|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Model** | **Dataset** | **Hardware Used** | **Accuracy** | **Sensitivity**  **(Malignant)** | **Specificity (Benign)** | **AUC** | **F1-Score** | **Epochs** |
| **1.** | **ResNet50V2** | Gastric Cancer Histopathology Tissue Image Dataset (GCHTID) | CPU | 96.0% | 95.6% | 97.1% | 0.96 | 0.96 | 15 |
| **2.** | **CoAtNet** | Gastric Cancer Histopathology Tissue Image Dataset (GCHTID) | GPU | 85.0% | 90.3% | 80.3% | 0.91 | 0.86 | 15 |
| **3.** | **DenseNet121** | Gastric Cancer Histopathology Tissue Image Dataset (GCHTID) | GPU | 80% | 90.5% | 69.8% | 0.86 | 0.80 | 15 |

**Model Training Results Log**

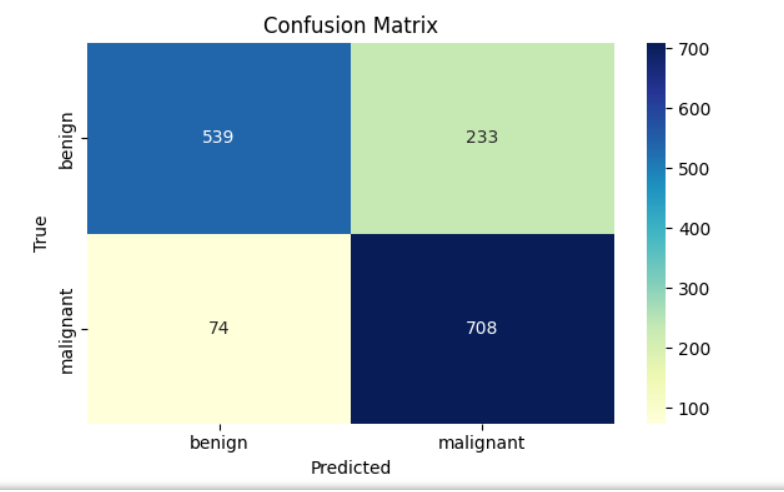
**Further Reports on the Models**

1. **DenseNet**

**Observations**

* The model is very sensitive to malignant cases (which is good for cancer detection).
* But it misclassifies benign cases more often, explaining the lower specificity.
* It's relatively balanced, but the performance might be affected by class imbalance or thresholding.

