping to overcome the data scarcity in radiogenomics studies for lung cancer.

Proposal in the area of Ultrasound imaging:

Right heart function assessment in echocardiography

Cardiovascular diseases account for more deaths than any other cause and are projected to remain the single leading cause of death[1]. Among the existing cardiac imaging modalities, echocardiography is especially attractive as it can be applied bedside, has good temporal resolution, is of relative low cost and does not make use of ionizing radiation. As such, echocardiography has come to play a crucial role in clinical cardiology with diagnostic, prognostic and interventional value. Echocardiography is used to derive numerous clinical parameters to assess cardiac morphology, deformation and function, which usually requires the delineation (i.e. segmentation) and/or tracking of the cardiac chamber walls or other structures.

In spite of the valuable clinical information that can be obtained from echocardiographic images, manual delineation/tracking is a time consuming task for clinicians, which has fuelled the need for semi-/fully automatic analysis tools. Echocardiographic image analysis poses however several challenges due to its low contrast-to-noise ratio, the presence of artefacts and the dependence on the acquisition conditions and settings. In spite of these challenges, numerous methods have been proposed for the segmentation/tracking of cardiac chambers/structures. Geometrical and shape-free models have been the most used approaches in literature and most studies have focused on the left ventricle and this is reflected on the available public datasets available.

The recent advent of deep learning has however changed the image analysis field, and deep learning has been shown to outperform traditional approaches in a wide array of problems. Deep learning approaches have been proposed for cardiac chamber segmentation in echocardiography but the dependence on large training datasets has hindered the development of analysis tools for chambers and structures other than the left ventricle, in spite of the clinical significance of these chambers.

As such this project aims to develop novel deep learning approaches designed for data-challenged settings in echocardiography, in specific, the right heart chambers. In this way, this project aims to retrieve crucial clinical information, currently difficult to assess by cardiologists.

Aortic function assessment in echocardiography

Cardiovascular diseases account for more deaths than any other cause and are projected to remain the single leading cause of death[1]. Among the existing cardiac imaging modalities, echocardiography is especially attractive as it can be applied bedside, has good temporal resolution, is of relative low cost and does not make use of ionizing radiation. As such, echocardiography has come to play a crucial role in clinical cardiology with diagnostic, prognostic and interventional value. Echocardiography is used to derive numerous clinical parameters to assess cardiac morphology, deformation and function, which

usually requires the delineation (i.e. segmentation) and/or tracking of the cardiac chamber walls or other structures.
In spite of the valuable clinical information that can be obtained from echocardiographic images, manual delineation/tracking is a time consuming task for clinicians, which has fuelled the need for semi-/fully automatic analysis tools. Echocardiographic image analysis poses however several challenges due to its low contrast-to-noise ratio, the presence of artefacts and the dependence on the acquisition conditions and settings. In spite of these challenges, numerous methods have been proposed for the segmentation/tracking of cardiac chambers/structures. Geometrical and shape-free models have been the most used approaches in literature and most studies have focused on the left ventricle, in spite of the clinical significance of these chambers.
One particular structure of interest is the aortic valve, given the high prevalence of aortic valve disease in developed countries.

Tissue characterization from gynaecological ultrasound data

Ovarian cancer is one of the cancers with the worst prognostic in adult women. More than half of the patients who present clinical signs such as abdominal bloating and a feeling of fullness already show advanced stages. The majority of ovarian cancers grow as cystic masses, and cancer cells easily spread into the pelvic cavity once the cysts rupture or leak. The early diagnosis of this pathology is very difficult due to the nonexistence of normality/abnormality ovarian tissue bio-markers and due to the fast progression of the pathology. This pathology can only be diagnosed by performing an anatomic pathology examination of an surgically removed ovary, which will impair the women's fertility during the rest of their lives. Unfortunately, it is known that the rate of false-positive detection in these cases can go up to 70%.
Ultrasonography is the typically used imaging modality used for the examination of the ovaries due to its affordability and easiness of use.
In the clinical practice, the diagnosis of these images is based on the visual assessment and depends on the expertise of the clinician.
This project aims at improving the characterization of the different tissues present within the ovary (stroma and follicles) with the aim to allow an early diagnosis of ovarian cancers in the future. For this, the investigation of the acoustic properties of the ovarian tissues must be done. Such investigation can be carried out by analyzing the radio-frequency (RF) and brightness-mode (B-mode) content of ultrasound data.

Proposal in the area of Endoscopic Capsule:

Abnormalities detection on videos of endoscopic capsules

Nowadays, capsule endoscopy is recognized as a potential main method of screening the entire intestinal tract. However each exam produces a colour video of the gastrointestinal track with 8 to 10 hours duration. The diagnosis process takes between 60 and 90 minutes typically and is difficult, tedious and prone to errors. Several difficulties make it difficult to detect abnormalities, e.g., the different aspects of colour, shape and textures of lesions, being influenced by the viewing angle and the distance from the capturing camera. Additionally, it is necessary to cope with the presence of intestinal content and turbidity can hinder the view of the mucosal surface, imaging burring due to moving unaided movement of the capsule and many saturated images. Machine learn approaches, such as deep learning, can be applied to for undoubtedly aid gastroenterologists in this task.

The main objective of this dissertation proposal is to use machine learning techniques, particularly deep learning, detect abnormalities in videos of endoscopic capsules.

In order to test and validate developed methods, it should be used the publicly available datasets, the MICCAI challenge dataset and the KID Dataset. These datasets include a diverse set of anomalies and regular video frames from various parts of the GI tract. Additionally, we have an anonymised private database with 449 endoscopic, made available by Centro Hospitalar do Porto (CHP).

Automatic polyp detection and segmentation in colon images

Colorectal cancer is the third leading cause of cancer death among adults younger than age 50, after breast and lung cancer (1.67 cases per 100,000). Polyps detection is vital to the prevention of colorectal cancers since, when early detected, it is highly curable. Initially, adenomatous polyps are benign; however, with time some of them can become malignant. The colonoscopy, as well as wireless capsule endoscopy (WCE), permit to perform visual inspection of the gastro-intestinal track (GIT). Human factors such as fatigue and insufficient attentiveness can lead to the miss-detection of polyps. Computer-aided automatic detection of polyps is a challenging task due to the variety of shape, size, colour, texture and size scale in the captured images. Additionally, the complex structure of the GI tract, similar colour between polyp and nonpolyp regions, poor image quality, and image variation of the same polyp caused by frequent camera angle changes creates further challenges. Although there many studied published for the automatic detection of polyps in colonic images the problem is not solved, and research still needed to reduce the miss-detection of polyps and the cost and time of screening in large number of colonoscopy or WCE frames. Furthermore, with significant advances in endoscopy imaging systems, clinicians are starting to use several visual classifications to provide a first prediction of polyp histology during the exploration that will need a first step to segment polyp region. Current trends in research have demonstrated that deep learning methods are very effective for automatic detection and segment patterns in vision applications.

The main objective of this dissertation proposal is to evaluate machine learning techniques, particularly deep learning, in the medical image processing domain, and research solutions for automatic detection and segmentation of polyps in colon images.

In order to train and test developed methods, the public datasets can be used (e.g. of "Endoscopic Video Challenge" [1-2]) as well the private database with 449 endoscopic, made available by Centro Hospitalar do Porto (CHP).