

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

Our team is composed of 4 members:

- **Samer EL ZANT:** He's the Associate Professor and Coordinator of the "Data Science and Artificial Intelligence" master's degree program and Computer Science bachelor's degree program at the University of Strasbourg (UFAZ project: French-Azerbaijani University). He's involved in various research activities and collaborations in the field of Artificial Intelligence (supervision of master's degree interns and PhD students).
- **Karim TOUT:** He's the Senior Computer Vision / Machine Learning Technical Lead at Uqodo managing the Computer Vision / Machine Learning department and is an expert consultant on topics related to AI and computer vision at Cetim France. He has several years of experience in various industries (automotive, aerospace, mechanical, ...) developing vision inspection systems for quality inspection using traditional and Deep Learning based solutions and provides expert consultancy and professional training on such topics for many companies in France. He's an invited professor at the French-Azerbaijani University (UFAZ) teaching a course on Deep Learning for Image Processing.
- **Emil ALIYEV:** He's a fresh master's degree graduate from the Data Science and Artificial Intelligence program at the French-Azerbaijani University (UFAZ). He's currently a Python Developer at SKYTECH TECHNOLOGY. He's also working on a collaborative project between Vijump and the CSTB team at the ICUBE laboratory in Strasbourg.
- **Hanin ZAIDAN:** She's a skilled dentist with a particular interest in using AI for diagnosing dental images. She started her career in 2020, opening her own clinic after completing a two-year residency at RHUC Clinics. Driven by her passion for leveraging advanced technologies, she practiced in research projects with CSTB team at ICUBE laboratory to detect and classify objects in panoramic x-ray images. Dr. Zaidan's commitment to excellence and patient-centered care has earned her a reputation as a forward-thinking and compassionate dental professional. Her dedication to staying updated with the latest advancements in dental technology allows her to provide the best possible care to her patients.

As can be seen, we are not part of the same organization. We met each other at UFAZ with a shared interest in AI and computer vision.

2. Explain why you participated in the Diagnodent challenge.

There are many reasons why we decided to participate in the Diagnodent challenge:

- Help our fresh graduates to experience their first real case study on a real dataset in a very important and high-risk field.
 - We previously worked on a small research project that aimed to detect certain anomalies in dental X-Rays and MRIs Deep Learning based approaches.
 - We are passionate about the applications of AI in the medical field, especially the field of dental medicine. We see a great potential of CNN-based models to help Dentists in their diagnosis.
3. Describe how you built your winning model and elaborate on the technical and modeling choices you made.

A full description of the reasoning behind our proposed approach is described in details in the submitted notebook. As a recap in bullet points:

- We first spent a lot of time exploring the dataset to better understand the different targets and how they correlate.
- In the training dataset we have 34 unique combinations of targets with huge imbalance between the different combinations. Therefore, creating a single predictive model for all 4 targets at the same time will be a challenge. **Hence we decided to split the task into separate models applied sequentially. By predicting each target independently, this will improve the overall robustness of the diagnosis.**
- Another reason for this choice is the practical side of the applications. In fact, even from a medical point of view, explainability is very important. A single model that predicts the 4 targets is more difficult to explain and interpret its results than a model that predicts a single target.
- The flow process goes like this:
 - a. First we develop a model for Cohort target only. If the model predicts Normal or Dentinogenesis Imperfecta we can identify the remaining targets without any other models and the flow stops here. If the model predicts Amelogenesis Imperfecta, we use other predictive models to predict the other targets.
 - b. The second step is to develop a model for the AI Type and another model for the isolated/syndromic target. Both these models will be trained only on images of AI patients.
 - c. The third step regarding the prediction of the Responsible Gene is the most difficult since the amount of images we have between the different classes is very low. Therefore, we decided to use a statistical based approach to predict the Responsible Gene based on the predicted AI Type and isolated/syndromic. We found that for many cases we only have one potential combination between these three targets or at least a majority class of Responsible Gene. One of the

cases where the statistical analysis was not deterministic was the case of AI_Hypoplastic and Isolated. For this specific case we can see that we don't have a clear majority Responsible Gene, but rather a class between AMELX, ENAM, COL17A1, COL7A1. Therefore, the fourth predictive model is trained for the Responsible_Gene_Name trained only on Amelogenesis Imperfecta patients, with AI_Hypoplastic and Isolated.