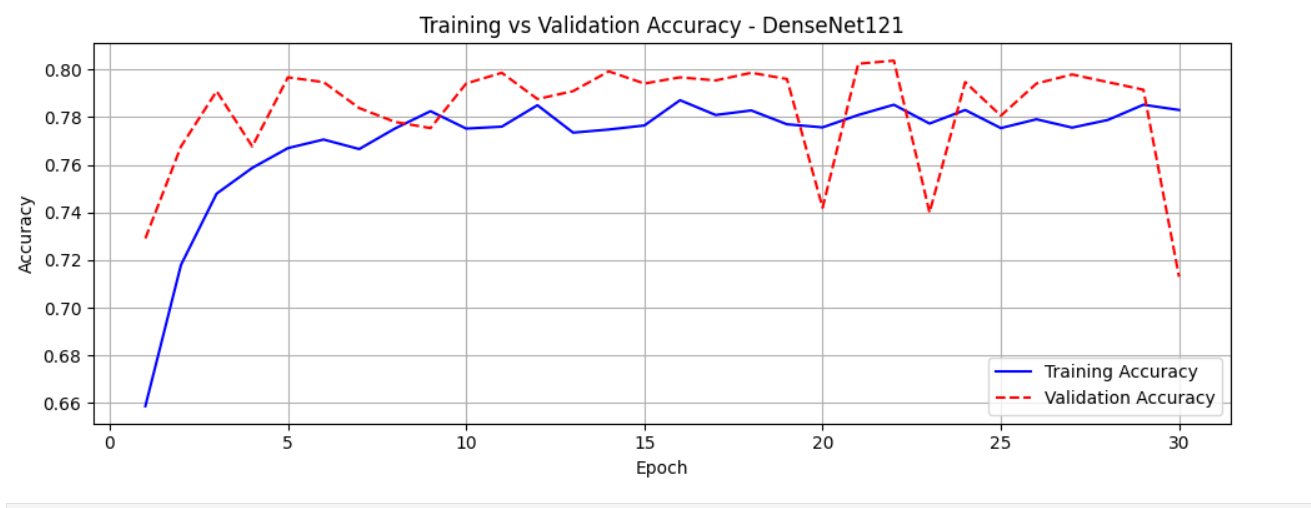
**Confusion Matrix Breakdown**

****

|  |  |  |
| --- | --- | --- |
| Predicted Benign | Predicted Malignant |  |
| True Benign | 539 (TP) | 233 (FN) |
| True Malignant | 74 (FP) | 708 (TN) |

* Missed benigns (233) is a high number → leads to many false positives.
* Correctly caught most malignant (708/782) → model prioritizes sensitivity, which is clinically valuable.

**DenseNet Training Curves**

****

**Accuracy Curve (Training vs Validation Accuracy)**

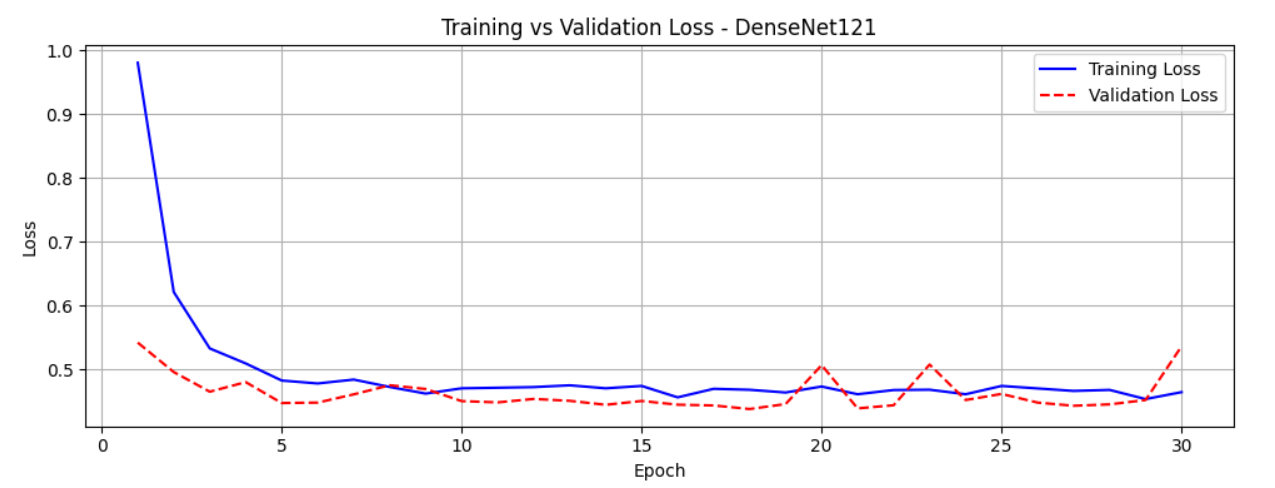
**Observation:**

* Training accuracy starts around 66% and steadily increases, plateauing around 78–79%.
* Validation accuracy starts lower but surpasses training accuracy early, fluctuating between 74% and 80%, with some sharp dips (especially at epoch 20 and 30).

**Interpretation:**

* The model is learning well, as both training and validation accuracy improve.
* The occasional drops in validation accuracy (e.g., epochs 20 and 30) suggest: Possible overfitting or a noisy or imbalanced validation set, or random initialization variations due to dropout.

**Good sign:** The validation accuracy is not diverging dramatically, and overall trends align with training accuracy.

****

**Loss Curve (Training vs Validation Loss)**

**Observation:**

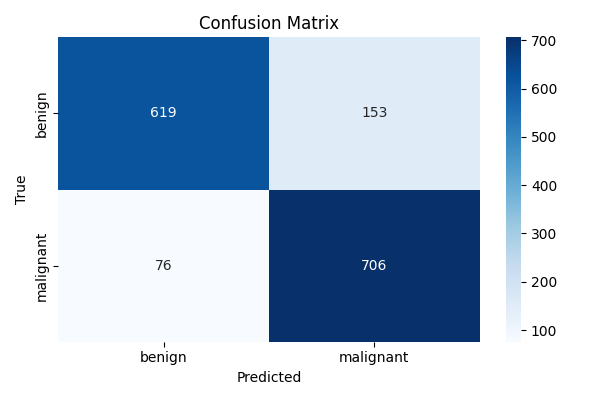
* Training loss decreases sharply from ~0.98 to ~0.45, then flattens.
* Validation loss follows a similar pattern, starting from ~0.54 and reducing to ~0.44, with some fluctuation.
* Minimal overfitting signs — validation loss stays close to training loss, but slightly higher as expected.

