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

Regarding the team working on this challenge, we are two data scientists with expertise in data engineering and deployment.
We are accustomed to working on projects of this nature and typically go beyond the proof-of-concept stage.
Our expertise extends to the entire project lifecycle, including deployment, maintenance, and more.

Within our organization, our team's objective is to develop tools and products for hospitals, pharmacies, doctors, and laboratories by leveraging data engineering and data science. Our overarching goal within the company is to enhance healthcare through the application of artificial intelligence.

IQVIA, our company, is a leading global provider of advanced analytics, technology solutions, and contract research services to the life sciences industry. With a primary focus on healthcare data and insights, IQVIA empowers clients to make informed decisions and drive innovation in pharmaceuticals, biotechnology, medical devices, and healthcare organizations. By combining deep industry knowledge with cutting-edge technologies, IQVIA delivers solutions that improve patient outcomes, optimize healthcare operations, and expedite the development of new therapies.

2. Explain why you participated in the Diagnodent challenge.

We decided to participate in the Diagnodent challenge for several reasons. First and foremost, we were eager to enhance our skills in machine and deep learning, particularly on imagery. Dealing with unstructured data and combining multi-modal images (photos and panoramic) has been a great challenge, and we have learnt a lot. This challenge provided us with an opportunity to apply and perfect our expertise on imagery and deep learning.

Additionally, we were highly motivated to the idea of working on a project that has practical implications and can contribute to the betterment of healthcare and patient care. By leveraging our expertise, we aim to tackle a meaningful problem and develop solutions that could positively impact dental healthcare.
Indeed, we hope that this challenge is not just providing a notebook and appearing on a leaderboard.
We are motivated to pursue and create effective and easy-to-use tool to assist dentists in their everyday job, by minimizing misdiagnoses and facilitating early detection, thus enabling patients to receive timely and appropriate treatment.

In fact, improving healthcare through new technologies and AI is the main reason we are part of the IQVIA R&D team right now.

In summary, our participation in the Diagnodent challenge was driven by our desire to enhance our skills, explore cutting-edge technologies and above all make a meaningful impact in the field of dental healthcare.

3. Describe how you built your winning model and elaborate on the technical and modeling choices you made.
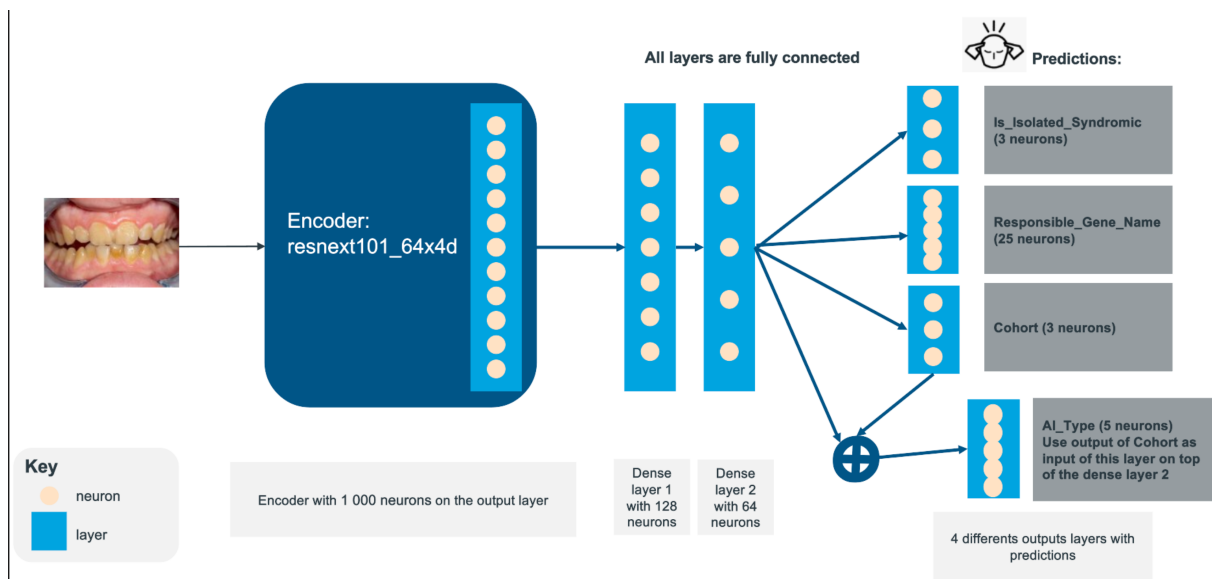
In the notebook, we provide detailed information on this question, including preprocessing techniques. However, it's important to note that the process to reach this model was not straightforward. We explored other approaches before arriving at the current one, which are also explained in detail at the end of our notebook.

Due to the limited number of available pictures, we decided to explore state-of-the-art encoders that were pre-trained on the ImageNet dataset. On images, we opted for a ResNext model with 1000 output neurons as our encoder. We then incorporated custom layers, which we referred to as our classifier. The outputs were divided into four categories: Cohort, AI_Type, Is_Isolated_Syndromic, and Responsible_Genes. Each category took 64 neurons as input for prediction, except for AI_Type, which additionally utilized 4 neurons from the Cohort output.

For Panoramics, the architecture is almost exactly the same except the encoder is different: DEIT encoder seem to have better performances on very small datasets like the panoramics one.

To provide a visual representation of our model, we have attached a diagram to this email, which can also be found in our notebook.

Furthermore, we made the decision to employ two models, one for photos and one for panoramics. This approach enables us to provide the dentist with predictions on a patient's condition even if he/she does not have both types of images available.

If you are interested, we can delve deeper into the parameters and techniques we used. Alternatively, you can find detailed explanations in our notebook.

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

Yes, from our perspective, the GPU resources provided in the challenge notebook were adequate. Although we had to make some adjustments and restart the kernel multiple times, we were able to load and train our data and models without significant issues. Even if the K80 isn't the most powerful GPU on the market, this forced us to remain frugal and design lightweight, portable models that can be used in real-life conditions. While we couldn't run our final submission notebook from top to bottom without restarting the kernel, the GPU resources fulfilled the requirements of the challenge effectively. We appreciate the provision of the GPU, which enabled us to perform the necessary computations efficiently.