- We developed our own data generator during the CNN training to have full control of the whole augmentation pipeline. There are in total 10 different augmentations applied to the training images randomly that were chosen after a thorough examination of the dataset:
 - a. Brightness change (simulate variation in brightness in real use case)
 - b. Contrast change (simulate variation in contrast in real use case)
 - c. Gamma change (simulate variation in gamma correction in real use case)
 - d. Flip up-down (Teeth affected by AI will be on the upper or bottom part)
 - e. Flip right-left (Teeth affected by AI will be on the right or left part)
 - f. Random crop (simulates a zoom-in effect)
 - g. Random cutout (simulates a zoom-in effect)
 - h. Random crop (simulates occlusions over the image)
 - i. Random rotation (with and without cropping)
 - j. Random noise (gaussian, speckle, or s&p applied only on the rgb photos)
- We used class weights to reduce the effect of class imbalance for all the models.
- As a CNN model, We chose to use the DenseNet121 model pretrained on ImageNet dataset as our predictive model for several reasons:
 - a. DenseNet121 has a unique architecture that connects each layer with every other layer in the network. This connection is done through feature maps concatenation. This allows the low level features such as corners, edges, angles, and colors to be concatenated with the high level features in the last layers of the model. These low level features are very important in this particular case study since according to the Data_Dictionnary PDF file all these low level features (corners, edges, angles, and colors) are important features for the identification of the targets, mainly the AI_Type.
 - b. Since we have a limited amount of images, a model trained from scratch risks not being able to learn important features, especially low level features, from the dataset. Therefore, using the DenseNet model pretrained on ImageNet will help the model learn better and faster since these low level features are already learned by the

pretrained model because they are common between many different tasks.

- c. DenseNet121 is one of the smallest models of the DenseNet family to prevent high overfitting.
- The models were trained with dropout to reduce the effect of overfitting.
- **For the AI_Type model, we preferred to use the sigmoid activation function and the binary_crossentropy loss function. The reason behind this decision is taken from a practical point of view. Since the model is trained with sigmoid activation function and the binary_crossentropy loss function, it will consider each class as an independent binary classification case, hence it can predict multiple classes at the same time. Since it is not easy to classify the AI_Type, the proposed prediction model can be used as an assistant to the Doctor giving him a list of prediction scores for each AI_Type which we believe is much more helpful for the Doctor in his diagnosis. The same concept was also used for the Is_Isolated_Syndromic and Responsible_Gene_Name models.**
- We started the challenge late and didn't have the time to experiment more and make more submissions. We have many ideas that could improve the overall robustness of the prediction flow. So as final notes:
 - a. Work collaboratively with dentists and medical professionals to better understand the different targets and identify their characteristics. Tools such as SHAP and GradCam will help explain why the model made a specific decision which will be examined by dentists to provide feedback. This feedback/improvement loop is very important at an initial stage of development.
 - b. Not include the None class in the training for AI cases. As we mentioned the None values could be a real class so training the model that this image is a separate class None is confusing. That's why we asked a question in the challenge Forum if we might have None class for AI cases in the test set and the answer was yes, hence we did it. In our submission we included the None class in the training to achieve a better result in the challenge, but for us it does not make sense in practice and is bad for the model training.
 - c. Currently the trained models can predict on both rgb images and X-Ray images. Dedicating a separate model per image type may improve the performance of the model since these two image types have different visual features.
 - d. Creating a model that predicts both AI_Type and is_isolated_syndromic might be a better approach since there might be some correlations between these two targets that will be learned by the model.

- e. Creating other Responsible Gene model for other AI Type and isolated/syndromic cases will also improve the overall performance. We didn't have time to spend on this task and to be honest we decided not to prioritize it since the Responsible Gene target has some major limitations. An answer from the organizers in the challenge Forum shows that the Responsible Gene is not something that can be determined by the photos but rather by a genetic test. We understand that the purpose of the competition is also to check if there are any patterns or features in the photos that will help identify the Responsible Gene. The predictive model trained on these photos might be helpful in identifying this pattern if it exists. However, in our opinion, the number of patients in the separate Gene types is very low to allow the development of a robust predictive model based on a supervised learning model. In addition, according to the mentioned answer, some None values may actually have one of the listed genes but the genetic test has not been done, and this will be confusing for the model for sure since the associated class is wrong. Another reason is the fact that the Responsible Gene types that we might have in reality are more than the classes provided in the training dataset. A trained model on only these classes will not be able to predict a new class (Gene) and might give a wrong prediction since it can only predict these particular classes which is misleading for the User/Doctor.
- f. Hyperparameter optimization is also an important tool to be used to push the model training towards the optimal solution. In our submission the tuning was done manually. Automated tuning will ensure better results.
- g. Having more images is always better.

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

Overall we think that the GPU resources were sufficient with some notes:

- Higher GPU resources will allow us to train with a bigger batch size which will help the training be more stable and might lead to better results.
- We were not able to train two models at the same time in two different notebooks/kernels. This was a major limitation since our approach consists in training 4 independent models that can be trained at the same time. We had to train the models sequentially which consumed a lot of time.