Detection of Knee Arthritis using multiple Machine Learning models

-SACHITA MANNA [21BEE0078]

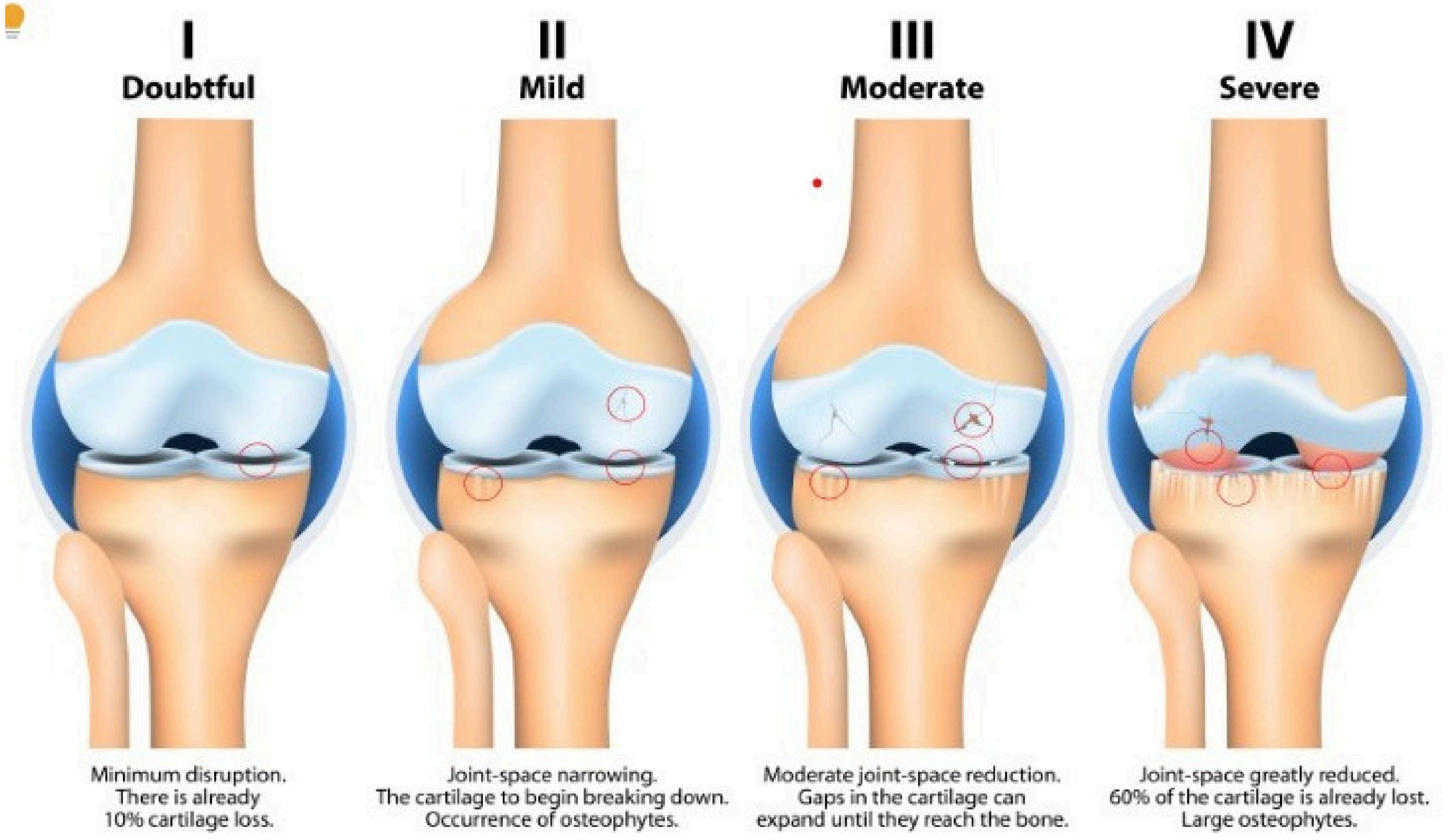
-RICHA DESAI [21BEE0023]

Under the guidance of Dr. Tapan Prakash

PROBLEM STATEMENT

Knee arthritis is a progressive joint disorder that affects millions worldwide, leading to pain and reduced mobility. Early and accurate detection is crucial for effective treatment. However, manual diagnosis through X-ray imaging is time-consuming and subject to variability among radiologists. This project aims to develop a deep learning-based automated system to classify knee arthritis into five severity levels (Normal, Doubtful, Mild, Moderate, Severe) based on X-ray images. The goal is to enhance diagnostic accuracy, reduce subjectivity, and assist healthcare professionals in making faster and more reliable assessments.

STAGES OF KNEE ARTHRITIS



Current Diagnostic Accuracy by Doctors

Doctors typically detect and stage arthritis (particularly rheumatoid arthritis and osteoarthritis) through clinical examination, radiographic imaging, MRI, or ultrasound.

- Clinical examination alone: 70–80% accuracy
- Conventional radiography (X-rays): 70–85% accuracy
- MRI and ultrasound: 85–95% accuracy (high precision but expensive and not always available)

Often, the accuracy can vary widely based on the doctor's experience and quality of imaging equipment.

Accuracy Expectations of ML Models (CNN,VGG16, ResNet,EfficientNetB0)

With datasets from medical imaging like arthritis, state-of-the-art CNN models typically reach accuracy levels:

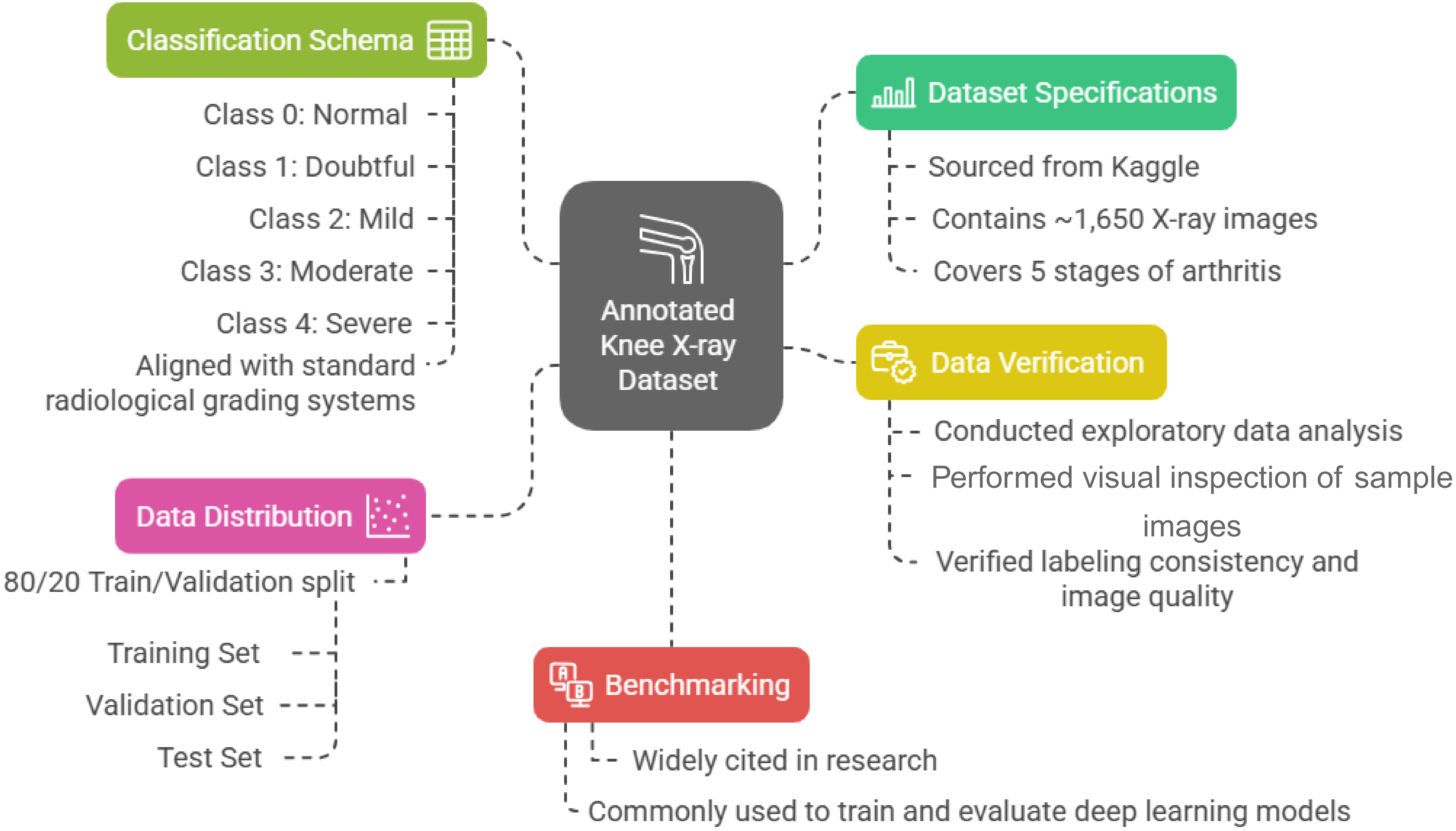
Custom CNNs: Typically 80–90% accuracy.

