International Journal of Innovative Computing, Information and Control
Volume **21**, Number **2**, April **2025**

ICIC International ©2025 ISSN 1349-4198
pp. 565–576

CONVOLUTIONAL NEURAL NETWORKS FOR IDENTIFYING PAPILLARY THYROID CANCER HISTOPATHOLOGICAL IMAGE

Nabila Husna Shabrina[1], Dadang Gunawan[1,*] and Agnes Stephanie Harahap[2]

[1]Department of Electrical Engineering
Universitas Indonesia
Depok, Jawa Barat 16425, Indonesia
nabila.husna31@ui.ac.id; *Corresponding author: guna@eng.ui.ac.id

[2]Department of Anatomical Pathology
Faculty of Medicine
Universitas Indonesia/Dr. Cipto Mangunkusumo Central General Hospital
Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta 10430, Indonesia
agnes.stephanie01@ui.ac.id

Received August 2024; revised December 2024

Abstract. *Artificial intelligence advancements have significantly sped up the development of specialized algorithms for diagnosing Papillary Thyroid Cancer from digital images. Several studies demonstrate that AI-based approaches provide highly satisfactory performance, with one of them being Convolutional Neural Networks (CNN). However, there is a noticeable gap in research regarding head-to-head comparisons of CNN architectures for identifying thyroid cancer histopathological imaging. This study seeks to address this gap by providing a thorough evaluation of 13 well-known CNN architectures performance using transfer learning. The selected CNN architecture includes CoAtNet-0, ConvNeXt Tiny, DenseNet121, DenseNet201, InceptionV3, InceptionResNetV2, EfficientNetV2B0, ResNet50, ResNet101, ResNet50V2, VGG19, VGG16, and Xception. The model will be assessed using a comprehensive set of metrics widely employed in medical applications, including accuracy, specificity, sensitivity, F1-score, negative predictive value, and positive predictive value in two different patch sizes of 512 × 512 and 256 × 256. ConvNeXt Tiny, ResNet50, and ResNet101 were proven as the leading models and demonstrated optimal performance across all metrics in both image patch sizes. The results emphasize the significance of model selection and appropriate patch size in identifying thyroid histopathological images.*

Keywords: Artificial intelligence, Convolutional neural networks, Histopathological image, Patch size, Papillary thyroid cancer, Transfer learning

1. **Introduction.** Papillary Thyroid Carcinoma (PTC) represents 80%-85% of all thyroid cancer cases [1]. PTC can be identified through a range of diagnostic tests, with histopathological diagnosis as the gold standard for identifying PTC [2]. However, traditional histopathological procedures present significant challenges, primarily because of their time-consuming nature [3]. As the number of tissue samples increases, pathologists' workloads also increase, which can lead to potential diagnostic errors [3]. Additionally, conventional histopathology is plagued by inconsistencies and disagreements among experts [4]. The rise in Digital Pathology (DP) has enabled pathologists to efficiently analyze and distinguish various tissue sample patterns associated with PTC using digital images. This advancement has also fostered the development of Computer-Aided

DOI: 10.24507/ijicic.21.02.565

Pathology (CAP), which employs computer-based methodologies to diagnose PTC. Artificial Intelligence (AI) advancements have significantly accelerated the development of specialized CAP algorithms for diagnosing PTC from digital images.

Numerous studies have employed AI-based algorithms to develop CAP systems for automatic identification of cancer histopathological images [5,6]. These studies have demonstrated that AI-based approaches yield satisfactory performance [7-16]. The majority of this research has focused on deep learning methods, specifically Convolutional Neural Networks (CNN) [7,9,11,13,15,17]. CNN has gained popularity due to their ability to automatically extract key features from image data, eliminating the requirement for manual feature extraction [18]. Some studies have explored hybrid models that combine CNN with traditional machine learning algorithms. In these approaches, CNN are utilized for feature extraction, while classifiers like K-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Gaussian Naïve Bayes, Decision Tree, Logistic Regression, and Random Forest are employed for the final classification [8]. This approach enables researchers to leverage the feature extraction capabilities of CNN alongside the computational efficiency of traditional machine learning. However, these hybrid approaches often require careful fine-tuning to ensure compatibility with specific datasets. Moreover, the exploration of Vision Transformer (ViT)-based methods introduces a cutting-edge alternative to traditional CNN architectures [19]. ViT can capture global context and long-range dependencies more effectively due to their attention mechanisms. Although promising, ViT faces challenges such as high computational requirements and the need for extremely large datasets to train effectively.

Despite the emergence of hybrid methods and novel architectures such as ViT for analyzing thyroid histopathological images, CNN remain the primary tool in AI-driven pathology owing to its straightforward implementation and satisfactory performance. However, the diversity of CNN architectures necessitates head-to-head evaluation to understand their advantages and limitations. This comparison is essential for establishing robust benchmarks, identifying the most effective architectures for identifying histopathological images of thyroid cancer, and guiding future research in this area. This study addresses this need by comprehensively evaluating the performances of 13 well-known CNN architectures using transfer learning. The selected CNN architecture includes VGG16, VGG19 [20], ResNet50, ResNet101, ResNet50V2 [21], DenseNet121, DenseNet201 [22], InceptionV3 [23], InceptionResNetV2 [24], Xception [25], EfficientNetV2B0 [26,27], ConvNeXt Tiny [28], and CoAtNet-0 [29]. These models were chosen for their established reputations and proven success in various biomedical image classification tasks. Unlike prior research that relied solely on conventional metrics, the present study will assess the model using a comprehensive set of metrics widely employed in medical applications, including accuracy, specificity, sensitivity, F1-score, Negative Predictive Value (NPV), and Positive Predictive Value (PPV) [30].

