CHEST X-RAY PNEUMONIA DETECTION

Complete Deep Learning & CNN Training Guide

8 Architectures · Transfer Learning · AUC-Weighted Ensemble · TensorFlow 2.x / Keras · 2025–2026

An AI-powered Computer-Aided Detection (CAD) system studying **5,856 pediatric chest X-rays** to classify **NORMAL vs PNEUMONIA**. This guide covers CNN architectures, transfer learning, data engineering, visual diagnosis, and the role of the physician. Current status: literature review and preprocessing complete — model training is the next step.

Note — Multi-Dataset Combination (Planned): We are planning to combine the Kermany dataset with NIH ChestX-ray14 and CheXpert to build a larger, more generalised model. This multi-dataset pipeline will be developed after initial training is complete.

8	5,856	In Prog.	1–5
CNN Models	X-Rays	Training	Age (yrs)

Dataset (Pediatric Ages 1–5): Kaggle Chest X-Ray Pneumonia — Kermany et al., 2018
https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

PROJECT TEAM — University of Saida, Algeria · AI & Medical Imaging 2025–2026

Role	Name
Data Engineer	**Bouhmidi Amina Maroua**
ML Engineer	**Labani Nabila Nour Elhouda**
Project Manager	**Kassouar Fatima**
Business Model	**Miloudi Meroua Amira**
Academic Supervisor	**Dr. Abderrahmane Khiat — University of Saida, Algeria**
Medical Advisor	**Dr. Aimer Mohammed Djamel Eddine**

GitHub: github.com/AminaMar/pediatric-pneumonia-detection | Python 3.8+ · TensorFlow ≥ 2.10 · Keras ≥ 2.10 · Jupyter Notebooks

■ *Note v1.0 — Models, pipeline and architecture choices may be updated as experimental results progress.*

TABLE OF CONTENTS

PART I — PROJECT OVERVIEW

1. Introduction & Project Goal

Pneumonia kills over 2.5 million people annually (WHO 2023) and is the leading infectious cause of death in children under five. Chest X-ray (CXR) is the standard first-line diagnostic tool, yet accurate interpretation requires expert radiological training — scarce in low-income regions. This project studies **8 deep learning architectures** on 5,856 pediatric CXRs from the Kermany dataset to automatically classify **NORMAL vs PNEUMONIA**. The literature review and data preprocessing are complete. Model training has not yet started — it is the next step. Work is also planned to combine this dataset with NIH ChestX-ray14 and CheXpert for a broader multi-dataset model.

Primary Goal	Secondary Goal	Future Work
NORMAL vs PNEUMONIA classification — training in progress	Compare all 8 architectures and identify the best for CXR	AUC-weighted ensemble + multi-dataset + patient dossier integration

2. Pneumonia — Medical Background

Definition: Acute inflammatory infection of lung parenchyma causing consolidation of alveoli, impaired gas exchange, and radiographic opacity. **Causes:** (1) Bacterial — S. pneumoniae, lobar consolidation. (2) Viral — Influenza, COVID-19, bilateral infiltrates. (3) Fungal — P. jirovecii, ground-glass opacities. **X-Ray Signs:** White opacity vs dark lung fields — lobar consolidation, interstitial infiltrates, pleural effusion. AI detects subtle early-stage changes that junior clinicians may miss.

Statistic	Value	Source
Annual deaths from pneumonia	**> 2.5 million**	WHO 2023
Leading infectious child killer under 5	**Pneumonia**	UNICEF
Missed diagnosis rate — junior clinicians	**~ 20–30%**	Literature
CheXNet F1 vs radiologist F1	**0.435 vs 0.387**	Rajpurkar 2017

3. Why AI for Medical Imaging?

Sub-Saharan Africa has fewer than **1 radiologist per 100,000 people** (WHO). AI is scalable, consistent, and always available — bridging a critical diagnostic gap in under-resourced regions.

Advantage	Detail
Speed	Prediction in < 100 ms per CXR vs 5–20 min for a radiologist. Critical for emergency triage.
Consistency	Same result 24/7. Human readers show up to 20% intra-reader variability across shifts.
New Biomarkers	Measures opacity density and lobe involvement at scale — impossible manually.
Equity	One model reaches millions of rural patients globally via a mobile or web application.
Explainability	Grad-CAM heatmaps highlight the exact region driving each prediction — building clinical trust.

