Evaluation of CNN Architectures for Kidney Stone Classification in Ultrasound Image

Zuriati[1*], Sriyanto[2], Agiska Ria Supriyatna[3], Nurul Qomariyah[4], Dian Ayu Afifah[5], Zarnelly[6]

[1,3,4,5]*Department of Internet Engineering Technology, Politeknik Negeri Lampung, Indonesia*

[2]*Department of Informatics, Institute of Informatics and Business Darmajaya, Indonesia*

[6]*Department of Information System, State Islamic University of Sultan Syarif Kasim Riau, Indonesia*

*corr-author: zuriati@polinela.ac.id

Abstract - Kidney stone diagnosis requires fast and reliable evaluation, yet ultrasound interpretation still largely depends on clinical expertise. This study evaluates four Convolutional Neural Network (CNN) architectures, VGG16, ResNet50, MobileNetV2, and EfficientNetB0 for classifying kidney ultrasound images into Normal and Stone categories. Using a public dataset of 9,416 images, the models were assessed in terms of predictive performance and computational efficiency. MobileNetV2 achieved perfect classification performance, recording 100% accuracy, precision, recall, and F1-score, while maintaining the smallest parameter size (≈3.6M) and fastest training time (~44 s/epoch). VGG16 and ResNet50 also delivered near perfect accuracy (99.79% and 99.89%) with full recall for Stone cases. In contrast, EfficientNetB0 failed to generalize, yielding only 51.62% accuracy due to severe misclassification of Normal images. These results demonstrate that MobileNetV2 provides the most reliable and efficient solution for ultrasound based kidney stone classification, highlighting its strong potential for practical clinical deployment.

Keywords: CNN architectures; image classification; kidney stone; transfer learning; performance evaluation.

I. INTRODUCTION

Kidney stone disease has become an increasingly prominent contributor to kidney-related morbidity worldwide. Chronic kidney disease affects more than 800 million individuals, more than ten percent of the global population [1], and recent global estimates reported approximately 673 million CKD cases with 1.5 million deaths in 2021 [2]. Epidemiological studies continue to highlight the growing prevalence of kidney stones, with population based research from Iran reporting lifetime prevalence rates between 6% and 20% and clinical observations in China showing similar patterns [3-5]. Dietary factors have also been associated with stone formation, where higher carbohydrate and copper intake contributes to increased risk [6]. These trends emphasize kidney stone disease as a persistent public health concern that demands improved approaches to early detection and monitoring.

Early identification of kidney stones is essential to prevent complications such as obstruction, infection, or progressive renal impairment. However, existing diagnostic practices remain constrained by notable limitations. Laboratory assessments cannot reliably detect structural abnormalities, and radiological procedures including ultrasonography and CT scanning depend heavily on the radiologist's interpretation, introducing inter observer variability and the potential for diagnostic inconsistency [7]. Such limitations underscore the need for automated diagnostic support tools capable of offering objective, reproducible, and accurate clinical assessment.

Rapid advancements in artificial intelligence have accelerated the adoption of deep learning methods for medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated strong performance in diverse applications such as tumor diagnosis, pneumonia detection, diabetic retinopathy, and kidney-related imaging tasks [8], [9]. The widespread success of CNNs has motivated researchers to explore more complex or specialized architectures and optimization strategies tailored to medical imaging requirements.

Recent studies on kidney disease detection have evaluated ensemble networks, hybrid CNN frameworks, model-fusion strategies, and lightweight architectures designed to balance accuracy and computational efficiency [10-15]. Additional research has investigated MRI based staging techniques, interpretable CNN models for CT imaging, and various transfer learning or fine tuning approaches [16-22]. Beyond architectural considerations, several works have emphasized data quality, image resolution, and training strategies as key determinants of CNN generalization capability [23-27]. Recent findings also show that employing lightweight

CNN models together with appropriate resampling and class balancing strategies can substantially improve classification reliability for kidney related imaging tasks [27], highlighting the importance of efficient architectures and balanced data distributions.

Several recent studies have further reported the effectiveness of convolutional neural network based approaches for disease classification and medical image analysis, reinforcing the potential of deep learning models for reliable diagnostic support in clinical imaging applications [28].

Despite intensive research activity, comparative evaluations of different CNN architectures using ultrasound, rather than CT or MRI, remain limited. Ultrasound imaging is the most common modality for kidney evaluation due to its accessibility, non invasiveness, safety, and low cost. However, many existing CNN studies continue to rely on CT or MRI data, which may not reflect real world clinical constraints. Moreover, past ultrasound based studies typically focus on predictive accuracy alone, whereas performance stability, convergence behaviour, misclassification patterns, and computational efficiency are rarely examined together. This lack of an integrated performance efficiency analysis represents a substantial gap in the current literature, especially considering the increasing need for deployable models in resource constrained clinical environments.