The primary goal of this study was to assess the effectiveness of 13 well-known CNN architectures in accurately identifying the histopathological images of thyroid cancer. By conducting this comprehensive comparison, this study seeks to identify the baseline CNN architectures most effective for classifying histopathological images of thyroid cancer. The outcomes of this study are expected to deliver valuable insights into the strengths and weaknesses of different CNN models, guiding future research and practical applications in the field of PTC histopathological imaging.

The remainder of this paper is organized as follows. Section 2 reviews the related work and summarizes previous studies in the field. Section 3 details the proposed methodology, including the datasets, CNN architectures, and model implementation, as well as the evaluation metrics used in this study. Section 4 presents the results and discussion. Finally,

Section 5 concludes the paper and summarizes the key contributions and outcomes of the research.

2. **Related Studies.** In recent years, various studies have demonstrated the effectiveness of image and data analysis by utilizing deep learning approaches using different datasets and methods. Zhu et al. adopted Otsu thresholding and patch sampling to process institutional datasets. Using a Patch U-Net model, the results underscoring the significance of patch-based methods and advanced neural network architectures for improving image analysis tasks [10]. Chen et al. focused on patch cropping techniques combined with transfer learning CNN models for their analysis of institutional datasets, further validating the effectiveness of patch-based processing methods [11]. Building upon this foundation, our study adopts a similar approach to harness its benefits for precise and efficient image analysis.

Dee et al. utilized the Tharun and Thompson, TCGA, and Nikiforov datasets, employing StyleGAN2-ADA for data augmentation and a transfer learning CNN for classification, increasing accuracy by 5%-9% [12]. However, the application of StyleGAN2-ADA for data augmentation can be computationally intensive for implementation. Dolezal et al. combined institutional datasets and TCGAs using patch extraction and filtering with transfer learning CNN models [13]. However, integrating multiple datasets can introduce challenges related to data heterogeneity and increased computational complexity.

Tsou and Wu employed the TCGA dataset with patch cropping and data augmentation, achieving high AUC using CNN Inception V3 model [14]. Nojima et al. focused on patch extraction and data augmentation with transfer learning CNN models, achieving a highest AUC while implementing VGG14 model [15]. The use of data augmentation techniques helps prevent overfitting and improves model performance. Nonetheless, the effectiveness of these techniques heavily depends on the quality and diversity of the augmented data.

Do and Khanh used the Tharun and Thompson dataset, as well as the Nikiforov dataset, with cell crop and patch crop techniques, achieving an increase in precision of up to 4% with a modified Inception V3 model [16]. While this method enhances precision, it may also lead to increased computational demands and necessitates meticulous preprocessing to ensure data quality. Liu et al. applied color transformation and data augmentation and more than 85% accuracy using inception residual CNN and SVM [31]. This approach enhances feature diversity and model robustness. Nonetheless, the combination of multiple preprocessing steps and complex models increases computational demands and may lead to overfitting.

Yin et al. employed patch extraction with a Pyramid Token-to-Token Vision Transformer on institutional datasets and achieved precision of more than 84% [19]. This method effectively captures both local and global features. However, Vision Transformers require substantial computational resources and large datasets for effective training. Sharma et al. focused on patch extraction and transformer feature extraction with machine learning, achieving superior performance when utilizing Swin transformer, PCA, and RF-FOX optimizer [32]. This approach can capture detailed patterns within the data but requires advanced preprocessing techniques. The ongoing development of algorithms using diverse datasets and methodologies highlights the critical importance of refining these techniques in thyroid histopathological image identification research.

3. **Materials and Methods.** Figure 1 illustrates the workflow of this study. The research utilized the Tharun and Thompson dataset, which contains two categories: non-PTC-like and PTC-like. Initially, the original images from the dataset were extracted into smaller patches of two distinct sizes: 256×256 and 512×512 pixels. The dataset

