University Of Sunderland

MSc Computer and Data Science

Giulio Leopaldi

**PROM04 - Research Proposal**

**Transfer learning and fine-tuning of MobNet V3 architecture for Melanoma detection**

*Total word count: 3233*

1. **Introduction**

Melanoma (ME) is the deadliest type of skin cancer, but it can have a high rate of survival when diagnosed on time. However, national health systems and clinical practices worldwide are often overwhelmed and unable to visit and screen patients with short-term appointments. Softwares supported by artificial intelligence (AI) that are able to automate diagnosis of ME can be very helpful to support clinical practice and save lives. In recent years, technological advancement in computer vision have permitted to develop softwares that can outperform clinicians in the diagnosis of ME (Brinker et al., 2019). Particularly, the progress of Neural Networks (NN) has brought huge improvements in this sector. The latest trend has seen a shift from deep learning models for maximum statistical performance, to more light-weight models that can be applied in a broader and practical manner. Currently, smartphone applications with embedded algorithms are gaining popularity. This work aims to create an ME detection model by deploying a MobileNet architecture in its newest version, V3. This will be compared to a baseline model designed with MobileNet V2, and then optimised by fine-tuning it in the attempt of improving its performance. Accuracy, sensitivity and specificity will be the designated metrics to measure the success of this experiment, as well as the computational timing for training the models. Section 5 will describe in details the hardwares and softwares that will be used and possible constraints of the project. Section 6 will be dedicated to social, ethical, professional and legal considerations.

1. **Project Description**

Skin cancer cases are growing at an alarming rate globally and ME represents the most dangerous form of it as it can spread to other organs. Early detection of ME is of primary importance since it can hugely increase survival rate, however national health systems and clinical practices worldwide are often overwhelmed and unable to visit and screen patients with short-term appointments (Melanoma skin cancer, 2022). Softwares and applications that can support early detection of ME are gaining popularity, thanks to the advancements of technology and the promising results so far obtained (Udrea et al., 2019). This project aims to solve a binary classification problem to distinguish ME from other types of skin lesions from a dataset of dermoscopy images with a supervised NN. A light-weight architecture based on MobileNet V3, which is designed specifically for smartphone applications, will be deployed and fine-tuned for the purpose (Jane, 2019). This project aspiration is to determine the best possible model in terms of performance, determined by the accuracy, sensitivity and specificity achieved, as well as the computational time for training the models. However, it is out of scope of this research to create an operative application for smartphones with the embedded algorithm, for which a broader team with different expertise would be needed. There are numerous people who could benefit from it. Researchers can use the results obtained in this project to keep exploring and optimise the model. Software developers could embed the algorithm in a mobile application for ME self-screening, using the phone camera to take a photo of a skin lesion, detect its nature, a degree of risk and monitor it in time. Clinicians could use the algorithm as a support tool for their ME diagnosis. Mostly and more importantly, this work hopes to be a further step for helping everyone who suffers or could potentilly suffer from ME, for detecting it on time and save lives.

1. **Preliminary Literature Review**

Skin cancer is one of the most common types of cancer and its incidence is increasing worldwide each year. It often develops as an abnormal growth of melanocytes skin cells that produce and contains a unique pigment called melanin, responsible of the protective effect of the darkening of skin following sun exposure. It also determines the skin, hair and eyes colour. Skin lesions (SL), areas of the skin that have different characteristics from the surrounding skin, are very common and usually don’t pose a threat. A clustering of melanocytes can generate moles (nevi), which are a common type of skin growth and appear as small, dark spots on the surface of the skin. Most often they result as benign, therefore harmless. However, some moles can be a sign of malignant melanoma (ME), the most dangerous type of skin cancer. ME is the 5th most common type of cancer in the UK and around 16,000 new cases are diagnosed each year. The primary treatment is surgical removal of the tumor and, if detected early, it has a high chance of being effective. However, if not detected in time, ME is a great threat for health and life as it can spread systemically to other parts of the body and organs, which makes treatment very complicated and leads to high mortality rates. Sunburn is the factor that poses the highest risk of developing ME. Early detection and immediate action where ME is localized and removed promptly, can yield a 5-year survival rate of 99%, therefore it is crucial (Melanoma skin cancer, 2022). Computer vision, a branch of AI that develops models for enabling computers to recognise and interpret objects like human eyes, has evolved hugely in the recent years. The substantial innovation in this field is consistently supporting the medical sector, allowing for automatic detection, classification and diagnosis of medical images with high accuracy, often surpassing experienced clinicians performance (Brinker et al., 2019). A lot of research and experiments have been carried to create performant models over time.