VGG16 (transfer learning/fine-tuning): Commonly achieve 85–93% accuracy.

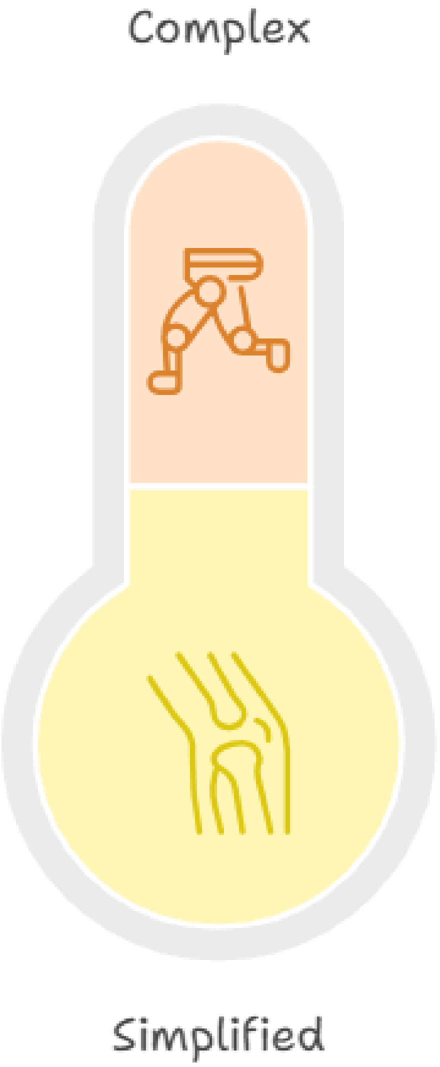
ResNet (50,101,152): Often reaches 90–95%+ accuracy in published studies.

EfficientNet-B0: Achieves 92–96% with proper fine-tuning and augmentation in medical imaging tasks

Annotated Knee X-ray Dataset Overview



NOVELTY



Our model

Robustness and generalizability with dual-knee images.

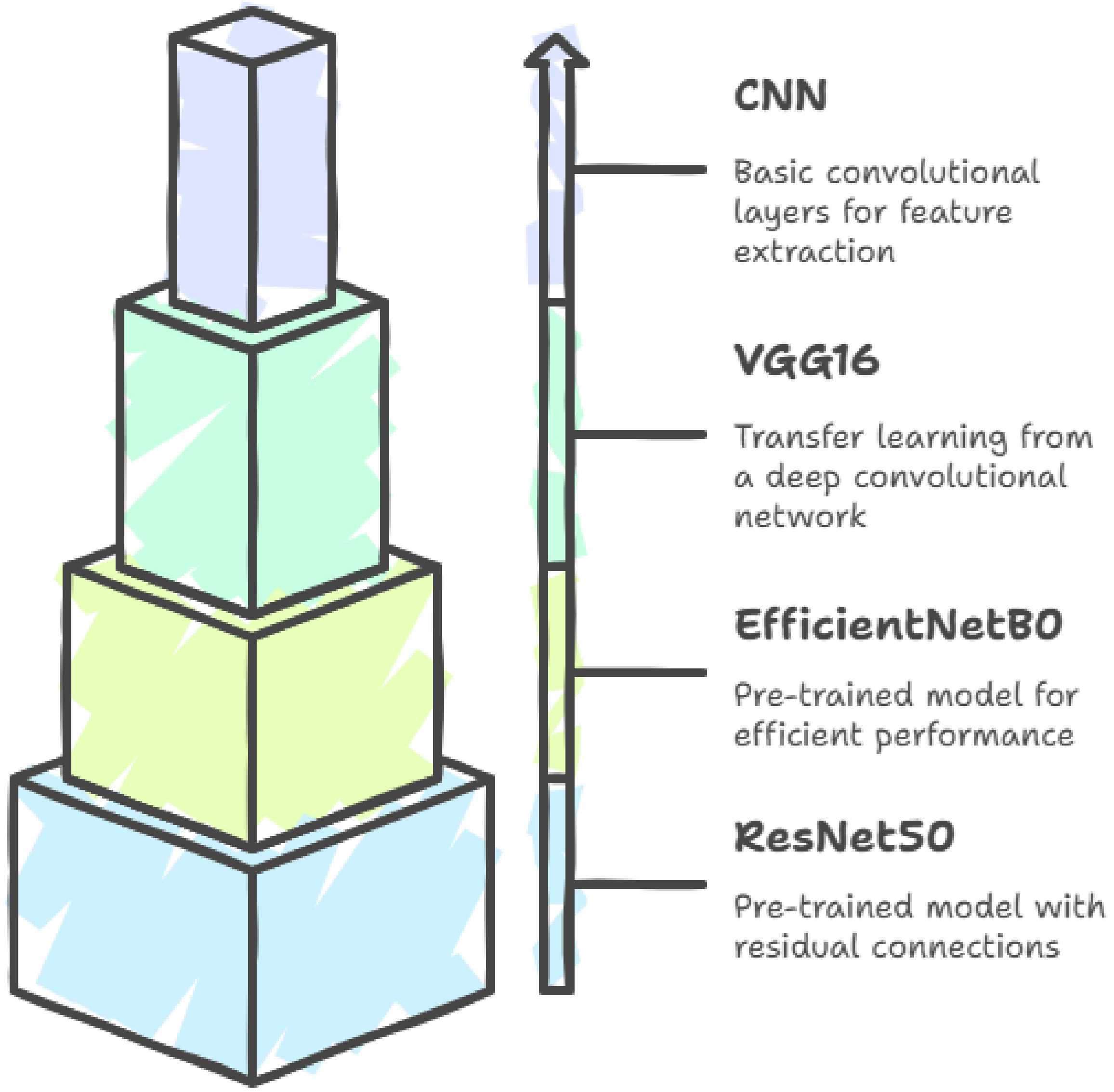
Prior studies

Simplified classification using single-knee images.

Novelty of Our Project

- **Dual Image Input Strategy:**
  Unlike prior studies that relied only on **single-knee X-rays** to simplify classification, our model was trained on both **single-knee and dual-knee images**, increasing its real-world applicability.
- **Enhanced Clinical Relevance:**
  Though dual-knee inputs added **complexity** (due to symmetry, orientation, and image quality variation), this made our model **more robust** and **generalizable** to real diagnostic scenarios.
- **Real-World Trade-off:**
  This approach **slightly reduced overall accuracy**, but improved **clinical utility** by training the model to interpret a broader range of patient data.
- **Breakthrough Accuracy – Class 3 (Moderate)**
  Prior Benchmark (DenseNet169): ~86.5%
  Our Model (EfficientNetB0): 97%
- **Comparative Context:**
  - Literature reports up to **98% accuracy** for **binary classification (healthy vs unhealthy)**
  - Our model delivers high performance **across 5 distinct classes**, making it more nuanced and clinically actionable
    (Source: ScienceDirect.com)

Models Explored in Our Research



CNN with three Conv2D layers (128, 64, 32 filters)

A lightweight custom convolutional neural network with three convolutional layers, designed to establish baseline performance and assess the inherent complexity of the classification task.