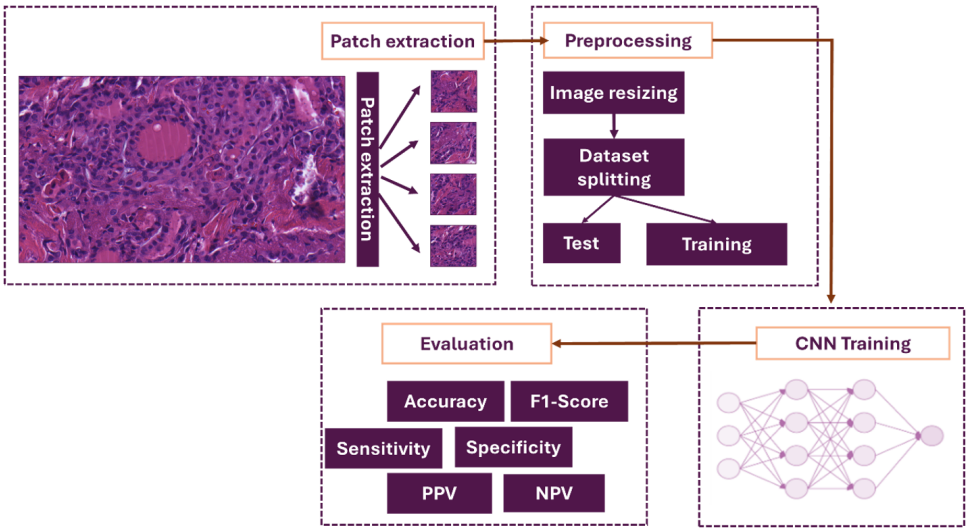

FIGURE 1. Research workflow

was then separated into testing and training sets with a split ratio of 0.1 for testing and 0.9 for training purposes. All images were scaled down to 224 × 224 pixels to match the optimal input size of the CNN models. Following this preprocessing, the data were trained using 13 CNN models through a transfer-learning approach. The CNNs model's performance was assessed based on widely used medical application statistical metrics, including accuracy, specificity, sensitivity, F1-score, NPV, and PPV.

3.1. **Datasets.** The Tharun and Thompson dataset, which comprises histopathological images of PTC obtained from 156 patients, was employed in this study. The dataset initially comes from the Whole Slide Images format, captured at $40x$ magnification with a resolution of 0.23 µm/px, and labeled by two pathologists. Each whole-slide image had representative images extracted from neoplastic areas, resulting in non-overlapping images of size 1916 × 1053 pixels. Due to computational limitations, the image was partitioned into smaller patches of two different sizes using a technique developed by Dee et al. [12]. This technique also mimicked the process of diagnosing histopathological samples by individually observing small areas within the images. The patch sizes were selected based on their frequent use in similar histopathological studies, with the goal of determining the optimal patch size for training thyroid histopathological images in this task. Subsequently, the dataset was partitioned into training, validation, and test sets using a 90 : 10 split ratio, with 10% of the training set designated for validation. The total counts for each image patch size are displayed in Table 1, while Figure 2 provides examples from the dataset for each category.

TABLE 1. Details of the dataset

Label	**Patch size of 512 × 512**			**Patch size of 256 × 256**		
	Train	**Test**	**Total**	**Train**	**Test**	**Total**
Non-PTC	6183	687	6870	21573	2397	23970
PTC	4651	517	5168	16279	1809	18088
Total	**10834**	**1204**	**12038**	**37852**	**4206**	**42058**

3.2. **CNN architecture.** In this study, various CNN architectures were selected for classification. We selected 13 established pre-trained CNN models known for their diverse architectural designs and strengths to ensure a thorough evaluation and reliable

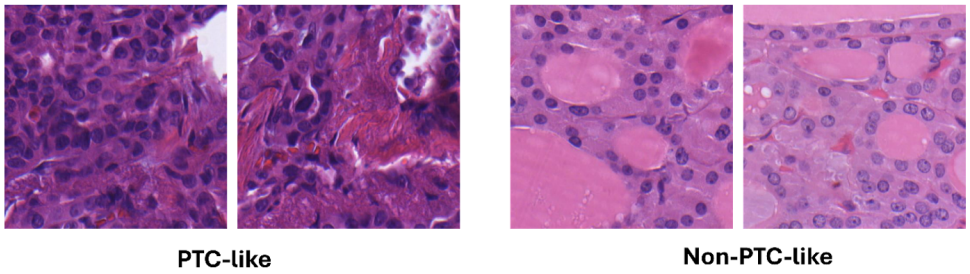

FIGURE 2. Dataset sample [8]

TABLE 2. Model parameters

Model	Size (MB)	Parameters	Depth
CoAtNet-0	86	25 M	–
ConvNeXt Tiny	109.42	28.6 M	–
DenseNet121	27	8.1 M	242
DenseNet201	71	20.2 M	402
InceptionResNetV2	208	55.9 M	449
EfficientNetV2B0	23	7.2 M	–
ResNet50	98	25.6 M	107
ResNet101	163	44.7 M	205
ResNet50V2	90	25.6 M	103
InceptionV3	84	23.9 M	189
VGG16	528	138.4 M	16
VGG19	549	143.7 M	19
Xception	80	22.9 M	81

performance assessment. Our approach uses transfer learning to implement CNN models [33]. Transfer learning enables the capitalization of pre-trained models trained on massive datasets, which serves as a solid foundation for our specific classification task. This approach mitigates the demand for extensive computational resources and large annotated datasets. It expedites the training process, allowing the models to achieve optimal performance within a relatively small number of training epochs. Each architecture was initialized using pre-trained weights from the ImageNet dataset, followed by fine-tuning to the Tharun and Thompson dataset. The different parameters of each architecture are provided in Table 2.

3.3. **Model implementation.** The training employed a batch size of 32 and a learning rate of 10^{-4} utilizing Adam optimizer. Binary cross-entropy served as the loss function with the output layer is activated using sigmoid. The training spanned 50 epochs with early stopping monitored for 40 epochs according to the validation loss. The learning rate decayed by 0.9 when the performance plateaued to optimize the training process. The training process for all CNN models was conducted using Google Colab Pro.

