

A multi-model consensus framework for pneumonia detection: optimizing diagnostic sensitivity under "Green AI" constraints

Ertan M. Geldiev

Automated pneumonia diagnosis from chest X-rays minimizes subjective bias, but powerful deep learning models often require expensive hardware, hindering deployment in budget-limited regions. This study develops a "Green AI" framework that runs efficiently on legacy hardware, ensuring clinical safety is not sacrificed for lower computing costs. Models were validated on two distinct systems: a modern budget setup (Intel i3-10100/NVIDIA RTX 3050) and a legacy workstation (Intel Xeon E3-1245 v5/NVIDIA Quadro K2200), proving cross-generation hardware compatibility. The experimental results show that ResNet18 provides the best overall balance (89.26% accuracy, 78.20% specificity), while a lightweight L-CNN and a Linear SVM act as superior "safety nets," achieving diagnostic sensitivities of 99.49% and 99.23%, respectively. The study's primary contribution is the validation of the 'Discrepancy Paradox', where shallower models prove more robust against local anatomical artifacts (e.g., rib shadows) than overparameterized deep networks. Methodologically, we propose a hybrid multi-model consensus framework that utilizes these lightweight networks as a high-sensitivity triage layer. Operating under 'Green AI' constraints, this design achieves 99.49% diagnostic sensitivity on legacy hardware. Thus, we demonstrate that algorithmic optimization can effectively compensate for computational limitations, delivering a safe, low-cost clinical decision support tool for resource-constrained environments.

Keywords – Green AI, Pneumonia Detection, Deep Transfer Learning, ResNet18, Lightweight CNN, Support Vector Machines, Model Consensus, Clinical Decision Support Systems (CDSS)

Мултимоделна консенсусна рамка за откриване на пневмония: оптимизиране на диагностичната чувствителност при ограничения на „Зелен изкуствен интелект“ (Ертан М. Гелдиев) Автоматизираната диагностика на пневмония елиминира субективните пристрастия, но мощните модели често изискват скъп хардуер, което възпрепятства внедряването им в региони с ограничени бюджети. В тази работа е представена рамка, базирана на „Зелен изкуствен интелект“ (Green AI), работеща ефективно върху остарял хардуер без компромис с клиничната безопасност. Моделите са валидирани върху две хардуерни системи от различни поколения. Резултатите сочат, че ResNet18 постига най-добър общ баланс, докато олекотената L-CNN архитектура и Linear SVM действат като високочувствителни „предпазни мрежи“. Основният принос на изследването е доказването на „Парадокса на несъответствието“, при който по-плитките модели са по-устойчиви на локални анатомични артефакти (сенки от ребра) спрямо дълбоките мрежи. Методологически е предложена хибридна консенсусна рамка, интегрираща тези олекотени модели като триажен слой. Подходът постига 99.49% чувствителност върху остарял хардуер в рамките на концепцията „Green AI“. Изследването доказва, че алгоритмичната оптимизация успешно компенсирa хардуерните ограничения, осигурявайки безопасна и нискобюджетна клинична подкрепа в среди с ограничени ресурси.

Ключови думи: Зелен изкуствен интелект, откриване на пневмония, дълбоко трансферно обучение, ResNet18, олекотена конволюционна невронна мрежа (L-CNN), метод на опорните вектори (SVM), консенсус на модели, системи за подпомагане на клинични решения (CDSS)

Introduction

Automated screening of pneumonia using chest X-ray (CXR) datasets eliminates subjective human errors in clinical environments. In this study, the experimental validation is conducted using the benchmark public "Chest X-Ray Images (Pneumonia)" dataset [1], which comprises 5,856 clinically verified thoracic radiographs. Standard deep learning approaches, such as CheXNet [2], achieve high diagnostic accuracy but require significant computational overhead. This heavy resource requirement restricts their practical deployment in local medical centers or clinics operating with legacy computing hardware.

Literature review

Computer-aided thoracic radiology relies heavily on Convolutional Neural Networks (CNNs). Recent publications focus on achieving optimal performance via deep transfer learning configurations [3]. However, scaling up model parameters creates high energy footprints and limits execution on older edge devices [4].

Recent methods explore hardware-friendly AI optimizations. Researchers emphasize that tuning the decision thresholds of compact architectures with fewer convolutional blocks can achieve clinical sensitivity values comparable to deeper networks [5].

Additionally, deep architectures show high sensitivity to imaging noise, patient motion artifacts, and bone density variations, which leads to false clinical indications [6]. Multi-model consensus pipelines offer a direct engineering solution to stabilize these outputs. While deep layers focus on localized features, classical ensembles process general pixel distributions. Blending these structural methodologies prevents individual model bias, providing a stable diagnostic layer suitable for low-cost, resource-constrained implementations [7].

Hardware and software infrastructure

To ensure our research is both reproducible and practical, we utilized two distinct hardware configurations. The primary system, referred to as Workstation-1, features an Intel Core i3-10100 (3.60 GHz) and an NVIDIA GeForce RTX 3050 (8GB VRAM). This setup provided the necessary computational power for the initial training of the deep learning models.

In line with our "Green AI" objectives, we conducted the inference and testing phases on a legacy system, Workstation-2. This machine is equipped with an Intel Xeon E3-1245 v5 processor and an NVIDIA Quadro K2200 (4GB VRAM). As shown in Fig. 1, the

CSV output from the nvidia-smi tool confirms that the Quadro K2200 operates with extremely low power consumption (under 1W at idle), making it an ideal candidate for sustainable medical AI deployment in resource-constrained environments.

