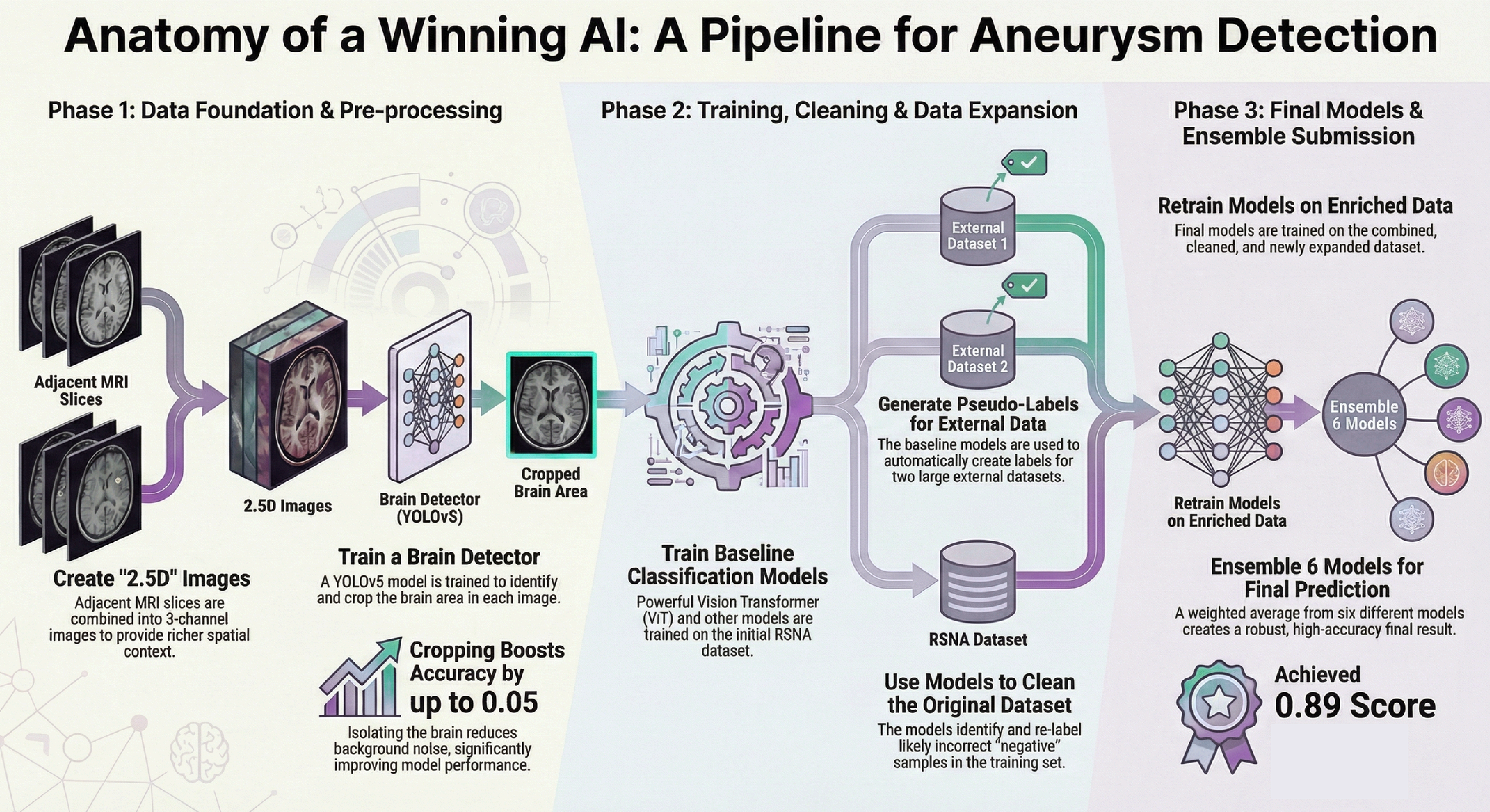
Engineering an Advanced Pipeline for Intracranial Aneurysm Detection