3.4. **Evaluation metrics.** To evaluate the performance of the CNN model in classifying PTC histopathological images, we employed six key metrics: accuracy, PPV, NPV, Sensitivity, Specificity, and F1-score with each mathematical equation given in Equations (1) to (6), where *TP*, *TN*, *FP*, and *FN* refer to True Positives, True Negatives, False Positives, and False Negatives.

$$Test\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP + FP} \tag{2}$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN + FN} \tag{3}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{4}$$

$$Specificity = \frac{TN}{TN + FP} \tag{5}$$

$$F1\text{-}score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{6}$$

Accuracy is a metric that evaluates the model's overall correctness by comparing the number of accurate predictions to the overall count of the generated prediction. PPV is the proportion of accurate PTC images among all images identified as PTC, including those mistakenly identified as PTC. NPV measures the proportion of true non-PTC images among all identified as non-PTC, including those incorrectly identified as PTC. Sensitivity assesses how effectively the model identifies actual PTC cases by comparing true PTC identifications to the total number of true PTC and PTC images mistakenly identified as non-PTC. Specificity measures the model's capability to correctly identify non-PTC cases by comparing true non-PTC identifications to the total number of true non-PTC and non-PTC images incorrectly identified as PTC. The F1-score is utilized in this research due to the imbalanced datasets and provides a single performance measure.

4. **Results and Discussion.** The heatmap plot highlights the performance of each model, with green representing the highest metric value and red indicating lower values.

4.1. **Performance results for image size of 512 × 512.** Figure 3 depicts the heatmap results of training with an image patch size of 512 × 512. The ResNet50 model achieved the highest performance, with an accuracy rate of 0.867. This performance was consistent across various metrics, demonstrating its reliability in accurately identifying PTC and non-PTC histopathological images. Similarly, the ResNet101 model demonstrated robust classification abilities, with a consistent metric of over 0.78 for all metrics and an accuracy of 0.857. ConvNeXt Tiny and VGG16 models also demonstrated impressive performance, with respective accuracies of 0.817 and 0.813. Those two models also show their ability to maintain a delicate balance between sensitivity and specificity effectively. The following models, EfficientNetV2B0, DenseNet121, and DenseNet201, also achieved noteworthy results with accuracies ranging from 0.777 to 0.838. Those results demonstrate their capacity to process and learn from histopathological images efficiently. These models exhibited high sensitivity, ensuring that most PTC cases were accurately identified while maintaining an acceptable level of specificity of higher than 0.7.

In contrast, the CoAtNet0 and ResNet50V2 models exhibited limited effectiveness, achieving an accuracy of only 0.571. This result suggests potential challenges in their architecture for this task. The InceptionResNetV2 model achieved a perfect sensitivity of 1.000 yet had a poor PPV of 0.002. This indicates many false positives, which could result in unnecessary follow-up tests or treatments for PTC patients. The results of the Xception model, which achieved an accuracy of 0.355 and a specificity of 0.268, demonstrate a lack of success in differentiating PTC from non-PTC images. This may be attributed to insufficient feature extraction using the Xception architecture.