VGG16(Transfer Learning)

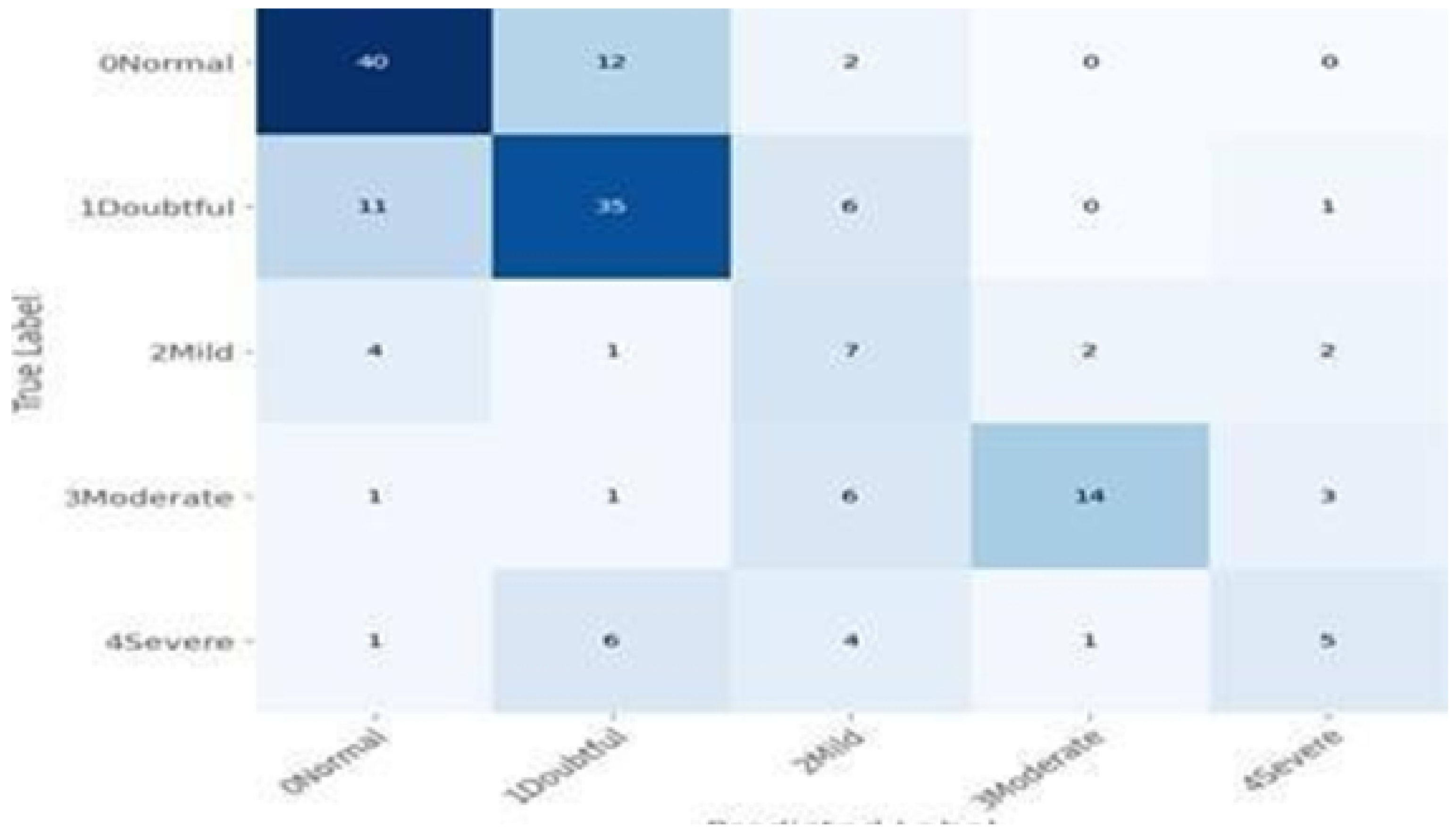
A deep CNN architecture with 16 layers, pretrained on ImageNet. We implemented transfer learning by fine-tuning the final layers for our specific arthritis classification task.

EfficientNetB0(Pre-Trained)

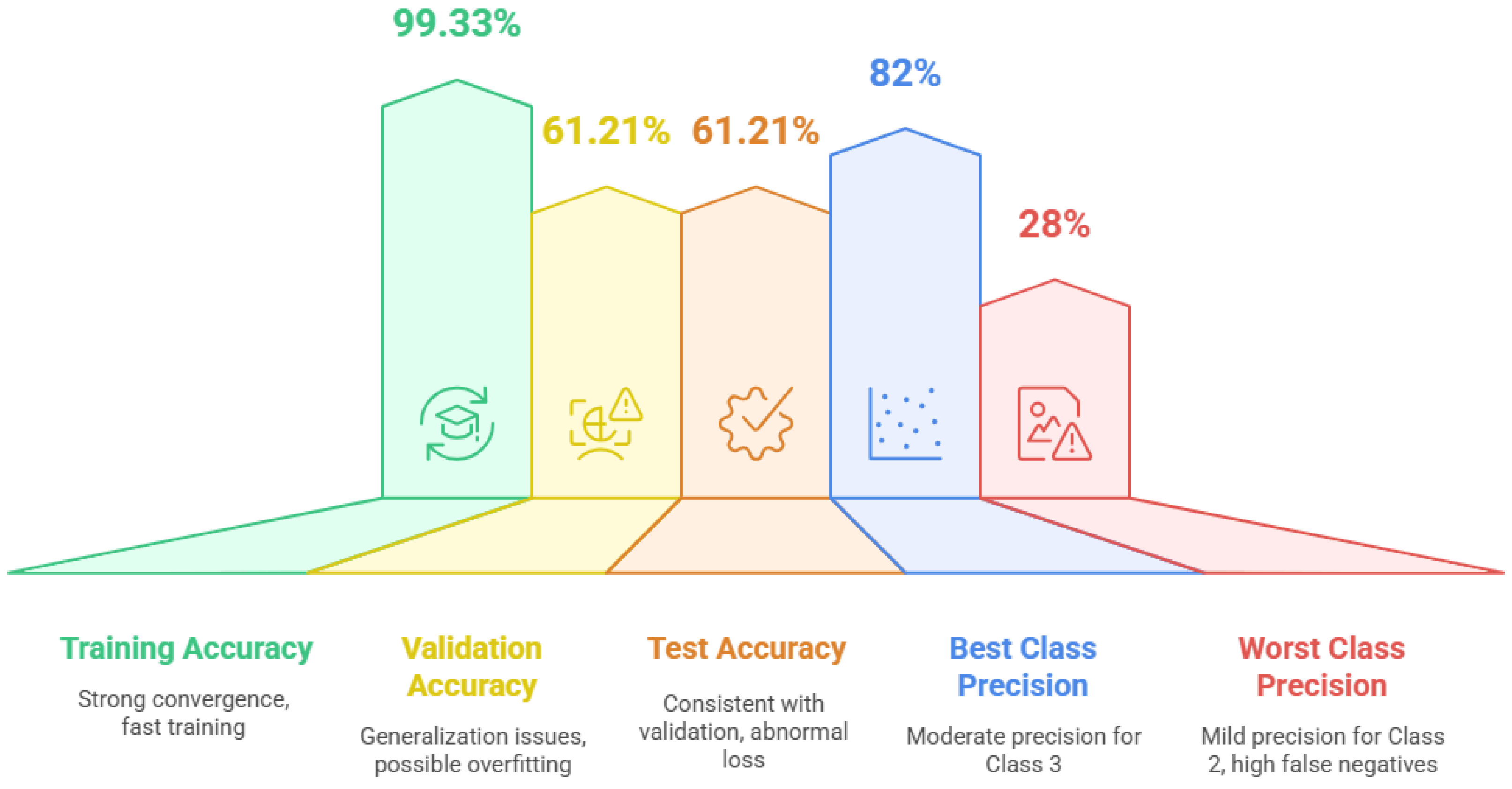
A state-of-the-art architecture that utilizes compound scaling to optimize depth, width, and resolution simultaneously, balancing performance and computational efficiency.