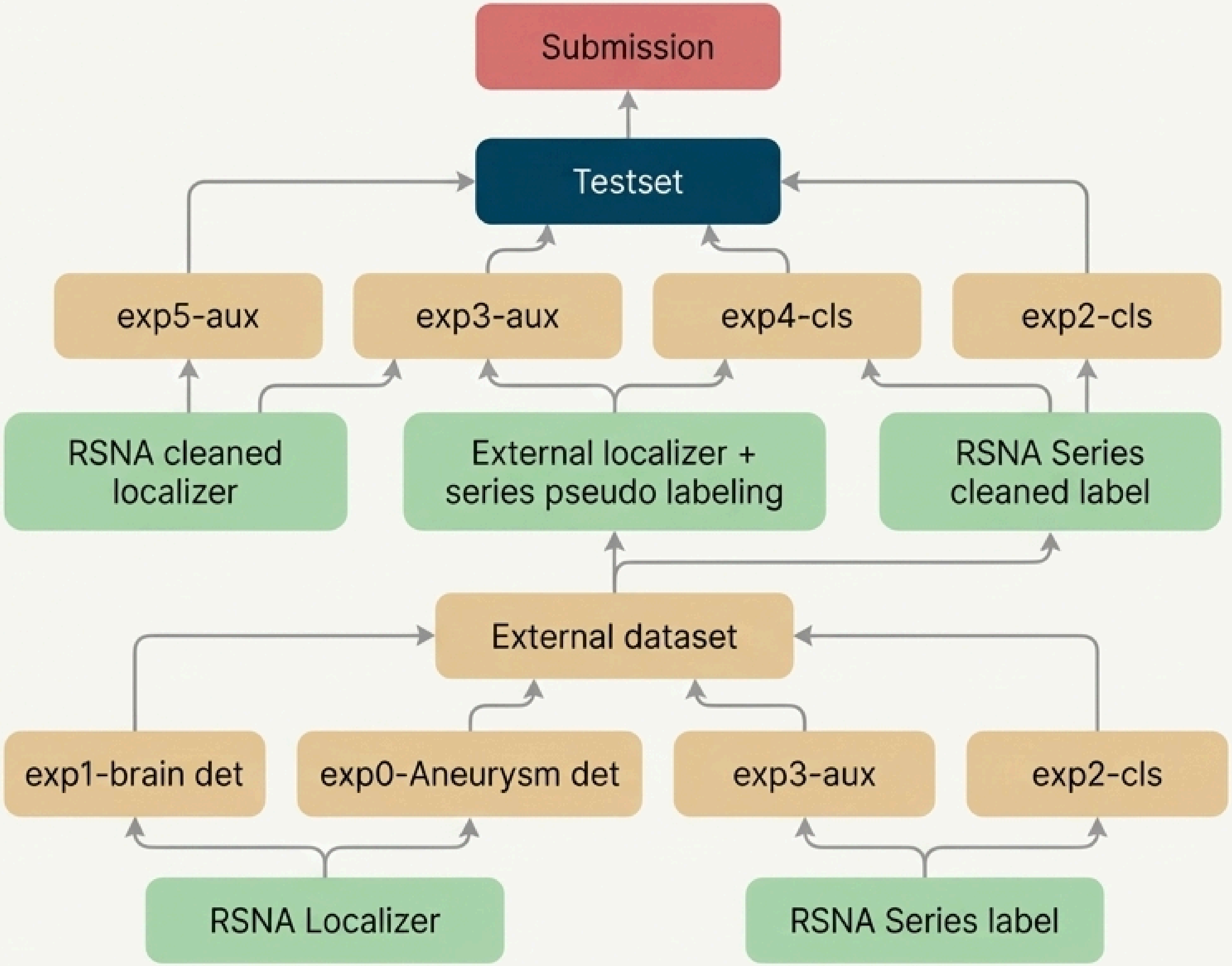
A Strategic Walkthrough of Our Deep Learning Architecture

Pugazhendhi J (230066)
J Amathziah (230139)
Meesala Sree Sai Nath (230103)


Anatomy of a Winning AI: A Pipeline for Aneurysm Detection
Phase 1: Data Foundation & Pre-processing
Phase 2: Training, Cleaning & Data Expansion
Phase 3: Final Models & Ensemble Submission
Adjacent MRI Slices
2.5D Images
Brain Detector (YOLOvS)
Cropped Brain Area
Create "2.5D" Images
Adjacent MRI slices are combined into 3-channel images to provide richer spatial context.
Train a Brain Detector
A YOLOv5 model is trained to identify and crop the brain area in each image.
Cropping Boosts Accuracy by up to 0.05
Isolating the brain reduces background nolse, significantly improving model performance.
Train Baseline Classification Models
Powerful Vision Transformer (ViT) and other models are trained on the initial RSNA dataset.
External Dataset 1
External Dataset 2
Generate Pseudo-Labels for External Data
The baseline models are used to automatically create labels for two large external datasets.
RSNA Dataset
Use Models to Clean the Original Dataset
The models identify and re-label likely incorrect "negative" samples in the training set.
Retrain Models on Enriched Data
Final models are trained on the combined, cleaned, and newly expanded dataset.
Retrain Models on Enriched Data
Ensemble 6 Models
Ensemble 6 Models for Final Prediction
A weighted average from six different models creates a robust, high-accuracy final result.
Achieved 0.89 Score

Our Solution: More Than a Model, an Evolving System

Our success was built not on a single model, but on a series of strategic experiments. Each stage was designed to solve a specific problem, creating a system that learns from its own predictions to continuously improve data quality and model accuracy.



