

Uncertainty Estimation in Deep Neural Networks for Dermoscopic Image Classification

Marc Combalia *
Hospital Clínic de Barcelona
Barcelona, Spain
mcombalia@clinic.cat

Ferran Hueto
Massachusetts Institute of Technology
Boston, EEUU
ferranhueto@gmail.com

Susana Puig
Hospital Clínic de Barcelona
Barcelona, Spain
spuig@clinic.cat

Josep Malvehy
Hospital Clínic de Barcelona
Barcelona, Spain
jmalvehy@clinic.cat

Verónica Vilaplana *
Universitat Politècnica de Catalunya
Barcelona, Spain
veronica.vilaplana@upc.edu

Abstract

The high performance of machine learning algorithms for the task of skin lesion classification has been shown over the past few years. However, real-world implementations are still scarce. One of the reasons could be that most methods do not quantify the uncertainty in the predictions and are not able to detect data that is anomalous or significantly different from that used in training, which may lead to a lack of confidence in the automated diagnosis or errors in the interpretation of results. In this work, we explore the use of uncertainty estimation techniques and metrics for deep neural networks based on Monte-Carlo sampling and apply them to the problem of skin lesion classification on data from ISIC Challenges 2018 and 2019. Our results show that uncertainty metrics can be successfully used to detect difficult and out-of-distribution samples.

1. Introduction

Machine learning and specifically deep learning, have dramatically improved the state-of-the-art in many areas of research, including computer vision, speech recognition, and natural language processing [27]. These advances are now seeing an application in the medical field, where deep neural networks are being used for a wide range of different purposes [29], including tumor segmentation [3], diabetic retinopathy detection [18], and cancer classification from

histological tissue images [15].

Skin lesion classification by deep neural networks into different cancer sub-types has also experienced significant progress in the last years. A breakthrough moment occurred when a convolutional neural network (CNN) was trained on a dataset of 129450 images of skin lesions of different diseases in [13]. The neural network achieved the same accuracy as expert dermatologists on two binary classification cases: keratinocyte carcinomas versus benign seborrheic keratoses and malignant melanomas versus benign nevi. Since then, many other deep learning models have been proposed for the same purpose [5, 38, 16]. A key player in the evolution of the field is the International Skin Imaging Collaboration (ISIC) [1]. This expert consortium has been developing digital imaging standards for skin cancer imaging and has created a public archive containing the most extensive publicly available collection of quality controlled dermoscopic images of skin lesions. Moreover, since 2016, ISIC organizes yearly artificial intelligence challenges presenting problems in lesion segmentation and lesion classification, promoting the growth of automated diagnostic systems for skin cancer [6, 2, 7, 31].

Despite the rapid acceleration of deep learning research in healthcare, with potential applications being demonstrated across various domains, there are currently limited examples of these techniques being successfully deployed into clinical practice [22]. The challenges and limitations for the deployment of such systems into real-world envi-