4. Dataset Overview

Source: Kaggle Chest X-Ray Pneumonia Dataset — Kermany et al., 2018 — 5,856 pediatric CXRs (ages 1–5) from Guangzhou Women and Children's Medical Center, China. All scans graded by two expert physicians and verified by a third.

Note — Multi-Dataset Work in Progress: The current model is trained exclusively on the Kermany Kaggle pediatric dataset (ages 1–5). We are actively working on combining it with NIH ChestX-ray14 and CheXpert to extend coverage across adult populations and multiple pathologies.

Split	NORMAL	PNEUMONIA	Total	Imbalance Ratio
Train	1,341	3,875	**5,216**	1 : 2.89

Split	NORMAL	PNEUMONIA	Total	Imbalance Ratio
Validation	8	8	**16**	1 : 1.0 (too small — fixed in §16)
Test	234	390	**624**	1 : 1.67

■ *Class imbalance of 2.9× more PNEUMONIA than NORMAL. Mitigation: class-weighted loss, AUC as primary metric.*

5. The 8 CNN Architectures

Model	Year	Params	Size	CPU/ep	Top-1	Key Innovation
MobileNetV2	2018	3.4 M	14 MB	~5 min	72.0%	Inverted residuals — CPU champion
AlexNet *	2012	60 M	240 MB	~35 min	—	First deep CNN — trained from scratch
ResNet50V2	2016	25 M	98 MB	~25 min	75.6%	Skip connections: output = F(x) + x
InceptionV3	2015	23 M	92 MB	~20 min	77.9%	Parallel 1×1, 3×3, 5×5 convolutions
DenseNet121	2017	8 M	33 MB	~12 min	74.7%	Every layer connects to all prior layers
EfficientNetB3	2019	12 M	48 MB	~14 min	81.6%	Compound scaling: depth + width + resolution
ViT	2020	86 M	330 MB	~40 min	81.8%	Image patches processed as transformer tokens
VGG16	2014	138 M	528 MB	~45 min	71.5%	Deep uniform 3×3 stacks — strong baseline

** AlexNet has no ImageNet pre-training — it is trained from scratch on CXR data only.*

PART II — CNN FUNDAMENTALS

5a. What is a CNN?

A Convolutional Neural Network uses small learnable filters that slide across the image, detecting local patterns at every position. **Early layers** → **edges**, **middle layers** → **textures**, **deep layers** → **lung shapes and pneumonic opacities**. Far more efficient than a fully-connected network, which would require 150,528 inputs for a single 224×224 image.

Operation	Formula / Behaviour	Role in Chest X-Ray
Conv2D	Output[i,j,k] = Σ(Input × Filter_k)	Detects edges, opacities, lung boundaries
BatchNorm	x' = (x–mean)/std × γ + β	Stabilises training across X-ray exposures
ReLU	f(x) = max(0, x)	Adds non-linearity; prevents vanishing gradient
MaxPooling	2×2 maximum of 4 neighbours	Down-samples; adds translation invariance
GlobalAvgPool	[B,H,W,C] → [B,C]	Bridges feature maps to the classifier head
Dense	y = W · x + b	Combines all features for binary decision
Dropout(p)	Zeros p% of neurons at training time	Prevents memorising specific patient scans
Sigmoid	1 / (1 + e^–x)	Outputs PNEUMONIA probability in [0, 1]

5b. Full Pipeline: Raw Pixel → Diagnosis

X-Ray Input	Normalise & Augment	CNN Backbone	Dense Head	Sigmoid p∈ [0,1]	Decision Normal/ Pneumonia

Figure: End-to-end pipeline from image loading to ensemble decision.