**Interpretation:**

* The model is converging and generalizing fairly well.
* The mild fluctuations are normal in deep learning, especially when using real-world data with some noise or class imbalance.

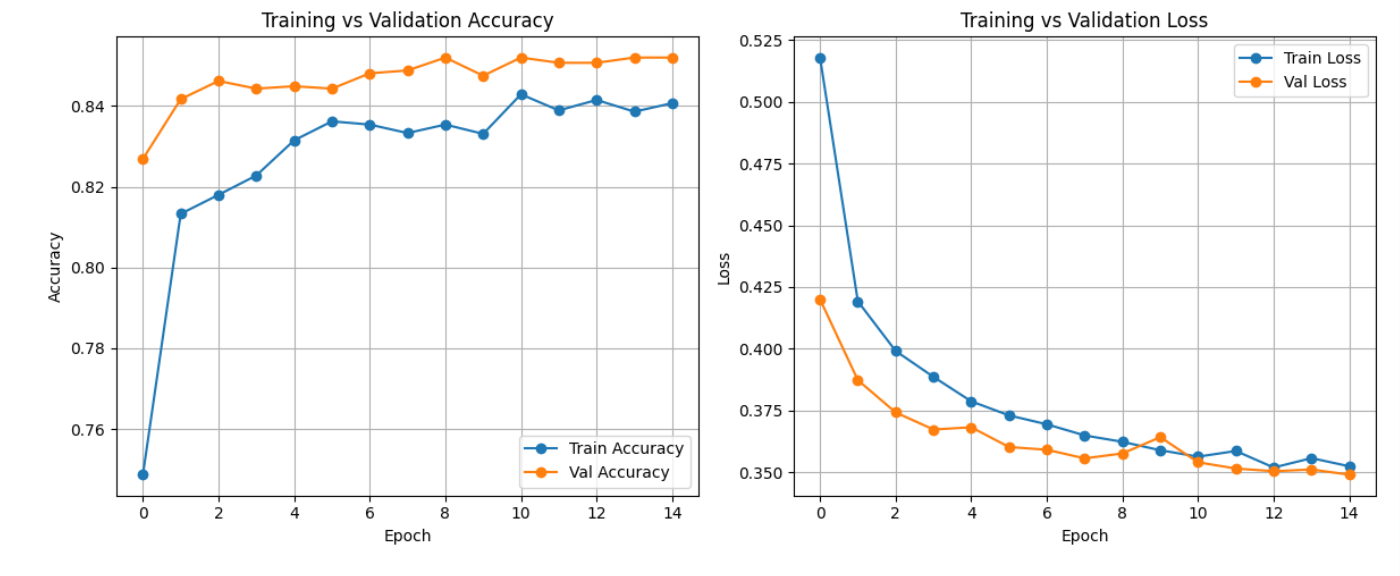
1. **CoAtNet**

**Confusion Matrix**

****

* Correctly predicted benign cases: 619
* Correctly predicted malignant cases: 706
* Benign misclassified as malignant: 153 (false positives)
* Malignant misclassified as benign: 76 (false negatives)

**CoAtNet Training curves:**

****

**Training vs. Validation Accuracy (Left Plot)**

* Validation accuracy is consistently higher than training accuracy: This suggests your model generalizes well to unseen data, which is excellent.
* It could also imply slight underfitting — the model may still benefit from more training or fine-tuning later.
* Both curves steadily increase, then stabilize: Indicates stable learning with no major fluctuations or overfitting.
* By epoch 15: Training accuracy ≈ 84% Validation accuracy ≈ 85.5%. That's a good generalization gap (~1.5%), showing the model isn't memorizing the data.