To address these limitations, the present study provides a thorough comparative evaluation of four widely used CNN architectures VGG16, ResNet50, MobileNetV2, and EfficientNetB0 for kidney stone classification using ultrasound images. Unlike prior works that emphasize accuracy alone, this research incorporates quantitative metrics, training dynamics, misclassification behaviour, and computational indicators such as parameter size and training time. By analyzing both predictive performance and practical efficiency, the study offers a more realistic and clinically meaningful assessment of model suitability.

Furthermore, the study relies exclusively on ultrasound images, enabling the experimental findings to align closely with routine diagnostic workflows. This design choice enhances the clinical relevance of the results and supports the development of lightweight and reliable decision-support systems suitable for real-world implementation. Ultimately, this research contributes a structured, evidence-driven comparison that helps clarify the strengths, limitations, and deployability potential of four well-established CNN architectures in the context of kidney stone classification.

II. METHOD

The research workflow is outlined in Fig. 1, consisting of dataset preparation, preprocessing, model development, evaluation, and comparative analysis. Four established Convolutional Neural Network (CNN) architectures: VGG16, ResNet50, MobileNetV2, and EfficientNetB0 were examined using a transfer learning approach to enable efficient feature extraction from ultrasound images.

A. Dataset Collection

The dataset employed in this study was sourced from the publicly available Kidney Stone Classification and Object Detection dataset on the Kaggle platform [29]. It comprises 9,416 ultrasound images, including 4,414 Normal and 5,002 Stone cases. For model development, the dataset was divided into 80% for training (7,533 images) and 20% for validation (1,883 images) using a fixed random seed (123) to ensure reproducibility. Representative samples of the two classes are shown in Fig. 2.

B. Dataset Preprocessing

All images were resized to 224 × 224 pixels to ensure compatibility with the CNN architectures. Pixel intensity values were normalized to the [0,1] range to support stable optimization. To improve generalization, augmentation techniques including horizontal flipping, slight rotations, and limited zooming were applied dynamically to the training subset. The validation subset underwent only resizing and normalization to maintain an unbiased evaluation protocol.

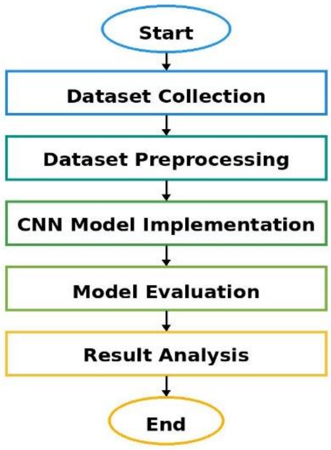


Fig. 1 CNN workflow for kidney ultrasound classification

C. CNN Model Implementation

Each architecture was initialized with ImageNet pretrained weights through transfer learning. A unified classification head was added, comprising a global average pooling layer, dense layers with ReLU activation, dropout regularization, and a softmax output layer. Early convolutional layers were frozen to retain general visual feature representations, while upper layers were fine tuned to learn ultrasound specific characteristics. All models were trained under a consistent optimization setup to ensure fair comparison.

D. Model Evaluation

Model performance was assessed using standard metrics: accuracy, precision, recall, and F1-score. Discriminative capability across thresholds was evaluated using ROC-AUC. Confusion matrices were analyzed to observe misclassification patterns between the Normal and Stone classes. Training and validation curves were examined to assess convergence behaviour, learning stability, and potential overfitting.

E. Comparative Analysis

The comparative analysis integrated quantitative metrics with computational indicators, including parameter size and training time. This holistic evaluation enabled identification of the most suitable architecture by balancing predictive accuracy, computational efficiency, and misclassification tendencies in the context of ultrasound-based kidney stone classification.

III. RESULT AND DISCUSSION

This section presents the main findings of the study, covering dataset characteristics, model training behaviour, quantitative evaluation, and comparative analysis across the four CNN architectures.

A. Training Performance

The learning curves in Fig. 3 and Fig. 4 show that VGG16, ResNet50, and MobileNetV2 achieved rapid and stable convergence during fine-tuning. MobileNetV2 and VGG16 reached high accuracy after only a few epochs, with validation loss stabilizing near zero, indicating excellent generalization. ResNet50 exhibited temporary fluctuations but ultimately converged to high performance in later epochs. EfficientNetB0 showed unstable patterns, with inconsistent accuracy and fluctuating loss values, suggesting failure to adapt effectively to the dataset.

(a)

(b)

Fig. 2 Sample ultrasound images of two classes (a) normal kidney, (b) kidney with stone