Step	Stage	What Happens	Output Shape
1	**Load & Resize**	JPEG decoded, resized to model input, RGB	[B, 160, 160, 3]
2a	**Normalise**	÷255 → [0,1] | MobileNetV2: ÷127.5–1 → [–1,1]	[B, 160, 160, 3]
2b	**Augment**	Flip, rotate, zoom, translate (training phase only)	[B, 160, 160, 3]
3	**CNN Backbone**	Convolutional layers extract hierarchical features	[B, H', W', C]
4a	**GlobalAvgPool**	Spatial averaging per channel	[B, C]
4b	**Dense Head**	Dense(512) → Dropout → Dense(256) → Dropout → Dense(1) logit	[B, 1] logit
5	**Sigmoid**	Logit → probability (must use dtype=float32)	[B, 1] ∈ [0,1]
6a	**Ensemble**	AUC-weighted average across all 8 models	1 scalar
6b	**Decision**	p > 0.5 → PNEUMONIA | p ≤ 0.5 → NORMAL	Class label

5c. Architecture Deep Dives

ResNet50V2 — Skip Connections: Solves vanishing gradient. Each block: **output = F(x) + x**. V2 uses pre-activation (BN + ReLU before conv) for even cleaner gradient flow. Enables stable training of 50–152 layer networks.

InceptionV3 — Multi-Scale Convolutions: Applies 1×1, 3×3, and 5×5 convolutions in parallel, then concatenates. Ideal for chest X-rays where pneumonia appears at multiple scales — from tiny alveolar opacities to large lobar consolidations.

DenseNet121 — Dense Connectivity (CheXNet Backbone): Every layer receives feature maps from all preceding layers. 8 M params for 121 layers. Backbone of Stanford CheXNet (2017) — first model to surpass radiologist F1.

ViT — Vision Transformer: Divides image into 196 non-overlapping 16×16 patches embedded as 768-dim tokens. Twelve Transformer encoder layers apply Multi-Head Self-Attention globally — captures bilateral lung patterns from layer 1.

MobileNetV2 — CPU Champion: Depthwise Separable Convolution reduces FLOPs ~8–9×. Inverted Residual Bottleneck: narrow → wide (×6) → depthwise 3×3 → narrow (linear, no activation). Linear projection preserves manifold information.

CRITICAL — MobileNetV2 requires pixel values in [–1, 1] (÷127.5 – 1). All other models use [0, 1] (÷255). Mixing these silently reduces accuracy by **5–15%**.

Step	Operation	Channels	Note
1 — Expand	1×1 Conv + ReLU6	C → C×6	Expansion factor t = 6
2 — Depthwise	3×3 DW Conv + ReLU6	C×6	One filter per channel; 8–9× fewer FLOPs
3 — Project	1×1 Conv (NO activation)	C×6 → C	Linear projection preserves manifold
Skip Add	Input + Output	C	Only when stride=1 and C_in=C_out

PART III — TRAINING METHODOLOGY

6. Transfer Learning Strategy

With only ~5,216 training images, training from scratch is insufficient. Transfer learning uses models pre-trained on ImageNet (1.2 M images, 1,000 classes). Low-level features (edges, textures, gradients) are universal visual primitives that transfer to X-ray images. Only the final classification head is replaced.

Parameter	Phase 1 — Head Training	Phase 2 — Fine-Tuning
Backbone	ALL layers frozen	Last 20 layers trainable
Learning Rate	1e–3	1e–5 (100× smaller than Phase 1)
Batch Size	64 (GPU) / 16 (CPU)	32 (GPU) / 8 (CPU)
Max Epochs	25 with early stopping patience=8	15 with early stopping patience=6
Objective	Train classification head on pre-learned features	Adapt backbone to X-ray domain
Main Risk	LR too small → slow convergence	LR too large → catastrophic forgetting

Classification Head — identical for all 8 models:

```
Backbone → GlobalAveragePooling2D → BatchNorm → Dense(512,relu) → Dropout(0.5) → Dense(256,relu)
→ Dropout(0.3) → Dense(1,dtype=float32) → Activation('sigmoid',dtype=float32)
```

■ The final Dense and Sigmoid layers MUST use **dtype=float32**. FP16 overflows on large logits and produces NaN loss or accuracy stuck at 50%.

7. Data Augmentation

With ~5,000 training images, overfitting is the primary risk. Augmentation ensures the model never sees the exact same pixel pattern twice across epochs.