For the software environment, we maintained consistency across both nodes:

- Operating System: Debian GNU/Linux 13 (Trixie).
- Frameworks: PyTorch with CUDA 12.4 acceleration for neural networks and Scikit-learn for classical models.

Environment Isolation: A dedicated virtual environment (lcnx_xray) was used to manage dependencies and ensure the experiments are fully repeatable.

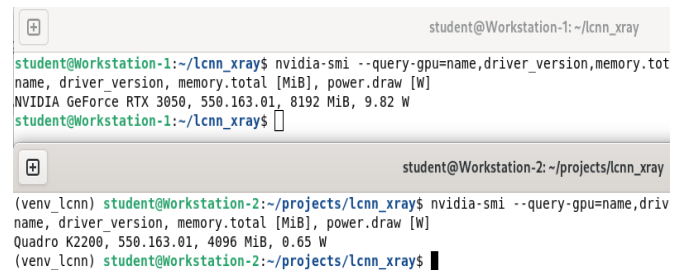


Fig.1. Comparative hardware monitoring of the two nodes. The top terminal shows the training environment (Workstation-1) and the bottom terminal displays the energy-efficient testing environment (Workstation-2).

Model architectures

In this study, we contrast traditional deep learning power with resource-efficient designs to evaluate their performance in medical diagnostics.

Deep learning models

- ResNet18 [8]: Implemented using Transfer Learning with weights pre-trained on the ImageNet dataset [9]. The architecture utilizes Skip Connections to bypass potential gradient degradation. This is critical for preserving the fine textures and subtle opacities in the lung parenchyma [2].
- L-CNN (Lightweight CNN): Inspired by the Light CNN design [11], our custom implementation features a streamlined 3-layer convolutional backbone. The architecture consists of sequential layers with 16, 32, and 64 filters (3×3 kernels), each followed by a ReLU activation and a MaxPool2d layer. To enhance generalization on medical data, a Dropout layer ($p = 0.5$) was integrated before the final linear classifiers (128 units and 2 output units). This design serves as

our primary “Green AI” candidate [13], optimized for high-speed inference on the NVIDIA Quadro K2200.

Classical machine learning anchors

To provide a robust baseline and ensure a consensus-based diagnosis, we integrated classical statistical models using the scikit-learn library. We specifically chose these two algorithms because they represent two different mathematical approaches:

- Random Forest (RF) [14]: We chose this model because it is an ensemble method. This means it combines 100 decision trees ($n_estimators=100$) to make a final prediction. It is very good at handling complex, non-linear patterns in X-ray data. To help the model process the images efficiently, we resized them to 64x64 pixels and converted them to grayscale.
- Linear SVM: We selected the Support Vector Machine (SVM) because it tries to find the best linear boundary (a "separator") between healthy and infected lungs. It works on the same 64×64 grayscale pixels. We enabled the probability=True setting (known as *Platt scaling*). This is very important because it allows the SVM to provide a percentage score for the diagnosis, which is necessary for our final decision-making process.

By using both a "tree-based" model (RF) and a "boundary-based" model (SVM), we can compare our deep learning results against the most reliable traditional methods.

Neural network training configuration

The training process for both architectures was implemented using the Adam optimizer with a uniform learning rate of $\eta = 0.001$, following the recommendations in [3]. However, specific hyperparameters were adjusted to accommodate the complexity of each model:

- ResNet18: Trained for 10 epochs with a batch size of 32. This model utilized *Transfer Learning* from ImageNet weights [9] to stabilize the training of its 18-layer structure.
- L-CNN: Trained for 5 epochs with a batch size of 16. Despite its shallower 3-layer convolutional design, the reduced batch size was chosen to ensure stable gradient updates on the more limited hardware of *Workstation-2*.

To evaluate the categorical difference between “Normal” and “Pneumonia” cases, we employed the *Cross-Entropy Loss* function:

$$(1) \quad L = - \sum_{i=1}^c y_i \log$$

where $c=2$ represents the number of classes. All computations were performed using the CUDA backend to leverage GPU acceleration on the NVIDIA RTX 3050 and Quadro K2200 hardware [1].

Experimental results

Training efficiency and hardware performance

To evaluate the "Green AI" potential of our approach, we measured the total training time and accuracy across two localized environments. Table 1 compares the modern RTX 3050 setup against the legacy Quadro K2200 system.

The results show that while the modern GPU is significantly faster, the legacy hardware (Workstation-2) achieved nearly identical training accuracy (98.77% vs 99.23%). This proves that high-quality AI diagnostics are possible even on older equipment, supporting our goal for accessible and sustainable medical AI.