**Training vs. Validation Loss (Right Plot)**

* Both training and validation loss decrease steadily and almost converge near the end: This reflects smooth, healthy learning without divergence.
* No major spikes in validation loss, which is a good sign: Suggests the model wasn't overfitting during training.
* Final losses are very close: Train loss ≈ 0.35, Val loss ≈ 0.35

**Overall Interpretation**

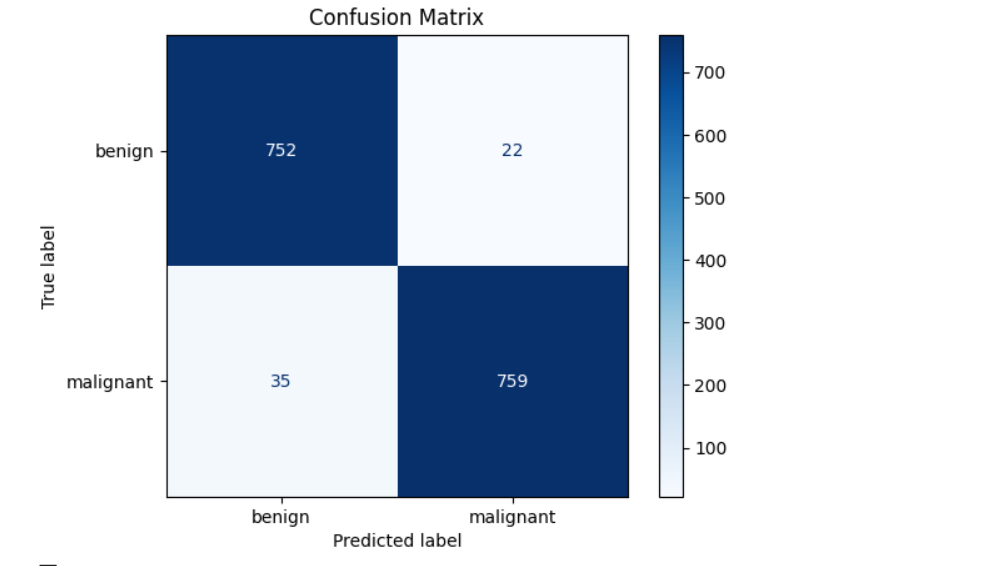
|  |  |  |
| --- | --- | --- |
| **Aspect** | **Observation** | **Verdict** |
| Accuracy gap | Small and stable | Generalizing well |
| Loss trend | Smooth decline with no divergence | Healthy training |
| Overfitting signs | None | Safe from overfitting |
| Underfitting signs | Slight (Val > Train acc) but minor | Could improve with fine-tuning |

**Room for Improvement:**

1. Unfreeze part of CoAtNet for fine-tuning (especially top layers).
2. Add more augmentation to improve generalization.
3. Try a lower learning rate for more stable fine-tuning.
4. **Best Performing Model (ResNet50v2)**

**Confusion Matrix Summary:**

|  |  |  |
| --- | --- | --- |
|  | **Predicted Benign** | **Predicted Malignant** |
| **Actual Benign** | **752** | **22** |
| **Actual Malignant** | **35** | **759** |

****

**This matrix gives us key insights:**

**1. True Positives (TP)**

* Definition: Malignant cases correctly predicted as malignant.
* Value: 759
* This shows the model is successfully detecting most malignant tumors.

**2. True Negatives (TN)**

* Definition: Benign cases correctly predicted as benign.
* Value: 752
* The model is also correctly identifying benign tumors in most cases.