Transform	Range	Medical Justification
Horizontal Flip	50%	Patients can be imaged from either side; dextrocardia exists
Rotation	±36°	Patient may be slightly tilted on the examination table
Zoom	±15%	Variable detector-to-patient distance across facilities
Translation	±10%	Patient not always perfectly centred on the film
Brightness	±15%	Varying X-ray exposure settings across different machines

8. Loss Function, Optimizer & Label Smoothing

Concept	Formula / Value	Purpose
Binary Cross-Entropy	L = –[y·log(p)+(1–y)·log(1–p)]	Penalises wrong confident predictions heavily
Label Smoothing	y_s = y×0.95+0.025 (α=0.05)	Prevents overconfidence; improves generalisation
AdamW Optimizer	Adam + Weight Decay λ=1e–4	Adaptive learning rate + L2 weight regularisation
Class Weighting	w_normal=2.89, w_pneumonia=1.0	Compensates 2.9× class imbalance in the dataset

9. Training Callbacks

Callback	Trigger	Action
ModelCheckpoint	val_accuracy improves	Save best model weights to .keras file
EarlyStopping	No improvement for N epochs	Stop training and restore best weights
ReduceLROnPlateau	val_loss plateau for 3 epochs	Multiply learning rate by 0.3

Callback	Trigger	Action
CSVLogger	After every epoch	Append all metrics to CSV for analysis

Training Diagnosis Guide

Symptom	Diagnosis	Fix
train_acc >> val_acc (gap > 10%)	**Overfitting**	Add dropout, more augmentation, early stopping
Both accuracies low and flat	**Underfitting / LR too small**	Increase LR or train more epochs
val_loss rises while train_loss falls	**Overfitting starting**	EarlyStopping triggers automatically
Accuracy stuck at 50–55%	**Predicts only one class**	Check class weights and imbalance handling
loss = NaN after epoch 1	**FP16 sigmoid overflow**	Use dtype=float32 on Dense and Sigmoid
Epoch 1 takes 14+ minutes	**Reading from Google Drive**	Copy data to /content/ before training

PART IV — REFERENCE & RESEARCH

11. Key Research Papers per Model

Model	Paper & Authors	Year	Key Contribution
MobileNetV2	Sandler et al. — CVPR arxiv.org/abs/1801.04381	2018	13× fewer FLOPs than ResNet50; 72% ImageNet with 3.4 M params
AlexNet	Krizhevsky, Sutskever, Hinton — NeurIPS	2012	Started the deep learning revolution; introduced ReLU and dropout
ResNet50V2	He et al. — CVPR Best Paper arxiv.org/abs/1512.03385	2016	Skip connections enable stable training of 50–152 layer networks
InceptionV3	Szegedy et al. — CVPR arxiv.org/abs/1512.00567	2016	Parallel multi-scale convolutions; 77.9% ImageNet
DenseNet121	Huang et al. — CVPR Best Paper arxiv.org/abs/1608.06993	2017	Dense connectivity; CheXNet backbone; 8 M params for 121 layers
CheXNet	Rajpurkar et al. — Stanford arxiv.org/abs/1711.05225	2017	First AI to surpass radiologist F1 on pneumonia: 0.435 vs 0.387
EfficientNetB3	Tan & Le — ICML arxiv.org/abs/1905.11946	2019	Compound scaling; 81.6% ImageNet with only 12 M params
ViT	Dosovitskiy et al. — ICLR arxiv.org/abs/2010.11929	2020	Pure transformer on image patches; global attention from layer 1
VGG16	Simonyan & Zisserman — ICLR arxiv.org/abs/1409.1556	2015	16-layer 3×3 conv architecture; 92.8% accuracy on Kaggle CXR

12. Common Errors & Fixes

Error / Symptom	Root Cause	Fix
ModuleNotFoundError: google.colab	Running outside Colab	Wrap import in try / except ImportError
No GPU devices available	CPU-only runtime assigned	Colab: Runtime → Change runtime type → T4 GPU
Unknown layer: 'CastToFloat32'	Model saved under FP16	Set mixed_precision policy before loading model
Epoch 1 takes 14+ minutes	Data read from Google Drive	Copy data to /content/ with shutil.copytree first
NaN loss or accuracy frozen at 50%	Sigmoid layer uses float16	Use Dense(1, dtype='float32') and Activation(..., dtype='float32')