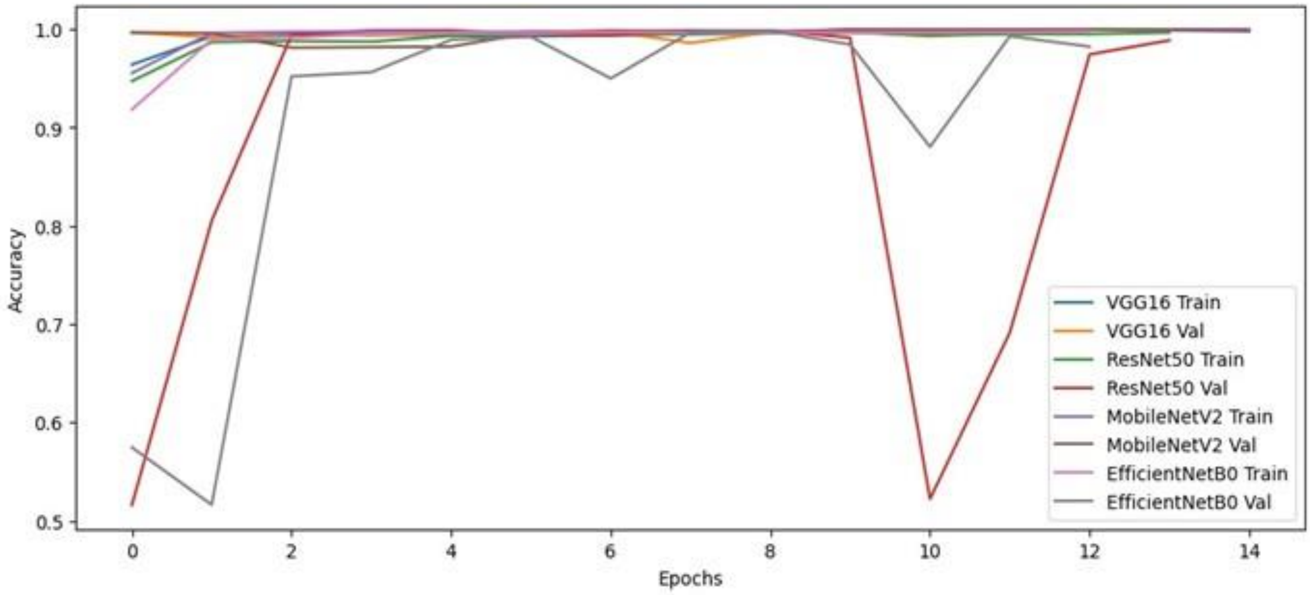


Fig. 3 Accuracy curves of the CNN models

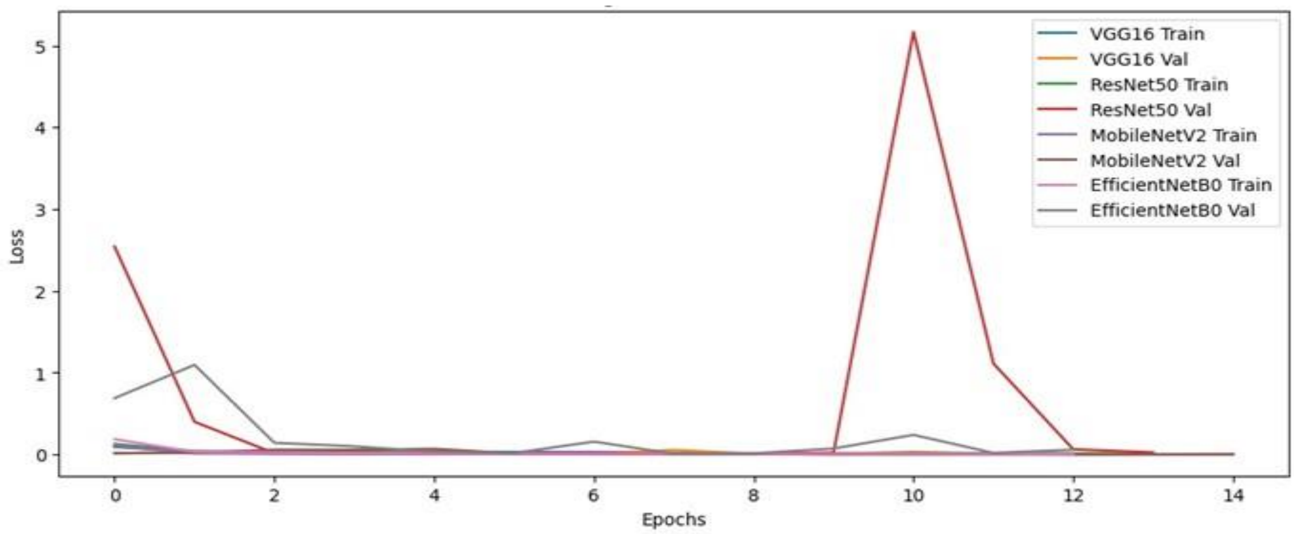


Fig. 4 Loss curves of the CNN models

B. Validation Set Evaluation

To provide a clear and consistent overview of model performance, the confusion matrices of all four CNN architectures are presented in Fig. 5, which illustrates the distribution of True Positives, True Negatives, False Positives, and False Negatives for each model. These matrices allow direct visual comparison of classification reliability and misclassification behaviour across the architectures.

