CXRML Classifier Architecture Overview

The frontend of the application is hosted on Amazon Web Services (AWS), specifically

utilizing AWS Elastic Beanstalk (EBN) for deployment. The frontend consists of multiple

components: the Home Page, About Page, and the Upload Image functionality. Elastic Beanstalk

automates the provisioning of the necessary infrastructure, such as load balancing, scaling, and

monitoring, ensuring high availability and performance of the application. The decision to use

AWS is driven by its reliability, scalability, and the vast ecosystem of services that seamlessly

integrate, ensuring that the frontend is robust and capable of handling fluctuating traffic

demands. Additionally, AWS's global infrastructure provides low latency and fast content delivery.

The backend of the application is deployed on Google Cloud Platform (GCP) using Cloud Run, which is a fully managed compute platform that automatically scales stateless containers. Cloud Run is ideal for this application as it allows for rapid deployment of containerized applications. The backend is powered by a Convolutional Neural Network (CNN) machine learning model that processes the uploaded images to provide predictions. The choice of GCP, and specifically Cloud Run, is based on its ability to integrate smoothly with machine learning workflows. GCP's Cloud Run provides the necessary scalability to handle spikes in processing requests, while also offering a pay-as-you-go pricing model, which is cost-effective. The combination of Cloud Run with the CNN model allows the application to deliver high performance in processing and generating predictions, ensuring a seamless experience for end-users. This architecture benefits from GCP's strengths in machine learning and container orchestration, making it the optimal choice for a backend that demands both power and flexibility.

The primary reason for using both AWS and GCP in this setup is to gain experience and knowledge about both cloud platforms. There is no specific technical reason for the use of two clouds; rather, this choice is driven by the desire to broaden my understanding and skills across multiple cloud environments. Learning and familiarizing myself with different cloud technologies is a valuable objective, which is why I opted to incorporate both AWS and GCP into my architecture.