透過遷移式學習偵測肺炎：應用貝式卷積神經網路分類器於胸腔 X 光影像之研究

[1*]李宗儒、[1,2]馬悅伶

[1]南臺科技大學資訊工程系，[2]菲律賓東南大學資訊電腦系

*michaelee@stust.edu.tw

摘　要

本研究聚焦於利用貝式卷積神經網路（Bayesian Convolutional Neural Networks, BCNN）在胸腔 X 光影像分類中的應用，重點探討肺炎檢測。研究採用遷移式學習和模型微調技術，將原先為多類別分類開發的預訓練 BCNN 模型，專門調整用於肺炎的二分類檢測。透過在新的胸腔 X 光影像資料集上進行微調，模型在測試資料上成功達到 91.19% 的準確率，優於 VGG16（86.54%）、InceptionV3（87.98%）和 ResNet50（74.04%）。此外，BCNN 在 AUC（0.98）、精確率（precision）、召回率（recall）與 F1-score 方面均展現卓越性能，證實其在區分正常（Normal）與肺炎（Pneumonia）病例方面的穩健表現。展現在醫學影像診斷中的顯著潛力。遷移式學習為模型帶來多重優勢：顯著降低運算成本，加速模型收斂，並保留預訓練資料中的關鍵特徵。這種知識遷移使模型能有效適應醫學影像的特定任務，即使在有限的資料集下，依然能維持高準確性和良好的通用性。貝式框架的導入進一步提升了模型的可靠性，利用信心值估算的方法，有效減少過度擬合的風險，這在醫療應用中尤為重要。為實現臨床實際應用，研究團隊將分類模型部署於基於 Flask 的網頁應用程式，實現從胸腔 X 光影像的即時肺炎檢測。此舉使醫療專業人員可將模型作為診斷輔助工具，提供快速且可靠的分析，從而協助臨床決策。研究結果充分展示了深度學習技術，特別是結合貝式卷積神經網路與遷移式學習方法在醫學影像領域的潛力，為提升診斷準確性和效率提供了嶄新的技術路徑。

關鍵詞：遷移式學習、貝式卷積神經網路、微調、胸腔 X 光影像

Efficient Pneumonia Detection via Transfer Learning: A Bayesian CNN Approach to Chest X-Ray Classification

[1*]Tsung-Lu Michael Lee, [1,2]Kyra Jyne A. Melendres

[1]Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, [2]College of Information and Computing, University of Southeastern Philippines

Abstract

This study explores the application of Bayesian Convolutional Neural Networks (BCNN) in chest X-ray image classification, specifically for pneumonia detection. Employing transfer learning and fine-tuning, we adapted a pre-trained BCNN model—originally designed for multi-class classification—to perform binary pneumonia classification. The model achieved a highest accuracy of 91.19%, outperforming VGG16 (86.54%), InceptionV3 (87.98%), and ResNet50 (74.04%). Additionally, BCNN demonstrated superior AUC (0.98), precision, recall, and F1-score, confirming its robust performance in distinguishing between NORMAL and PNEUMONIA cases. Transfer learning provided multiple advantages, including reduced computational costs and accelerated model convergence by leveraging pre-trained feature representations. The Bayesian framework further enhanced model reliability by providing uncertainty estimates and mitigating overfitting risks, which is crucial in

Received: Nov. 11, 2024; first revised: Feb. 6, 2025; second revised: Mar. 6, 2025; accepted: Mar. 2025.
Corresponding author: T.-L.M., Lee, Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan 710301, Taiwan.

medical applications with limited labeled data. We deployed the classification model in a Flask-based web application, enabling real-time pneumonia detection and providing healthcare professionals with an effective diagnostic support tool. This research highlights the promising role of deep learning technologies, particularly BCNN with transfer learning, in improving diagnostic accuracy, efficiency, and clinical usability. Future work could extend this approach to multi-class disease classification and explore its applicability in other medical imaging modalities.