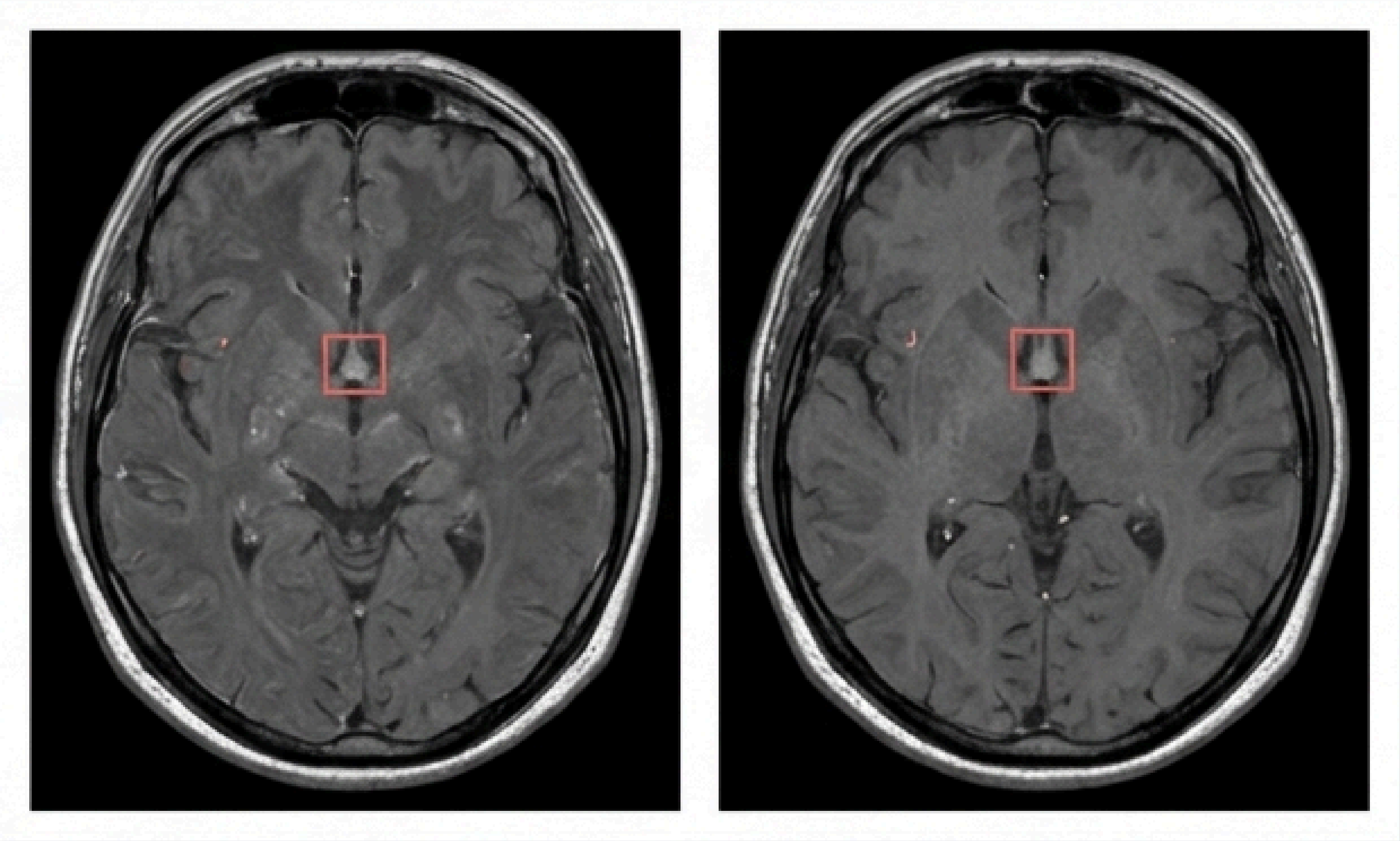
Step 1: Building a Foundational Aneurysm Detector (Exp0)

Problem

Simple classification is insufficient. We needed precise localization to guide later models and improve our data.