The most common medical image to classify ME is obtained through dermoscopy. This is a non-invasive tool that captures characteristics of the epidermis not visible to the naked eye.

Several databases of dermoscopy images of different skin lesions are publicly available. Traditional machine learning (ML) models have been developed over the years, usually adopting a standard workflow that include the following steps: pre-processing, segmentation, feature extraction and classification.

In 2009, Alcon et al. (Fernandez Alcon et al., 2009) after the pre-processing and segmentation steps, adopted a feature extraction method following the ABCD rule of dermatoscopy (Nachbar et al., 1994), based on the criteria of asymmetry (A), border (B), colour (C) and the differential structure (D). 55 features have been extracted from the SL images to discriminate between benign and malignant lesions, but not all had the same relevance. The best result for the classification step was obtained with a logistic model decision tree. A decision support system tool was created, obtained by combining the outcome of the image classification model with patient-related data known to pose different degrees of risk for ME such as skin type, age, gender, part of the body. An accuracy of 86%with 94% sensitivity and 68% specificity was achieved.

Karimian et al. (Karimian, Ramezani and Moallem, 2014) in 2014 attempted classifying ME from images obtained by conventional digital cameras . They adopted principal component analysis to restrict the feature extraction to 13 relevant ones out of the 137 initially detected and support vector machines (SVM) for the classification step. They achieved an accuracy of 82.2%, sensitivity of 77% and specificity of 86.93%.

It can be noticed how in the latest years the focus has shifted from traditional ML models to the implementation of NN, as technological advancement and enhanced computational power has allowed to set new standards of accuracy and performance. This approach automatises the steps of segmentation and feature extraction, making it more accurate, reliable and less tedious. NN is an interconnected group of nodes that resembles the neurons structure of the human brain, allowing the recognition of hidden patterns to solve a specific problem. In computer vision, one of the most effective models is an evolution of NN, named Convolutional Neural Network (CNN), specifically designed to process pixel data. For some time, the main trend has seen researchers focusing on optimising performance by increasing the size of the network, including several deep layers in order to achieve maximum statistical results at the expense of computational timing and resources. Szegedy et al. (Szegedy et al., 2015) built the GoogLeNet architecture in 2015 to achieve a classification error of 6.67% with a network which is 22 layers deep. Taking in as an input an RGB coloured image of 224x224 pixels, it uses ReLU as an activation function for all intermediate convolutional layers. It finalises its outcome through a softmax layer of one thousand neurons.

In 2019 Kassani and Kassani (Hosseinzadeh Kassani and Hosseinzadeh Kassani, 2019) elaborated a comparative study of different deep CNNs for ME classification. After the necessary pre-processing operations on the ISIC 2018 dataset, 4 architectures have been compared: Xception, AlexNet, VGGNet and ResNet. Using transfer learning from the ImageNet pretrained weights for an enhanced performance, the network that achieved the best performance after the pre-processing has been ResNet50, with an accuracy of 92.08 %, a precision of 93.73 %, recall of 92.53 % and F-score of 92.74%. It outperformed in accuracy the other networks by a value between 2 and 12%, with AlexNet having obtained the lowest result (80.45%).

In 2019 Brinker et al. (Brinker et al., 2019) compared the performance of CNN and dermatologist. The automated system outperformed 136 out of 157 participating dermatologists of all categories of experience in ME detection, achieving a sensitivity of 84.2% and a specificity of 69.2%, versus a result from the clinicians of, respectively, 60 % and 74.1%. It suggests that computer vision models can positively assist dermatologist in clinical practice for ME detection.