Keywords: Transfer Learning, Bayesian CNN, Fine-tuning, Chest X-ray images

I. Introduction

Deep learning has emerged as a transformative technology in medical imaging analysis, offering unprecedented capabilities in automated diagnosis and clinical decision support. Deep learning has become a state-of-the-art artificial intelligence technology due to its promising performance across multiple disciplines [1]. In the field of radiology, particularly chest X-ray interpretation, artificial intelligence applications are addressing critical challenges faced by healthcare providers. Traditionally, interpreting chest X-rays has relied on human radiologists manually reviewing and analyzing images, a process that is prone to interobserver variability and time restrictions. The application of deep learning models offers a transformative solution by enabling automated image analysis and classification. This innovation has the potential to significantly improve efficiency and accuracy in disease diagnosis [2]. In this study, the BCNN model of our current thesis project which is multi-class crack classification will be utilized and implemented in our new dataset through transfer learning and fine-tuning. The new model will be deployed in a web application for chest X-ray images Pneumonia classification.

II. Background and Literature Review

Medical image processing (MIP) is a crucial field that involves the development of algorithms and software for extracting and analyzing information from medical images. It plays a significant role in the diagnosis, treatment, and early detection of diseases, and has seen advancements through the application of deep learning and machine learning techniques [3]. Bargagna et al [4] developed a Bayesian convolutional neural network (BCNN) framework for classifying cardiac amyloidosis (CA) subtypes. The Bayesian model exhibited a comparable performance of 78.28% test accuracy to deterministic models. Furthermore, the Bayesian approach provided confidence estimates, enhancing classification reliability, and mitigating issues observed in deterministic models. The findings suggest that Bayesian CNNs hold promise in overcoming challenges associated with data scarcity in medical imaging classification tasks. A study [5] also utilizes the BCNN model for classifying chest X-ray images.

III. Methodologies

1. Dataset

This study utilizes a publicly available chest X-ray dataset sourced from Data Mendeley[6]. The dataset includes RGB images of both normal and pneumonic cases, encompassing viral and bacterial infections. All images were resized to 227x227 pixels for consistent processing. The Normal and Pneumonic Chest X-rays example images are shown in Figure 1 and Figure 2. The dataset was subsequently divided into training, validation, and test sets comprising 4185, 1047, and 624 images respectively, ensuring an approximately equal distribution of normal and pneumonic cases in each subset. This balanced distribution across the subsets is crucial for training robust and unbiased machine learning models for accurate pneumonia detection. Here's a concise summary:

To address challenges in medical image classification, we implemented comprehensive data preprocessing and augmentation strategies. The approach included multiple data transformation techniques for the training set: image rotation (±20°), translation (±20%), shearing (20%), zooming (±20%), and horizontal flipping. These methods help expand dataset diversity, reduce overfitting, and enable the model to learn more robust features. Pixel values were normalized to the [0, 1] range to improve compatibility with activation functions and optimization algorithms. Validation and test sets were simply normalized without augmentation to maintain real-world data representation. A batch size of 32 was used, with training data shuffling to prevent data order memorization and promote more efficient model learning. The goal was to create a more generalized and accurate model for medical image classification by addressing data variability and model performance challenges.