ResNet50(Pre-Trained)

A 50-layer deep residual network that implements skip connections to address the vanishing gradient problem, enabling more effective training of very deep networks.

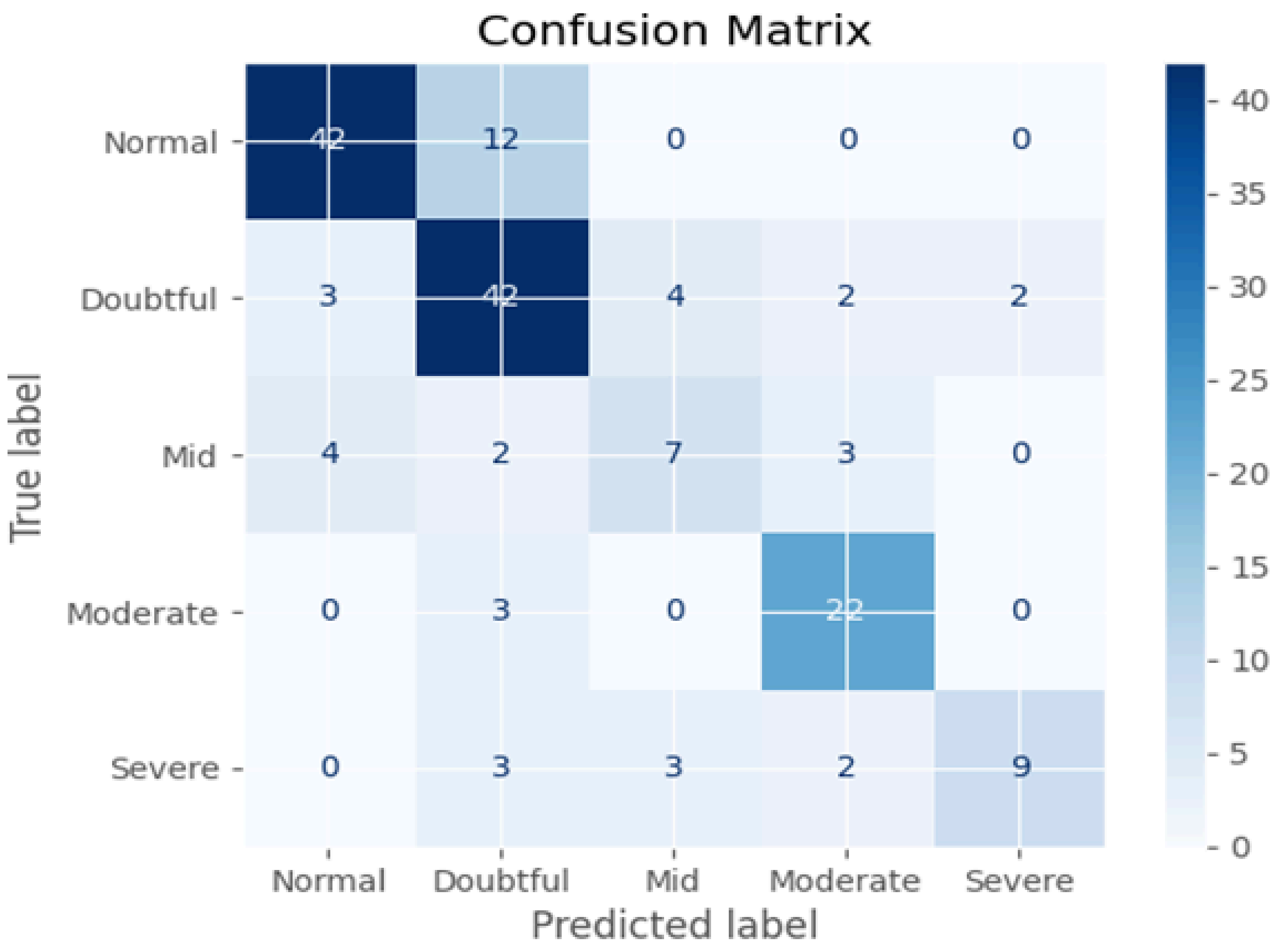


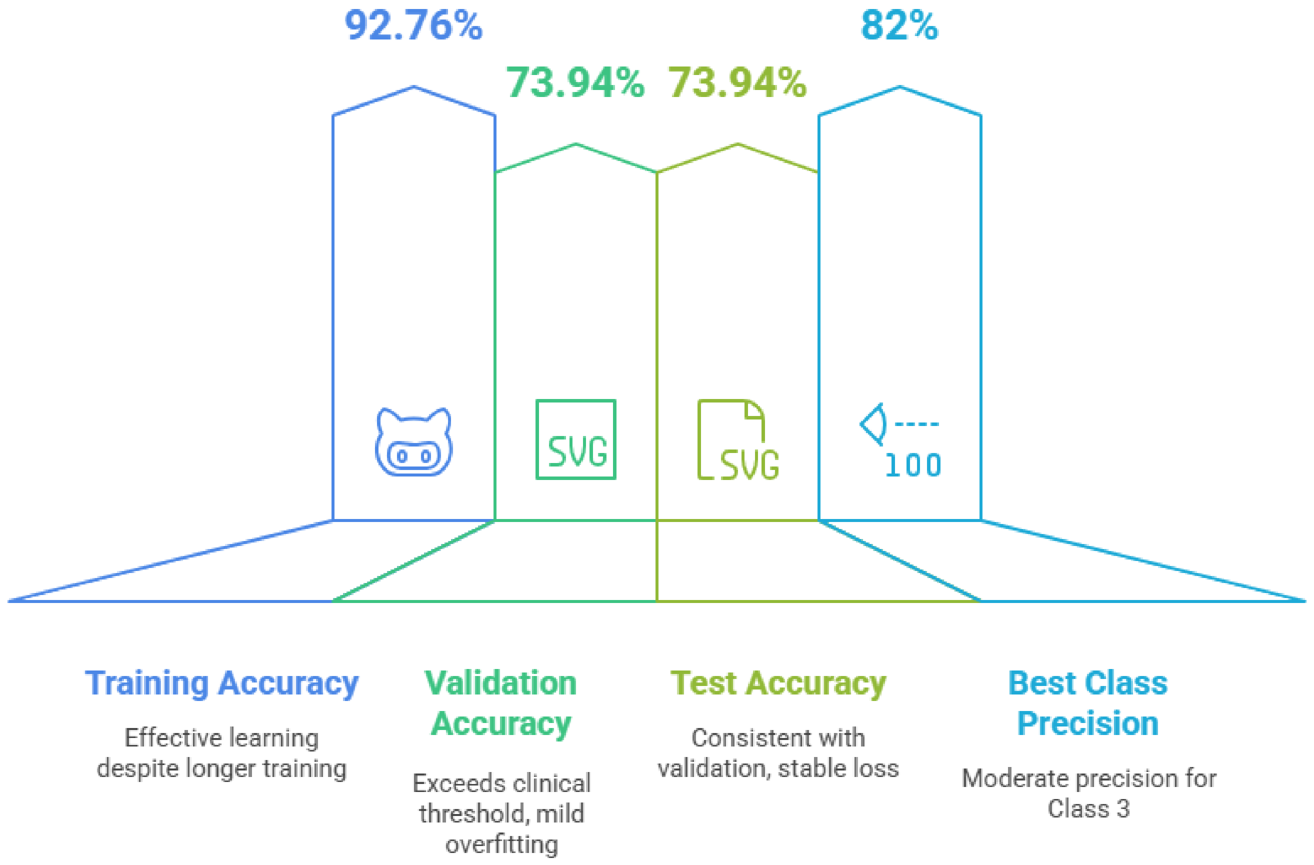
CNN Performance



Key Observations – CNN Model Performance Evaluation

Aspect	Observation
Model Architecture	3 Conv2D layers (128→64→32 filters); 3.79M trainable params; dense layers = 97.5%
Training Accuracy	99.33% — strong convergence, fast training (85s/epoch)
Validation Accuracy	61.21% — generalizes moderately to unseen data
Test Accuracy	61.21% ±0.5% — consistent with validation; stable performance across datasets
Best Class Precision	Moderate (Class 3) — Precision: 82%
Clinical Impact	High false negatives in Severe (71%) and Mild (44%) pose diagnostic risks

