

OncoVision

Rocco Iuliano, Simone Delle Porta

ACM Reference Format:

Rocco Iuliano, Simone Delle Porta. 2023. OncoVision. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 PROJECT CONTEXT

Addressing health-related issues represents the next frontier of artificial intelligence. Several attempts have been indeed already conducted to develop or experiment with machine learning, genetic algorithms, and others, in health contexts. For instance, in the Aravind Eye Care System in India, ophthalmologists and computer scientists are working together to test and deploy an automated image classification system to screen millions of retinal photographs of diabetic patients. Since the mid-twentieth century, researchers have proposed and developed many clinical decision support systems for help the physicians. Rule-based approaches was proposed in 1970s that allow us to:

- (1) diagnose diseases;
- (2) choose appropriate treatments;
- (3) provide interpretations of clinical reasoning;
- (4) assist physicians in generating diagnostic hypotheses in complex patient cases.

It seem a good approach but it present several problems, namely:

- (1) it is costly to build;
- (2) it require explicit expressions of decision rules and require human-authored updates;
- (3) it is difficult to encode higher-order interactions among different pieces of knowledge authored by different experts;
- (4) the performance of the systems is comprehensible only by a medical knowledge;
- (5) it was difficult to implement a system that integrates deterministic and probabilistic reasoning to narrow down relevant clinical context, prioritize diagnostic hypotheses, and recommend therapy.

Instead the first generation of AI systems, which relied on the curation of medical knowledge by experts and on the formulation of robust decision rules, recent AI research has leveraged machine-learning methods, which can account for complex interactions, to identify patterns from the data. Moreover, machine-learning methods enable the development of AI applications that facilitate the discovery of previously unrecognized patterns in the data without the need to specify decision rules for each specific task [3]. Therefore, the main goal of AI system in Medicine is help the physicians

in the clinical process, namely detect something, cluster the patient on some criteria, etc.

In this project we will focus on skin cancer, namely the melanoma. Melanoma is a malignancy of melanocytes, which are pigment-producing cells of neuroectodermal origin that can be found throughout the body (including in the skin, iris and rectum). The cutaneous form of the disease is common in the Western world and causes the majority (75%) of deaths related to skin cancer; its global incidence is 15–25 per 100,000 individuals. Survival rates in patients with melanoma (cumulative for all forms) have shown a huge differences between countries in Europe, ranging between <50% in Eastern Europe to >90% in northern and central Europe for 5-year survival after primary diagnosis [2]. In the follow image is shown the incidence and mortality of cutaneous melanoma in the world.

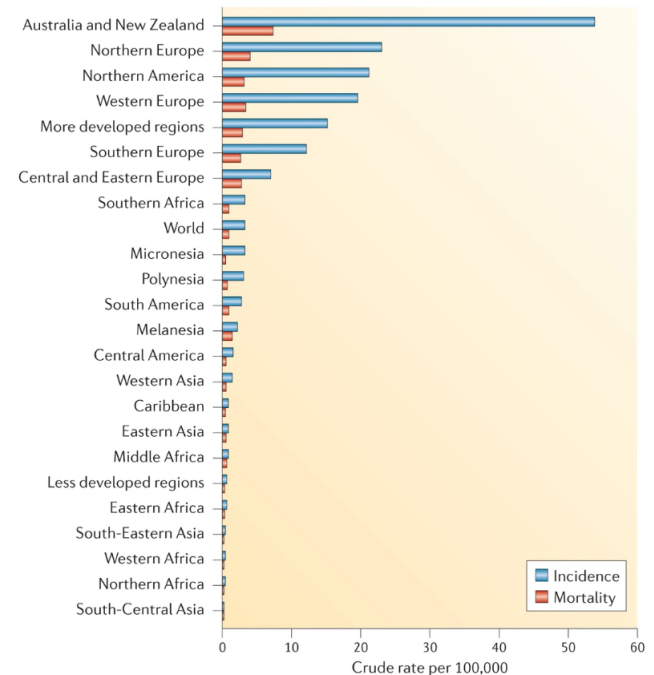


Figure 1: Incidence and mortality of cutaneous melanoma [2]

The biggest problem with this cancer is a lack of early detection that could aim treatments to treat the disease in a timely manner. Therefore, for this reason we decided to develop a melanoma detection AI system in order to help physicians in treatment of this disease. We compared performance obtained by our model with the proposed one by Di Biasi et al [1].

1.1 Related project

Di Biasi et al. proposed a system that combine the Genetic Algorithms (GAs) with Convolutional Neural Networks (CNNs) to detect

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2023 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

the melanoma. They used the GA for improve the architecture of CNN not to improve the network's hyperparameters. Indeed, they defined a population of neural networks (NN) that are codified in vectors where each vector elements represents a type of layer or pre-processing routine. Whereas for the hyperparameters of each individual, they fixed them to this values: 'sgdm', 'MaxEpochs', 16, 'MiniBatchSize', 12, 'Shuffle', 'every-epoch', 'InitialLe-arnRate', 0.0001, 'ExecutionEnvironment', 'auto'.

