mlangles CloudAngles

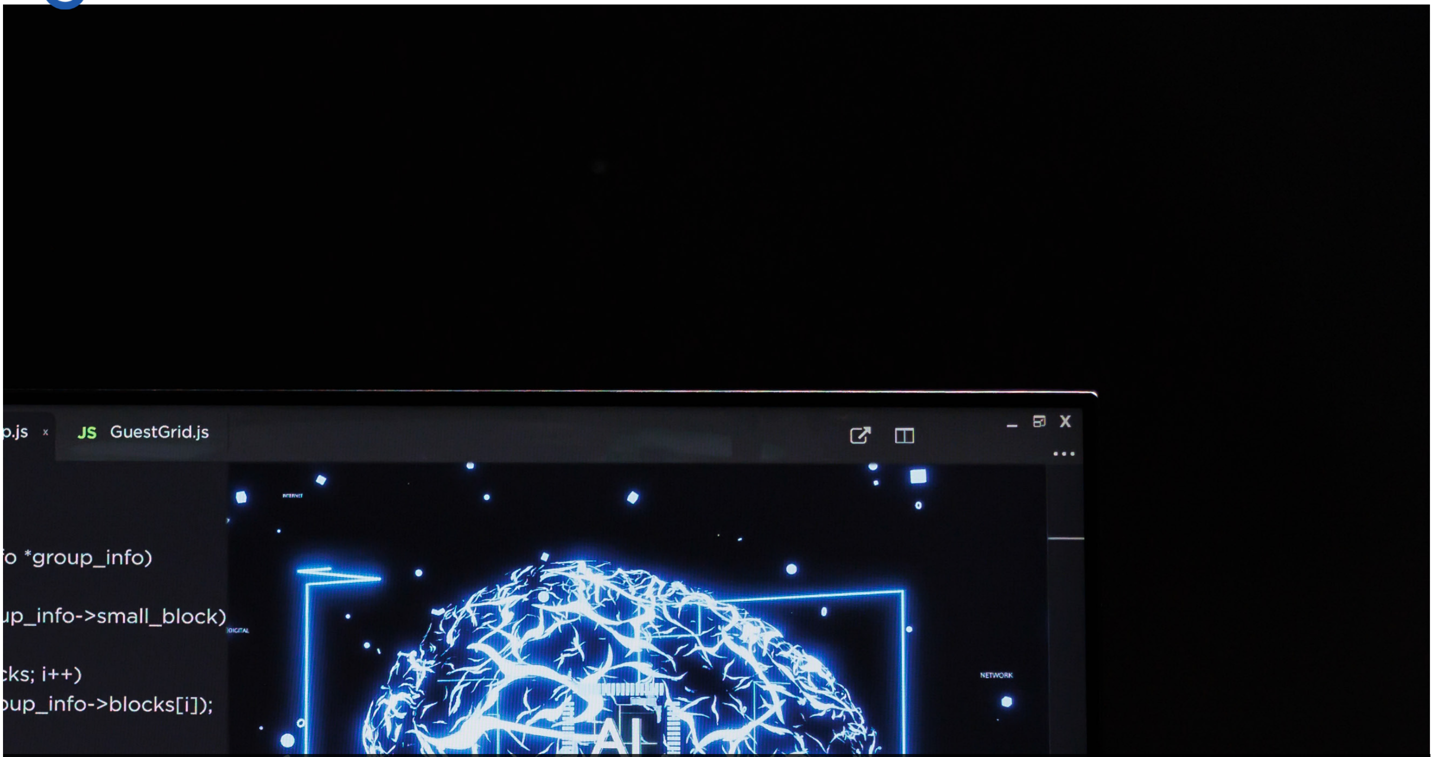
mlangles Predictive AI

Harmful Brain Waves Prediction for early Neurological disorder diagnosis

About mlangles Predictive AI

mlangles is a comprehensive AI platform designed to manage the lifecycle of data and models, offering streamlined solutions for every stage of the process.

Through its Predictive AI component, mlangles provides a suite of tools to navigate efficiently through each phase of AI project development, encompassing data engineering, development, deployment, and monitoring. It facilitates continuous integration, continuous deployment, continuous training, continuous monitoring (CI-CD-CT-CM), enabling enterprises to effectively manage their AI initiatives.



Objective of the Use Case

The goal of this use case is to detect and classify seizures and other types of harmful brain activity and develop a model trained on electroencephalography (EEG) signals recorded from critically ill hospital patients.

It may help rapidly improve electroencephalography pattern classification accuracy, unlocking transformative benefits for neurocritical care, epilepsy, and drug development. Advancement in this area may allow doctors and brain researchers to detect seizures or other brain damage to provide faster and more accurate treatments.

EXPLAINATION OF THE USECASE:

EEG (electroencephalogram) is a test that measures electrical activity in the brain using small, flat metal discs (electrodes) attached to the scalp. The channels (e.g., Fp1, F3, C3, P3, etc.) represent specific locations on the scalp where the electrical activity is being measured. Each number in your data represents the amplitude of the electrical activity recorded at a specific electrode location during a certain time. This is to detect and classify seizures and other types of harmful brain activity in electroencephalography (EEG) data.