Haghighi et al. (Haghighi, Danyali, Helfroush and Karami, 2020) proposed a method where data augmentation is performed to balance the dataset. It then utilises a CNN architecture of 25 layers, 5 of which convolutional for the feature extraction step with a SVM as a classifier, achieving an accuracy of 89.5%, a sensitivity of 87.7% and a specificity of 91.5%.

The latest trend has seen a gradual shift of researchers from very deep learning models to more computationally efficient ones, where the challenge is to achieve the best balance between performance and computational timing and resources. In this context, several architectures have been developed. MobileNet architectures have captured the attention of several researchers as it provides an efficient system with a lightweight design. They have been created specifically thinking of mobile usage and embedded vision applications, which means they can be handed with low computational power. They are based on depth-wise separable convolutions, which divides the channel and spatial computation in 2 steps by applying a single convolutional layer filter per each input channel and then a pointwise convolution to create the output of the depthwise convolution. The splitting of these operations in 2 steps, makes it lighter than standard convolutions (Howard et al., 2017). Indraswari et al. (Indraswari, Rokhana and Herulambang, 2022) proposed a model where MobileNet V2 is used as the base model for transfer learning, and then by adding a global pooling and 2 fully connected layers on 4 different datasets. This method achieved an accuracy, sensitivity, specificity and precision between 83 and 85% on the ISIC archive dataset.

Early this year, the third version of MobileNet has been released. The proposed method will use MobileNet V3 with transfer learning and fine-tuning in the attempt of achieving the best performance possible. Since early detection of ME is crucial for survival, a system that aids classificatoin and diagnosis of the disease can be extremely useful to improve the patient’s outcome. Mobile phone apps with embedded algorhithms that can allow self-screening and detect ME can help towards the scope, therefore increasing survival rate.

1. **Research Question, Design and Methodology**

Early diagnosis of ME plays a vital role for patients, but often medical screening and visits have long waiting lists. AI supported softwares for ME detection can significantly help clinical practice by relieving the workload from the clinicians and guaranteeing excellent results. Computer vision has seen a substantial progress in recent years, especially with the increasing evolution of NN models. Now, more than ever, it is possible to build light-weight models for ME classification which can be embedded on smartphone apps to allow patients to self-screen and detect ME on time. Several researches can be found in literature that have deployed MobileNet V2, an architecture built purposely for mobile applications, with promising results. However, to the best of our knowledge, ME classification has not been attempted yet with the newer MobileNet V3, released earlier this year. The research question posed is “Does MobileNet V3 architecture improve performance and computational timing compared to MobileNet V2?”

The hypothesis to verify is that a fine-tuned MobileNet V3 architecture should improve both performance (identified in accuracy, specificity and sensitivity) and computational timing in seconds for ME detection.

This research is based on a binary classification for detecting malignant ME from benign lesions using supervised learning. It proposes a comparison between MobileNet V2 and MobileNet V3 in order to verify whether the results are improved in terms of performance and computational timing. The dataset which will be used is the HAM 10000, a large collection of dermatoscopic images from different populations (The HAM10000 dataset, 2018). The first step will be pre-processing the dataset. HAM10000 is a representation of a variety of skin lesions, but our focus lies in a binary classification which distinguishes ME from not ME. For this purpose we will convert the labels into 0 and 1 for classifying positively or negatively the ME. Data augmentation techniques will be applied to balance the dataset, since ME labelled images are singnificantly a minority (about 10%) compared to the whole dataset. These will include flipping, rotating, zooming. The training and testing dataset will be then split with a 70/30% ratio, a commonly used ratio. The next step will be to create a baseline model. For this purpose, MobileNet V2 with transfer learning from ImageNet dataset will be used. Transfer learning is a technique used to allow the model to maintain knowledge acquired from a previous training on a dataset of a similar task. The weights of the network are transferred to the new dataset, without needing to start from scratch, which usually results in better performance. The criteria to evaluate the success will be measuring the accuracy, specificity and sensitivity achieved, as these are the main metrics commonly used for this type of problem. The computational timing will be measured in seconds taken to undergo the training process, utilising a common laptop (specifics in Section 5). The model will then be trained in the same way with MobileNet V3 architecture, and the results recorded. Finally, this wil be fine-tuned in the attempt of improving the performance. A comparison table will then display the accuracy, sensitivity, specificity and computational timing of all models, as well as a confusion matrix to visualise exactly how many instances have been classified correctly or incorrectly, to allow discussion and drawing of conclusions from the experiment.