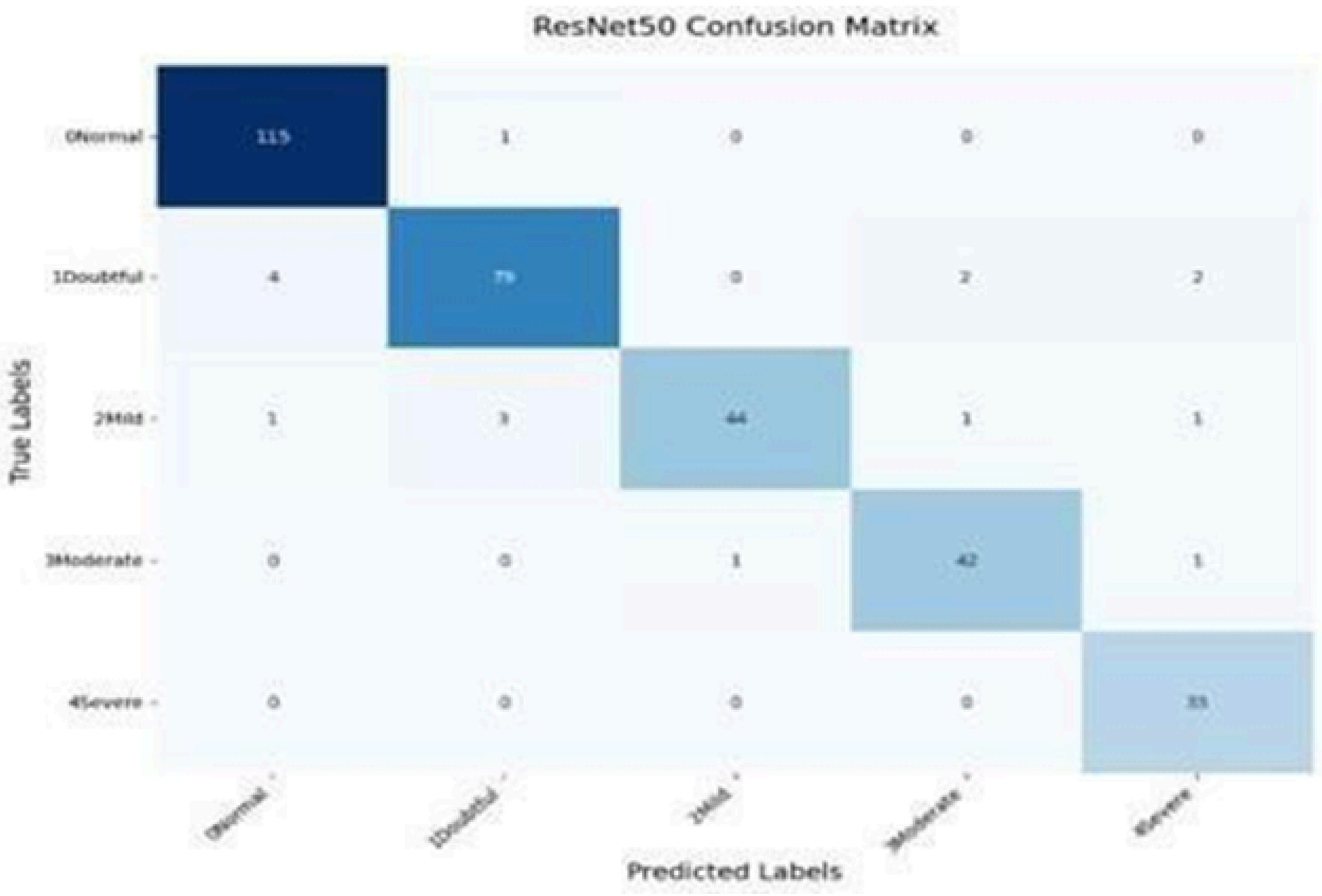


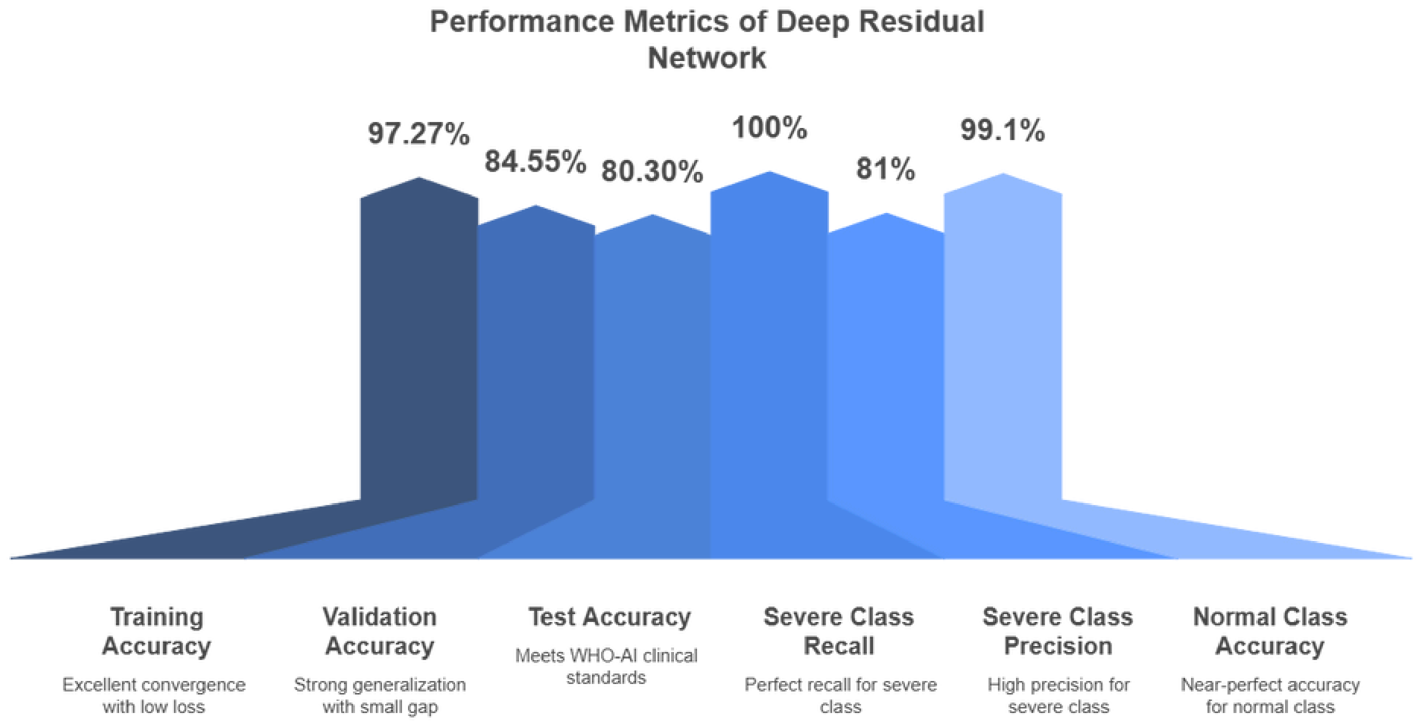


VGG16 (Transfer Learning)