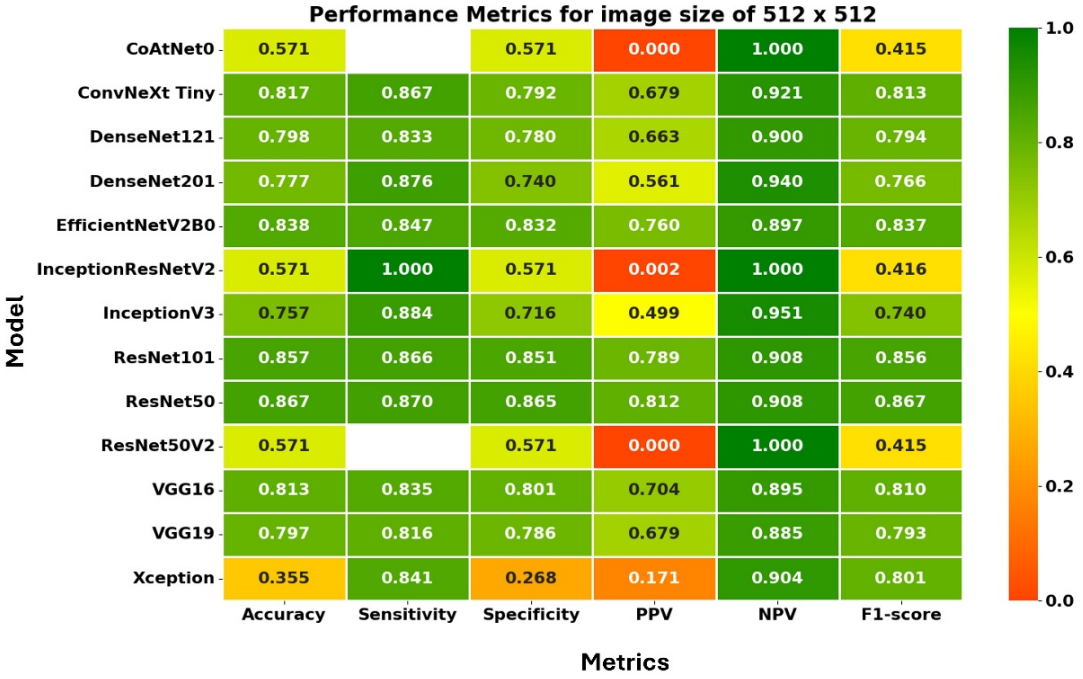

FIGURE 3. Performance metric results for image size of 512 × 512

In summary, when training thyroid histopathological images using a patch size of 512 × 512, models such as ResNet50, ResNet101, VGG16 and ConvNeXt Tiny demonstrated the optimal combination of accuracy and other key metrics, including sensitivity, specificity, and F1-score. A balance between sensitivity and specificity is essential in diagnosing PTC to guarantee accurate identification and proper elimination of non-PTC cases. EfficientNetV2B0 also emerged as a strong contender, performing well across these metrics and making it an alternative option following those four models.

4.2. **Performance results for image size of 256 × 256.** Figure 4 illustrates the heatmap results of training with image patch size of 256 × 256. The ConvNeXt Tiny model exhibited superior performance across all metrics, achieving an accuracy of 0.844, sensitivity of 0.855, specificity of 0.836, PPV of 0.766, NPV of 0.902, and an F1-score of 0.808. The VGG19 model also performed satisfactorily with an accuracy of 0.845, sensitivity of 0.886, specificity of 0.822, PPV of 0.734, NPV of 0.929, and an F1-score of 0.803. This suggests that both ConvNeXt Tiny and VGG19 are particularly effective at differentiating between PTC and non-PTC images when trained using an image patch size of 256 × 256. ResNet101 and ResNet50 models also exhibited a strong performance, achieving accuracy of 0.832 and 0.829, respectively. These models also demonstrated a noteworthy balance between sensitivity and specificity. EfficientNetV2B0, following with a respectable accuracy of 0.821 and a moderate F1-score of 0.771, delivered the next strong performance.

The Xception model demonstrated moderate overall performance, achieving an accuracy of 0.764, sensitivity of 0.828, specificity of 0.738, PPV of 0.571, NPV of 0.910, and an F1-score of 0.676. These results indicate that the model effectively detects PTC cases and produces higher false positive results. In contrast, CoAtNet0 and DenseNet121 demonstrated lower performance levels, achieving accuracy scores of 0.550 and 0.570, respectively. These models particularly struggled with PPV, indicating difficulties in accurately identifying the PTC class. InceptionResNetV2, InceptionV3, and ResNet50V2

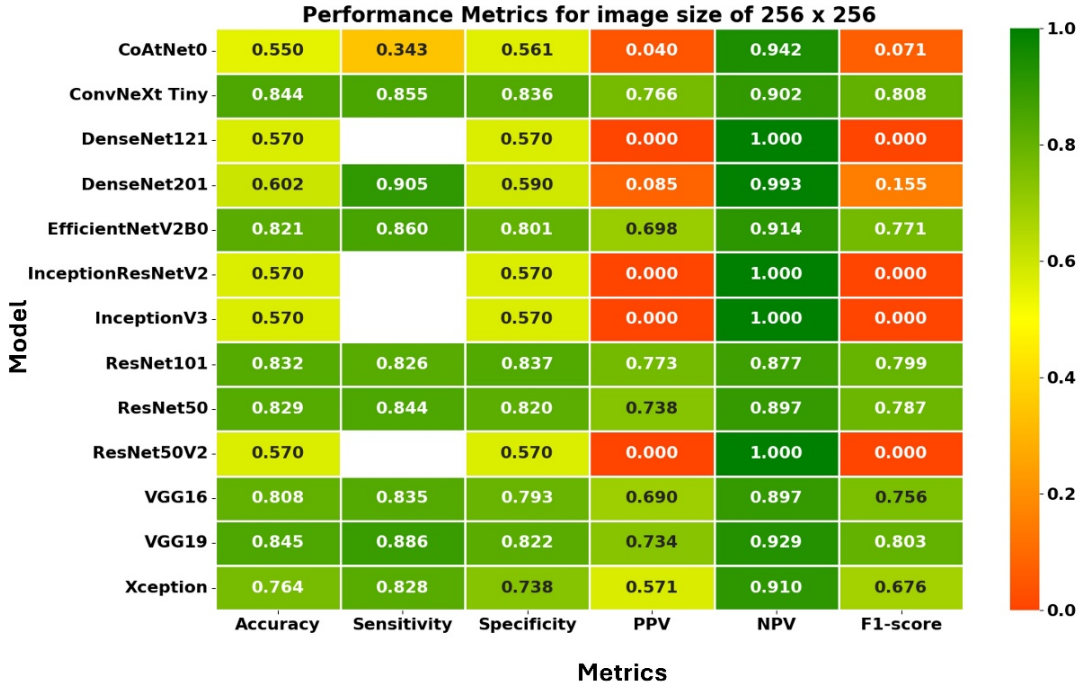

FIGURE 4. Performance metric results for image size of 256 × 256

showed inadequate PPV and F1-scores of 0.000 despite having a high NPV of 1.000, suggesting a high level of false positives, which is inevitable in clinical settings. These findings suggest that ConvNeXt Tiny, VGG19, ResNet50, and ResNet101 architectures are well-suited for identifying PTC using a patch size of 256 × 256 due to their leading performance across all metrics.