EXPLAINATION ON TARGETS:

The target columns are seizure (SZ), generalized periodic discharges (GPD), lateralized periodic discharges (LPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), or “other”.

Lateralized Periodic Discharges (LPD):

Refers to periodic discharges which are often associated with brain abnormalities

Generalized Periodic Discharges (GPD):

Refers to periodic discharges which are often associated with brain abnormalities

Lateralized Rhythmic Delta Activity (LRDA):

Rhythmic delta activity refers to regular, repetitive patterns of slow-wave delta waves in the EEG.

Generalized Rhythmic Delta Activity (GRDA):

Like LRDA, GRDA is characterized by rhythmic patterns of slow-wave delta waves but is not confined to one side of the brain.

Metadata for the train set. The expert annotators reviewed 50-second-long EEG samples plus matched spectrograms covering 10-minute window centered at the same time and labeled the central 10 seconds. Many of these samples overlapped and have been consolidated. train.csv provides the metadata that allows you to extract the original subsets that the raters annotated.

Metadata for the test set. As there are no overlapping samples in the test set, many columns in the train metadata don't apply.

Spectrograms are assembled using exactly 10 minutes of EEG data. Use the metadata in train.csv to select specific annotated subsets. The column names indicate the frequency in hertz and the recording regions of the EEG electrodes. The latter are abbreviated as LL = left lateral; RL = right lateral; LP = left parasagittal; RP = right parasagittal.

The EEG parquet files are longer than 50 seconds. One EEG parquet file has multiple rows of time windows inside it. Similarly, the Spectrogram parquet files are longer than 600 seconds. One Spectrogram parquet file has multiple rows of time windows inside it.

mlangles CloudAngles

WORKING OF THE USECASE

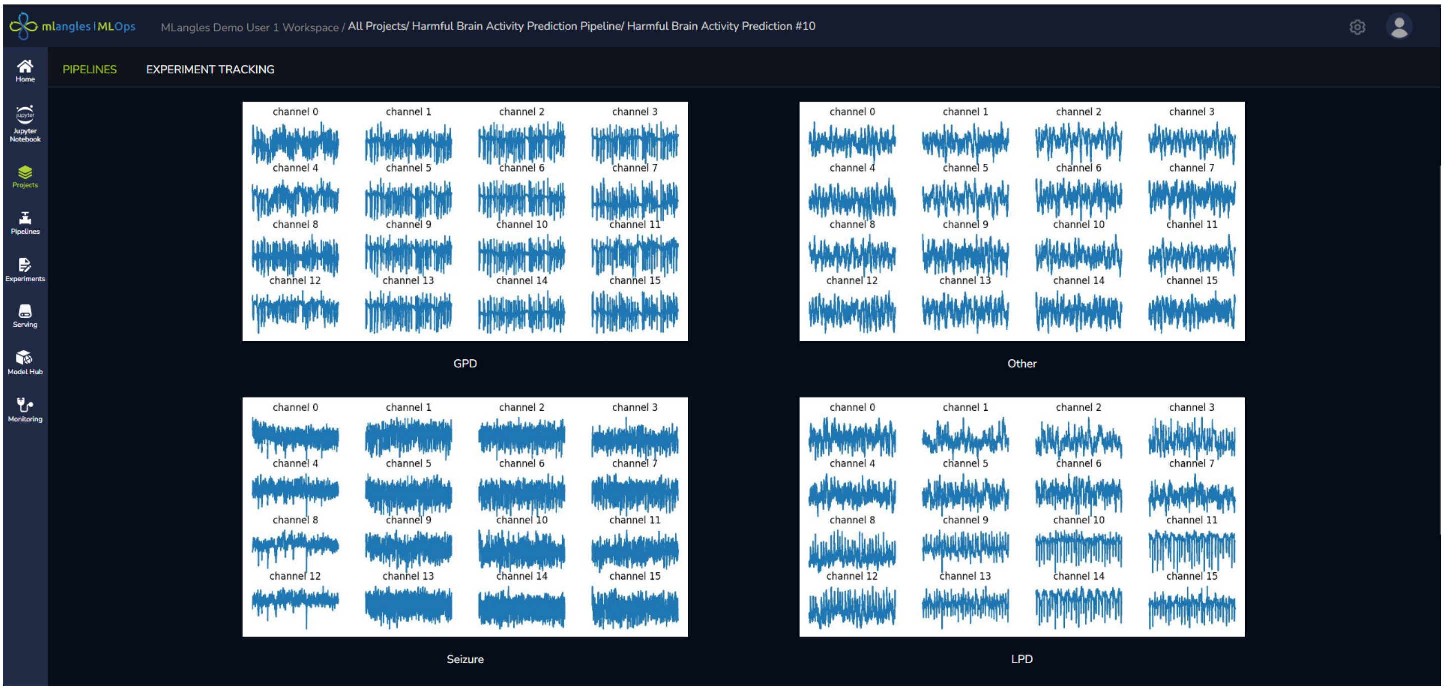
STEP 1: DATA ENGINEERING AND PIPELINE CREATION:

Install Dependencies: Essential packages and libraries have been installed.