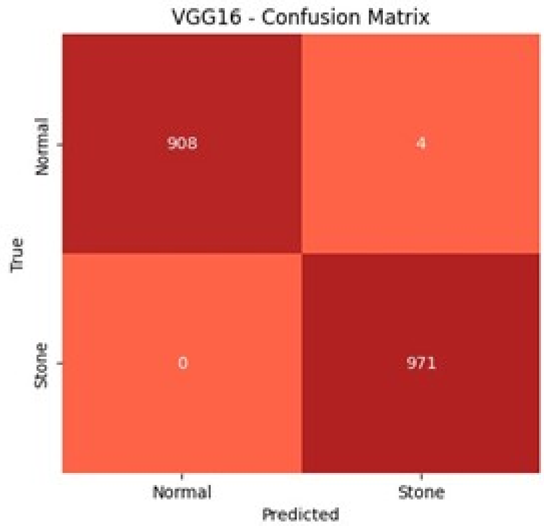
As shown in Fig. 5, the confusion matrices reveal noticeable variations in classification quality across the four CNN models. VGG16 demonstrates strong and balanced performance, correctly identifying nearly all Normal and Stone cases while producing only a small number of False Positives. This indicates that VGG16 effectively learns the structural and textural differences between normal kidney tissue and stone-containing regions, resulting in high sensitivity with no missed Stone cases.

ResNet50 displays an even more refined classification pattern, with fewer misclassifications than VGG16. Its confusion matrix shows strong diagonal dominance, reflecting the benefit of residual connections in capturing deeper and more discriminative features. This architectural advantage allows ResNet50 to achieve perfect recall for the Stone class and maintain very low error rates for the Normal class, confirming its consistent learning behaviour.

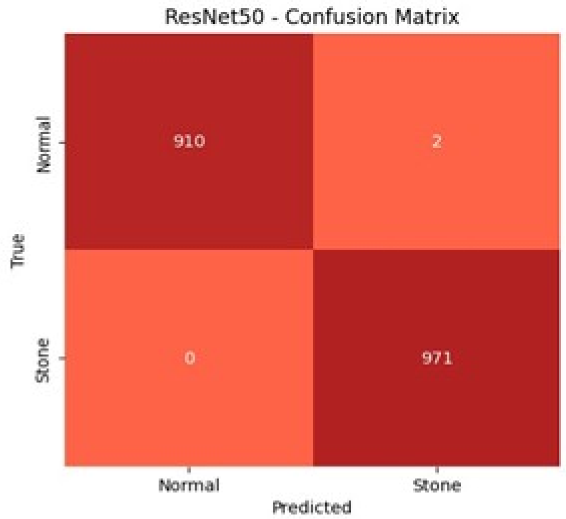
MobileNetV2 achieves the most ideal and clinically desirable outcome among all architectures. Every input sample is classified correctly, producing a perfect diagonal confusion matrix with no misclassifications. This flawless separation between Normal and Stone cases demonstrates highly reliable discrimination despite the model's lightweight design, highlighting its exceptional generalization capability and computational efficiency.

In contrast, EfficientNetB0 exhibits a distinctly problematic confusion matrix. The model predicts nearly all samples as Stone, resulting in an extreme imbalance where True Negatives collapse to zero. Although this yields a perfect Recall for the Stone class, the absence of correctly identified Normal cases indicates that the model fails to learn the distinguishing characteristics of healthy kidney images. This pattern aligns with the unstable training behaviour observed earlier and suggests underfitting or ineffective feature extraction on this dataset.

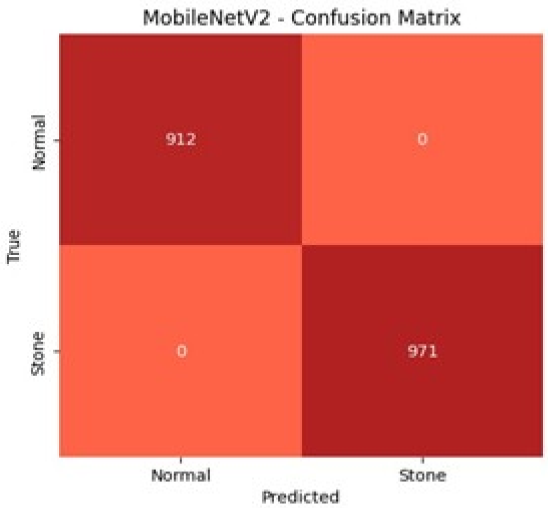
Overall, the comparative view in Fig. 5 clearly shows that MobileNetV2, ResNet50, and VGG16 provide strong and dependable classification, while EfficientNetB0 fails to generalize, misclassifying nearly all Normal cases. These findings further reinforce the superiority of lightweight yet well-optimized architectures particularly MobileNetV2 for ultrasound-based kidney stone classification.