4.3. **Different results between image patch size.** When using the image size of 512 × 512 pixels, the model maintains high accuracy and specificity though with greater variability, as depicted in Figure 5. The sensitivity value shows small variability with a high median of approximately 0.8. PPV resulted in high variability, with the smallest value being nearly zero. NPV and F1-score demonstrated small variability with a high median of approximately 0.9 and 0.8, respectively. The presence of an outlier was only found in sensitivity and F1-score. The small number of outliers within the results of this image size shows that the majority of the models performed reliably using the image patch size of 512 × 512.

For the smallest image size of 256 × 256 pixels, the models exhibited a slight decrease in performance consistency, as can be seen in Figure 6. Accuracy remained high but with increased variability. PPV and F1-score show high variability, indicating that more models face challenges when training using this image size. Sensitivity continues to perform high with the lowest variability compared to other metrics, followed by NPV. There is one outlier, which shows that only one model did not result in good sensitivity.

Summarizing the findings, the effect of image size on the performance of CNN models in classifying thyroid histopathological images is significant. The larger image size of 512 × 512 pixels provides more comprehensive results, although it also comes with high variability in specific measures, such as PPV and accuracy. The small patch size of 256 × 256 pixels, while still providing a high median, results in excessive variability compared to larger patch size. Based on these findings, we recommend using a larger patch size of 512 × 512 for optimal performance.