13. Improvement Roadmap

Priority	Improvement	Expected Gain
1 — HIGH	Scale to NIH ChestX-ray14 (112,120 images) — multi-dataset combination in progress	+ 5–10% accuracy
2 — HIGH	Grad-CAM heatmap visualisation	Clinical explainability and trust
3 — MED	Test-Time Augmentation (TTA — 8 augmented copies)	+ 0.5–1.5%
4 — MED	Class-weighted loss (penalise missed PNEUMONIA)	+ 2–4% sensitivity
5 — MED	Replace ImageDataGenerator with tf.data pipeline	2–5× CPU speedup

Priority	Improvement	Expected Gain
6 — LOW	LR Warm-Up + Cosine Decay (CosineDecayRestarts)	+ 0.5–1%
7 — LOW	Focal Loss: FL(p) = –α(1–p)^γ · log(p)	+ 1–2% on hard examples

14. Best Datasets to Scale Up

Multi-Dataset Combination (In Progress): We are currently working on combining the Kermany dataset with NIH ChestX-ray14 and CheXpert to build a more generalised model covering adult and pediatric populations across multiple pathologies. Results will be published as the project progresses.

Dataset	Images	Labels	Best Use
Kaggle CXR — Kermany ★ Current	5,856	Normal / Pneumonia	This project. Pediatric ages 1–5. CPU-friendly.
NIH ChestX-ray14 ← Next	112,120	14 pathologies incl. pneumonia	Best first upgrade; original CheXNet training set. Combination in progress.
CheXpert — Stanford ← Next	224,316	14 conditions + uncertainty	State-of-the-art; uncertainty label handling. Combination in progress.
MIMIC-CXR	377,110	13 labels + radiology reports	Largest public CXR; enables report generation
VinBigData CXR	18,000	14 pathologies + bboxes	Object detection; lesion localisation tasks
PadChest	160,000	174 radiological findings	European cohort; rare pathologies coverage

15. Quick-Reference Cheat Sheet

Concept	Formula / Value
Binary Cross-Entropy	L = –[y·log(p) + (1–y)·log(1–p)]
Label Smoothing (α=0.05)	y_s = y×0.95 + 0.025
MobileNetV2 normalisation	pixel ÷ 127.5 – 1 → range [–1, 1]
All other models	pixel ÷ 255 → range [0, 1]
LR reduction rule	LR_new = LR_old × 0.3 (after 3 plateau epochs)
Weighted ensemble	p_final = Σ(AUC_i × p_i) / Σ(AUC_i)
Decision threshold	p_final > 0.5 → PNEUMONIA | ≤ 0.5 → NORMAL
Phase 1 LR	1e–3 (head only — backbone frozen)
Phase 2 LR	1e–5 (last 20 layers — risk of catastrophic forgetting)
Dropout schedule	0.5 after Dense(512) | 0.3 after Dense(256)
Weight Decay	λ = 1e–4 (AdamW)

PART V — STUDENT PROJECT CONTEXT ★

This part covers (16) our team's preprocessing pipeline and GitHub repository, (17) a practical visual guide for reading chest X-rays, and (18) why the physician's role is irreplaceable in any AI-assisted diagnostic system. Project based at the University of Saida, Algeria.

16. Team Dataset — AminaMar GitHub Project

Our team built a complete data engineering pipeline on top of the Kaggle Kermany dataset, hosted at **github.com/AminaMar/pediatric-pneumonia-detection**. The central contribution is fixing the original Kaggle validation set — which contained only **16 images** — replacing it with a proper **stratified 70/15/15 split** producing 878 validation images. We are also working in parallel on combining this dataset with NIH ChestX-ray14 and CheXpert to build a more generalised multi-dataset model — this combination pipeline is currently in progress.

GitHub Repository: github.com/AminaMar/pediatric-pneumonia-detection
Source dataset: Kermany et al., 2018 · Guangzhou Women and Children's Medical Center, China
Note: We are not limited to this single dataset — multi-dataset combination work with NIH ChestX-ray14 and CheXpert is currently in progress.