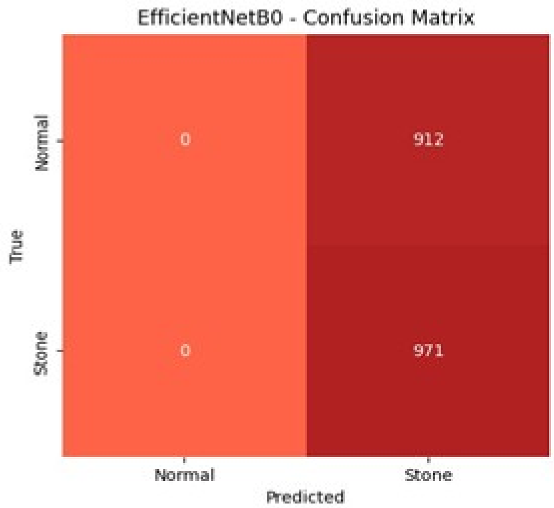
(a)



(b)



(c)



(d)

Fig. 5 Confusion matrix of (a) VGG16, (b) ResNet50, (c) MobileNetV2, (d) EfficientNetB0

C. ROC Curve Analysis

Fig. 6 presents the ROC curves of all evaluated architectures, providing a concise comparison of their ability to distinguish between Normal and Stone classes. The results show that VGG16, ResNet50, and MobileNetV2 achieve equally strong and ideal discrimination, each producing an ROC curve that aligns closely with the upper-left boundary and yielding a perfect AUC value. This indicates that all three models are able to separate the two classes across all thresholds without degradation in sensitivity or specificity.

EfficientNetB0, however, displays an almost linear ROC curve with an AUC of 0.50, reflecting performance equivalent to random classification. This finding is consistent with the model's misclassification pattern and confirms its inability to learn effective decision boundaries from the ultrasound dataset. Overall, the ROC analysis reinforces the reliability of VGG16, ResNet50, and MobileNetV2, while highlighting the substantial limitations of EfficientNetB0.

D. Comparative Analysis

A comparative analysis was conducted to evaluate the predictive performance and computational characteristics of the four CNN architectures. The quantitative comparison of predictive performance and computational efficiency across the four CNN architectures is summarized in Table I. The combined results from the quantitative metrics, confusion matrices, and ROC curves show consistent performance patterns across the models.

MobileNetV2 demonstrates the strongest overall performance, achieving perfect predictions while maintaining the lowest parameter count and fastest training time. This combination of accuracy and efficiency positions MobileNetV2 as the most suitable architecture for deployment in real time or resource limited clinical environments. ResNet50 also performs exceptionally well, achieving near-perfect accuracy and perfect recall, although its deeper architecture results in longer training time and higher computational cost. VGG16 provides similarly high accuracy with stable learning behaviour, offering a balance between performance and computational demand.

EfficientNetB0 performs the weakest among the four models. Despite achieving high recall, the model fails to correctly identify Normal cases, resulting in low overall accuracy. The confusion matrix and ROC curve confirm that EfficientNetB0 does not learn effective decision boundaries for this dataset. These findings highlight that compound scaling strategies may require larger or more diverse datasets to function optimally, and may not generalize well to focused medical imaging tasks such as ultrasound based kidney stone classification.

Overall, the comparative evaluation shows that VGG16, ResNet50, and MobileNetV2 consistently deliver reliable performance, with MobileNetV2 offering the best trade off between accuracy and computational efficiency. EfficientNetB0, however, exhibits significant limitations and requires further optimization to be viable for clinical use.