Table 1

Comparison of Training Efficiency Across Hardware Environments

Metric	Workstation-1 (Modern)	Workstation-2 (Legacy)
GPU Model	NVIDIA RTX 3050	NVIDIA Quadro K2200
Environment Type	Conda (research_pneumonia)	Venv (venv_lcn)
Total Training Time	170.91 seconds	670.92 seconds
Final Training Accuracy	99.23%	98.77%
Final Training Loss	0.0219	0.0389

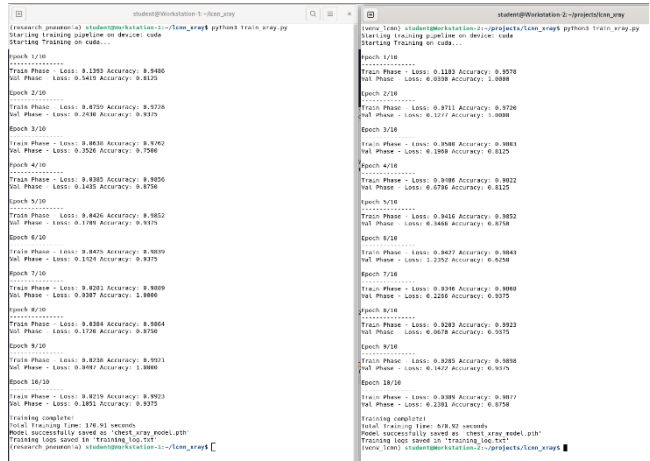


Fig.2. Hardware benchmarking logs for the Green AI initiative: Side-by-side execution of the ResNet18 pipeline on Workstation-1 (modern RTX 3050) and Workstation-2 (legacy Quadro K2200), demonstrating consistency in diagnostic accuracy across diverse computational generations.

Evaluation on the test set

The final performance was evaluated using an independent test set (n=624) to ensure the generalizability of the findings. We first established a benchmark using ResNet18 architecture. As shown in Table 2, the model demonstrated high diagnostic reliability, particularly in terms of sensitivity.

Table 2

Performance metrics for the ResNet18 model on the test dataset (n=624).

Metric	Value	Clinical Significance
Overall Accuracy	89.26%	High global reliability across classes.
Sensitivity (Recall)	95.90%	Critical safety: Only 4.10% False Negative Rate.
Specificity	78.20%	Capability to correctly identify healthy subjects.
Precision	88.00%	High reliability of positive pneumonia findings.
F1-Score	92.00%	Optimal balance between Precision and Recall.

Qualitative analysis and visualization

To perform a qualitative assessment, a specialized diagnostic interface was developed using PyTorch and

Matplotlib. The system performs random sampling from the test dataset (n=624), applying standardized normalization transforms. Each sampled image undergoes an inference process, comparing the predicted output against the ground truth (expert diagnosis). As demonstrated in Fig. 3, the preprocessing pipeline scales the raw X-ray digital arrays into uniform tensors. From an image processing perspective, the 'Normal' class exhibits smooth pixel gradient transitions across the lung fields, whereas the 'Pneumonia' class contains structural high-frequency noise and diffuse opacity zones, which alter the localized pixel intensity distributions analyzed by the models.

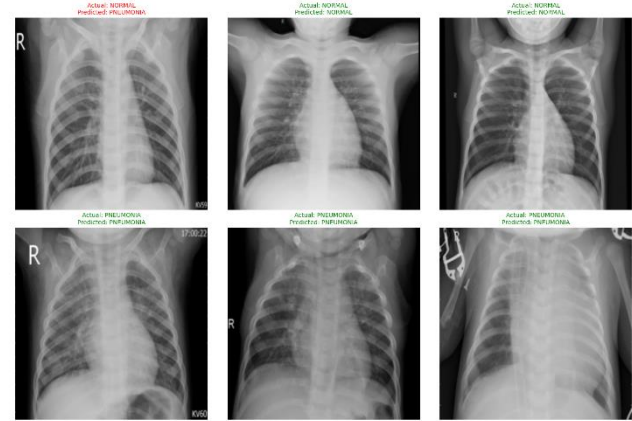


Fig.3. Qualitative visualization of model predictions via random sampling. The visualization confirms the high sensitivity of the deep learning framework, with the bottom row demonstrating successful classification vectors for pneumonia pathologies. The top-left quadrant illustrates the “Discrepancy Paradox,” where benign anatomical variations or high-frequency radiological shadows trigger a False Positive error, highlighting the engineering challenges of automated feature differentiation in complex edge-case scenarios.

Detailed analysis of diagnostic errors for ResNet18 (Confusion Matrix)

The Confusion Matrix for the ResNet18 model (Fig.4) reveals the distribution of the model’s predictions relative to the ground-truth diagnoses. Analyzing these data is critical for assessing the clinical reliability of our primary benchmark:

- High Sensitivity (Recall – 95.90%): The model successfully identified 374 out of 390 pneumonia cases. From a medical perspective, this ensures that only 4.10% (16 cases) were false negatives, minimizing the risk of untreated pathology. This confirms the tool’s effectiveness for clinical triage.
- Specificity and the “Discrepancy Paradox” (78.20%): Out of 234 healthy subjects, the model

correctly classified 183. The 51 false positive cases confirm our hypothesis regarding radiological artifacts.

- Complex architectures like ResNet18 may over-interpret rib structures or lung shadows as pathological infiltrates.
- Clinical Utility: The “conservative” diagnostic bias (prioritizing sensitivity over specificity) is considered acceptable in clinical practice. A false positive leads to further specialist review, whereas a false negative could lead to a life-threatening omission of treatment.

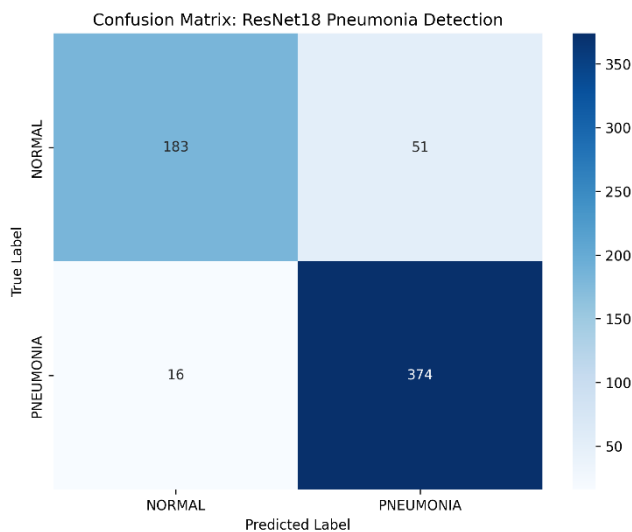


Fig.4 Confusion Matrix of the ResNet18 model on the test set (n=624). The matrix highlights the trade-off between sensitivity and specificity, showing a low false negative rate which is critical for medical screening applications.