Solution

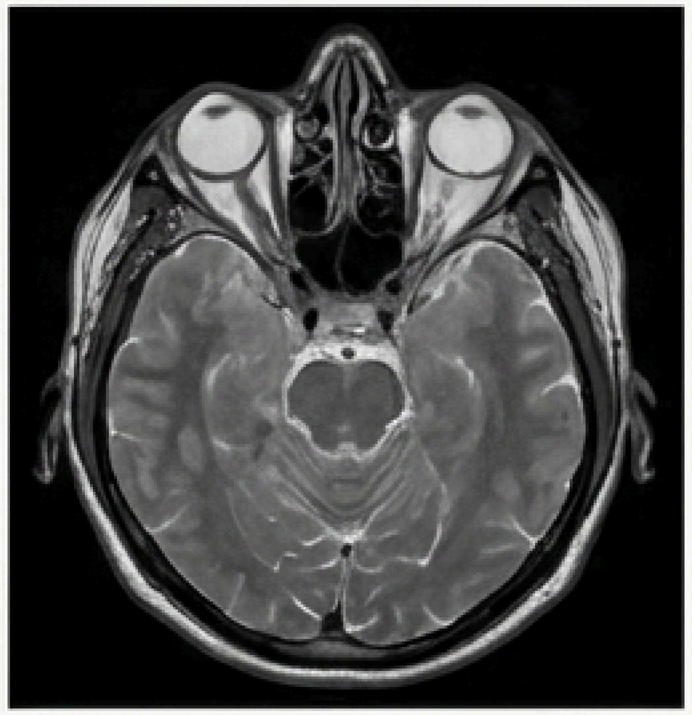
- Used provided aneurysm centroids as a starting point.
- Manually annotated bounding boxes in ±10 neighboring slices using LabelImg.
- Trained five YOLOv11x models (one for each data fold) to create a robust initial detector.



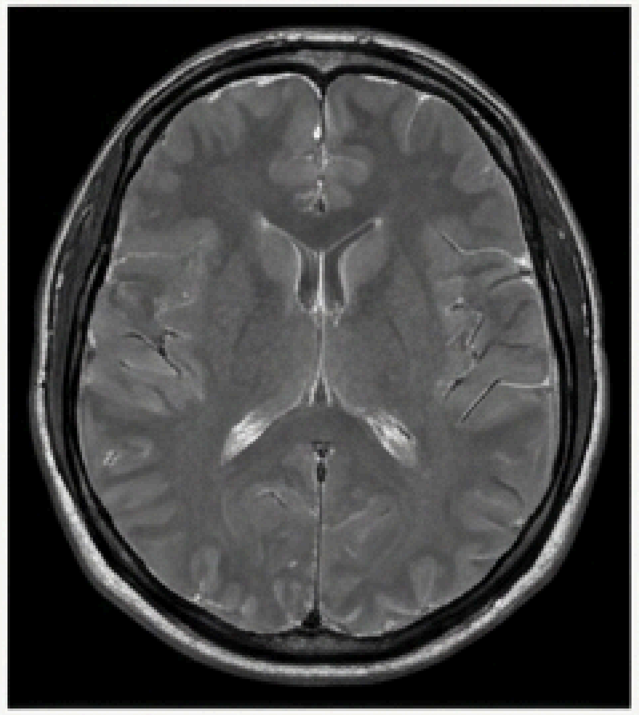
Performance		
Fold	**mAP50**	**mAP50-95**
Fold 0	0.705	0.460
Fold 1	0.647	0.429
Fold 2	0.766	0.504
Fold 3	0.702	0.482
Fold 4	0.691	0.449

Step 2: Isolating the Signal by Removing Background Noise (Exp1)

Before: Raw Slice



After: Cropped Slice



Problem

Problem

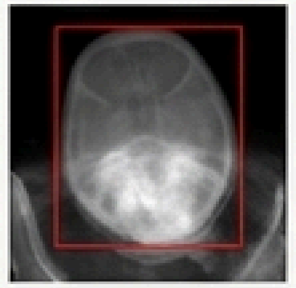
Raw images contain significant background noise (e.g., lung regions) that can confuse classification models.