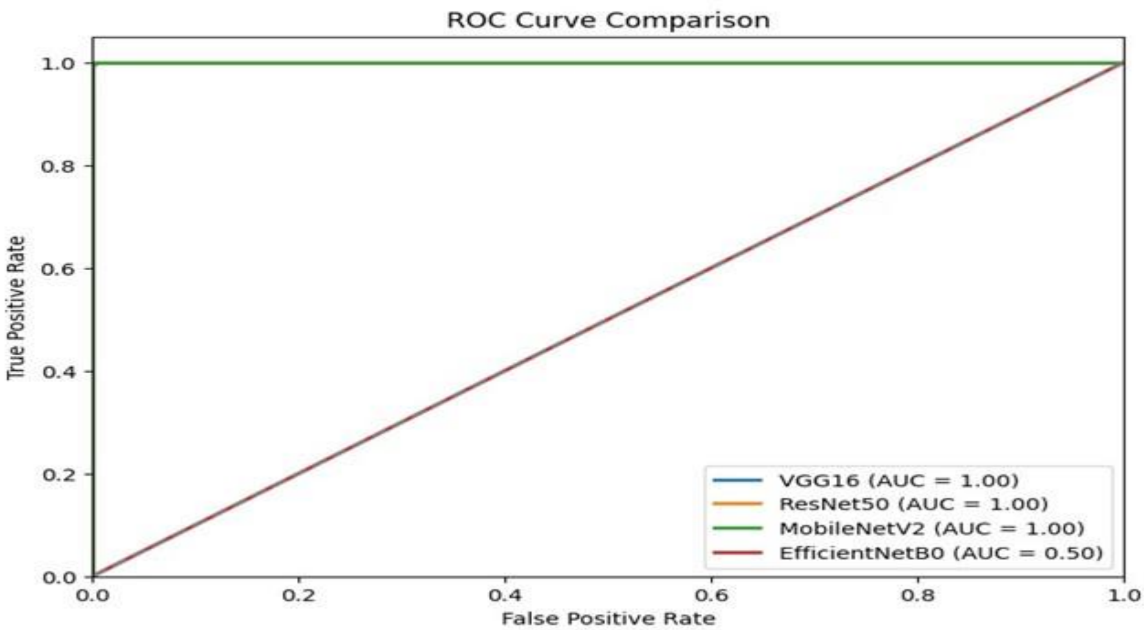


Fig. 6 ROC curve

TABLE I
PERFORMANCE COMPARISON OF CNN ARCHITECTURES FOR KIDNEY STONE

Model	Accuracy	Precision	Recall	F1-Score	Parameters (Million)	Average Training Time
VGG16	99.79	99.59	100.00	99.79	~14.7	~59
ResNet50	99.89	99.79	100.00	99.89	~23.6	~114
MobileNetV2	100.00	100.00	100.00	100.00	~3.6	~44
EfficientNetB0	51.62	51.62	100.00	68.12	~5.4	~73

The comparative evaluation of the four CNN architectures shows substantial differences in predictive accuracy, feature generalization, and computational efficiency when applied to kidney ultrasound images. MobileNetV2 achieved the strongest performance, recording perfect accuracy, precision, recall, and F1-score. This result aligns with prior findings indicating that lightweight CNN architectures can provide competitive diagnostic results with significantly reduced computational requirements [30].

VGG16 and ResNet50 also demonstrated near-perfect accuracy, supported by their strong feature extraction capabilities. The deeper structure of these models enabled them to capture discriminative texture patterns in the ultrasound images. However, their higher number of parameters and longer training times make them less suitable for fast or resource-constrained clinical environments compared with MobileNetV2.

In contrast, EfficientNetB0 showed severe limitations, achieving only 51.62% accuracy and misclassifying all Normal cases. This behaviour is consistent with earlier evidence that model performance can be highly sensitive to image resolution and dataset characteristics [26]. Additional studies further reported that architectures with compound scaling, such as EfficientNet, often struggle to generalize effectively when trained on datasets with limited diversity or domain-specific imaging characteristics [31,32]. Such observations are also supported by recent findings showing that lightweight CNN architectures often outperform deeper or compound-scaled models when applied to focused and low-variability medical imaging datasets [27]. These factors help explain the underfitting and unstable learning patterns observed in EfficientNetB0.

The confusion matrices further highlight the diagnostic implications of each architecture. MobileNetV2 demonstrated perfect separation between Normal and Stone cases, offering the highest level of reliability for clinical decision support. VGG16 and ResNet50 likewise showed strong and consistent classification performance with only minimal misclassifications. EfficientNetB0, however, exhibited a collapsed decision boundary, leading to a high number of false positives and making it unsuitable for practical deployment without further optimization.

Although MobileNetV2 produced outstanding results, certain limitations must be acknowledged. The study used a single 80/20 train–validation split, which does not fully replace multi-fold validation. The dataset was obtained from a single public repository, limiting its variability in terms of imaging equipment and acquisition settings. Future work should involve external validation using multi-institutional datasets and a broader range of ultrasound imaging conditions. Overall, MobileNetV2 provides the best balance of predictive accuracy, efficiency, and practical deployability. Supported by findings related to lightweight CNN performance [30], image resolution sensitivity [25], and generalization challenges in scaled architectures [31, 32], as well as empirical evidence that simpler CNN backbones can offer better stability on constrained medical datasets [27], the results indicate that MobileNetV2 is the most suitable model for real-world ultrasound-based kidney stone classification.

IV. CONCLUSION

This study conducted a comprehensive comparison of four widely used CNN architectures VGG16, ResNet50, MobileNetV2, and EfficientNetB0 for kidney stone classification using ultrasound images. The evaluation incorporated classification metrics, confusion matrix patterns, ROC-AUC performance, and computational efficiency to identify the most suitable architecture for practical clinical use. The results clearly establish MobileNetV2 as the most effective and reliable model. It achieved perfect diagnostic performance, recording 100% accuracy, 100% precision, 100% recall, and a 100% F1-score. This exceptional result is particularly notable given its lightweight design (~3.6 million parameters) and fast training time (~44 seconds per epoch), making it highly suitable for real-time and resource-constrained clinical environments. Both VGG16 and ResNet50 also delivered outstanding results. VGG16 achieved 99.79%

accuracy with 100% recall, while ResNet50 reached 99.89% accuracy with 100% recall. The ability of these models to maintain perfect sensitivity is crucial in medical diagnosis, ensuring that no kidney stone cases are missed. Their high performance confirms that deep CNN architectures can effectively extract discriminative features from ultrasound images when fine-tuned appropriately. In contrast, EfficientNetB0 demonstrated a severe inability to generalize. Although it achieved 100% recall, it recorded only 51.62% accuracy due to misclassifying all Normal cases as Stone. This catastrophic failure highlights that compound-scaled architectures like EfficientNetB0 may require larger, more diverse datasets or more complex optimization strategies to perform reliably in specialized medical imaging tasks. Overall, the experimental evidence shows that MobileNetV2 is the most optimal architecture for ultrasound-based kidney stone classification. It offers perfect diagnostic performance, rapid convergence, and minimal computational cost, distinguishing it as the most practical and deployable option for clinical decision-support systems. These findings provide meaningful insights for selecting CNN models in real-world healthcare scenarios, where both diagnostic accuracy and computational efficiency are of paramount importance.