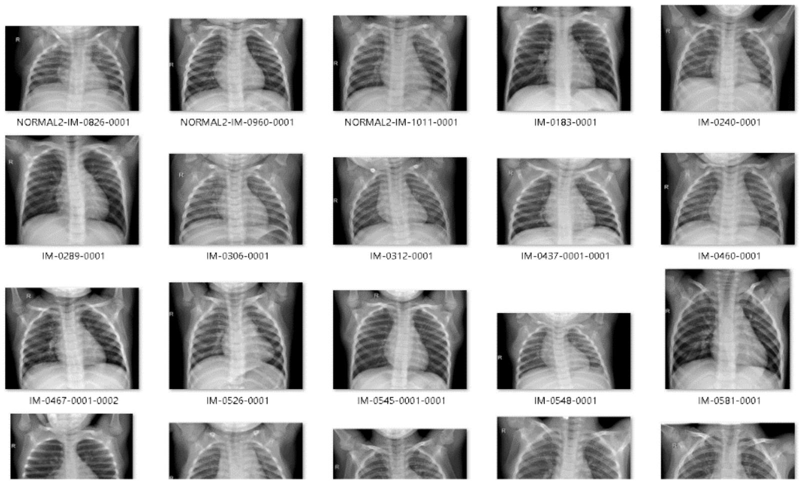

Figure 1. *Normal Chest X-rays example images*

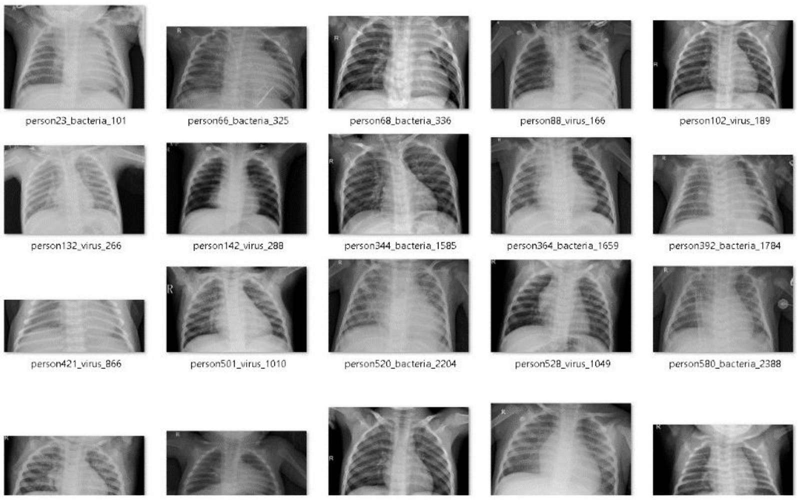

Figure 2. *Pneumonic Chest X-rays example images*

2. Bayesian CNN

The Bayesian Convolutional Neural Network (BCNN) implementation employs a sophisticated architecture that integrates uncertainty quantification through Monte Carlo (MC) Dropout, strategically applied after each convolutional layer before pooling operations [7].

As depicted in Figure 3, our pre-trained model comprises 16 carefully structured layers, with MC Dropout systematically implemented throughout the network, enabling robust probabilistic inference and uncertainty estimation. This architectural design proves particularly advantageous in medical imaging applications, where model confidence quantification is crucial for clinical decision support. The network, initially trained on 150,000 instances, establishes a robust foundation for feature extraction and pattern recognition. The implementation involves carefully transferring the pre-trained BCNN model's weights and biases, followed by systematic parameter adjustments to optimize performance for chest X-ray analysis. This adaptation process is particularly advantageous as it enables the model to leverage low-level feature detectors (such as edge and texture patterns) learned from the original dataset while fine-tuning higher-level features specific to medical image characteristics. The combination of MC Dropout's regularization effects and the model's deep architecture enables reliable uncertainty estimation in predictions, a critical feature for medical diagnostic applications where understanding model confidence is as important as the prediction itself.

