

**Provide your team background and organization description (if applicable).**

The EISBM (European Institute for Systems Biology and Medicine) is a non-for-profit association based in Lyon (France) created in 2017 by Charles Auffray (EISBM president).

We are focused on developing and applying informatics methods for data analysis, knowledge and information management. We participate in biomedical research projects financed by the European Commission such as PIONEER (<https://prostate-pioneer.eu/>) or OPTIMA (<https://www.optima-oncology.eu>) while offering our services and expertise to private and academic partners.

EISBM is specializing in medical research, dedicated to leveraging cutting-edge technologies to advance scientific knowledge and improve healthcare outcomes. Our core activities revolve around data analysis, systemic approaches, data management, and artificial intelligence.

We currently employ 5 persons, with backgrounds in biology, bioinformatics, mathematics, and physics.

The team answering the challenge is composed of:

Albert Saporta, PhD. Albert is a researcher, expert in AI and applied computer vision. He obtained a PhD in theoretical physics from the University Claude Bernard in Lyon in 2019 and has worked as a post-doctoral researcher in medical physics and deep learning at the Centre Léon Bérard in Lyon before joining the EISBM in 2022.

Johann Pellet, MSc. Johann is our engineering director, expert in data management, omics data analysis and bioinformatics. He obtained his MSc in Bioinformatics from the University of Montpellier (France) in 2005 and has joined the EISBM in 2022.

Bertrand De Meulder, PhD. Bertrand is our scientific director, expert in bioinformatics, omics data analysis and biostatistics. He obtained his PhD in biological sciences from the University of Namur (Belgium) in 2013 and has joined the EISBM as a researcher at the creation of EISBM in 2017.

**Explain why you participated in the Diagnodent challenge.**

There were several reasons why we decided to take part in the Diagnodent challenge:

- an opportunity to evaluate our expertise in AI against expert participants responding to the challenge;
- to start a new collaboration with research teams (Strasbourg Hospital) requiring our services and expertise in the field of AI and medical research;
- to have access to a high-quality dataset and a computational infrastructure;
- the interest in developing a diagnostic aid for a rare oral disease and implementing it in an hospital setting;
- to raise awareness of our association and get media coverage of our work, following the outcome of the challenge;
- to obtain a financial reward to further develop our activities.

**Describe how you built your winning model and elaborate on the technical and modeling choices you made.**

We chose to use widespread libraries skLearn, pyTorch, OpenCV and GRADcam as we had expertise with them, and they ensure reproducibility.

We first analyzed the targets columns in the dataset. We studied the genes and considered that the genes appearing for less than four patients were unpredictable and set these genes to None. We thus kept five genes and the label None.

Then, we considered two classes of models: ResNet and EfficientNet that are state of the art for image classification, following the articles

<https://ai.googleblog.com/2019/05/efficientnet-improving-accuracy-and.html> and <https://arxiv.org/pdf/2104.00298.pdf>. We replaced the final fully connected layer, also called a classification head by four heads. Each of the four heads are dedicated to a class (Cohort, AI\_Type, Responsible\_Gene\_Name and Is\_Isolated\_Syndromic). We thus transformed the multi label classification task to four multi class classification tasks. From experience, this allows for easier prediction of each class separately and allows for faster convergence of the models.

We choose the cross-entropy loss function for each of the four heads and minimized the global cross entropy loss (average of the four losses). We normalized the images pixel values between 0 and 1 and we used one hot encoding on each of the four target columns.

We split the dataset into a training set and a validation set (90% -10%).

We identified that our model was overfitting as the validation loss started to increase after several epochs while the training loss was still decreasing. We then preformed augmentation on all patients except on the Normal Cohort. We applied two set of transformations to the augmented images: horizontal flip of the image and alteration of the brightness and contrast of the images. We choose this set of augmentation in order to increase the robustness of the model against various image quality and lightning condition. This improved the performance of the model by reducing overfitting and increasing generalizability, thus leading to higher F1 score.

We tested several versions of ResNet and EfficientNet models and obtained the best F1 score with EfficientNetV2L. We also tested different image resolutions and batch sizes and found that a resolution of 400\*400 with a batch size of six images gave the best F1 score with the provided GPU and the training time requirement.

We also tried several learning rates and found that the value 0.0001 led to best results as the model converged smoothly in less than six hours.

In addition, we evaluated each head separately, and we found that the losses of the heads dedicated to the classes Cohort and AI\_Type were much lower than the losses of the two other heads. We thus multiplied the losses of the Cohort and AI\_Type heads by two. We chose this factor after testing different values between 1 and 10, and the factor two gave the best performance.

**Were the GPU resources provided in the challenge notebook sufficient from your point of view?**

No, the GPU resource were not sufficient, for several reasons.

- The GPU provided do not allow to use high resolution images, we had to lower the resolution (400\*400).
- Moreover, the batch size had to be low (five or six images per batch). These limiting factors can really affect the performance of the model as a higher resolution would lead to lower loss of information. A higher batch size would allow the model to improve its generalizability.
- Finally, the 6 hours compute time puts serious limitations to the number of epochs available to train our model, which means that our model is most likely not taking full advantage of the dataset available and has suboptimal performances.