REFERENCES

[1] C. P. Kovesdy, “Epidemiology of chronic kidney disease: an update 2022,” *Kidney International Supplements*, vol. 12, no. 1, pp. 7–11, 2022, doi: 10.1016/j.kisu.2021.11.003.

[2] L. Wang, Y. He, C. Han, P. Zhu, Y. Zhou, and R. Tang, “Global burden of chronic kidney disease and risk factors, 1990–2021: an update from the global burden of disease study 2021,” *Frontiers in Public Health*, vol. 13, no. July, pp. 1–12, 2025, doi: 10.3389/fpubh.2025.1542329.

[3] L. Moftakhar, F. Jafari, M. Ghoddusi Johari, R. Rezaeianzadeh, S. V. Hosseini, and A. Rezaianzadeh, “Prevalence and risk factors of kidney stone disease in population aged 40–70 years old in Kharameh cohort study: a cross-sectional population-based study in southern Iran,” *BMC Urology*, vol. 22, no. 1, pp. 1–9, 2022, doi: 10.1186/s12894-022-01161-x.

[4] A. Basiri, A. H. Kashi, H. S. Omran, N. Borumandnia, and S. Golshan, “National lifetime prevalence and demographic factors of urolithiasis in Iran,” *Urology Journal*, vol. 20, no. 2, pp. 102–108, 2023, doi: 10.22037/uj.v20i.7576.

[5] L. Guo, L. Liu, Y. Sun, and L. Xue, “Prevalence and related factors of nephrolithiasis among medical staff in Qingdao, China: a retrospective cross-sectional study,” *BMC Nephrology*, vol. 25, no. 1, pp. 1–8, 2024, doi: 10.1186/s12882-024-03651-6.

[6] B. Cheraghian, A. Meysam, S. J. Hashemi, and S. A. Hosseini, “Kidney stones and dietary intake in adults: a population-based study in southwest Iran,” *BMC Public Health*, vol. 24, no. 1, pp. 1–7, 2024, doi: 10.1186/s12889-024-18393-1.

[7] R. Ahmad and B. K. Mohanty, “Chronic kidney disease stage identification using texture analysis of ultrasound images,” *Biomedical Signal Processing and Control*, vol. 69, no. April, p. 102695, 2021, doi: 10.1016/j.bspc.2021.102695.

[8] S. A. Hannan and P. Pal, “Detection and classification of kidney disease using convolutional neural networks.,” *J Neurol Neurorehab Res*, vol. 8, no. August, p. 1, 2023, doi: 10.35841/aajnnr-8.2.136.

[9] A. Altalbe and A. R. Javed, “Applying customized convolutional neural network to kidney image volumes for kidney disease detection,” *Computer Systems Science and Engineering*, vol. 47, no. 2, pp. 2119–2134, 2023, doi: 10.32604/csse.2023.040620.

[10] S. Asif, X. Zheng, and Y. Zhu, “An optimized fusion of deep learning models for kidney stone detection from CT images,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 7, p. 102130, 2024, doi: 10.1016/j.jksuci.2024.102130.

[11] K. L. Kohsasih, M. D. Agung Rizky, R. Rosnelly, and W. W. Widjaja, “A deep learning model to detect the brain tumor based on magnetic resonance images,” *Jurnal Infotel*, vol. 14, no. 3, pp. 203–208, 2022, doi: 10.20895/infotel.v14i3.793.

[12] S. Verma, P. K. Singh, G. Kaur, A. Vashistha, and S. Pansari, “Non-Invasive Kidney Stone Prediction using Machine Learning: An Extensive Review,” *Biomedical and Pharmacology Journal*, vol. 18, no. March, pp. 45–58, 2025, doi: 10.13005/bpj/3072.

[13] A. Saber, E. Hassan, S. Elbedwehy, W. A. Awad, and T. Z. Emara, “Leveraging ensemble convolutional neural networks and metaheuristic strategies for advanced kidney disease screening and classification,” *Scientific Reports*, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-93950-1.

[14] G G. Katkar and S. Shinde, “Enhanced chronic kidney disease detection using deep learning: A comparative analysis of CNN and LSTM models.,” *Journal of Information Systems Engineering and Management*, vol. 10, no. 30s, pp. 316–323, 2025, doi: 10.52783/jisem.v10i30s.4837.

[15] S. N. Almuayqil, S. Abd El-Ghany, A. A. Abd El-Aziz, and M. Elmogy, “KidneyNet: A novel CNN-based technique for the automated diagnosis of chronic kidney diseases from CT scans,” *Electronics (Switzerland)*, vol. 13, no. 24, 2024, doi: 10.3390/electronics13244981.