Solution

Solution

- Trained a dedicated, lightweight YOLOv5 model to detect only the brain.
- Manually annotated two classes: "brain" (axial view) and "abnormal" (other views).
- Used this model to crop every slice in every series, forcing subsequent models to focus on relevant anatomy.

brain



abnormal

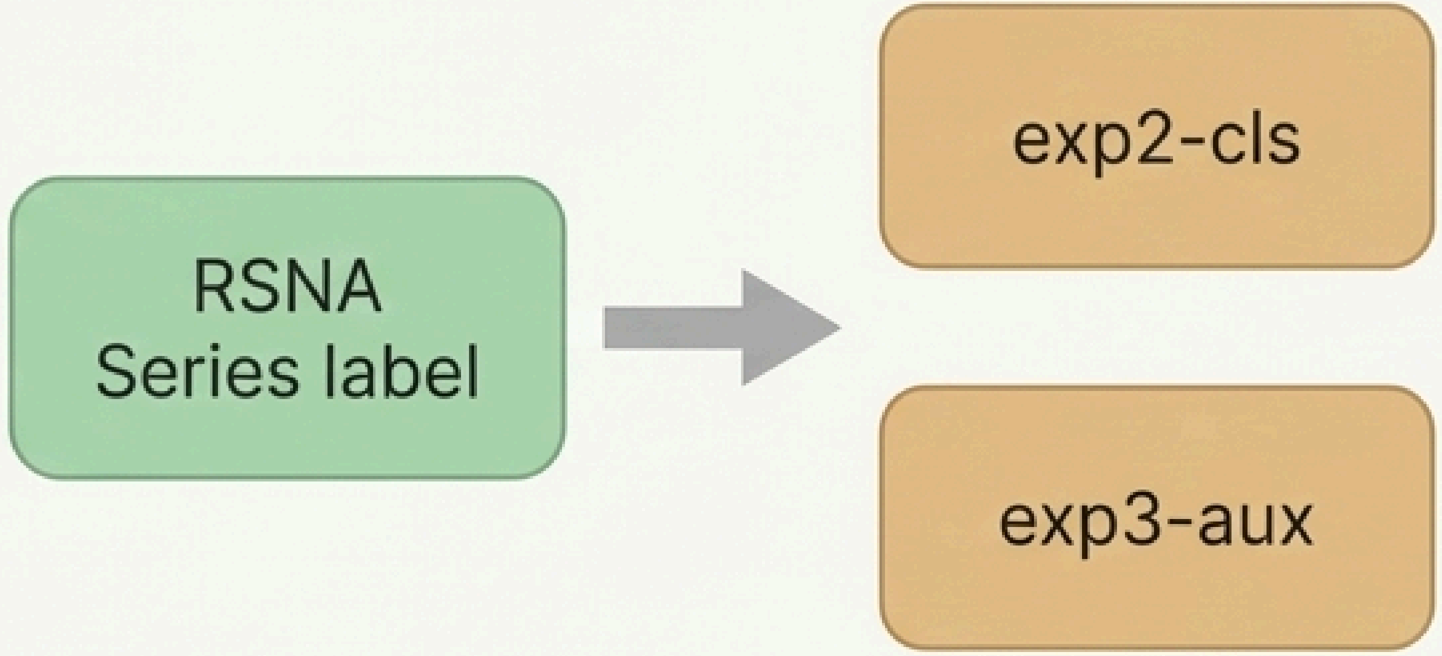
Impact

This single step improved overall accuracy by a remarkable 0.03–0.05.

Step 3: Establishing a Strong Baseline (Exp2 & Exp3)