Data Extraction: The data is loaded from the source (in this case the source is s3 bucket)

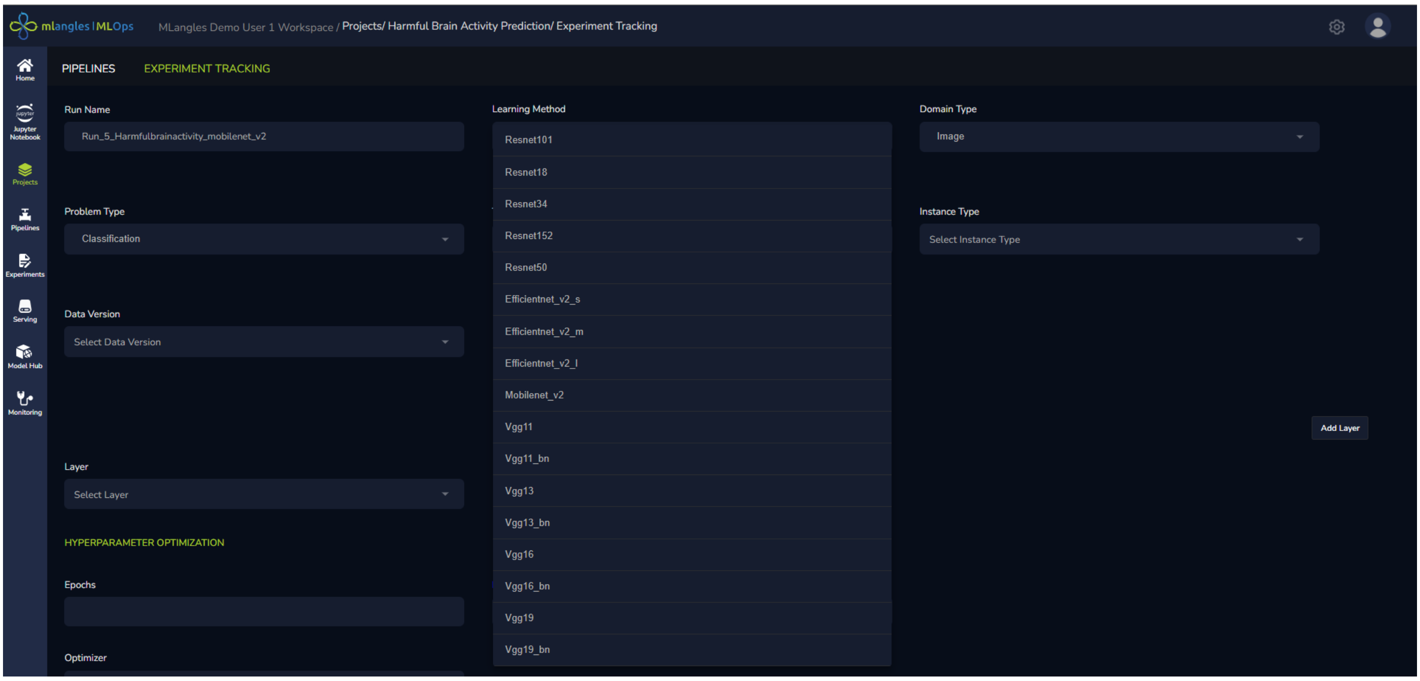
Data Visualization: One sample image from each of the 6 types of brain activities which are seizure (SZ), generalized periodic discharges (GPD), lateralized periodic discharges (LPD), lateralized rhythmic delta activity (LRDA), generalized rhythmic delta activity (GRDA), or "other".

Data Transformation: A series of image transformations is performed on the input images. These transformations include resizing the image to 256x256 pixels, cropping the center to 224x224 pixels, converting the image to a PyTorch tensor, and normalizing the image with predefined mean and standard deviation values. The transformed images are used to create the dataset which is stored to perform the training.


mlangles | MLOps
MLangles Demo User 1 Workspace / All Projects/ Harmful Brain Activity Prediction Pipeline/ Harmful Brain Activity Prediction #10
Home
Jupyter Notebook
Projects
Pipelines
Experiments
Serving
Model Hub
Monitoring
PIPELINES
EXPERIMENT TRACKING
GPD
Other
Seizure
LPD


mlangles | MLOps
MLangles Demo User 1 Workspace / All Projects/ Harmful Brain Activity Prediction Pipeline/ Harmful Brain Activity Prediction #10
Home
Jupyter Notebook
Projects
Pipelines
Experiments
Serving
Model Hub
Monitoring
PIPELINES
EXPERIMENT TRACKING
Seizure
LPD
GRDA
LRDA