Why This Project Matters for Algeria

Challenge in Algeria	Scale
Daily pediatric pneumonia cases in emergency departments	**200–300 cases per day**
Children who travel > 50 km to access an X-ray facility	**40% of cases**
Pediatric radiologist availability	**~ 1 per 500,000 children**
Average diagnostic delay in rural hospitals	**6–24 hours**
Planned integration target	**EMR systems via HL7 / FHIR standards**

Original Kaggle Dataset vs Team Preprocessed Version

Property	Kaggle Original — Kermany	Team Version — AminaMar GitHub
Total images	5,856	5,856 (same source)
NORMAL count	1,583 (27%)	1,583 (27%)
PNEUMONIA count	4,273 (73%)	4,273 (73%)
Class imbalance	2.7 : 1	2.7:1 — corrected with class weights 1.850/0.685
Training split	5,216 images (89%)	4,099 images (70%) — proper balance
Validation split	**16 images only (!) too small**	**878 images (15%) — FIXED ★**
Test split	624 images	879 images (15%)
Image resolution	Variable — original resolution	Standardised to 224×224 px
Format	RGB JPEG	Grayscale, normalised to [0–1]
Split strategy	Pre-defined — not stratified	Stratified — equal class ratio in each split
Augmentation	Not standardised	Rotation ±15°, Shift 10%, Zoom 10%, Horizontal Flip
EMR integration	None	Planned: HL7/FHIR for Algerian hospital systems

★ The validation set fix (16 → 878 images) is the most critical preprocessing step. Training with only 16 validation images produces unreliable val_accuracy and prevents meaningful early stopping or model selection.

Pipeline Phases

Phase	Name	Status	Key Deliverable
1	**Literature Review**	■ **Complete**	Study of 8 CNN architectures, transfer learning, and related medical AI papers
2	**Data Exploration**	■ **Complete**	EDA report — class imbalance analysis — validation set issue identified
3	**Preprocessing**	■ **Complete**	Stratified 70/15/15 split — class weights — augmentation pipeline — 224×224 px
4	**Data Loaders**	■ **Complete**	TF ImageDataGenerator — batch size 32 — performance < 0.5 s/batch
5	**Model Training**	■ **Not Started**	VGG16, ResNet50, DenseNet121 and all 8 architectures — to begin next
6	**Multi-Dataset Combination**	■ Planned	Combining Kermany + NIH ChestX-ray14 + CheXpert for a generalised model
7	**EMR Integration**	■ Planned	HL7/FHIR interface for Algerian hospital information systems

17. Student Visual Guide — How to Read a Chest X-Ray

As AI students we are not trained radiologists, but by studying thousands of labeled X-rays and examining what our CNN models learn through Grad-CAM, we can identify visual patterns that consistently distinguish Normal from Pneumonia. This section serves as a practical reference when reviewing model predictions.

Key Anatomical Regions and Their Normal vs Pneumonia Appearance

Region	Normal Appearance	Pneumonia Signal
Lung fields	Mostly dark — air-filled alveoli are radiolucent	Bright white patches — alveoli filled with fluid or pus
Lung symmetry	Both sides visually equal in density	Asymmetry — one or both lungs appear brighter
Heart border	Sharp, clearly defined left cardiac margin	Blurred border — fluid or consolidation obscuring the edge
Diaphragm	Crisp dome-shaped line at the lung base	Blurred or elevated — consolidation pressing from above
Costophrenic angles	Sharp V-shape at both lung bases	Blunted angle — pleural effusion (fluid) pooling there
Lung texture	Fine, homogeneous vascular markings	Dense focal area or diffuse haze — consolidation zone
Air bronchograms	Not visible in a normal lung	Dark air-filled bronchi visible within a white opacity

Bacterial vs Viral Pneumonia — Visual Differences

Feature	Bacterial Pneumonia	Viral Pneumonia
Pattern	Lobar consolidation — dense focal area in one lobe	Bilateral interstitial infiltrates — diffuse haze in both lungs
Location	Usually one lobe or one segment	Both lungs, distributed throughout
Appearance	Dense, opaque white patch	Bilateral ground-glass haziness
Air bronchograms	Often present within the opacity	Less commonly seen
Common pathogen	Streptococcus pneumoniae	Influenza, RSV, COVID-19
Typical context	Any age, acute onset, high fever	Children, elderly, immunocompromised patients

What CNN Models Learn at Each Depth (verified via Grad-CAM)