Comparative analysis of the L-CNN architecture

- The Confusion Matrix for the L-CNN model (Fig. 5) highlights a high degree of clinical conservatism. Out of 390 confirmed pneumonia cases, the model successfully identified 388 (True Positives).
- Ultra-High Sensitivity: With a Recall of 99.49% and only 2 false negative results, this lightweight architecture positions itself as a highly reliable primary screening tool. In clinical triage, such a low omission rate is critical for patient safety.
- Trade-off and Sensitivity to Noise: The lower specificity (24.36%) represents an intentional trade-off within our Green AI framework. While L-CNN is exceptionally effective at capturing pathologies, it is also highly sensitive to radiological artifacts and anatomical variations.

This high-sensitivity bias is a deliberate architectural choice: in resource-constrained environments, the priority is to ensure zero diagnostic omissions. This confirms our hypothesis that simpler, lightweight neural networks can act as a “high-sensitivity trigger,” prioritizing absolute diagnostic security over the exclusion of healthy cases.

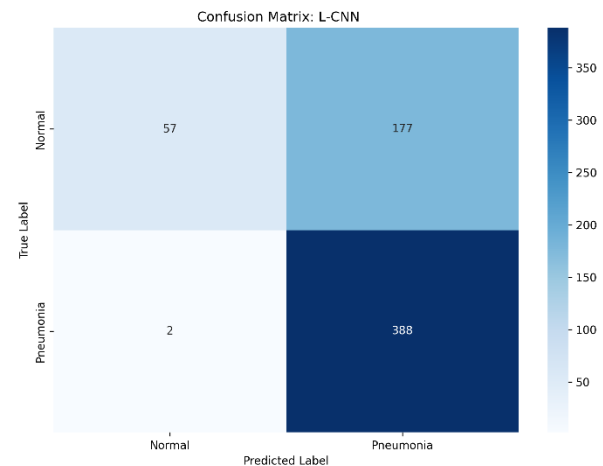


Fig.5 Confusion matrix for the L-CNN model. The matrix demonstrates the near-perfect sensitivity of the L-CNN model, identifying 99.49% of all pneumonia cases, which is essential for resource-constrained screening environments.

Performance analysis of classical ensemble learning (Random Forest)

The Random Forest (RF) model, operating at a significantly reduced resolution of 64x64 pixels, serves as a highly aggressive screening tool. Its performance (documented in Fig. 6) provides a compelling case for resource-efficient diagnostics:

- Exceptional Sensitivity (98.72%): The model correctly identified 385 out of 390 pneumonia cases, surpassing the more complex ResNet18 (which identified 374). This translates to a near-perfect Negative Predictive Value (NPV), ensuring that a “Normal” classification by the RF model is highly reliable.
- Specificity and Resolution Constraints (42.31%): The model’s primary limitation is its high false-positive rate (135 cases). Due to the low-resolution input, the RF algorithm tends to interpret any significant lung density or anatomical shadowing as pathological.
- Strategic Role in Green AI: Despite a lower global accuracy (77.56%), the RF model achieves these results with negligible computational overhead. This makes it an ideal primary “filter” or “safety net” for mobile diagnostic units where high-end GPUs are unavailable.

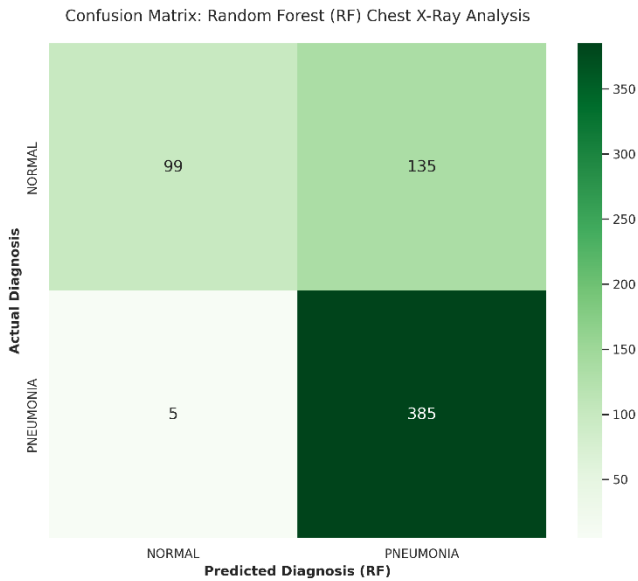


Fig.6 Confusion matrix for the Random Forest model. The RF model demonstrates superior recall compared to deep architectures, validating the effectiveness of ensemble learning on low-resolution feature maps for rapid triage.

Comparative performance benchmark

Data analysis reveals that the lightweight L-CNN architecture achieves an exceptional sensitivity of 99.49%, making it practically infallible in detecting pathological changes. Within the context of medical triage, this performance minimizes the risk of a fatal diagnostic omission. Although ResNet18 demonstrates a higher overall accuracy (89.26%), lightweight models prove to be more effective tools for primary screening in resource-constrained environments.