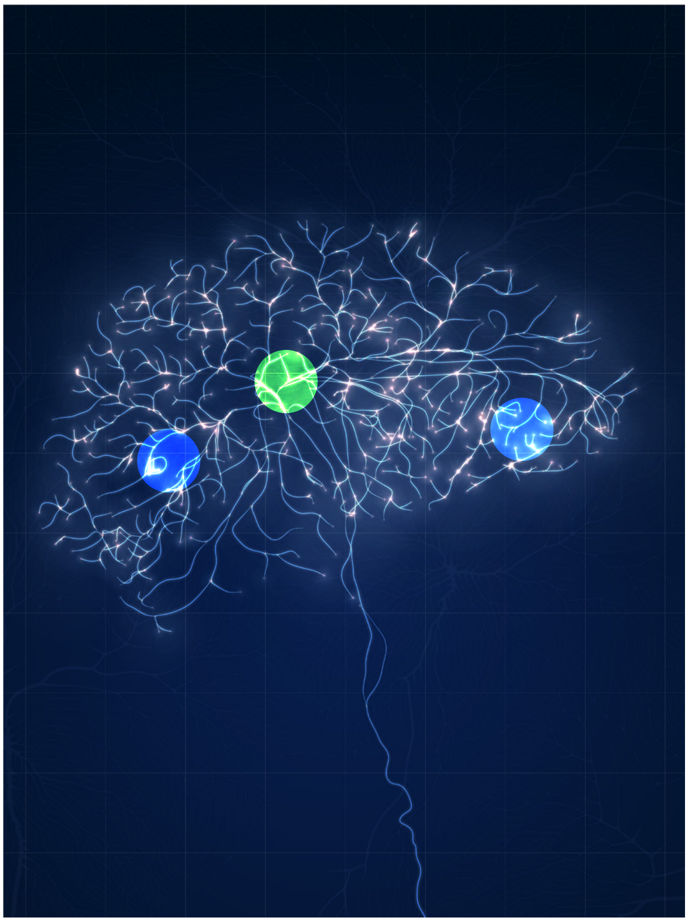
STEP 2: EXPERIMENT TRACKING- MODELLING

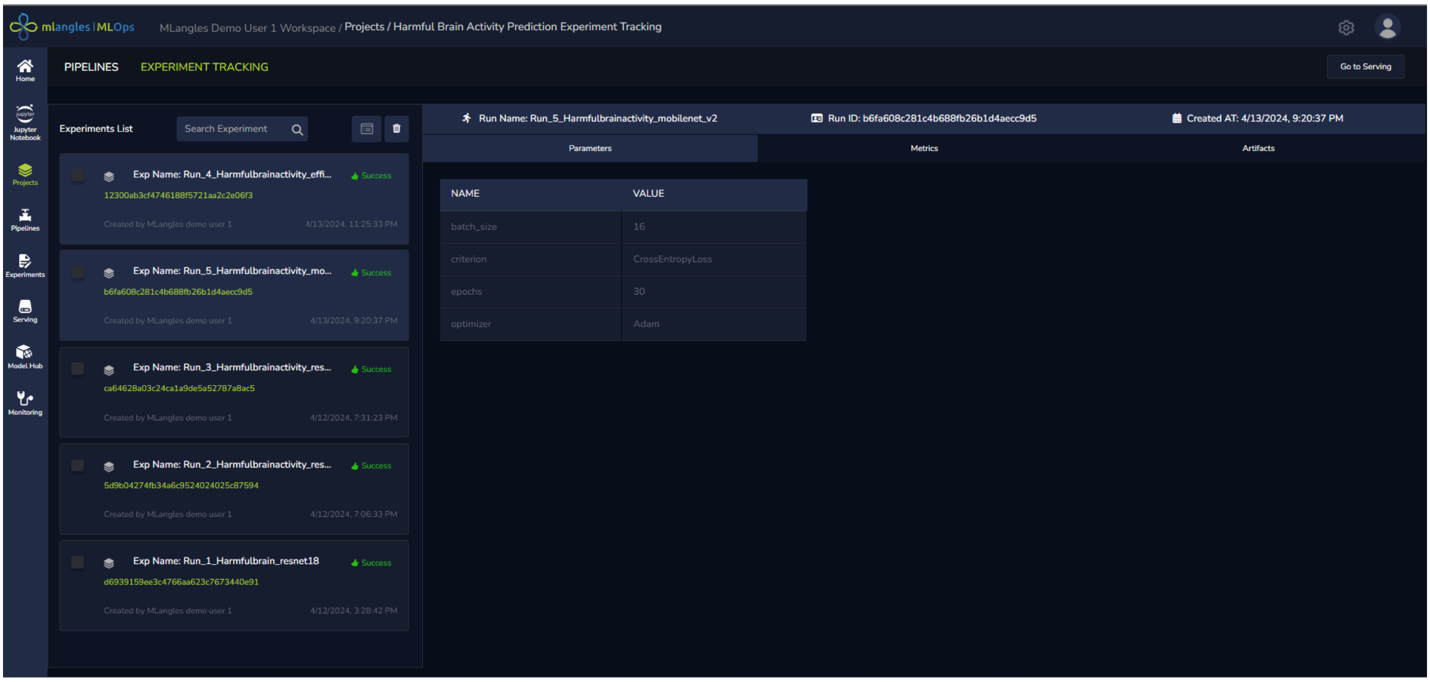


Once the dataset is created from the pipelines, it is then used to finetune any of the various pretrained models such as ResNet, VGG, Efficient Net. The idea of using pretrained models involves utilizing knowledge gained from training a model on one task to improve performance on a different but related task. In the context of using pretrained models for predicting tumors, transfer learning entails taking a model that has been pre-trained on a large dataset (such as ResNet, pretrained on ImageNet), and fine-tuning it on a smaller dataset containing images of brain scans with four classes: glioma, pituitary, meningioma, and no tumor. By leveraging the features learned during the initial training on ImageNet, the model can quickly adapt to the new