1. **Resources and Constraints**

This project will be carried out with a laptop with the following specifications:

* Operating system: Windows 10 Home
* Processor: Intel Core i7 - 8550 1.80 - 1.99 GHZ
* Ram: 8GB
* Graphic card: Intel UHD graphics 620

The softwares that will be used are:

* Python 3: this is the programming language of choice for the project. Python is one of the most widely used languages within the Data Science community and it is supported by a number of open-source libraries for the creation and deployment of the models. The main one’s for this project will be:
  + NumPy: a library for scientific computing, used for working with arrays, matrices and performing operations of linear algebra.
  + Pandas: this library allows data manipulation and analysis. Useful for exploring the dataset and performing pre-processing operations.
  + Matplotlib: a versatile and intuitive library for creating static, animated, and interactive visualizations in Python.
  + Tensorflow: a collection of workflows to develop and deploy machine learning models.
  + Keras: a high-level NN library developed in Python that runs on top of TensorFlow.
* Google Colaboratory (Colab): it is an interactive notebook based on a Jupyter interface that allows to write and execute Python code in one’s own browser. It is an excellent tool for deep learning projects, as it gives access to Google computing resources.

HAM 10000 will be the dataset used (The HAM10000 dataset, 2018). It is publicly available and used for benchmarking and research purposes. It consists of 10,015 labeled dermatoscopic images from different populations, of which about 10% representing ME.

A possible constraint could be an underestimation of the computational power needed to deal with the models, that would lead to unsustainable number of hours for model training and fine-tuning. However, this should not be case as the architectures selected are light-weight, even though the laptop in use is not of the latest generation.

1. **Social, Ethical, Professional and Legal Considerations**

The progress of efficient automated systems for diagnosing ME have been helping clinicians to deal with this dangerous disease. However, AI can not guarantee a 100% rate of success for ME detection, therefore misdiagnosis can happen and have serious consequences for patients. Especially concerning are false-negative cases (sensitivity) which could leave undetected a life-threatening ME. False-negatives (specificity) instead could lead to unnecessary surgical removal of a benign skin lesion, mistaken for ME. This could have a series of consequences such as post-surgery complications, aesthetical issues due to scars, financial burden to the patient and the health system. From an ethical, social and professional point of view, all professionals involved in the development, deployment and use of an AI supported technology should minimise the risks of over and under-diagnosis. The developer and distributor must take responsible care in the creation process of the device, as well as the supply and marketing of it, providing clear information about the performance and warnings about risks. A clinician should take all necessary measurements to benefit from AI assistance, but also informing the patient of the risks and limitations. This is also a legal responsibility in order to enable patients to take an informed decision whether to use or not AI softwares (Wu and Negbenebor, 2022).

Since AI software will collect, store and use personal information, privacy issues need to be addressed as a legal and ethical consideration. Personal health information is strictly private and confidential. It is protected by rigid regulations, such as the Data Protection Act (DPA) 2018 in the United Kingdom (Data protection, 2022). It is standard procedure for applications to require user agreement where risks, limitations and private data handling are stated. Often, softwares improve and update their AI algorithms by using the new data from the user in a de-identified form. The National Statement on Ethical Conduct in Human Research 2007 (Updated 2018) regulates the use of de-identified data requiring express written consent from the user (National Statement, 2022). However, most times the patient agrees without giving proper attention to the contract and without being aware of what will happen with their personal data. Companies take advantage of users negligence to make them agree to unethical usage of their personal information, such as selling it to third parties (Jobson, Mar and Freckelton, 2021).