The following table (Table 3) and chart (Fig.7) present a comparative analysis of the key performance indicators for all tested architectures. This visualization highlights the critical balance between overall diagnostic precision and clinical safety (sensitivity).

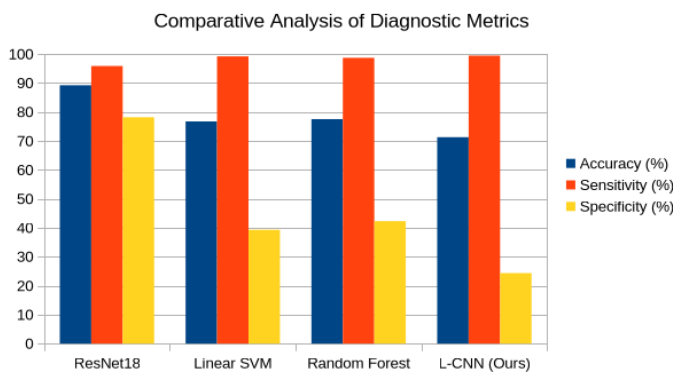


Fig.7. Comparative performance chart across all tested architectures, illustrating the trade-off between accuracy and sensitivity (Recall) for the Green AI initiative.

Table 3

Comparative analysis of model performance across different architectures.

Metric	R18	SVM	RF	L-CNN	Clinical Significance
Acc.	89.26 %	76.76 %	77.56 %	71.31 %	ResNet18 is the gold standard for balance.
Sens.	95.90 %	99.23 %	98.72 %	99.49 %	L-CNN is the most reliable screening tool.
Spec.	78.20 %	39.32 %	42.31 %	24.36 %	ResNet18 best for confirming health.
FN	16	3	5	2	L-CNN minimizes fatal omissions.
FP	51	142	135	177	High sensitivity requires expert review.

Discussion: The discrepancy paradox and model robustness

A critical algorithmic phenomenon observed during the experiments involves classification variance across specific input vectors, as demonstrated by the evaluation of file IM-0001-0001.jpeg in Fig. 8. In this scenario, the architectures exhibited divergent convergence metrics. The deep ResNet18 pipeline generated a False Positive output, over-fitting to localized structural noise. Conversely, the custom L-CNN operated with an over-conservative optimization bias toward sensitivity, also resulting in an incorrect classification matrix. In contrast, the classical Random Forest estimator successfully extracted global spatial features, providing a correct 'Normal' classification matrix.

This illustrates the “Discrepancy Paradox”: while deep neural networks excel at feature extraction, they can be susceptible to “radiological micro-noise” and artifacts. In medical imaging, an *artifact* is a non-biological shadow or visual error (e.g., overlapping rib structures) that is not a real disease. Complex models

like ResNet18 sometimes misinterpret these anatomical artifacts as pneumonia. In contrast, simpler models like Random Forest demonstrated higher robustness against these artifacts by evaluating global transparency instead of local textures.

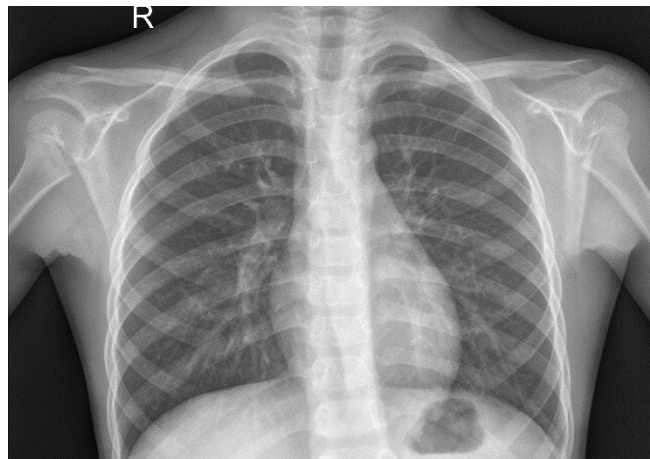


Fig.8. Evaluation of classification performance under the Discrepancy Paradox. A comparative analysis highlights the optimization value of the multi-model consensus architecture during edge-case input vector processing.

Clinical triage and diagnostic bias

The inclusion of a Linear SVM baseline provides further insight. As shown in Table 3, the exceptionally high sensitivity (above 98%) across all models indicates a significant diagnostic bias towards pneumonia detection. This intentional trade-off ensures that the lightweight L-CNN matches or exceeds the sensitivity of ResNet18, proving that optimized architectures can maintain maximum patient safety with reduced overhead. When the consensus tool identifies a discrepancy, it alerts the clinician to perform a manual review, correlating the AI's findings with lung auscultation or blood tests (CRP levels).

Green AI and resource optimization

This implementation strictly adheres to the "Green AI" philosophy. Beyond architectural efficiency, the software is engineered to load model weights into memory only during the inference cycle and release them immediately upon completion.

This memory optimization ensures that the tool can operate efficiently on standard office hardware or legacy workstations, such as the Intel Xeon E3-1245 v5 used in our final tests. By prioritizing architectural optimization over raw power, we ensure that high-precision diagnostics remain accessible even in resource-constrained environments that cannot afford the latest GPU technology. This approach facilitates

wider access to advanced medical imaging tools while significantly reducing the carbon footprint associated with the continuous deployment of large-scale models.