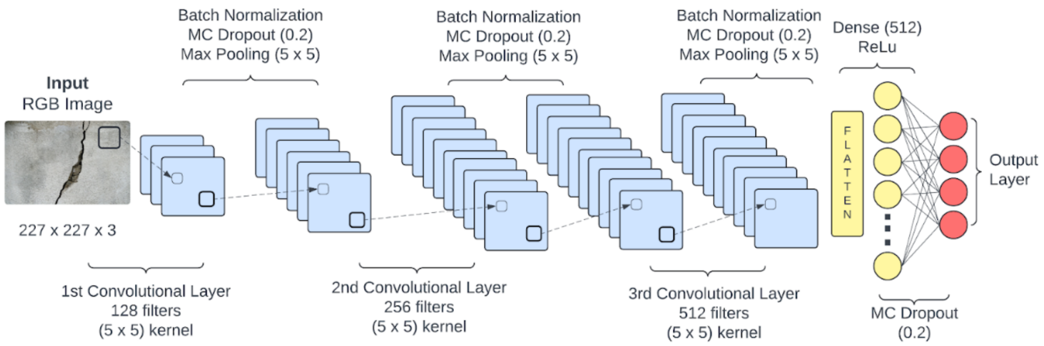

Figure 3. *BCNN Model Architecture*

3. Transfer Learning

Our approach implements a transfer learning methodology to leverage the robustness of a pre-trained BCNN model, originally trained on an extensive dataset of 150,000 instances, for the specific task of chest X-ray classification. This strategic application of transfer learning significantly optimizes computational resources and accelerates model convergence by retaining the learned feature hierarchies from the source domain while adapting them to the target medical imaging domain [8]. The process not only reduces the computational overhead typically associated with training deep learning models from scratch but also improves the model's generalization capabilities by building upon previously learned robust feature representations. Figure 4 illustrates how this pre-trained architecture is leveraged through transfer learning, where the model's learned representations are adapted for chest X-ray classification while maintaining the Bayesian characteristics that address overfitting concerns and enhance performance on limited medical datasets. However, the last dense layer was removed to add a binary dense layer for chest x-ray prediction.

```
[ ] xray_model = Sequential()

[ ] # Copy the convolutional layers from BCNN Multiclass
    for layer in bcnn_model.layers[:-1]:  # Exclude the last dense layer
        xray_model.add(layer)

[ ] xray_model.add(Dense(2, activation='softmax', name='binary_output'))
```

Figure 4. *Implementation of Transfer Learning*

4. Fine-tuning

The fine-tuning phase of our methodology addresses the domain shift between the original training data and the target chest X-ray dataset through a strategic layer unfreezing approach. As illustrated in Figure 5, our fine-tuning architecture selectively unfreezes the last nine layers of the network while maintaining the learned features in earlier layers, a decision empirically determined through extensive experimentation with various layer configurations.

```
[ ] # Unfreeze layers for fine-tuning
    for layer in xray_model.layers[-9:]:
        layer.trainable = True
```

```
[ ] # Compile the model
    xray_model.compile(optimizer=tf.keras.optimizers.Adam(learning_rate=1e-5),
                       loss='sparse_categorical_crossentropy',
                       metrics=['accuracy'])
```

```
[ ] # Set path to your X-ray dataset zip file
    xray_zip_path = '/content/drive/MyDrive/AI/chest_xray.zip'
    extracted_path = '/content/chest_xray'
```

Figure 5. *Implementation of fine-tuning architecture*

This selective unfreezing strategy optimally balances the preservation of fundamental feature extractors from the pre-trained model while allowing sufficient adaptability for medical image-specific features. The approach is particularly effective as the earlier layers typically capture generic features (such as edges and textures) that remain valuable across domains, while the unfrozen later layers can adapt to the specific characteristics of chest X-ray images [9]. The fine-tuned architecture maintains the robust foundational feature extraction capabilities of the original model while enabling specialized learning for medical image classification, resulting in a model that effectively combines transfer learning benefits with domain-specific optimization. This carefully calibrated balance between frozen and trainable layers proved crucial in achieving optimal performance in chest X-ray classification tasks while maintaining computational efficiency.

5. Web Application Deployment using Flask

The deployment phase leverages Python Flask [10], a powerful micro-web framework chosen for its lightweight architecture and seamless integration capabilities with machine learning models. The implementation, as shown in Figure 6, demonstrates the Flask deployment, incorporating essential routing mechanisms, image processing utilities, and model prediction integration.

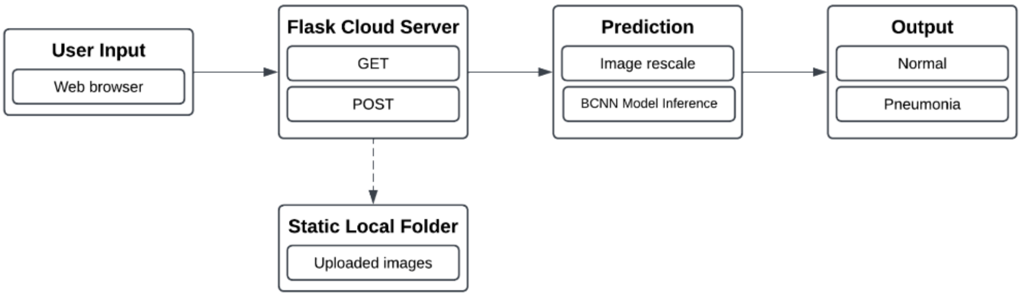

Figure 6. *System Architecture*

The deployment architecture features a responsive web interface, as shown in Figure 7, that enables healthcare professionals to upload chest X-ray images through an intuitive portal, while implementing secure file handling in a designated 'static' directory. Upon image upload, the system executes real-time preprocessing and model inference, delivering immediate visual feedback through a clear, clinically relevant format.