[16] K. Nagawa, Y. Hara, K. Inoue, Y. Yamagishi, M. Koyama, and H. Shimizu, “Three-dimensional convolutional neural network-based classification of chronic kidney disease severity using kidney MRI,”

Scientific Reports, vol. 14, no. 1, pp. 1–10, 2024, doi: 10.1038/s41598-024-66814-3.

[17] A. Pimpalkar, D. K. J. B. Saini, N. Shelke, A. Balodi, G. Rapate, and M. Tolani, "Fine-tuned deep learning models for early detection and classification of kidney conditions in CT imaging," *Scientific Reports*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-94905-2.

[18] D. Alzu'Bi, M. Abdullah, I. Hmeidi, R. Alazab, M. Gharaibeh, and M. El-Heis, "Kidney tumor detection and classification based on deep learning approaches: A new dataset in CT scans," *Journal of Healthcare Engineering*, vol. 2022, 2022, doi: 10.1155/2022/3861161.

[19] M. Badawy, A. M. Almars, H. M. Balaha, M. Shehata, M. Qaraad, and M. Elhosseini, "A two-stage renal disease classification based on transfer learning with hyperparameters optimization," *Frontiers in Medicine*, vol. 10, 2023, doi: 10.3389/fmed.2023.1106717.

[20] T. Zuo, Y. Zheng, L. He, and T. Chen, "Automated classification of papillarrey nal cell carcinoma and chromophobe renal cell carcinoma based on a small computed tomography imaging dataset using deep learning," *Frontiers in Oncology*, vol. 11, no. November, pp. 1–10, 2021, doi: 10.3389/fonc.2021.746750.

[21] E. N. Yildiz, E. Cengil, M. Yildirim, and H. Bingol, "Diagnosis of Chronic Kidney Disease Based on CNN and LSTM," *Acadlore Transactions on AI and Machine Learning*, vol. 2, no. 2, pp. 66–74, 2023, doi: 10.56578/ataiml020202.

[22] N. Sulaksono, K. Adi, and R. Isnanto, "Utilization of convolutional neural network in image interpretation techniques for detecting kidney disease," *IAES International Journal of Artificial Intelligence*, vol. 14, no. 1, pp. 602–613, 2025, doi: 10.11591/ijai.v14.i1.pp602-613.

[23] J. Jose and S. Sheeja, "Synergistic excellence: CNN-LSTM hybrid model for Improved CKD diagnosis," *SSRG International Journal of Electronics and Communication Engineering*, vol. 11, no. 3, pp. 12–23, 2024, doi: 10.14445/23488549/IJECE-V11I3P102.

[24] D. Saif, A. M. Sarhan, and N. M. Elshennawy, "Deep-kidney: an effective deep learning framework for chronic kidney disease prediction," *Health Information Science and Systems*, vol. 12, no. 1, pp. 1–22, 2024, doi: 10.1007/s13755-023-00261-8.

[25] S. Tang, C. Jing, Y. Jiang, K. Yang, Z. Huang, and H. Wu, "The effect of image resolution on convolutional neural networks in breast ultrasound," *Heliyon*, vol. 9, no. 8, p. e19253, 2023, doi: 10.1016/j.heliyon.2023.e19253.

[26] D. A. Zebari, "Kidney disease segmentation and classification using firefly sigma seeker and magweight rank techniques," *Bioengineering*, vol. 12, no. 4, 2025, doi: 10.3390/bioengineering12040350.

[27] A. Larasati, S. Surono, A. Thobirin, and D. A. Dewi, "Performance Analysis of Resampling Techniques for Overcoming Data Imbalance in Multiclass Classification," *Jurnal Informatika, JUITA*, vol. 13, no. 1, pp. 57–66, 2025, doi: 10.30595/juita.v13i1.25270.

[28] N. Helmawati and E. Utami, "Analysis for Detecting Banana Leaf Disease Using the CNN Method," *Jurnal Infomatika JUITA*, vol. 13, no. 1, pp. 29–36, 2025, doi: 10.30595/juita.v13i1.24514.

[29] M. Momeni, "Kidney Stone Classification and Object Detection." [Online]. Available: https://www.kaggle.com/datasets/imtkaggleteam/kidney-stone-classification-and-object-detection

[30] G. Maçin, F. Genç, B. Taşcı, S. Dogan, and T. Tuncer, "KidneyNeXt: A lightweight convolutional neural network for multi-class renal tumor classification in computed tomography imaging," *Journal of Clinical Medicine*, vol. 14, no. 14, pp. 1–22, 2025, doi: 10.3390/jcm14144929.

[31] K. Kawadkar, "Comparative analysis of vision transformers and convolutional neural networks for medical image classification," pp. 1–9, 2025, doi: 10.48550/arXiv.2507.21156.

[32] M. Zhang, Z. Ye, E. Yuan, X. Lv, Y. Zhang, and Y. Tan, "Imaging-based deep learning in kidney diseases: recent progress and future prospects," *Insights into Imaging*, vol. 15, no. 1, 2024, doi: 10.1186/s13244-024-01636-5.