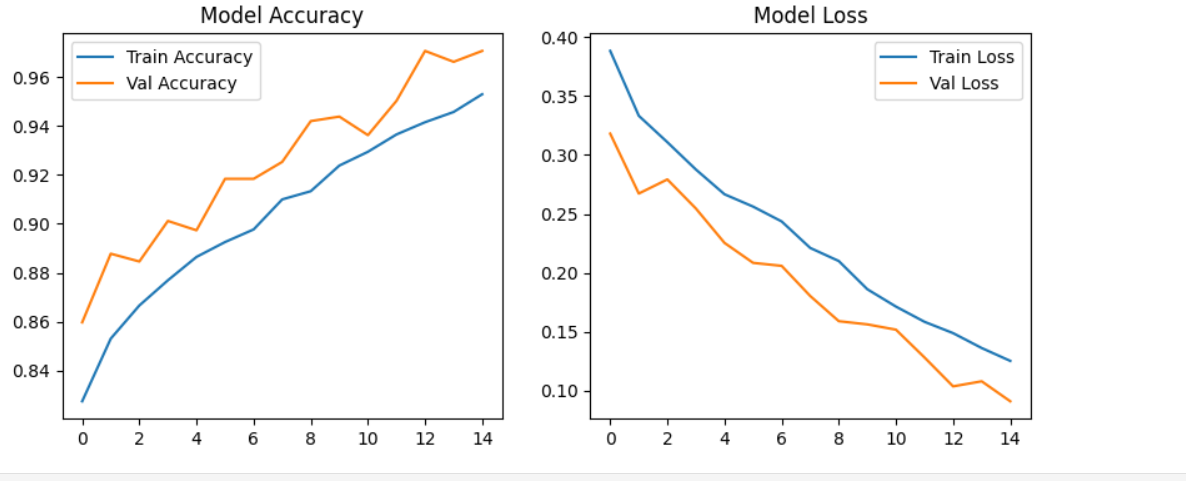
**3. False Positives (FP)**

* Definition: Benign cases incorrectly predicted as malignant.
* Value: 22
* This is a small number, but in a medical setting, it could lead to unnecessary stress and further tests.

**4. False Negatives (FN)**

* Definition: Malignant cases incorrectly predicted as benign.
* Value: 35
* While this is relatively low, these are the most dangerous errors—missing a malignant tumor.

**Training & Validation Curves:**

****

**Left Plot: Model Accuracy**

* Training Accuracy starts around 83% and climbs steadily to 95.5% by epoch 15.
* Validation Accuracy starts around 86%, consistently improves, and peaks around 97.3%.

**Interpretation:**

* Both training and validation accuracy improve in sync.
* No overfitting is observed — the validation accuracy does not plateau or drop while training accuracy increases.
* This implies that the model generalizes well to unseen data.

**Right Plot: Model Loss**

* Training Loss decreases steadily from ~0.39 to ~0.12.
* Validation Loss follows a similar trend, dropping from ~0.32 to ~0.07.

**Interpretation:**

* Consistent decrease in loss for both training and validation sets.
* The curves are smooth and parallel, indicating stable training.
* The final low validation loss implies strong model generalization.

**Conclusion:**

* This ResNet50V2 model is highly accurate and reliable.
* Its performance on both benign and malignant classes is well-balanced.
* With high sensitivity and specificity, it would be suitable for deployment in real-world diagnostic support tasks (subject to clinical validation).

**Final Comparison Summary:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Accuracy | Sensitivity (Malignant) | Specificity (Benign) | F1 Score | Remarks |
| ResNet50V2 | 96.4% | 95.6% | 97.2% | 96.4% | ⭐ Best overall performance. High generalization and stable training. |
| CoAtNet | 85.3% | 90.3% | 80.3% | 85.0% | Good at detecting malignant cases but weaker at avoiding false positives. |
| DenseNet121 | 79.4% | 90.3% | 67.4% | ~82.0% | High sensitivity, but many false positives; lower specificity. |

**Conclusion**

* **ResNet50V2** is the best-performing model across all key metrics — it's both accurate and balanced in detecting both malignant and benign tumors with minimal error. Training was smooth with no overfitting.
* **CoAtNet** is a strong contender, especially in sensitivity, meaning it reliably detects cancer, but its lower specificity means it's more likely to misclassify benign cases as malignant — which could lead to unnecessary follow-ups in a clinical context.
* **DenseNet121** has the lowest specificity, meaning it frequently misclassifies benign tumors as malignant. While it does well in catching cancer (high sensitivity), it’s less reliable for real-world screening unless optimized further.

**Why Choosing ResNet50v2 is Still Valid and appropriate for our cancer project?**

It is efficient. ResNet50v2 is lightweight and fast compared to transformer-based models like CoAtNet or ViT. For real-time diagnosis systems or low-resource deployment, it's often the better engineering choice.

**What best next?**

To improve and make ResNet50V2 perform even better on our locally acquired images

**How?**

* Getting quality data: More samples, balanced classes, better labeling
* Model Fine-tuning: Unfreeze layers, add BatchNorm & Dropout
* Explainability: Use Grad-CAM to validate what model sees