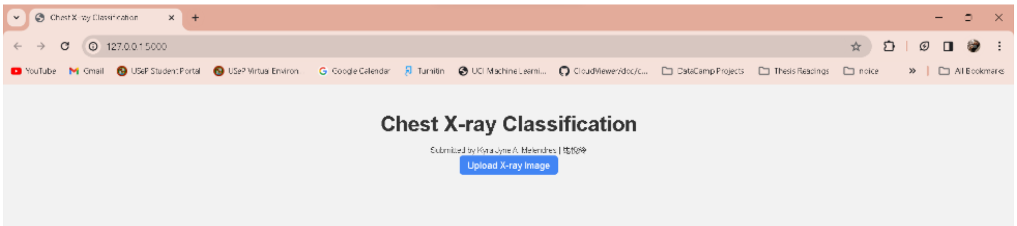

Figure 7. *The interface of the web portal*

Figure 8 showcases the application's prediction interface, where both the uploaded X-ray and corresponding classification results are displayed simultaneously. Testing with previously unseen chest X-ray images from our test dataset demonstrated the system's practical reliability, successfully maintaining consistent performance in classifying between normal and pneumonic cases. This Flask-based deployment architecture proved robust and efficient, effectively translating our BCNN model's capabilities into a production-ready medical imaging tool suitable for clinical environments.

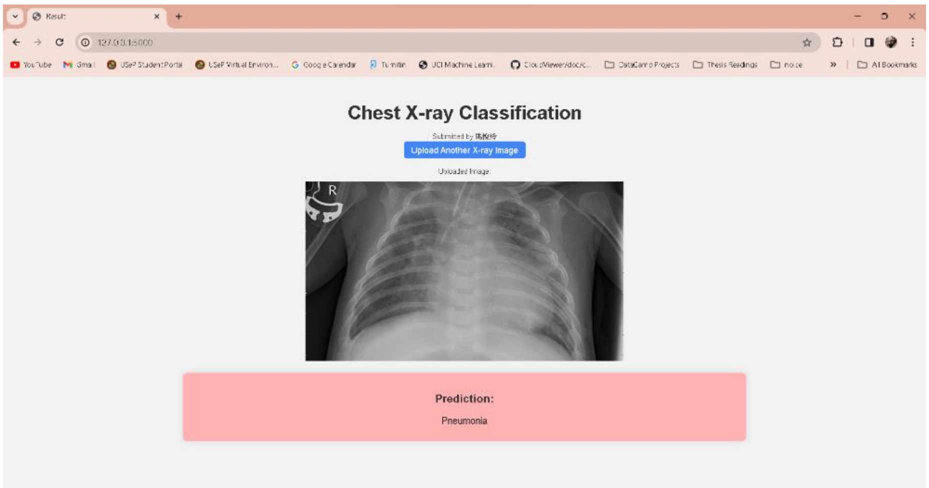

Figure 8. *The web application's prediction interface*

IV. Results and discussions

This section presents a comprehensive analysis of the empirical results obtained from applying BCNN transfer learning and fine-tuning to chest X-ray classification. The narrow gap between training and validation losses is especially significant, as it indicates robust generalization capabilities and effective regularization through the Bayesian framework. Figure 9, which presents the Comparison of ROC Curves, further reinforces the model's strong performance. BCNN achieves the highest Area Under the Curve (AUC) value of 0.98, outperforming VGG16 (AUC = 0.95), InceptionV3 (AUC = 0.97), and ResNet50 (AUC = 0.82). The superior AUC score highlights BCNN's exceptional ability to distinguish between classes, with minimal false positives and false negatives. This behavior strongly suggests that the model has achieved an optimal balance between learning the underlying patterns in the chest X-ray data and maintaining generalization capability, effectively avoiding the common pitfall of overfitting. The consistent performance across different evaluation metrics validates the effectiveness of our fine-tuning strategy and supports the model's potential reliability in clinical applications.