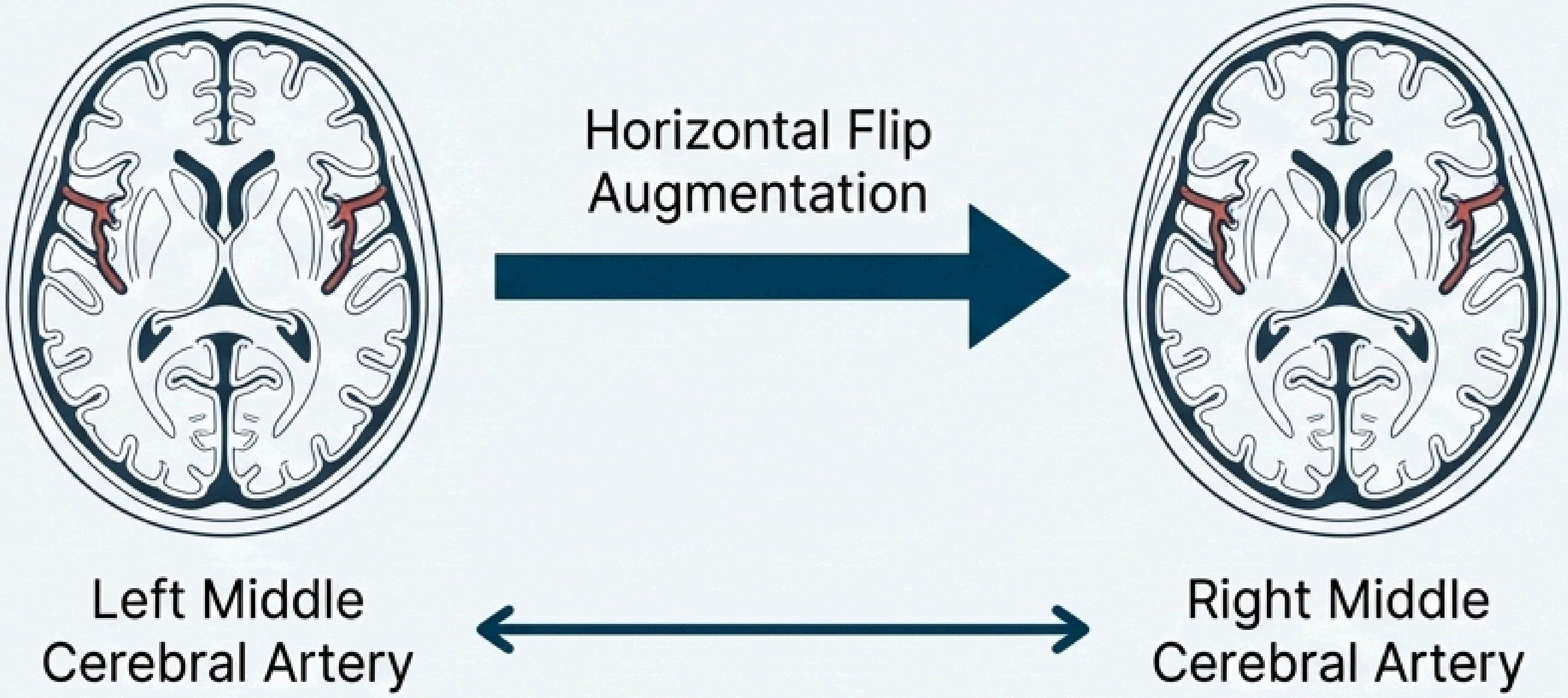
With cleaned, cropped data from the previous step, we trained a suite of initial models on the original dataset. This included both pure classification and multi-task (classification + segmentation) approaches to create a robust performance baseline.

- ViT large 384 (Exp2): **0.8503**
- eva large 384 (Exp2): **0.8551**
- mit-b4 FPN (Multi-task, Exp3): **0.8549**



An Unconventional Edge: Exploiting Anatomical Symmetry

Insight: We hypothesized that we could leverage the brain's bilateral symmetry to effectively double our training data.