Network Depth	Features Learned	Clinical Equivalent
Early layers (1–5)	Edges, lines, intensity gradients	Lung borders, rib outlines, cardiac silhouette
Middle layers (6–15)	Textures, local intensity patterns	Lung texture, vascular markings, interstitial pattern
Deep layers (16–25+)	Semantic regions and complex shapes	Lobar opacity, consolidation zone, effusion area
Grad-CAM output	Heat map overlaid on the original X-ray	Highlights the specific region driving the PNEUMONIA prediction

Student Decision Checklist

Question to Ask	NORMAL	PNEUMONIA
Are the lung fields mostly dark?	**YES**	**NO — white or bright patches present**
Are both lungs symmetric in density?	**YES**	Often NO — one or both sides appear brighter

Question to Ask	NORMAL	PNEUMONIA
Are heart and diaphragm borders sharp?	**YES**	May be blurred or obscured
Are costophrenic angles sharp V-shapes?	**YES**	May be blunted — pleural effusion likely
Is there a dense white patch in one lobe?	**NO**	**YES — bacterial lobar pattern likely**
Is there bilateral diffuse haziness?	**NO**	**YES — viral or atypical pneumonia likely**
Does Grad-CAM highlight lung tissue?	N/A	Should — if it highlights bone or device, verify the prediction

✔ *Use this checklist when reviewing model predictions to verify the model attends to clinically meaningful regions — not image artifacts or equipment.*

18. The Irreplaceable Role of the Doctor

Our model achieves **>90% accuracy** on the test set — but test set accuracy is not clinical safety. A percentage on a labeled dataset does not capture rare presentations, patient history, comorbidities, or the full complexity of a real consultation.

What Our AI System Does Well

Capability	Performance	Clinical Value
Prediction speed	< 100 ms per X-ray	Rapid triage flag in crowded emergency departments
Consistency	Same result 24/7	No fatigue, no shift-to-shift variability
Pneumonia sensitivity	To be determined after training	Aim: rarely miss a true positive case
Scale	Thousands of CXRs/hour	Mass screening programs; rural deployment
Explainability	Grad-CAM available	Guides radiologist attention to the right region
Accessibility	Mobile and web deployable	Patients far from specialist radiology centres

What Only the Doctor Can Do — 10 Irreplaceable Dimensions

Clinical Dimension	What the Doctor Does	Why AI Cannot Do This
Patient history	Asks about fever onset, travel, contacts, vaccination	AI sees pixels only — no clinical context
Physical examination	Listens to breath sounds, checks O2 saturation	AI has no sensor beyond the single image
Clinical correlation	Combines X-ray with symptoms, history, and lab results	AI produces a probability — not a clinical judgement
Differential diagnosis	Considers TB, lung cancer, pulmonary oedema, others	Our model is binary: Normal / Pneumonia only
Age & comorbidities	Adjusts for neonate vs elderly vs immunocompromised	Model trained on pediatric ages 1–5 only
Treatment decision	Selects antibiotic type, dose, route, and duration	AI has no pharmacological knowledge
Patient follow-up	Reassesses response to treatment, orders repeat imaging	AI processes one static image at one time point
Legal accountability	Bears full moral and legal responsibility	Accountability cannot be delegated to a software tool
Edge cases	Recognises rare presentations, poor-quality films	AI may confidently misclassify out-of-distribution images
Communication	Explains diagnosis and treatment plan to the patient	AI produces a probability score — not a human conversation

Correct Workflow — AI and Physician Collaboration

Step	Actor	Action
1	**Radiographer**	Chest X-ray acquired and digitised
2	**AI System**	NORMAL / PNEUMONIA probability computed in < 100 ms
3	**System**	High-probability cases flagged for priority radiologist review
4	**Radiologist**	Reviews AI output and Grad-CAM heatmap alongside the full image

Step	Actor	Action
5	**Physician**	Correlates X-ray findings with patient history, symptoms, and lab results
6	**Doctor**	Confirms the final diagnosis — bears full clinical accountability
7	**Medical Team**	Prescribes treatment: antibiotics, oxygen therapy, hospital admission
8	**Feedback Loop**	Corrections and new cases fed back to retrain and improve the model