The authors defined the following constraints to the Genetic Algorithm in order to obtain a correct population of NN at each evolution step:

- the first gene of each entity must be an image input (II) or one of the pre-processing routines;
- if the gene q is a pre-processing routine, then the gene q + 1 must be a II layer or another pre-processing layer;
- the latest gene of an entity must be a classification layer.

Furthermore, they forced the possible values that the genes of an individual can take in this range: *Convolution, ReLu, Cross Channel Normalization, Max Pooling Grouped Convolution, Fully Connected Layer, Dropout and Softmax*.

The authors stopped the GA when no accuracy improvement was detected for ten consecutive evolution steps or after 100 evolutionary steps. In conclusion, their dataset is composed by skin images and it is divided in two classes: melanoma (positive class) and moles (negative class). The number of instances for positive class is 70 and 100 for negative class.

2 PROJECT GOALS

The project goals are:

- (1) Conduct a detailed investigation of the baseline approach selected from the literature in order to understand the performance of the approach and its limitation;
- (2) Understand the problems related to the datasets, namely lack of relevant features, few samples and so on;
- (3) Definition of an AI pipeline that might be used for cancer detection and does not affect by the problems which the baseline approach selected from the literature suffer.

Our project is available on GitHub at this link: <https://github.com/Rocco000/OncoVision>

3 METHODOLOGICAL STEPS TO CONDUCT TO ADDRESS THE GOALS

Based on the goals set, the methodological step that we will conduct to address it are:

- Define a survey for physicians expert in the cancer domain in order to understand some data problems and which features of data are relevant. We will conduct this step by using Google Forms;
- Re-implement the existing approach, because the source code of this one is not available. We need to do this in order to compare our model performance with the performance of the existing approach;
- Study some image processing techniques in order to improve the quality of the data;

- Search one or more new datasets in order to train our model and test the existing approach. We will conduct these researches by using Kaggle;
- Develop three genetic algorithm (GA) that produce a population of artificial neural networks (ANNs) that are optimized in three point of views:
 - The goal of the first one is to improve the hyperparameters of the network;
 - The goal of the second one is to improve the architecture of the network. After that we will select the best individual and we will apply the Grid Search algorithm on it in order to improve hyperparameters;
 - The goal of the last one is to improve both hyperparameters and architecture of the network.

At the end of each Genetic Algorithm we will select the best individual in the last population obtained based on the evaluation metrics. For developing these alternatives we will use Python and its libraries.

In the end we will compare the results obtained by our models with the results of the baseline approach selected. Therefore, the methodological steps that we will conduct will be the steps that the follow image represents:

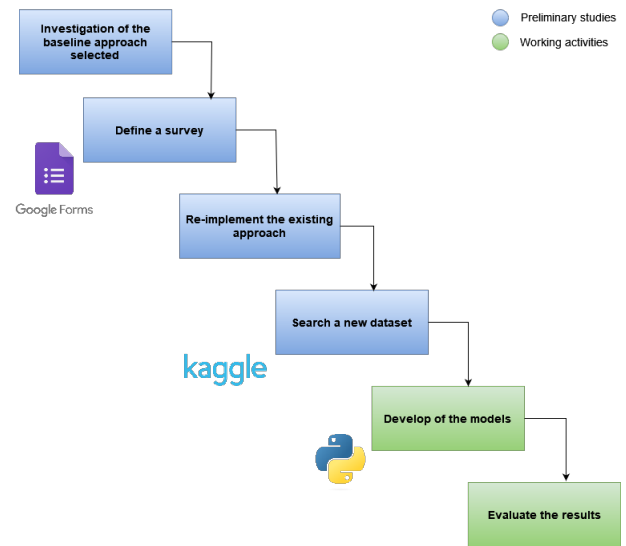


Figure 2: Methodology

REFERENCES

- [1] Luigi Di Biasi, Alessia Auriemma Citarella, Fabiola De Marco, Michele Risi, Genevieve Tortora, and Stefano Piotto. 2023. Exploration of Genetic Algorithms and CNN for Melanoma Classification. In *Artificial Life and Evolutionary Computation: 15th Italian Workshop, WIVACE 2021, Winterthur, Switzerland, September 15–17, 2021, Revised Selected Papers*. Springer, 135–138.
- [2] Dirk Schadendorf, David E Fisher, Claus Garbe, Jeffrey E Gershenwald, Jean-Jacques Grob, Allan Halpern, Meenhard Herlyn, Michael A Marchetti, Grant McArthur, Antoni Ribas, et al. 2015. Melanoma. *Nature reviews Disease primers* 1, 1 (2015), 1–20.
- [3] Kun-Hsing Yu, Andrew L Beam, and Isaac S Kohane. 2018. Artificial intelligence in healthcare. *Nature biomedical engineering* 2, 10 (2018), 719–731.