Universitat de Girona
Escola Politècnica Superior

Treball Final de Màster

Estudi: Màster en Ciència de Dades
Títol: Plataforma per Classificar Melanomes
Document: Memòria
Alumne: Wilber Eduardo Bermeo Quito
Tutor: Rafael Garcia Campos Departament: ARQUITECTURA I TECNOLOGIA DE COMPUTADORS Àrea: ARQUITECTURA I TECNOLOGIA DE COMPUTADORS
Convocatòria (mes/any): Setembre 2023

Universitat de Girona
Escola Politècnica Superior

MASTER'S THESIS

A Platform for Classifying Melanoma

WILBER EDUARDO BERMEO QUITO
September 2023

Master in Data Science

Advisors:

DR. RAFAEL GARCIA CAMPOS
University of Girona
Departament of Computer Architecture and Technology

SR. LUIS PLA LLOPIS
Accenture S.L.

Abstract

We present a platform for Melanoma Classification, leveraging a technical infrastructure based on Convolutional Neural Network (CNN) models. The models exclusively utilize image data, and no additional metadata is incorporated during the training process. Various training strategies were employed to enhance model performance.

The resulting models are accessible through a API, enabling users to interact with them via a straightforward web application. Users can submit batches of images to the API for classification, contributing to a user-friendly experience.

This platform demonstrates the efficacy of CNNs in melanoma classification, highlighting the importance of diverse training approaches. The API provides a practical interface for users to seamlessly integrate melanoma classification into their workflows.

1 Introduction

Skin cancer, including melanoma, is a significant global public health concern. Melanoma presents a considerable challenge due to its high mortality rate and the critical importance of early detection for successful treatment. Cancer begins when healthy cells undergo changes that cause them to grow and divide uncontrollably, forming tumors. These tumors can be classified as either cancerous (malignant) or non-cancerous (benign).

In recent times, there has been a growing focus on automating tasks in the medical field through Computer-Aided Diagnosis (CAD)[1]. Some studies have demonstrated that these systems can achieve results similar to those of professionals. However, the integration of CAD into the medical system remains a significant challenge.

The development of a CAD system necessitates the creation of models capable of effectively classifying melanoma. The SIIM-ISIC Melanoma Classification challenge specifically tasks participants with building models for identifying melanoma using skin lesion images and associated metadata. This thesis outlines our approach, wherein we leverage data from this challenge to train our models and subsequently expose them through our platform. By doing so, we contribute to the ongoing efforts to bridge the gap between cutting-edge medical imaging technology and practical clinical applications.

[1]CAD refers to the use of computer algorithms and technologies to assist healthcare professionals in the process of medical diagnosis.

2 Objectives

The final objective of this thesis is to craft a CAD infrastructure, focused on melanoma detection using deep learning vision models capable of detecting melanoma on dermoscopy images. To this end, the gradual achievements that must be accomplished are:

- Gaining a comprehensive understanding of the theory behind deep learning vision models and its practical applications.
- Select a base transfer model. Figuring out why, is this base model enough for our thesis?, the selection of this model is given by the technical limitations?, or any other justification.
- Study different approaches to train the models and select a good evaluate metric given the dataset distribution of dermoscopy images.
- Develop the CAD infrastructure. It should contain the trained models, a simple web UI[2], an API[3] and finally a mechanism using Docker to create the images of these services making it ease to deploy in any based Linux System.

3 Development Process

The project methodology employed in this endeavor follows a continuous process. The project incorporates the concept of utilizing idle time effectively. For instance, during the training of models, there are periods of idle time, which we exploited by concurrently working on other tasks related to developing the entire infrastructure. This approach allows for maximizing productivity throughout the project (see Figure 1).

[2]User Interface. Is the point of human-computer interaction and communication in a device.

[3]Application Programming Interface. Is a set of protocols, routines, tools, and definitions that allow different software applications to communicate with each othe