A detailed evaluation of precision, recall, and F1-score (as shown in Table 1) further supports these findings. BCNN demonstrates the highest classification performance, with a precision of 0.972 for NORMAL cases and 0.869 for PNEUMONIA cases. Its recall values of 0.752 (NORMAL) and 0.987 (PNEUMONIA) indicate strong sensitivity to positive cases, leading to an overall F1-score of 0.848 for NORMAL and 0.924 for PNEUMONIA. InceptionV3 and VGG16 also achieve competitive results, with InceptionV3 yielding an F1-score of 0.818 for NORMAL and 0.910 for PNEUMONIA, while VGG16 attains 0.794 and 0.900, respectively. ResNet50, however, lags behind with an F1-score of 0.599 for NORMAL and 0.808 for PNEUMONIA, suggesting lower overall effectiveness in classification. These results further validate BCNN's superior performance, demonstrating its effectiveness in distinguishing between NORMAL and PNEUMONIA cases with high precision and recall. The model's robust performance across multiple evaluation metrics reinforces its potential for reliable deployment in clinical applications.

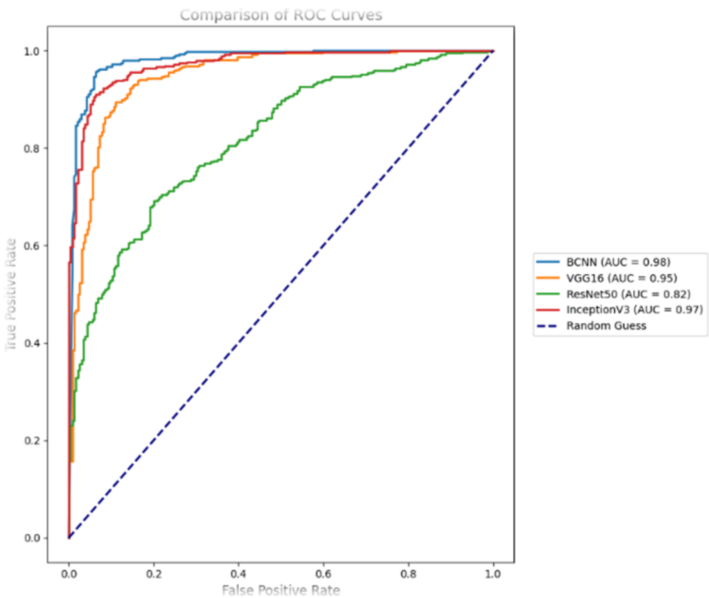

Figure 9. *Comparison of the ROC Curves and AUC*

Table 1. *This table shows the precision, recall, and F1-score of the compared models.*

		BCNN	
	Precision	Recall	F1-Score
NORMAL	**0.972**	**0.752**	**0.848**
PNEUMONIA	**0.869**	**0.987**	**0.924**
		ResNet50	
	Precision	Recall	F1-Score
NORMAL	0.712	0.517	0.599
PNEUMONIA	0.751	0.874	0.808
		InceptionV3	
	Precision	Recall	F1-Score
NORMAL	0.949	0.718	0.818
PNEUMONIA	0.852	0.977	0.910
		VGG16	
	Precision	Recall	F1-Score
NORMAL	0.931	0.692	0.794
PNEUMONIA	0.840	0.969	0.900

Table 2 presents the overall model accuracy on the test set, further demonstrating BCNN's superior performance. BCNN achieves the highest accuracy at 91.19%, outperforming InceptionV3 (87.98%), VGG16 (86.54%), and ResNet50 (74.04%). These results indicate that BCNN consistently provides the best balance between sensitivity and specificity, ensuring high reliability in clinical applications. Inference time analysis revealed that BCNN (0.4604s) was slower than InceptionV3 (0.2120s) and ResNet50 (0.2005s) but still performed within an acceptable range for real-time medical applications. These findings indicate that while BCNN achieves superior accuracy, further optimization of computational efficiency may enhance its practicality in clinical deployment. The combined evaluation of AUC, precision, recall, F1-score, and overall accuracy validates BCNN as the most effective model for chest X-ray classification. Its robust performance across multiple evaluation metrics highlights its potential for real-world deployment in medical imaging diagnostics.

Table 2. *The overall comparison of the model accuracy and inference time on the test set*.

Model	Accuracy (%)	Inference Time (s)
BCNN	91.19	0.4604
VGG16	86.54	0.5897
ResNet50	74.04	0.2005
InceptionV3	87.98	0.2120

Table 3 presents a comparative analysis of Flask Web Inference performance focusing on the average time required to predict an image. In Chest X-Ray datasets, the models—including BCNN, VGG16, ResNet50, and InceptionV3—demonstrate varying prediction speeds. The variation in prediction times highlights the importance of model selection in medical imaging applications, where speed can be crucial for timely diagnostic insights. Different models show different performance characteristics across datasets, suggesting that no single model consistently outperforms others across all scenarios, and the choice of model may depend on specific computational constraints and dataset characteristics.