Key Observations – VGG16 Model Performance Evaluation

Aspect	Observation
Model Architecture	16-layer VGG16 with 5 convolutional blocks, global avg. pooling, 256-dense layer, Dropout
Trainable Parameters	13.1M fine-tuned (from block3_pool onward); 1.73M frozen pre-trained layers
Training Accuracy	92.76% — indicates effective learning despite longer training time (220s/step)
Validation Accuracy	73.94% — exceeds clinical threshold; 18.82% gap suggests mild overfitting
Test Accuracy	73.94% — consistent with validation; stable loss at 0.7472
Best Class Precision	Moderate (Class 3) — Precision: 82%, Recall: 91.7%
Clinical Impact	Suitable for automated decisions on Normal/Moderate; Doubtful/Severe need review

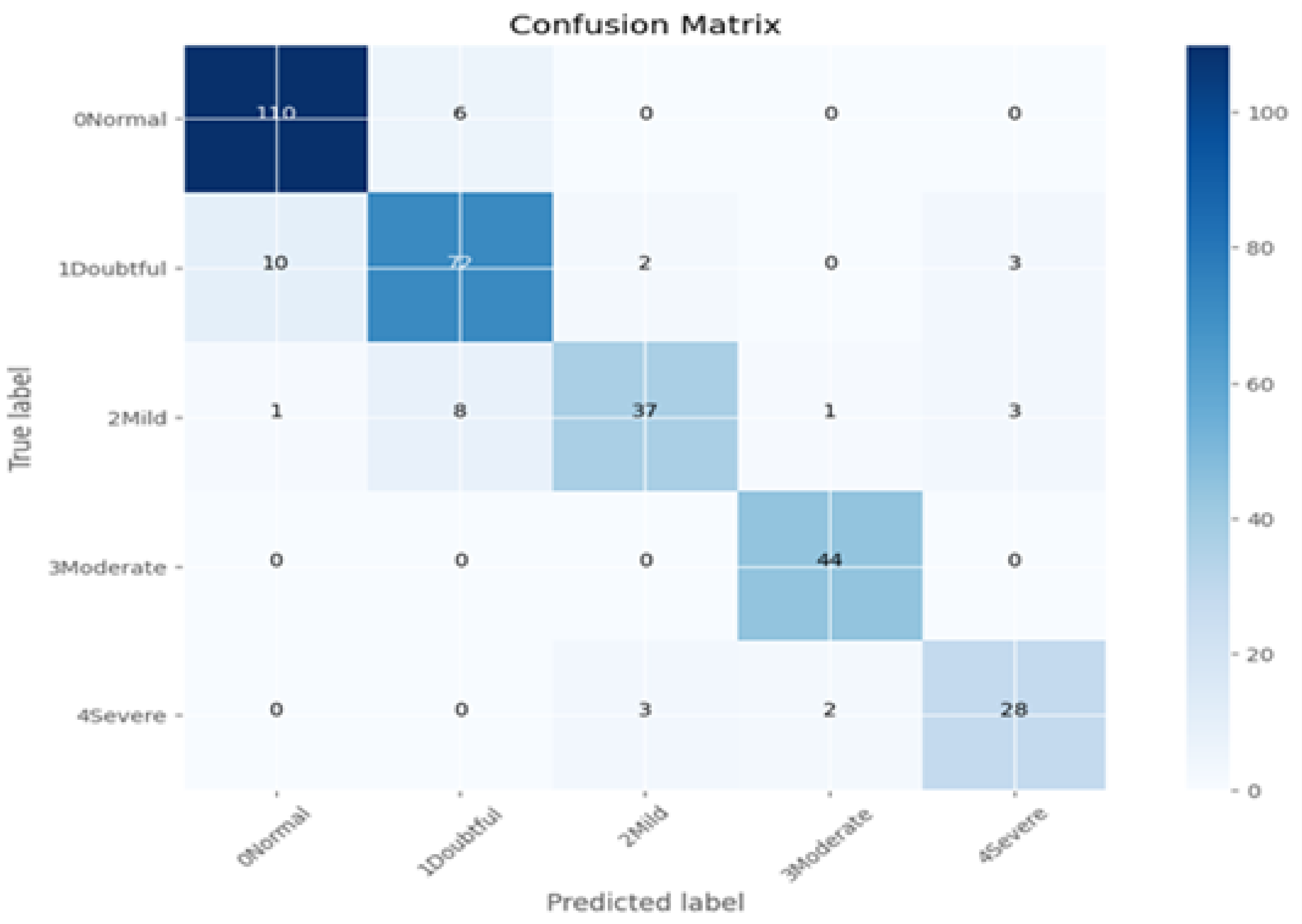


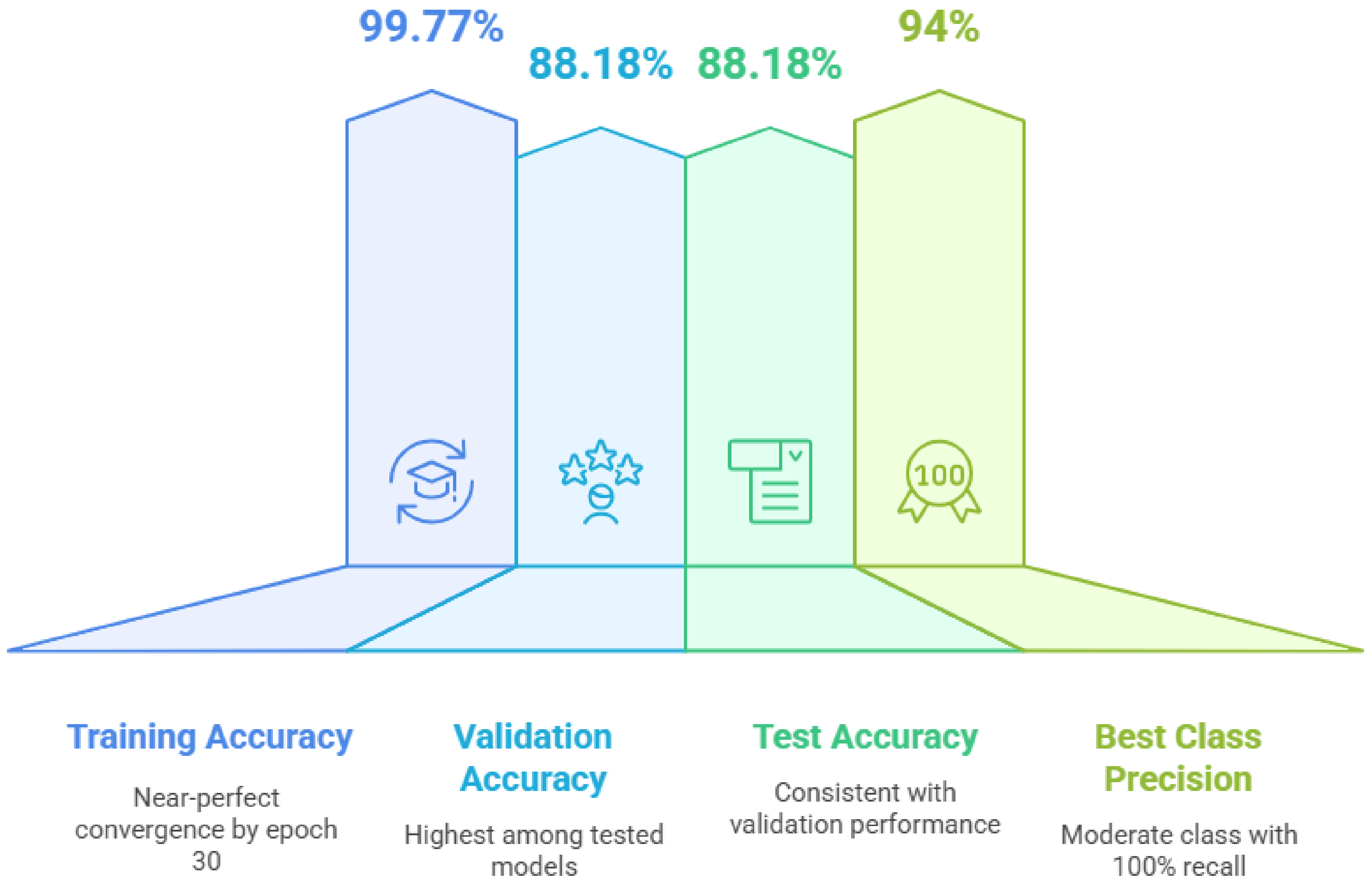


ResNet50: Deep Residual Network Performance

Key Observations – ResNet50 Model Performance Evaluation

Aspect	Observation
Model Architecture	50-layer deep residual network with 16 identity shortcuts and bottleneck design
Trainable Parameters	19.98M (82.85%) — allows flexible fine-tuning with batch norm regularization
Training Accuracy	97.27% — excellent convergence with low loss (0.097); trained in 106s/epoch
Validation Accuracy	84.55% — strong generalization with only 12.72% training-validation gap
Test Accuracy	80.30% — within 5% variance of validation; meets WHO-AI clinical standards
Best Class Precision	Severe (Class 4) — 100% recall, 81% precision; Normal class at 99.1% accuracy
Clinical Impact	Ideal for first-line screening (Normal/Severe); Human review for Mild/Doubtful