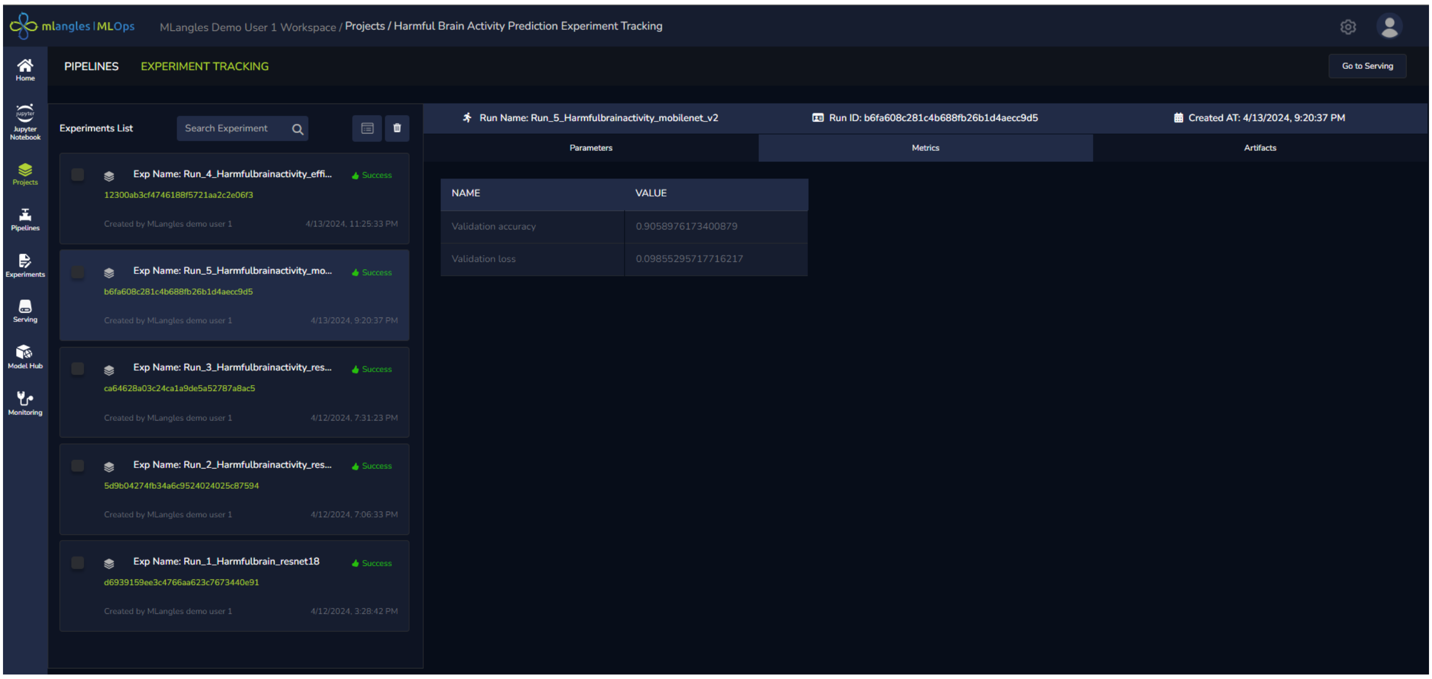
task of tumor classification, potentially achieving higher accuracy and efficiency compared to training from scratch. This approach reduces the need for extensive labeled data and computational resources, making it a powerful technique for medical image analysis tasks. Once the model is fine-tuned, the parameters used to finetune, the results of fine-tuning are displayed in the form of parameters, metrics and artifacts respectively.

The parameters would include the number of epochs that the model is finetuned for, the batch size, the optimizer and the loss function along with the respective sub-parameters.