Since caucasian population is the mostly affected by skin cancers, skin lesions datasets are most often gathered by predominantly caucasian countries. This can lead to a lack of representation of minority groups and ethnicities, contributing to inaccurate results for these groups and increased healthcare inequality and disparities. To overcome this issue, there is a need to create larger datasets with more representation of minorities. It is hoped that in time this issue will be limited with improved and more inclusive datasets (Wu and Negbenebor, 2022).

1. **References**

2022. [online] Available at: <https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-condhttps://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018uct-human-research-2007-updated-2018> [Accessed 21 August 2022].

Brinker, T., Hekler, A., Enk, A., Berking, C., Haferkamp, S., Hauschild, A., Weichenthal, M., Klode, J., Schadendorf, D., Holland-Letz, T., von Kalle, C., Fröhling, S., Schilling, B. and Utikal, J., 2019. Deep neural networks are superior to dermatologists in melanoma image classification. *European Journal of Cancer*, 119, pp.11-17.

Fernandez Alcon, J., Ciuhu, C., ten Kate, W., Heinrich, A., Uzunbajakava, N., Krekels, G., Siem, D. and de Haan, G., 2009. Automatic Imaging System With Decision Support for Inspection of Pigmented Skin Lesions and Melanoma Diagnosis. *IEEE Journal of Selected Topics in Signal Processing*, 3(1), pp.14-25.

GOV.UK. 2022. *Data protection*. [online] Available at: <https://www.gov.uk/data-protection#:~:text=The%20Data%20Protection%20Act%202018%20is%20the%20UK's%20implementation%20of,used%20fairly%2C%20lawfully%20and%20transparently> [Accessed 21 August 2022].

Haghighi, S., Danyali, H., Helfroush, M. and Karami, M., 2020. A Deep Convolutional Neural Network for Melanoma Recognition in Dermoscopy Images. *2020 10th International Conference on Computer and Knowledge Engineering (ICCKE)*.

Harvard Dataverse. 2018. *The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions*. [online] Available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T> [Accessed 21 August 2022].

Hosseinzadeh Kassani, S. and Hosseinzadeh Kassani, P., 2019. A comparative study of deep learning architectures on melanoma detection. *Tissue and Cell*, 58, pp.76-83.

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

Indraswari, R., Rokhana, R. and Herulambang, W., 2022. Melanoma image classification based on MobileNetV2 network. *Procedia Computer Science*, 197, pp.198-207.

Jane, V., 2019. *Everything you need to know about MobileNetV3 and its comparison with previous versions*. [online] Medium. Available at: <https://towardsdatascience.com/everything-you-need-to-know-about-mobilenetv3-and-its-comparison-with-previous-versions-a5d5e5a6eeaa> [Accessed 20 August 2022].

Jobson, D., Mar, V. and Freckelton, I., 2021. Legal and ethical considerations of artificial intelligence in skin cancer diagnosis. *Australasian Journal of Dermatology*, 63(1).

Karimian, A., Ramezani, M. and Moallem, P., 2014. Automatic Detection of Malignant Melanoma using Macroscopic Images. *Journal of Medical Signals &amp; Sensors*, 4(4), p.281.

Nachbar, F., Stolz, W., Merkle, T., Cognetta, A., Vogt, T., Landthaler, M., Bilek, P., Braun-Falco, O. and Plewig, G., 1994. The ABCD rule of dermatoscopy. *Journal of the American Academy of Dermatology*, 30(4), pp.551-559.

nhs.uk. 2022. *Melanoma skin cancer*. [online] Available at: <https://www.nhs.uk/conditions/melanoma-skin-cancer/> [Accessed 20 August 2022].

Szegedy, C., Wei Liu, Yangqing Jia, Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A., 2015. Going deeper with convolutions. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Udrea, A., Mitra, G., Costea, D., Noels, E., Wakkee, M., Siegel, D., Carvalho, T. and Nijsten, T., 2019. Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms. *Journal of the European Academy of Dermatology and Venereology*, 34(3), pp.648-655.

Wu, J. and Negbenebor, N., 2022. Melanoma Screening: The Ethics of Over- and Underdiagnosis. *Rhode Island Medical Journal*, (April issue).