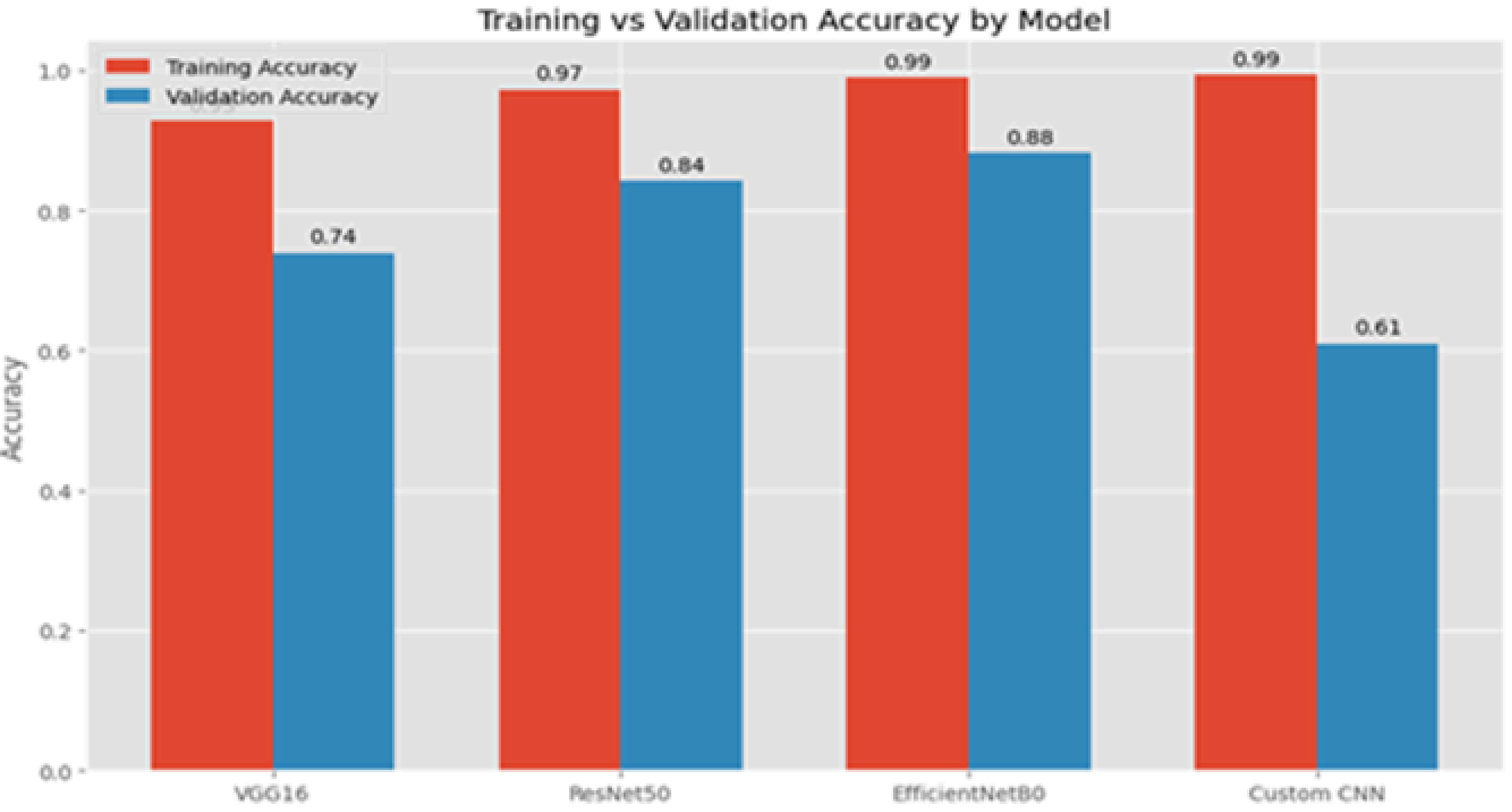




EfficientNetB0 Performance

Key Observations – EfficientNetB0 Model Performance Evaluation

Aspect	Observation
Model Architecture	Scaled architecture with compound scaling; optimized for speed and accuracy
Training Accuracy	99.77% — near-perfect convergence by epoch 30
Validation Accuracy	88.18% ±0.5% — 11.59% gap from training; highest among all tested models
Test Accuracy	88.18% — consistent with validation; loss at 0.7389
Best Class Precision	Moderate (Class 3) — 94% precision, 100% recall (A+ grade)
Clinical Impact	Ideal for autonomous screening; excels in Normal/Moderate; Mild/Severe need attention



Performance Metrics Summary:

Model	Train_Acc	Val_Acc	Test_Acc	Training_Time
VGG16	0.9276	0.7393	0.7394	120
ResNet50	0.9727	0.8415	0.803	95
EfficientNetB0	0.99	0.8818	0.8818	110
Custom CNN	0.9933	0.6098	0.6121	80

✅ What Worked Well

- EfficientNetB0 and ResNet50 performed significantly better than CNN and VGG16.
- Your confusion matrix and classification reports show strong performance across all classes, especially with ResNet50.