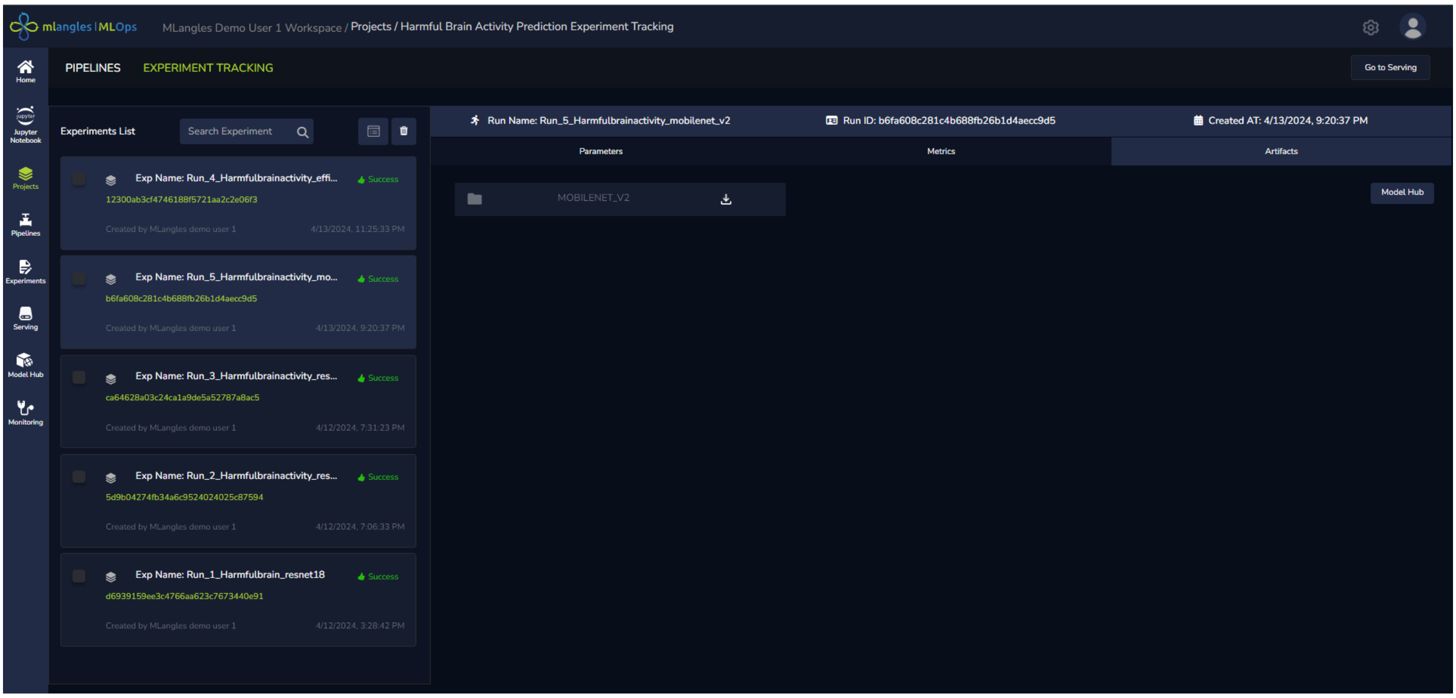




The metrics include the best validation accuracy and the least validation loss recorded during finetuning which measure how well the model generalizes to unseen data, often referred to as the validation set, which is distinct from the training data.