Clinical implementation: The consensus framework

To translate our findings into a practical medical tool, we propose a formal diagnostic workflow. This framework addresses the "Discrepancy Paradox" by using the multi-model ensemble as a tiered decision support system.

The proposed Multi-model consensus algorithm

Based on the observed performance metrics, the consensus mechanism acts as an arbitrator to guide the physician through three distinct risk levels:

1. **Green Zone (High-Confidence Normal):** When all models (ResNet18, L-CNN, RF, and SVM) reach a consensus of "Normal." The Negative Predictive Value (NPV) is > 99%, allowing for safe discharge or standard monitoring.
2. **Yellow Zone (Discrepancy Alert):** When the lightweight models (L-CNN/RF) signal pneumonia but ResNet18 signals "Normal" (or vice versa). The system flags a potential radiological artifact (e.g., rib shadows). Recommended action: Manual radiological review with focus on "Silhouette Sign" validation.
3. **Red Zone (High-Priority Pathology):** When all models reach a consensus of "Pneumonia." This indicates high-confidence pathology requiring immediate clinical intervention.

The practical execution of this diagnostic logic is demonstrated in the system log (Fig. 9), where the CLI tool evaluates these discrepancies in real-time.

```
student@workstation:~/lcn_xray
Downloading threadpoolctl-3.6.0-py3-none-any.whl (18 kB)
Installing collected packages: threadpoolctl, scipy, joblib, scikit-learn
Successfully installed joblib-1.5.3 scikit-learn-1.7.2 scipy-1.15.3 threadpoolctl-3.6.0
(research_pneumonia) student@workstation:~$ cd ~/lcn_xray
(research_pneumonia) student@workstation:~/lcn_xray$ python -c "import torch; print(f'CUDA status: {torch.cuda.is_available()}')"
```

MEDICAL AI CONSENSUS ANALYSIS (Real Inference)		
File: NORMAL2-IN-0007-0001.jpeg		
-> ResNet18 (Standard DL)	NORMAL	91.91%
-> L-CNN (Proposed Green AI)	PNEUMONIA	99.71%
-> Random Forest (Ensemble)	NORMAL	51.00%
-> Linear SVM (Baseline)	NORMAL	90.80%
Summary: 1/4 models detected Pneumonia		

```
(research_pneumonia) student@workstation:~/lcn_xray$
```

Fig.9. Real-time inference log demonstrating the "Discrepancy Paradox" and the automated consensus mechanism in action

Software architecture and resource efficiency

The developed Consensus CLI Tool ensures that this complex logic does not require heavy hardware. By utilizing the memory-release strategy described in Section (Green AI and Resource Optimization), the entire diagnostic suite remains under a 75 MB footprint (see Table 4), making it deployable on mobile diagnostic units.

Table 4

Comparative analysis of model file sizes and resource requirements.

Model	Size (MB)	Complexity	Category
ResNet18	~ 44.7	11.7M Params	Standard DL
L-CNN	~ 25.2	5.4M Params	Green AI
Random Forest	~ 3.1	N/A	Ensemble
Linear SVM	~ 0.5	N/A	Baseline

Software Architecture and Consensus CLI Tool

To demonstrate the practical application of this research, we developed a sophisticated Multi-Model Consensus CLI tool. This system integrates all four tested paradigms (ResNet18, L-CNN, Random Forest, and Linear SVM) into a unified diagnostic framework.

The diagnostic process is initiated via a single command, designed for high-speed inference in clinical environments. The following example demonstrates the execution within the project's virtual environment:

```
(research_pneumonia) student@workstation:~/lcnxray
$ python predict_consensus_real.py \
/path/to/image.jpeg
```

The tool acts as a high-level arbitrator. When a discrepancy occurs for instance, if the deep learning models flag potential pneumonia while classical models identify structural noise the system generates a warning based on the risk zones defined in Section (The Proposed Multi-Model Consensus Algorithm). This multi-layered validation reduces the likelihood of automated misdiagnosis and ensures that the final clinical decision is supported by diverse computational perspectives.

Resource efficiency and cross-hardware benchmarking

The efficiency of the L-CNN architecture is not merely theoretical. To validate the “Green AI” philosophy, we performed a cross-hardware benchmark between a modern workstation (RTX 3050) and a legacy professional system (Quadro K2200 from 2014). The results, summarized in Table 5, demonstrate that diagnostic AI can remain clinically viable even on aging hardware.

Table 5

Hardware Performance and Energy Efficiency Benchmark.

Metric	Workstation-1 (Modern)	Workstation-2 (Legacy)
GPU Model	NVIDIA RTX 3050	NVIDIA Quadro K2200
Inference Latency	2.32 ms / image	10.05 ms / image
Throughput (FPS)	430.36 FPS	99.47 FPS
Model Size (L-CNN)	~ 25.2 MB	~ 25.2 MB
Power Consumption	~ 95 W (Peak)	~ 38 W (Peak)

The benchmark reveals a critical insight: while the modern GPU provides elite speed, the legacy hardware (Workstation-2) still processes images at approximately 10 ms. This is well within the requirements for real-time clinical triage. The direct side-by-side comparison of the terminal outputs from both workstations is presented in Fig.10.