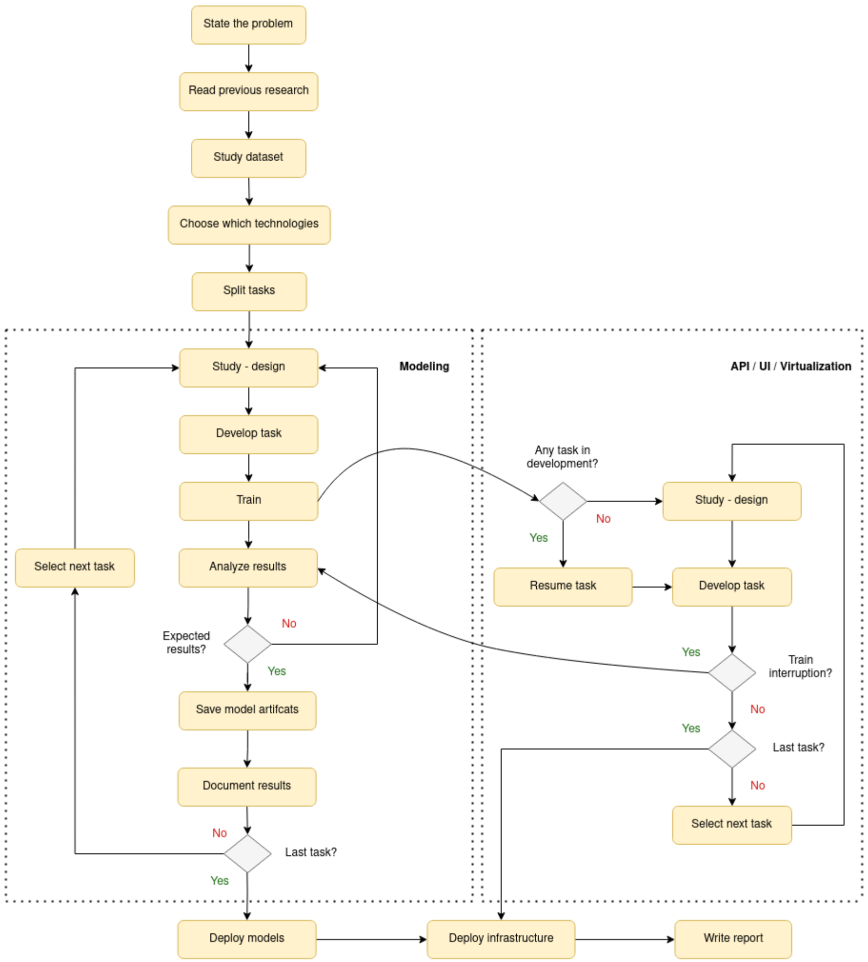

Figure 1: Activity diagram describing the methodology.

The process used to train, validate, test, and implement the models is illustrated in Figure 2. This sequence consists of several stages elaborated below.

The first stage involves cleaning and splitting the initial dataset into smaller datasets. This step ensures that the data is organized and ready for further processing.

The second stage focuses on training and validating the models using the training and validation datasets. During this stage, the system reads images and applies data augmentation techniques to train images and Test Time Augmentation (TTA) to validation images. These techniques enhance the model's performance by introducing variations in the data and improving its generalization ability.

The third stage involves analyzing the training results obtained from different training approaches. In this section*, we evaluate and analyze the model's performance by comparing the predicted results against the test dataset. This step helps us understand how well the models are learning and performing on unseen data.

The last stage revolves around exposing the trained models through an API's container image. This container image allows for easy deployment and integration of the models into other systems or applications, providing a convenient way to utilize the trained models for various tasks.

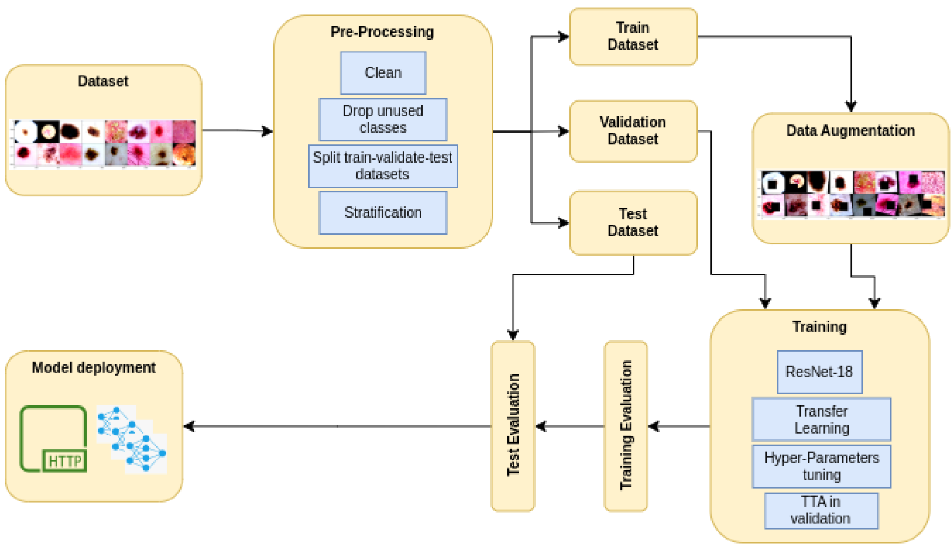

Figure 2: *CAD infrastructure pipeline.*

4 Results

The training phase ended with the development of eight models using an imbalanced dataset comprising eight classes. Various learning policies and Artificial Intelligence (AI) techniques were tested during the experimentation process.

These models were divided into two categories: one without any additional regularization and another with additional regularization techniques such as data augmentation and the inclusion of dropout layers.

Model	**Test AUC**	**Model**	**Test AUC**
M0	0.892	M4	0.858
M1 ⋆	0.892	M5 ⋆	0.843
M2 ∗	0.885	M6 ∗	0.848
M3 •	0.886	M7 •	0.849
Mean	88.875%	Mean	84.950%
SD	0.377%	SD	0.625%

Table 1: *Models metrics in test dataset.*

The initial group of models performed well on the test set with an average AUC of 88.875% and a small standard deviation of ±0.377%. However, they showed signs of overfitting on the validation set. In contrast, the second group of models, trained with additional regularization techniques, achieved lower results but did not suffer from overfitting. They had an average AUC of 84.950with a standard deviation of ±0.625%, which was influenced by more training epochs.

During model training, we also developed the necessary CAD infrastructure. For the API, we used a flexible approach with soft configurations that could be specified through file-based parameters, offering adaptability and simplified management. Additionally, we created an intuitive UI for seamless interaction between healthcare professionals and the models.

We also provided a Docker-based script for easy deployment of the infrastructure on any Linux operating system, ensuring efficient startup and operation.

Bibliography

[ISIC 2019] ISIC. *ISIC Challenge*, 2019. Available at https://challenge.isic-archive.com/. (Not cited.)

[Marr 2021] B. Marr. *What is Weak AI*, 2021. Available at https://bernardmarr.com/what-is-weak-narrow-ai-here-are-8-practical-examples. (Not cited.)