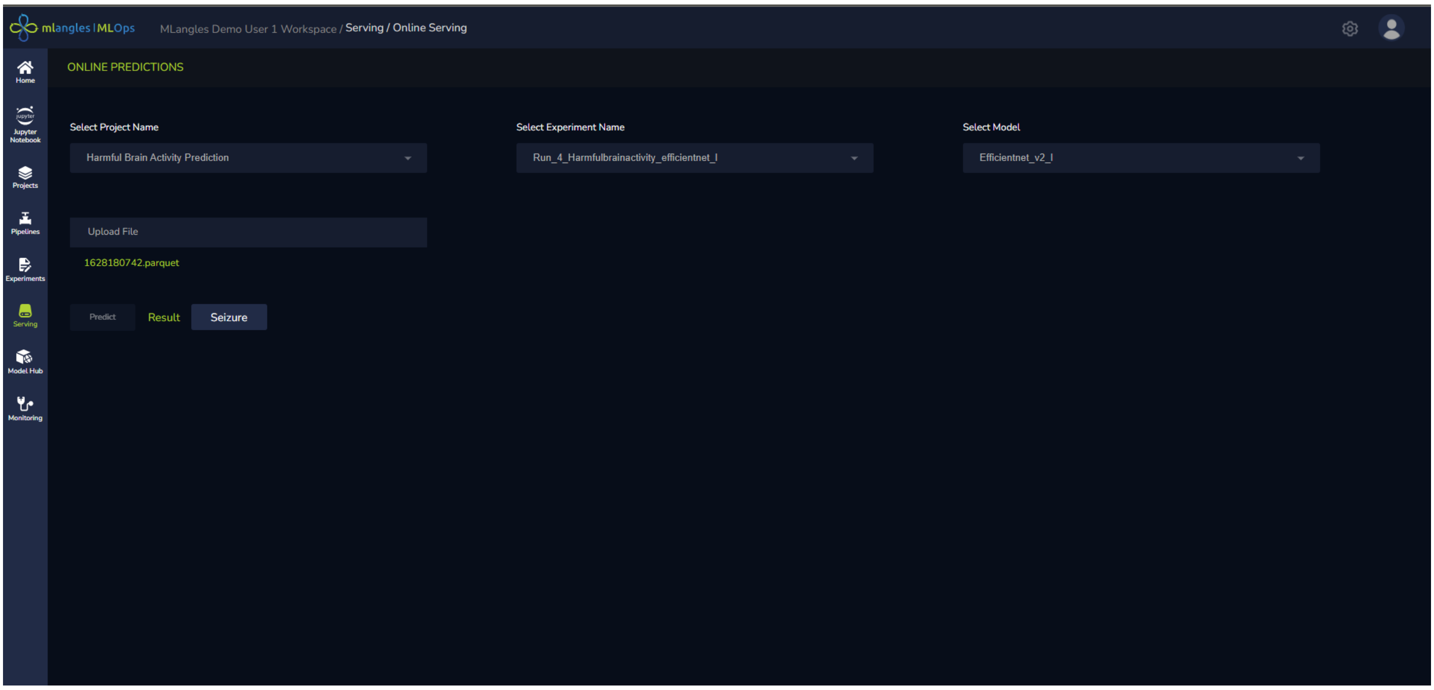


The trained model along with its dependencies could be downloaded from artifacts to perform predictions offline.



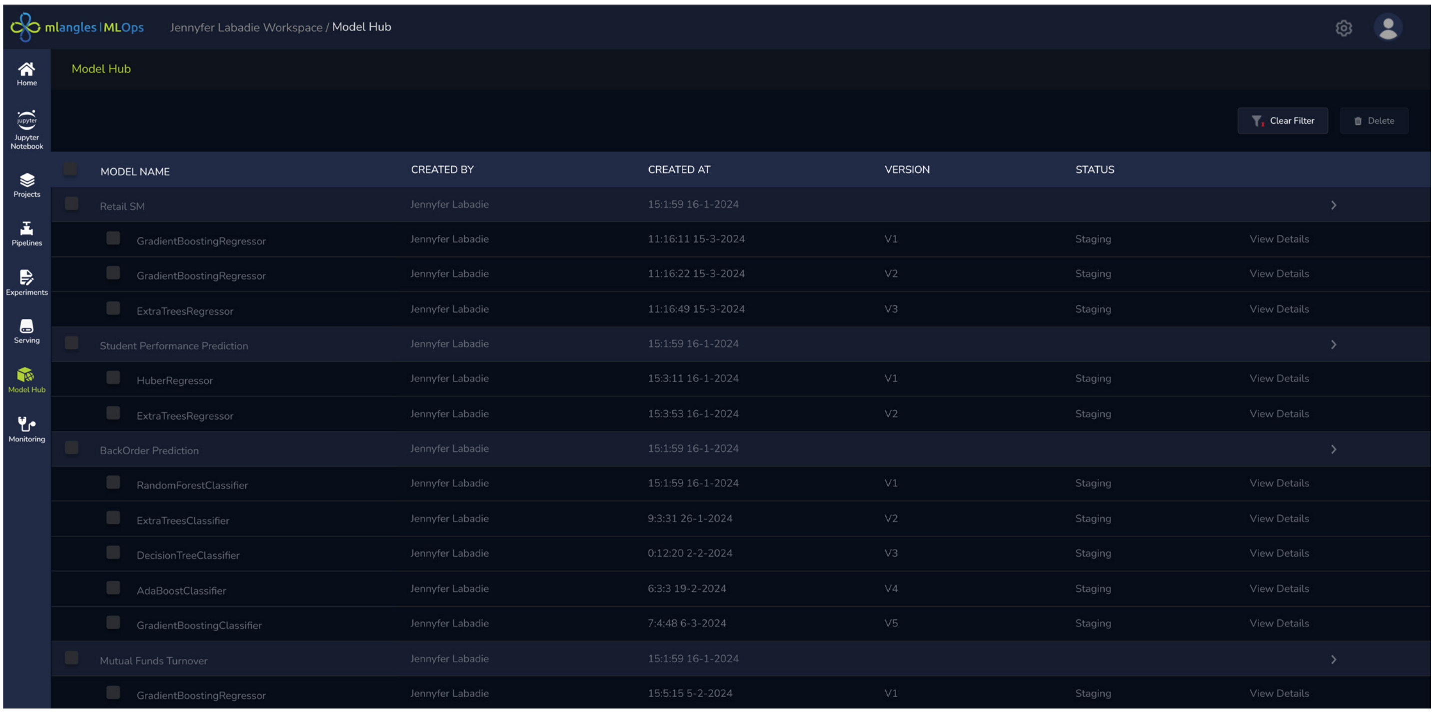
STEP 3: SERVING

Serving starts by uploading a Parquet file containing EEG (electroencephalogram) data. Once the Parquet file is uploaded, it preprocesses the EEG data to generate spectrogram data. A spectrogram is a visual representation of the spectrum of frequencies in a signal as it varies with time. In the context of EEG data, this means converting the time-domain EEG signals into the frequency-domain representation. The spectrogram data is then transformed into images. Finally, the generated images are used for prediction.