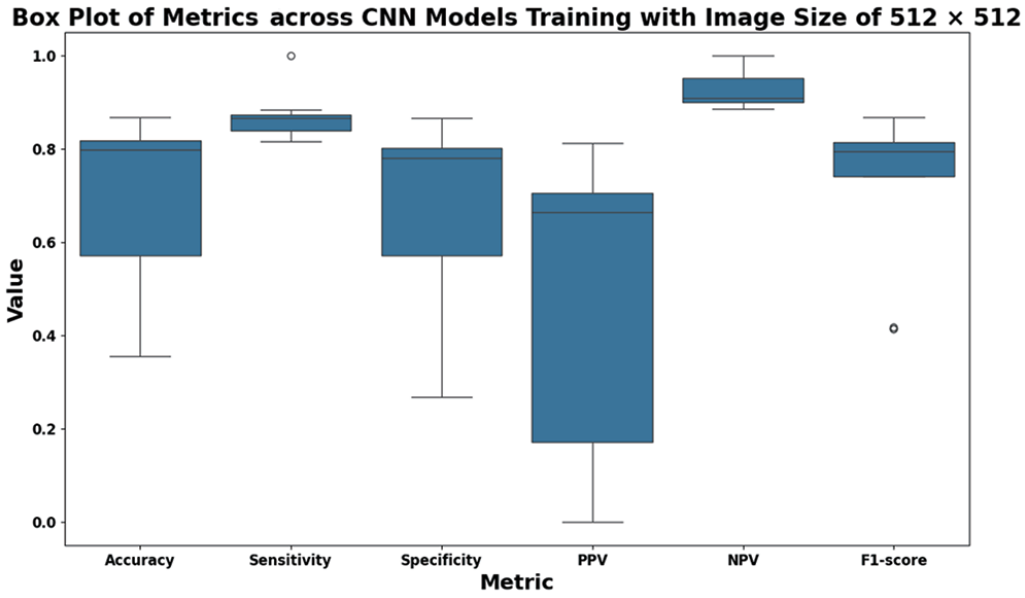

FIGURE 5. Boxplot of metric for image size of 512 × 512

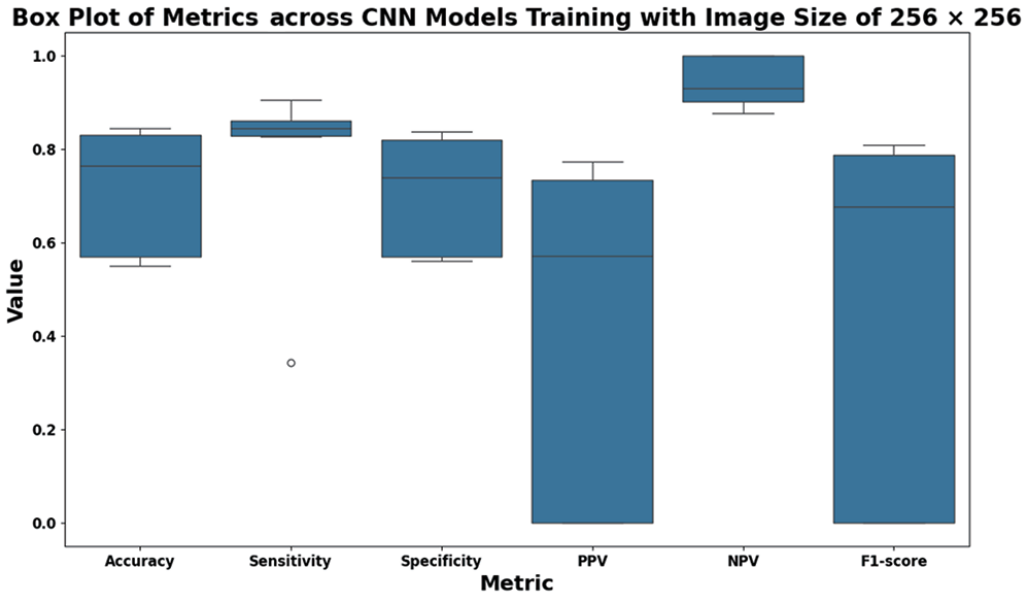

FIGURE 6. Boxplot of metric for image size of 256 × 256

5. **Conclusions.** The present study reveals that the choice of CNN architecture significantly impacts performance in identifying PTC and non-PTC histopathological imaging, highlighting the importance of model selection when using CNN. Furthermore, the findings underscore the significance of image resolution in influencing model performance and emphasize the importance of optimizing and choosing the appropriate image patch size alongside model selection. Future research should optimize the model by exploring layer modification or investigating the effect of adding new variables, such as expanding the dataset and preprocessing techniques, to enhance the robustness of the CNN in diagnosing PTC and non-PTC using histopathological image.

Acknowledgment. The first author would like to thank Universitas Multimedia Nusantara for partly supporting this study.

REFERENCES

[1] F. Limaiem, A. Rehman and C. Anastasopoulou, *Papillary Thyroid Carcinoma*, StatPearls Publishing, Treasure Island (FL), 2023.

[2] M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot and B. Yener, Histopathological image analysis: A review, *IEEE Rev. Biomed. Eng.*, vol.2, pp.147-171, 2009.

[3] L. He, L. R. Long, S. Antani and G. R. Thoma, Histology image analysis for carcinoma detection and grading, *Computer Methods and Programs in Biomedicine*, vol.107, no.3, pp.538-556, 2012.

[4] H. K. Su, B. M. Wenig, G. C. Haser, M. E. Rowe, S. L. Asa, Z. Baloch, E. Du, W. C. Faquin, G. Fellegara, T. Giordano, R. Ghossein, V. A. LiVolsi, R. Lloyd, O. Mete, U. Ozbek, J. Rosai, S. Suster, L. D. Thompson, A. T. Turk and M. L. Urken, Inter-observer variation in the pathologic identification of minimal extrathyroidal extension in papillary thyroid carcinoma, *Thyroid*, vol.26, no.4, pp.512-517, 2016.

[5] S. Prabhu, K. Prasad, A. Robels-Kelly and X. Lu, AI-based carcinoma detection and classification using histopathological images: A systematic review, *Computers in Biology and Medicine*, vol.142, 105209, 2022.

[6] A. A. Ahmed, M. Abouzid and E. Kaczmarek, Deep learning approaches in histopathology, *Cancers*, vol.14, no.21, 5264, 2022.

[7] C. Deng, D. Li, M. Feng, D. Han and Q. Huang, The value of deep neural networks in the pathological classification of thyroid tumors, *Diagn. Pathol.*, vol.18, no.1, 95, 2023.

[8] M. Böhland, L. Tharun, T. Scherr, R. Mikut, V. Hagenmeyer, L. D. R. Thompson, S. Perner and M. Reischl, Machine learning methods for automated classification of tumors with papillary thyroid carcinoma-like nuclei: A quantitative analysis, *PLoS One*, vol.16, no.9, e0257635, 2021.

[9] Y. Wang, Q. Guan, I. Lao, L. Wang, Y. Wu, D. Li, Q. Ji, Y. Wang, Y. Zhu, H. Lu and J. Xiang, Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: A large-scale pilot study, *Ann. Transl. Med.*, vol.7, no.18, 468, 2019.

[10] X. Zhu, C. Chen, Q. Guo, J. Ma, F. Sun and H. Lu, Deep learning-based recognition of different thyroid cancer categories using whole frozen-slide images, *Front. Bioeng. Biotechnol.*, vol.10, 857377, 2022.

[11] P. Chen, X. Shi, Y. Liang, Y. Li, L. Yang and P. D. Gader, Interactive thyroid whole slide image diagnostic system using deep representation, *Computer Methods and Programs in Biomedicine*, vol.195, 105630, 2020.

[12] W. Dee, R. A. Ibrahim and E. Marouli, Histopathological domain adaptation with generative adversarial networks: Bridging the domain gap between thyroid cancer histopathology datasets, *PLoS One*, vol.19, no.12, e0310417, 2024.

[13] J. M. Dolezal, A. Trzcinska, C.-Y. Liao, S. Kochanny, E. Blair, N. Agrawal, X. M. Keutgen, P. Angelos, N. A. Cipriani and A. T. Pearson, Deep learning prediction of BRAF-RAS gene expression signature identifies noninvasive follicular thyroid neoplasms with papillary-like nuclear features, *Modern Pathology*, vol.34, no.5, pp.862-874, 2021.

[14] P. Tsou and C.-J. Wu, Mapping driver mutations to histopathological subtypes in papillary thyroid carcinoma: Applying a deep convolutional neural network, *JCM*, vol.8, no.10, 1675, 2019.