Specific Risks of Using AI Without a Doctor

Risk	Scenario	Consequence
False positive	Normal CXR flagged as Pneumonia	Unnecessary antibiotics → resistance, side effects
False negative	Atypical pneumonia missed by the model	Delayed treatment → patient condition deteriorates
Wrong pathogen	Viral pneumonia classified as bacterial	Wrong antibiotic prescribed — no benefit, potential harm
TB missed	TB opacity classified as Pneumonia	Infectious disease spreads untreated
Artifact confusion	Rotated or underexposed X-ray	AI gives a confident but incorrect prediction
Age mismatch	Adult CXR evaluated by pediatric model	Out-of-distribution — results are unreliable

Summary — AI Role vs Doctor Role

Dimension	AI System	Doctor
Speed	Milliseconds	Minutes
Scale	Unlimited, 24/7	Limited by shift hours
Scope	Binary: Normal/Pneumonia only	All pathologies, all clinical context
Patient history	None	Full clinical and social context
Treatment	Cannot prescribe	Selects, adapts, and monitors therapy
Accountability	None — it is a tool	Full legal and ethical responsibility
Communication	A probability score	Empathetic dialogue with patient and family
Role in system	First-pass screening and support	**Final diagnosis and treatment authority**

★ As students, we built a powerful screening tool. As future professionals, we understand that deploying it safely requires doctors, regulators, ethicists, and patients working together. **AI raises the floor** — ensuring no case is silently missed. **The doctor raises the ceiling** — ensuring correct treatment, safety, and accountability.

REFERENCES

[1] Sandler et al. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. CVPR 2018. arxiv.org/abs/1801.04381

[2] Krizhevsky, Sutskever, Hinton (2012). ImageNet Classification with Deep CNNs (AlexNet). NeurIPS 2012.

[3] He et al. (2016). Deep Residual Learning for Image Recognition. CVPR 2016 Best Paper. arxiv.org/abs/1512.03385

[4] Szegedy et al. (2016). Rethinking the Inception Architecture. CVPR 2016. arxiv.org/abs/1512.00567

[5] Huang et al. (2017). Densely Connected Convolutional Networks. CVPR 2017 Best Paper. arxiv.org/abs/1608.06993

[6] Tan & Le (2019). EfficientNet: Rethinking Model Scaling for CNNs. ICML 2019. arxiv.org/abs/1905.11946

[7] Dosovitskiy et al. (2021). An Image is Worth 16x16 Words. ICLR 2021. arxiv.org/abs/2010.11929

[8] Simonyan & Zisserman (2015). Very Deep Convolutional Networks for Large-Scale Recognition. ICLR 2015. arxiv.org/abs/1409.1556

[9] Rajpurkar et al. (2017). CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays. arXiv 2017. arxiv.org/abs/1711.05225

[10] Irvin, Rajpurkar et al. (2019). CheXpert: A Large Chest Radiograph Dataset. AAAI 2019. arxiv.org/abs/1901.07031

[11] Wang et al. (2017). ChestX-ray8: Hospital-scale Chest X-ray Database. CVPR 2017. arxiv.org/abs/1705.02315

[12] Selvaraju et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. ICCV 2017. arxiv.org/abs/1610.02391

[13] Kermany et al. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell 172(5) 2018.

[14] WHO (2023). Pneumonia Fact Sheet. who.int/news-room/fact-sheets/detail/pneumonia

[15] Kaggle Dataset (Pediatric Ages 1–5): kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[16] Bouhmidi A.M., Labani N., Kassouar F., Miloudi M.A. (2025–2026). AI-Powered Pediatric Pneumonia Detection — Integration with EMR in Algeria. Supervisor: Dr. Abderrahmane Khiat. Medical Advisor: Dr. Aimer Mohammed Djamel Eddine. University of Saida. github.com/AminaMar/pediatric-pneumonia-detection

[17] EU Medical Device Regulation MDR 2017/745. Official Journal of the European Union.

Contact for more information, collaboration, or dataset access:

Kassouar Fatima (Project Manager) — kassouarfatima48@gmail.com

GitHub: github.com/AminaMar/pediatric-pneumonia-detection · University of Saida, Algeria · 2025–2026