**Report: RSNA Pneumonia Detection Model (V32)**

**Objective:** Develop a robust deep learning model to classify chest X-rays as Normal or Pneumonia, resolving data pipeline instability, overfitting, and class imbalance for the RSNA dataset.

**Dataset Overview**

**Source:** RSNA Pneumonia Detection Challenge.  
**Class Distribution:**

* **Normal:** 75% (4,135 samples)
* **Pneumonia:** 25% (1,282 samples)  
  **Preprocessing:**
* Images resized from 1024x1024 to 224x224 for DenseNet compatibility.
* Normalized pixel values to [0, 1].

**Model Architecture (V32)**

**Base Model:** DenseNet201 (pretrained on ImageNet).

* **Input Shape:** (224, 224, 3)
* **Fine-Tuning Strategy:**
  + **Frozen Layers:** First 200 layers (preserve generic feature extraction).
  + **Trainable Layers:** Last 53 layers (adapt to pneumonia-specific patterns).

**Custom Classification Head:**

python

Copy

Sequential([

SpatialDropout2D(0.7), # Combat overfitting in convolutional layers

GlobalAveragePooling2D(), # Reduce spatial dimensions

Dense(64, activation='relu', kernel\_regularizer=l2(0.05)), # L2 regularization (λ=0.05)

Dropout(0.6), # Further regularize dense layer

Dense(1, activation='sigmoid')

])

**Loss & Optimizer:**

* **Loss:** BinaryFocalCrossentropy(gamma=3.0)
  + Prioritizes hard-to-classify pneumonia cases.
* **Optimizer:** AdamW(learning\_rate=1e-5, clipnorm=1.0)
  + Combines adaptive learning rates with gradient clipping for stability.
* **Metrics:** AUC (primary), Accuracy.

**Training Configuration**

| **Parameter** | **Value** |
| --- | --- |
| Batch Size | 32 |
| Epochs | 30 (early stopping) |
| Class Weights | {0: 1, 1: 8} |
| Augmentation | Flips, Rotation, Random Crops |
| Steps per Epoch | len(train\_df) // 32 |

**Results (Post-Fixes)**

**Performance Metrics:**

| **Metric** | **Train** | **Validation** |
| --- | --- | --- |
| AUC | 0.88 | 0.87 |
| Loss | 0.22 | 0.17 |

**Classification Report:**

| **Class** | **Precision** | **Recall** | **F1-Score** | **Support** |
| --- | --- | --- | --- | --- |
| **Normal** | 0.94 | 0.75 | 0.83 | 4,135 |
| **Pneumonia** | 0.49 | 0.84 | 0.62 | 1,282 |
| **Accuracy** | - | - | **0.77** | 5,337 |

**Key Improvements:**

1. **Stable Training:** Eliminated erratic accuracy jumps (e.g., 0.33 → 1.0).
2. **Validation AUC:** Improved from **0.60 → 0.87**.
3. **Overfitting Mitigation:** Train/validation loss gap reduced to **0.05**.

**Key Fixes in V32**

1. **Data Pipeline Stability:**
   * Explicit shape enforcement: img.set\_shape((1024, 1024)) post-DICOM decoding.
   * Dedicated augmentation for training (flips/rotation) vs. none for validation.
2. **Regularization Enhancements:**
   * **SpatialDropout2D(0.7):** Reduces feature map co-adaptation.
   * **Dropout(0.6) + L2(0.05):** Penalizes large weights in dense layers.
3. **Class Imbalance Mitigation:**
   * **Focal Loss (γ=3.0):** Downweights well-classified samples.
   * **Class Weights (1:8):** Prioritizes minority class (Pneumonia).

**Remaining Challenges**

1. **Low Pneumonia Precision:** High false positive rate (49% precision).
2. **AUC Plateau:** Validation AUC stabilized at **0.87** (potential model capacity limits).

**Next Steps**

1. **Hyperparameter Tuning:**
   * Use **Keras Tuner** to optimize dropout rates, L2 penalty, and focal loss gamma.
2. **Test-Time Augmentation (TTA):**
   * Average predictions over augmented validation samples (e.g., multi-crop, brightness shifts).
3. **Model Interpretability:**
   * **Grad-CAM Analysis:** Visualize attention maps to debug false positives.
4. **Architecture Experimentation:**
   * Test EfficientNetV2 or Vision Transformers (ViTs) for improved feature extraction.

**Conclusion:** Version 32 achieves stable training and improved generalization, but precision for pneumonia cases remains a critical focus area. Future iterations will prioritize reducing false positives and breaking the AUC plateau.

While 77% accuracy shows progress, focus on **improving Pneumonia precision** (reduce false positives) and use **AUC (0.87)** as the primary metric, as it evaluates performance across all classification thresholds and is more robust to class imbalance.