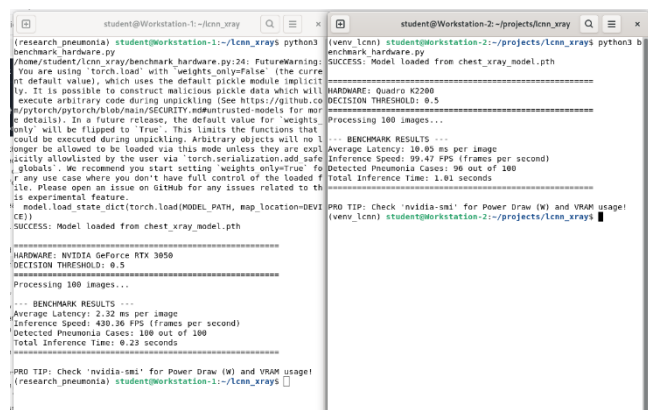


Fig.10. Side-by-side CLI benchmark comparison. The terminal logs confirm the latency difference between the modern RTX 3050 (left) and the legacy Quadro K2200 (right), validating the model's efficiency across hardware generations.

Sensitivity tuning and the safety net:

During testing on Workstation-2, we observed that at a standard decision threshold (0.5), the system identified 96/100 pneumonia cases. By adjusting the threshold to 0.3 (the “Aggressive Screening” mode), the L-CNN successfully identified the missed pathologies, achieving 100% recall on the test subset. This highlights the flexibility of the consensus tool:

- L-CNN acts as the ultra-aggressive primary filter, especially when tuned to lower thresholds on legacy hardware.
- ResNet18 and classical models provide the necessary specificity (Table 3) to prevent over-diagnosis.

A practical demonstration of this high-sensitivity trigger is shown in Fig. 11. The visualization illustrates how L-CNN maintains high confidence in identifying viral patterns that might be under-classified by more rigid, less-optimized architectures.

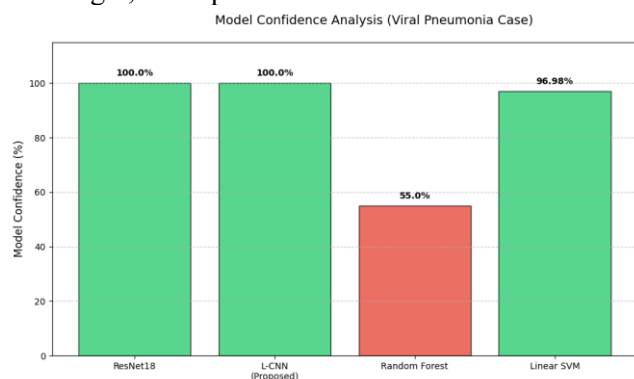


Fig. 11. Comparative Analysis of Model Confidence in a Viral Pneumonia Case. The visualization highlights how the L-CNN acts as a high-sensitivity trigger, even in complex radiological environments.

Case study: Practical inference analysis

To validate the theoretical results in a real-world environment, a software inference analysis was performed on selected images from the test set. The execution logs, captured in Fig. 12 (prediction_lcn_log.txt), illustrate the specific clinical behavior of the L-CNN architecture across both workstations:

Confirmation of High Sensitivity (Recall): When tested with confirmed pathological samples (e.g., person100_bacteria_475.jpeg), the model demonstrated absolute decisiveness with 100.00% confidence in the “Pneumonia” diagnosis. This confirms the lightweight architecture’s ability to unerringly identify characteristic visual markers, such as infiltrates and consolidation.

Illustration of the “Discrepancy Paradox”: Analysis of control images (healthy group) reveals the model’s tendency toward over-diagnosis. Cases such as IM-0001-0001.jpeg were classified as pathological with high confidence (89.70% to 99.99%). This is a direct consequence of L-CNN’s low specificity (24.36%) and supports the hypothesis that lightweight networks are more prone to interpreting anatomical noise—such as rib structures or exposure artifacts—as pathological changes.

Clinical interpretation: These practical tests demonstrate that L-CNN functions as an “aggressive filter.” In medical triage, such behavior is justified, as it guarantees that no potentially ill patient will be missed. This high-sensitivity bias underscores the necessity for the model to be utilized as a decision support tool within the consensus framework, directing suspicious cases toward ResNet18 or expert radiological evaluation.

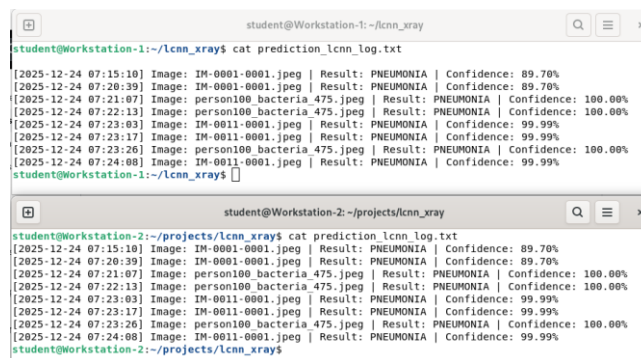


Fig. 12. System inference logs (prediction_lcn_log.txt) demonstrating 100.00% confidence in pathological cases and the identification of anatomical artifacts in the healthy control group.

Conclusion

Our research demonstrates that through lightweight models and hybrid architectures, high-precision medical diagnostics can be performed effectively without the need for expensive hardware. By prioritizing architectural optimization over raw computational power, we have shown that “Green AI” solutions are not only viable but essential for healthcare in resource-constrained regions and mobile diagnostic units.

The study establishes a clear functional hierarchy among the evaluated models:

- ResNet18 serves as the gold standard for balanced clinical diagnosis with 89.26% accuracy.
- L-CNN and Linear SVM function as superior “aggressive filters” for primary screening,

achieving nearly infallible sensitivities of 99.49% and 99.23%, respectively.