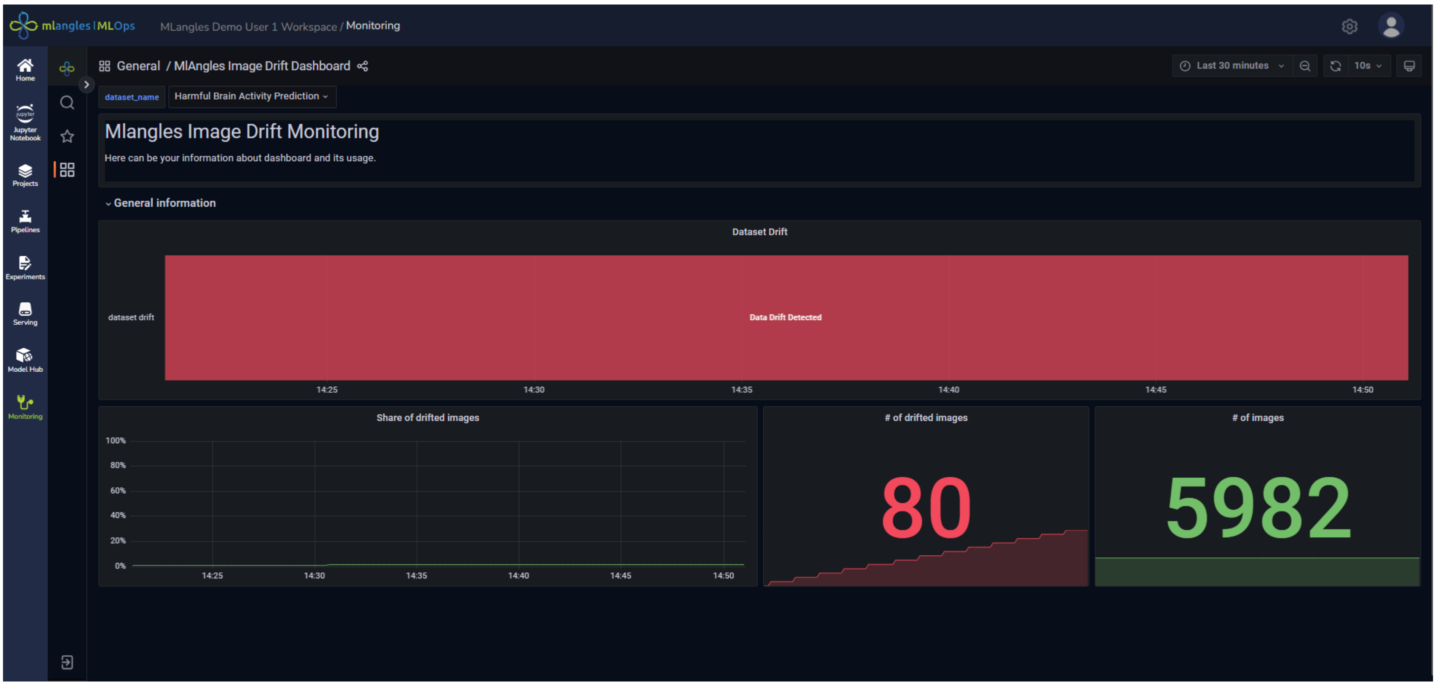
Model Hub:

- Trained models are uploaded to the model hub, whereupon deployment, a REST API endpoint is automatically generated.
- Data is transmitted to this endpoint as a request, triggering the model to execute a prediction and return the output as the response to the request.



STEP 4: MONITORING

Image drift refers to the phenomenon where the statistical properties of images in a dataset change over time, potentially leading to a decrease in the performance of machine learning models trained on that dataset. This change can occur due to various factors such as changes in imaging equipment, variations in lighting conditions, differences in image acquisition techniques, or shifts in the characteristics of the target population being imaged.



Conclusion

In conclusion, the development of machine learning models for EEG data analysis, particularly utilizing advanced architectures like ResNet, holds immense promise for revolutionizing the diagnosis and treatment of neurological disorders. By combining cutting-edge technology with clinical expertise, we can unlock transformative benefits for patients, clinicians, and researchers alike, ultimately advancing our understanding of the brain and improving healthcare outcomes.

To setup a demo

Info.mlangles@cloudangles.com

Visit: www.mlangles.ai