Table 3. *Comparative Performance of Neural Network Models in Flask Web Inference Speed*.

Datasets	Model	The average time of Flask web Inference (s)
Chest X-Ray	BCNN	0.4604
	VGG16	0.2153
	ResNet50	0.3895
	InceptionV3	0.5229

V. Conclusion

The implementation of Bayesian Convolutional Neural Networks (BCNN) through transfer learning and fine-tuning demonstrates remarkable effectiveness in medical image classification, achieving the highest accuracy of 91.19% on chest X-ray analysis. The model also outperforms other deep learning architectures, including VGG16 (86.54%), InceptionV3 (87.98%), and ResNet50 (74.04%), as evidenced by its superior AUC (0.98), precision, recall, and F1-score. These results confirm BCNN's strong generalization capability and robustness, effectively distinguishing between NORMAL and PNEUMONIA cases with minimal misclassification. This approach offers multiple advantages: substantially reduced computational overhead and training time, successful adaptation across disparate domains, and seamless deployment in a web-based application environment. The model's robust performance in both development and deployment phases, coupled with its computational efficiency, establishes BCNN as a promising framework for medical imaging applications where both accuracy and resource optimization are crucial. Our findings suggest that this methodology provides a scalable and adaptable foundation for various medical imaging tasks, particularly valuable in healthcare settings where reliable and efficient diagnostic tools are essential. Future research could extend this framework to other medical imaging modalities, explore multi-class disease classification, and integrate interpretability techniques to enhance clinical trust and usability. These directions will further validate BCNN's potential for broader clinical applications and real-world deployment in radiology and automated disease diagnosis.

References

[1] Sarker, I.H. (2021). Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions. *SN Comput. Sci.*, *2*(6), 420.

[2] Hussain, A., Amin, S.U., Lee, H., Khan, A., Khan, N.F., & Seo, S. (2023). An automated chest X-ray image analysis for COVID-19 and pneumonia diagnosis using deep ensemble strategy. *IEEE Access*, *11*, 97207–97220.

[3] Abumalloh, R.A., Nilashi, M., Yousoof Ismail, M., Alhargan, A., Alghamdi, A., Alzahrani, A.O., Saraireh, L., Osman, R., & Asadi, S. (2022). Medical image processing and COVID-19: A literature review and bibliometric analysis. *J. Infect. Public Health*, *15*(1), 75–93.

[4] Bargagna, F., De Santi, L.A., Martini, N., Genovesi, D., Favilli, B., Vergaro, G., Emdin, M., Giorgetti, A., Positano, V., & Santarelli, M.F. (2023). Bayesian convolutional neural networks in medical imaging classification: A promising solution for deep learning limits in data scarcity scenarios. *J. Digit. Imaging*, *36*(6), 2567–2577.

[5] Loey, M., El-Sappagh, S., & Mirjalili, S. (2022). Bayesian-based optimized deep learning model to detect COVID-19 patients using chest X-ray image data. *Comput. Biol. Med.*, *142*, 105213.

[6] Kermany, D., Zhang, K., & Goldbaum, M. (2018). Large dataset of labeled optical coherence tomography (OCT) and chest X-Ray images. *Mendeley Data*, V3.

[7] Gal, Y., & Ghahramani, Z. (2016). Bayesian convolutional neural networks with bernoulli approximate variational inference. *International Conference on Learning Representations*, 1–12.

[8] Krishna, S.J., & Kalluri, H. (2019). Deep learning and transfer learning approaches for image classification. *Int. J. Recent Technol.*, 7, 2277–3878.

[9] Fakhar, M. (2024). Unfreezing & transfer learning in deep learning. *Medium*, https://medium.com/@fakhar3534/unfreezing-and-transfer-learning-in-deep-learning-a31ef2ad9e8c.

[10] Flask Documentation Team. (2024). Introduction to Flask-Python for you and me 0.5.beta1 documentation. *Python for you and me*, https://pymbook.readthedocs.io/en/latest/flask.html