[15] S. Nojima, T. Kadoi, A. Suzuki, C. Kato, S. Ishida, K. Kido, K. Fujita, Y. Okuno, M. Hirokawa, K. Terayama and E. Morii, Deep learning-based differential diagnosis of follicular thyroid tumors using histopathological images, *Modern Pathology*, vol.36, no.11, 100296, 2023.

[16] T.-H. Do and H. N. Khanh, Supporting thyroid cancer diagnosis based on cell classification over microscopic images, *2022 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*, Phu Quoc, Vietnam, pp.1-5, 2022.

[17] N. H. Shabrina, D. Gunawan, M. F. Ham and A. S. Harahap, Papillary thyroid cancer histopathological image classification using pretrained ConvNeXt Tiny and Grad-CAM interpretation, *2023 IEEE 11th Joint International Information Technology and Artificial Intelligence Conference (ITA-IC)*, vol.11, pp.1836-1842, 2023.

[18] K. R. Ummah, T. Karlita, R. Sigit, E. M. Yuniarno, I K. E. Purnama and M. H. Purnomo, Effect of image pre-processing method on convolutional neural network classification of COVID-19 CT scan images, *International Journal of Innovative Computing, Information and Control*, vol.18, no.6, pp.1895-1912, 2022.

[19] P. Yin, B. Yu, C. Jiang and H. Chen, Pyramid Tokens-to-Token Vision Transformer for thyroid pathology image classification, *2022 11th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Salzburg, Austria, pp.1-6, 2022.

[20] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *Proc. of the 3rd International Conference on Learning Representations (ICLR2015)*, San Diego, CA, USA, 2015.

[21] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp.770-778, 2016.

[22] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp.2261-2269, 2017.

[23] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, Rethinking the Inception architecture for computer vision, *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp.2818-2826, 2016.

[24] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, Inception-v4, Inception-ResNet and the impact of residual connections on learning, *Proc. of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, pp.4278-4284, 2017.

[25] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp.1800-1807, 2017.

[26] M. Tan and Q. V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, CA, USA, vol.97, pp.6105-6114, 2019.

[27] M. Tan and Q. V. Le, EfficientNetV2: Smaller models and faster training, *Proceedings of the 38th International Conference on Machine Learning*, vol.139, pp.10096-10106, 2021.

[28] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell and S. Xie, A ConvNet for the 2020s, *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, pp.11966-11976, 2022.

[29] Z. Dai, H. Liu, Q. V. Le and M. Tan, CoAtNet: Marrying convolution and attention for all data sizes, *The 35th Conference on Neural Information Processing Systems*, pp.3965-3977, 2021.

[30] T. F. Monaghan, S. N. Rahman, C. W. Agudelo, A. J. Wein, J. M. Lazar, K. Everaert and R. R. Dmochowski, Foundational statistical principles in medical research: Sensitivity, specificity, positive predictive value, and negative predictive value, *Medicina*, vol.57, no.5, 503, 2021.

[31] Y. Liu, L. Han, H. Wang and B. Yin, Classification of papillary thyroid carcinoma histological images based on deep learning, *IFS*, vol.40, no.6, pp.12011-12021, 2021.

[32] R. Sharma, G. K. Mahanti, G. Panda, A. Rath, S. Dash, S. Mallik and R. Hu, A framework for detecting thyroid cancer from ultrasound and histopathological images using deep learning, meta-heuristics, and MCDM algorithms, *J. Imaging*, vol.9, no.9, 173, 2023.

[33] A. Hosna, E. Merry, J. Gyalmo, Z. Alom, Z. Aung and M. A. Azim, Transfer learning: A friendly introduction, *Journal of Big Data*, vol.9, no.1, 102, 2022.

Author Biography

Nabila Husna Shabrina received her bachelor's degree in Telecommunication Engineering and a master's degree in Electrical Engineering from Institut Teknologi Bandung. She is currently a doctoral student at Department of Electrical Engineering, Universitas Indonesia. Her research interest includes computer vision and image processing.

Dadang Gunawan received the bachelor's degree in Electrical Engineering from the University of Indonesia, in 1983, the master's degree from Keio University, Japan, in 1989, and the Ph.D. degree from the University of Tasmania, Australia, in 1995. He is currently a Professor with the Department of Electrical Engineering, Universitas Indonesia. He has published hundreds of academic articles as a first author or a co-author in proceedings and international journals. His research interests include wireless and signal-processing technology.

Agnes Stephanie Harahap currently serves as the Head of the Endocrine and Hematolymphoid Division at the Faculty of Medicine, Universitas Indonesia, and Dr. Cipto Mangunkusumo Central General Hospital, Jakarta, Indonesia. She completed her Pathology specialization, subspecialty, and doctorate degree in Universitas Indonesia. Dr. Harahap has undergone extensive training, including a fellowship in Bone Marrow/Hematopathology at Siriraj Hospital, Mahidol University, Bangkok Thailand, training in Endocrine Pathology at the University of Pittsburgh Medical Center, USA, and Thyroid Core Needle Biopsy at Seoul St. Mary's Hospital, Korea. Her research focuses on thyroid pathology, lymphoma, hematologic neoplasms, artificial intelligence, and molecular diagnostics.