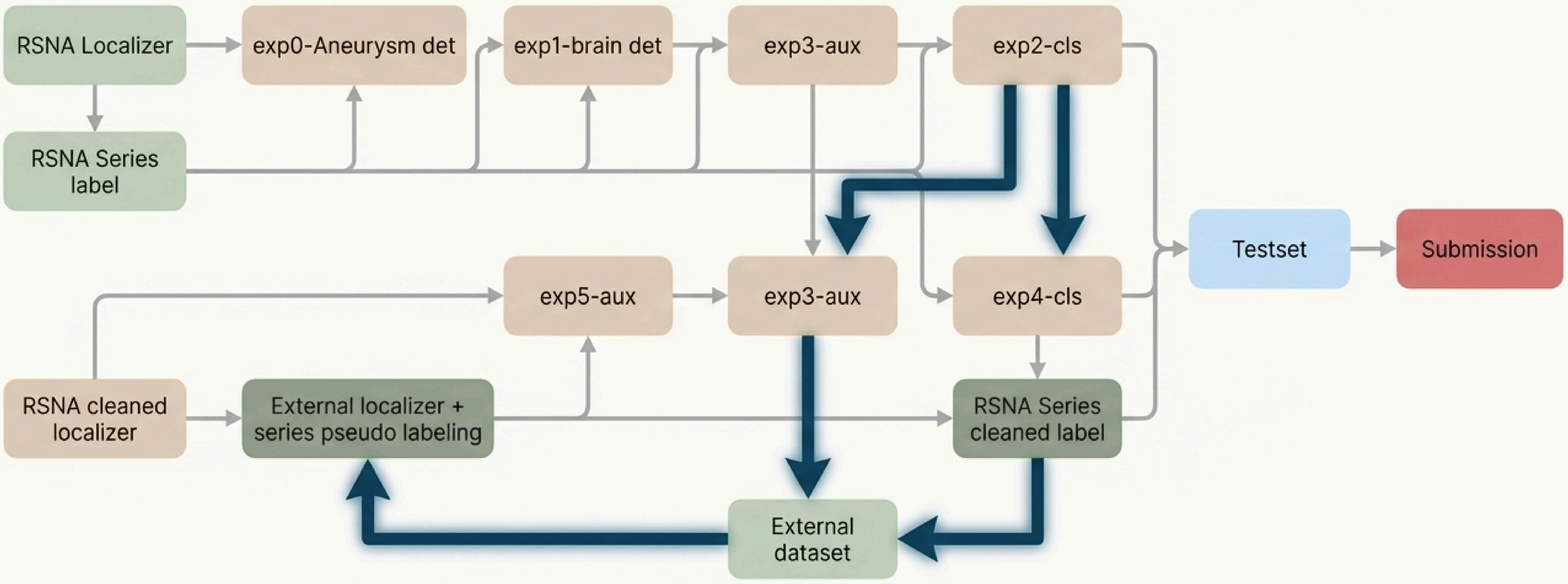
Technique

- We applied a standard horizontal flip augmentation.
- Crucially, we also swapped the corresponding anatomical labels (e.g., Left Middle Cerebral Artery <-> Right Middle Cerebral Artery).

Result: Improved model accuracy by an additional 0.01.

The Core Innovation: A Self-Improving Data Engine

Our baseline models were strong. Strong enough, we reasoned, to act as teachers. We decided to use their predictions to refine our existing training data and generate high-quality labels for new, external datasets.



The Engine in Action, Part 1: Cleaning the Training Set

Negative Series

Run baseline models (ViT large and mit-b4) on negative series.

High Prediction Score > 0.9

>0.9

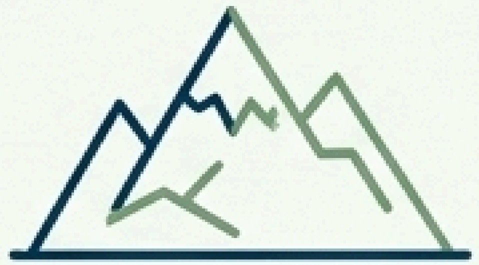
Flag series as likely mislabel.

Relabel as Positive

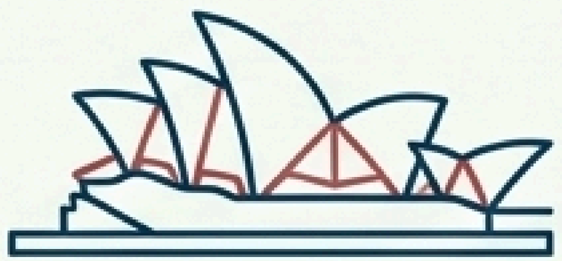
Generate BBox with Exp0

Outcome: A corrected, higher-fidelity training set.

The Engine in Action, Part 2: Pseudo-Labeling External Datasets



`Lausanne_TOFMRA`



`Royal_Brisbane_TOFMRA`

1. **Series Labels:** We used our baseline models to predict series-level labels for all external data.

2. **Localization Labels:** For positive series with a score > 0.5, we used our Exp0 Aneurysm Detector to generate bounding box pseudo-labels.

Outcome: A massive, diverse, and automatically labeled dataset to augment our original training data.