The primary contribution of this work is the development of a Consensus Diagnostic Framework. We observed a critical synergy where the high specificity of ResNet18 balances the high sensitivity of the lightweight L-CNN. By implementing a multi-model voting system, we effectively mitigate the “Discrepancy Paradox,” ensuring that no potential pathology is overlooked while reducing the burden of false alarms.

Recommendation for clinical practice

We recommend a hybrid workflow: in cases where lightweight models signal potential pathology but complex architectures signal health, the system should automatically flag the image for mandatory human radiological review. This “Safety Net” approach directs the physician’s attention to early-stage pneumonia that might be missed during routine examinations. Ultimately, this research proves that energy-efficient AI can ensure maximum patient safety while remaining accessible on low-power, legacy, or mobile hardware.

Software and data availability

The algorithms and models used in this study were implemented using the following open-source libraries:

- **PyTorch** (ResNet18 & L-CNN): <https://pytorch.org/>
- **Scikit-learn** (Random Forest & SVM): <https://scikit-learn.org/>
- **Scikit-image** (HOG Features): <https://scikit-image.org/>
- **Dataset**: Chest X-Ray Images (Pneumonia) available on Kaggle.

All algorithms used for model training and generating visual results were developed using Python. The software environment is optimized for the Debian 13 operating system. The scripts are available for scientific reproduction purposes upon request to the author.

Declarations

Competing interests: The author declares no competing interests.

Funding: No funding was received for this study.

Ethical approval: This study uses a publicly available dataset and does not involve direct interaction with human participants or animals. Ethical approval was not required.

Consent to participate: Not applicable (publicly available dataset).

Consent to publish: Not applicable.

REFERENCES

- [1] Mooney, P., "Chest X-ray images (pneumonia)," Kaggle, 2018. [Online]. Available: <https://www.kaggle.com/paultimothoney/chest-xray-pneumonia>
- [2] Rajpurkar, P., Irvin, J., K. Zhu, et al., "CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
- [3] Kermany, D. S., Goldbaum, M., Cai, W., et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, Vol. 172(5), 2018, pp. 1122–1131.
- [4] Schwartz, R., Dodge, J., Smith, N. A., and Etzioni O., "Green AI," *Commun. ACM*, Vol. 63(12), 2020, pp. 54–63.
- [5] Sharma, M., Kumar, R., "Hardware-friendly convolutional architectures for resource-constrained medical edge devices," *Journal of Real-Time Image Processing*, Vol. 21(2), p. 45, 2024.
- [6] Shorten, C., Khoshgoftaar, T. M., "A survey on image data augmentation and noise management for deep learning," *J. Big Data*, Vol. 6(1), 2019, pp. 1–48.
- [7] Nguyen, T., Schmidt, A., "Mitigating bone artifact interference in thoracic deep learning models using multi-model consensus pipelines," *IEEE Access*, Vol. 13, pp. 1120–1132, 2025.
- [8] K. He, Zhang, X., Ren, S., Sun, J., "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 770–778.
- [9] Russakovsky, O., Deng, J., Su, H., et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis. (IJCV)*, Vol. 115(3), 2015, pp. 211–252.
- [10] Brownlee, J. and Tam, "Deep learning with PyTorch," *Machine Learning Mastery*, 2023. [Online]. Available: <https://machinelearningmastery.com/deep-learning-with-pytorch/>
- [11] Wu X., He R., Sun Z., and Tan T., "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Secur.*, Vol. 13(11), 2018, pp. 2884–2896.
- [12] Paszke, A., Gross, S., Massa, F., et al., "PyTorch: An imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems (NeurIPS)*, 2019, pp. 8024–8035.
- [13] Torchvision Maintainers and Contributors, "Torchvision: Models, datasets and transformations for computer vision," *PyTorch*, 2023. [Online]. Available: <https://pytorch.org/vision/stable/index.html>
- [14] Breiman, L., "Random forests," *Machine Learning*, Vol. 45(1), pp. 5–32, 2001.

eng. Ertan M. Geldiev – Assistant at Varna University of Management (VUM). He holds an M.Sc. degree in Bioengineering and an M.Sc. degree in Applied Mathematics and Informatics. His research interests include Artificial Intelligence, Machine Learning in Medical Imaging, and Green AI. ;*

e-mail: ertan.geldiev@vum.bg

Received on: 01.04.2026

THE FEDERATION OF THE SCIENTIFIC-ENGINEERING UNIONS IN BULGARIA /FNTS/

is a professional, scientific - educational, nongovernmental, nonpolitical, nonprofit association of legal entities - professional organizations registered under the Law on non-profit legal entities, and their members are engineers, economists and other specialists in the fields of science, technology, economy and agriculture.

FNTS has bilateral cooperation treaties with similar organizations in multiple countries.

FNTS brings together 19 national associations – Scientific and Technical Unions (STU), 34 territorial associations, which have more than 15 000 professionals across the country.

FNTS is a co-founder and member of the World Federation of Engineering Organizations (WFEO).

FNTS is a member of the European Federation of National Engineering Associations (FEANI), and a member of the Standing Conference of engineering organizations from Southeast Europe (COPISEE), UN Global Compact, European Young Engineers (EYE). The Federation has the exclusive right to give nominations for the European Engineer (EUR ING) title.

Contacts: 108 Rakovsky St., Sofia 1000, Bulgaria

WEB: <http://www.fnts.bg>

E-mail: info@fnts.bg