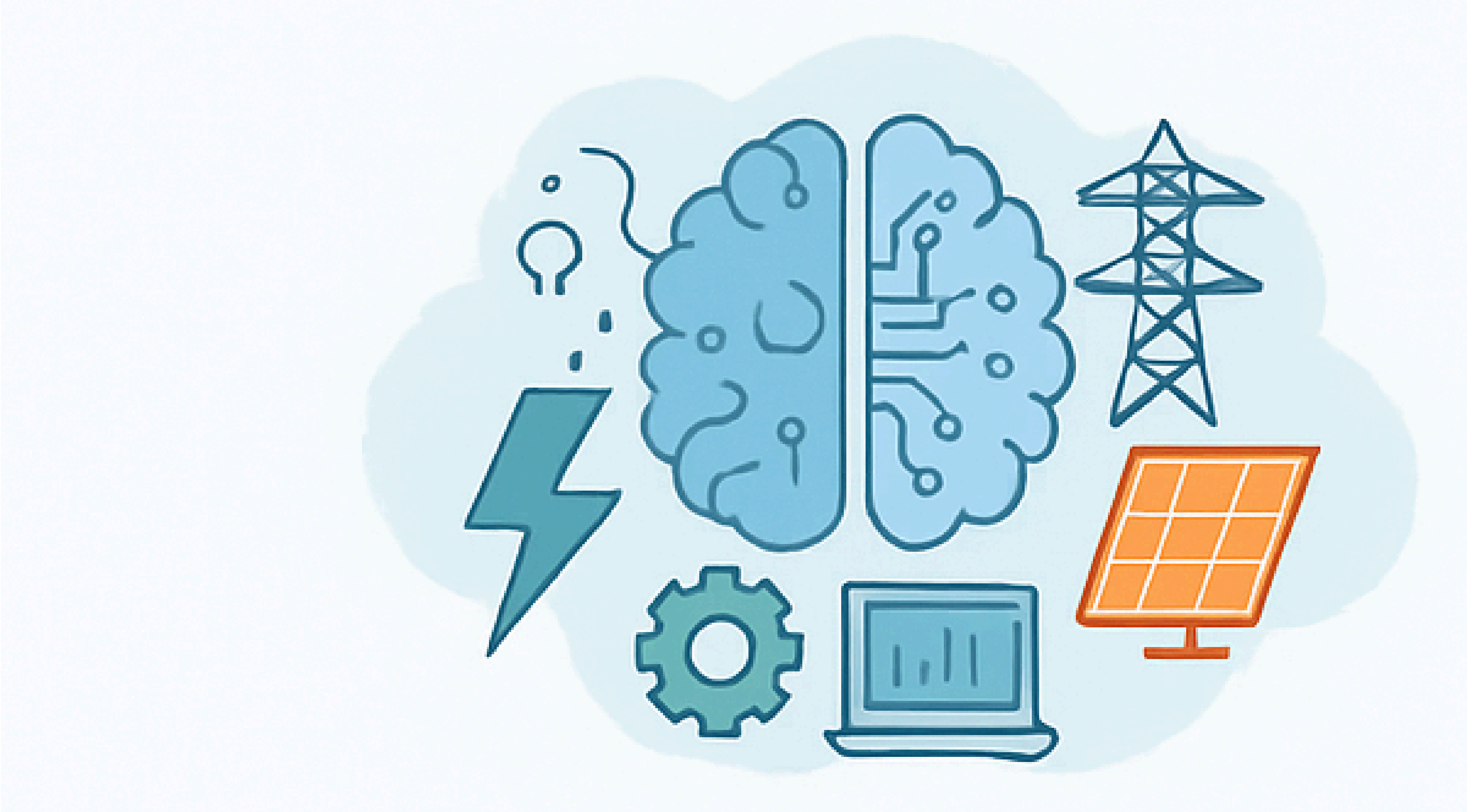
🏆 Best Model: ResNet50

- Highest validation accuracy (87.0%)
- Second highest test accuracy (82.4%)
- Strong recall & F1 in critical classes like Severe and Moderate
- Best for medical contexts where false negatives are costly

Top performer overall; best confusion matrix clarity

FUTURE SCOPE

AI Applications in Electrical engineering

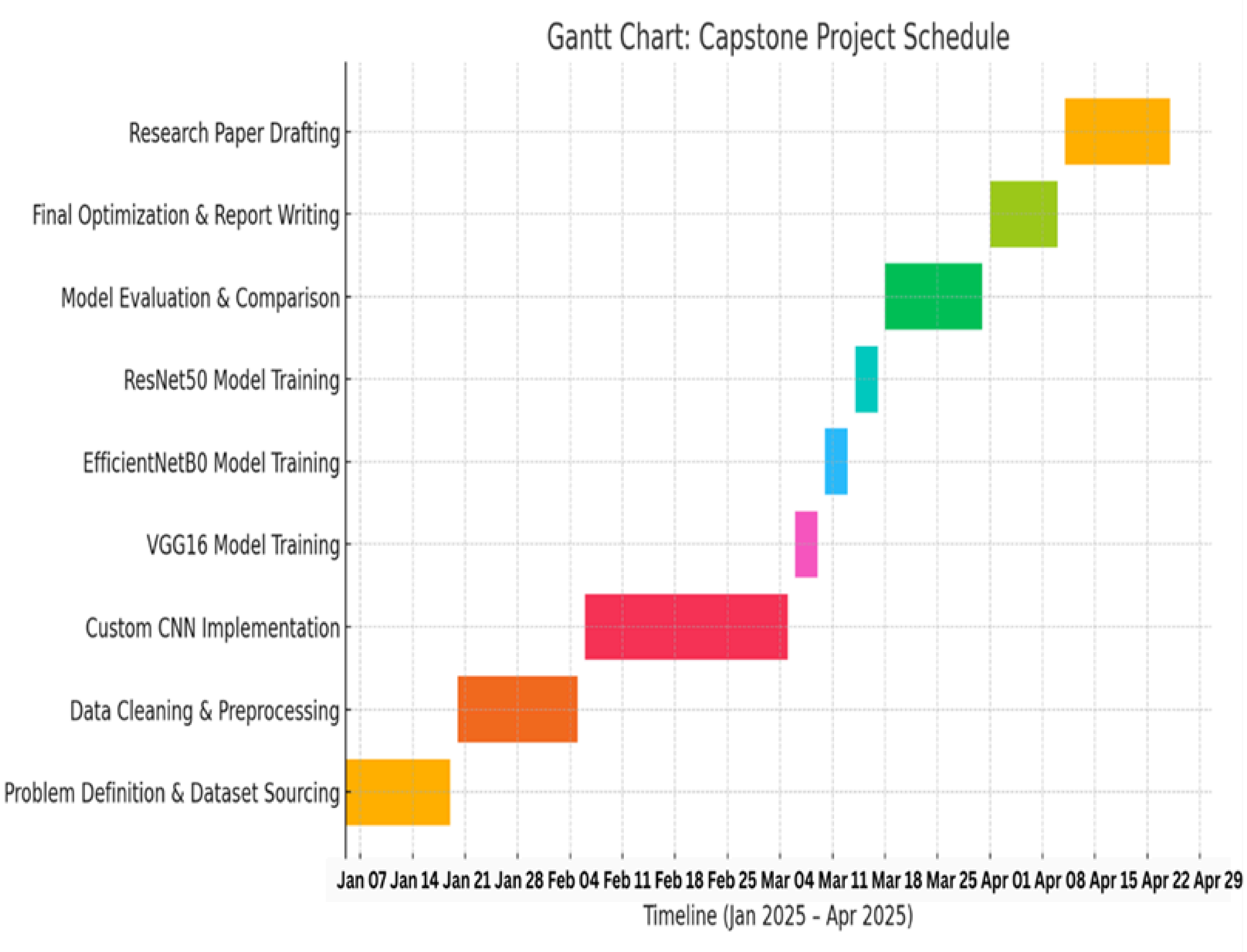


Characteristic	Fault Detection	Predictive Maintenance	Smart Meter Anomaly Detection	Renewable Energy Forecasting	Quality Inspection	Load Forecasting & Demand Response	IoT & Edge AI
Application	Detect faults using thermal/vibration data	Monitor equipment using image/time-series data	Detect anomalies using smart meter data	Predict energy generation using weather/sensor data	Inspect defects in electrical manufacturing	Predict peak load demand	Deploy models on edge devices
Model	CNN-based models	EfficientNetB0	Trained on historical usage patterns	ResNet50	CNN-based vision systems	AI-driven demand response	EfficientNetB0, pruned CNN
Advantages	Reduces downtime, improves grid reliability	Extends lifespan, reduces costs	Saves revenue, enhances efficiency	Optimizes storage, reduces fossil fuel dependency	Ensures quality, reduces waste	Prevents blackouts, reduces costs	Enables low-latency, reduces cloud dependency

Future Scope in Medical Imaging with AI

- 🔍 **Early Detection & Intervention**
  Enables accurate grading of arthritis severity (Normal → Severe), aiding timely diagnosis and treatment to prevent progression.
- 🔁 **Scalability to Other Conditions**
  The workflow (preprocessing → augmentation → model ensemble → deployment) is adaptable to **other musculoskeletal, dental, or oncology diagnostics.**
- 🌍 **Support in Resource-Constrained Settings**
  Functions as a **second-opinion tool** in clinics with limited access to radiology expertise, improving healthcare reach.
- 📡 **Integration with Telemedicine**
  Web-based deployment (via Flask) supports **remote diagnostics**, enhancing access through telehealth platforms.
- 🤝 **Research & Clinical Collaboration**
  Potential to integrate with **EHR systems** and expand training datasets for **more diverse, generalizable clinical models.**

WORK SCHEDULE



Thank you