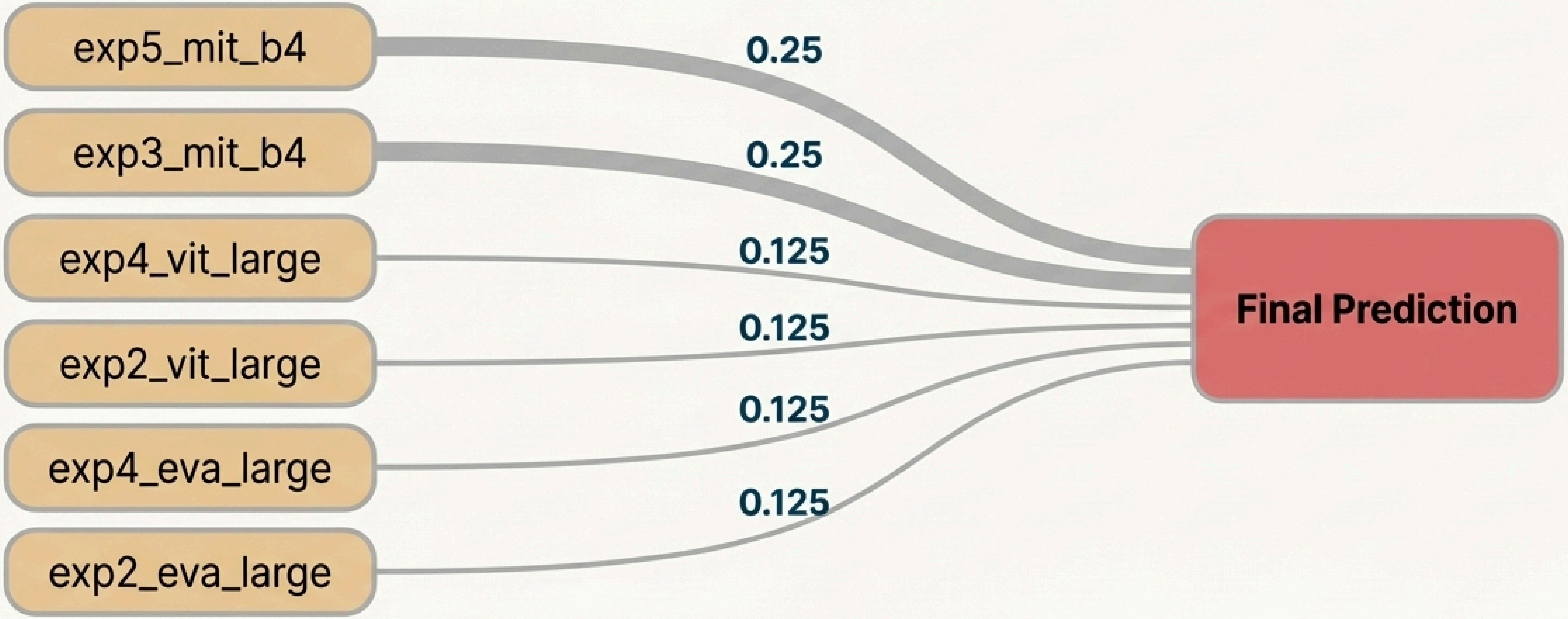
The Result: Quantifiable Lift from Superior Data (Exp4 & Exp5)

We retrained our key architectures on the combined, enhanced dataset (cleaned initial data + pseudo-labeled external data). The performance improvement was significant and **consistent across** all models.

Performance Comparison (OOF AUC)		
Model	**Baseline AUC (Original Data)**	**Final AUC (Enhanced Data)**
vit large 384	0.8503	**0.8558 ↗**
eva large 384	0.8551	**0.8579 ↗**
mit-b4 FPN	0.8549	**0.8629 ↗**

The Final Synthesis: An Ensemble of Six Specialized Models

To achieve maximum accuracy and robustness, we created a weighted ensemble of six models, blending predictions from models trained on both the original and the enhanced datasets. This leverages the unique strengths of each "expert."



Project Outcome: A New Benchmark in Detection Accuracy

Our final ensemble pipeline achieved an exceptional level of performance, demonstrating the power of our iterative, data-centric approach.

Local Cross-Validation (OOF AUC)

0.8823

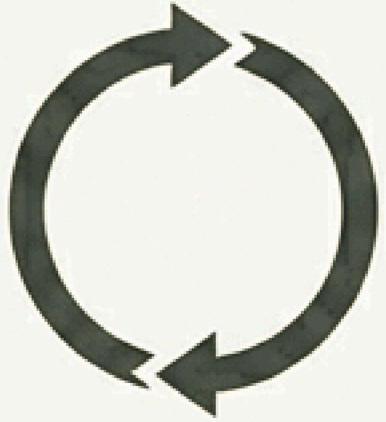
Performance on Test Data

0.89

Our Architecture of Success: Key Takeaways

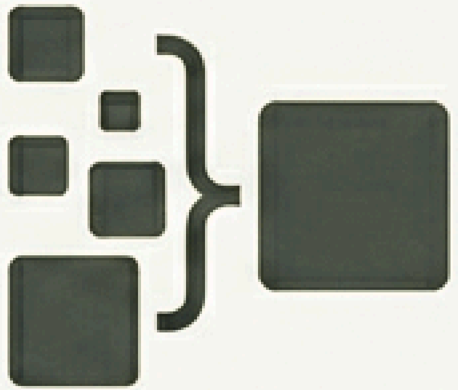
1. Strategic Pre-processing is Non-Negotiable

Isolating the brain with a dedicated detector provided a critical 0.03-0.05 accuracy boost that compounded through the entire pipeline.



2. Models Can Be Teachers

The self-improving loop, using strong models to clean existing data and pseudo-label new data, was the engine of our performance gains.



3. Ensemble Diversity is Key

Combining models trained on both original and enhanced datasets created a final solution more robust and accurate than any single component.

Thank You.

Pugazhendhi J
J Amathziah
Meesala Sree Sai Nath

Project Repository: github.com/pugazhjs9/RSNA-Intracranial-Aneurysm-Detection