Project Report: Advanced Chest X-Ray Analysis Using Hybrid Deep Learning Models

1. Introduction & Project Objective

Chest X-rays are a first-line diagnostic tool for detecting a wide range of thoracic diseases. Often, patients present with multiple conditions simultaneously, making multi-label classification a critical real-world problem. The objective of this project was to design, implement, and compare different AI models that can automatically and accurately identify the presence of the top five most common thoracic conditions:

- **Atelectasis**
- **Effusion**
- **Infiltration**
- **Nodule**
- **No Finding** (a significant and challenging class)

The goal was to explore whether combining handcrafted image features (OpenCV) with deep learning features could create a more robust and interpretable model.

2. Dataset Overview

- **Source:** A large public chest X-ray dataset.
- **Initial Metadata:** 111,010 entries.
- **Curated Dataset: 3,710 images** after filtering for available files.
- **Task:** Multi-label classification (each image can have more than one label).
- **Class Distribution:** Highly imbalanced, with "No Finding" being the dominant class. This imbalance is a central challenge addressed in the modelling process.

3. Methodology

3.1 Data Preprocessing & Augmentation

To ensure model robustness and generalization, a comprehensive preprocessing pipeline was implemented using OpenCV:

- **Resizing:** All images standardized to 224x224 pixels.
- **Colour Conversion:** Converted from BGR to RGB format.
- **Advanced Augmentation:** Applied in real-time to artificially expand the dataset and improve model generalization. Techniques included:
    - Gaussian Noise Injection

- Random Rotation (±15 degrees)
- Perspective Transformation
- Histogram Equalization (in YUV color space)

3.2 Feature Engineering

A two-stream feature extraction approach was investigated:

1. **Deep Learning Features:** Utilized pre-trained Convolutional Neural Networks (CNNs) like ResNet50, DenseNet121, and EfficientNetB0 as powerful feature extractors.
2. **Handcrafted Features:** Extracted **ORB (Oriented FAST and Rotated BRIEF)** features using OpenCV to capture traditional image characteristics like texture and keypoints.

3.3 Model Architectures

Three distinct model architectures were developed:

Model	Description	Key Features
Model 1: Baseline Hybrid	A proof-of-concept hybrid model.	ResNet50 (frozen) + Custom Classifier Layers + ORB features fused and classified with a Random Forest.
Model 2: Production Hybrid	An enhanced, more robust version of Model 1.	Incorporated data validation checks, used a larger sample (1,000 images), and improved data splitting logic.
Model 3: Advanced DL Exploration	A systematic exploration of state-of-the-art architectures and training strategies.	Tested ResNet50, DenseNet121, and EfficientNetB0 with various optimizers (Adam, SGD, RMSprop), learning rates, and layer freezing/fine-tuning strategies.

3.4 Training Configuration

- **Optimizer:** Adam, SGD, RMSprop (experimented)
- **Loss Function:** Binary Cross-Entropy (suitable for multi-label tasks)
- **Metrics:** Accuracy, Precision, Recall, F1-Score (macro)
- **Validation:** Early Stopping and Learning Rate Reduction on plateau were used to prevent overfitting.

4. Results & Comparative Analysis

The following table provides a comprehensive comparison of the final performance of all three models on the held-out test set.

Table: Model Performance Comparison on Test Set

Model	Architecture	Test Accuracy	Test Precision	Test Recall	**Test F1-Score**	Parameters
Model 1	ResNet50 + ORB (Hybrid)	66.67 %	16.00 %	16.00%	16.00%	~24.8M (DL) + RF
Model 2	ResNet50 + ORB (Improved Hybrid)	45.00 %	12.29 %	15.09%	13.54%	~24.8M (DL) + RF
Model 3 - Config A	ResNet50 (Frozen)	**71.00 %**	**61.89 %**	53.93%	14.35%	~24.8M (1.2M Trainable)
Model 3 - Config B	ResNet50 (Partial Fine-Tune)	**71.00 %**	60.21 %	**61.07%**	14.84%	~24.8M (More Trainable)
Model 3 - Config C	DenseNet121 (Frozen)	69.33 %	60.69 %	56.79%	14.52%	~7.1M (Trainable)
Model 3 - Config D	**EfficientNetB0 (Unfrozen)**	68.33 %	52.96 %	**63.93%**	**18.92%**	~5.3M (All Trainable)

5. Discussion & Insights

1. **Deep Learning Outperformed Hybrid Approaches:** Contrary to the initial hypothesis, the hybrid models (Model 1 & 2) underperformed. The ORB features, while rich in texture information, were not discriminative enough for this specific medical domain task and likely introduced noise when combined with the high-level features from the CNN.

2. **The Power of Fine-Tuning:** Among the pure deep learning models, **EfficientNetB0**, which was fully unfrozen and fine-tuned, achieved the highest **F1-Score (18.9%)**. This is a more important metric than accuracy for imbalanced datasets, as it balances precision and recall. This shows that allowing the entire network to adapt to the specific features of chest X-rays is highly beneficial.

3. **Addressing Class Imbalance:** The relatively lower F1-scores across all models highlight the significant challenge of class imbalance. The model's performance was heavily influenced by the over-represented "No Finding" class.

4. **Trade-off between Accuracy and F1:** Configurations A and B (ResNet50) achieved the highest accuracy, but EfficientNetB0 achieved a better F1-score by being more effective at identifying the positive disease classes (higher recall).

6. Conclusion

This project successfully demonstrates a full pipeline for medical image analysis, from data preprocessing to model evaluation. The key conclusion is that for this complex multi-label task, **sophisticated deep learning architectures fine-tuned end-to-end (like EfficientNetB0) yield better results than hybrid models combining deep and handcrafted features.**

The experiments provide a clear roadmap: future efforts should prioritize:

- **Advanced techniques to handle class imbalance** (e.g., weighted loss functions, oversampling).
- **Leveraging larger datasets** to improve model generalization.
- **Exploring attention mechanisms and vision transformers** to better capture global context in images.
- **Incorporating explainable AI (XAI) techniques** like Grad-CAM (which was implemented in the code) to build trust and provide diagnostic insights.