Hyperparameter Tuning Exploration to Maximize MobileNet Performance in Classification Kidney Tumor

Sandy Putra Siregar*[1], M. Safii[2], Sundari Retno Andani[3]
[1,2,3]Department of Informatics Magister's, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
e-mail: *[1]sandy@amiktunasbangsa.ac.id, [2]m.safii@amiktunasbangsa.ac.id, [3]sundari.ra@amiktunasbangsa.ac.id

Abstract

The main focus of this research is on developing the MobileNet architecture in order to produce a kidney tumor classification model that is accurate, resistant to overfitting, and remains consistent with variations in datasets and training parameters. This study aims to develop a MobileNet architecture to produce a software model that can precisely identify with high accuracy, perform kidney tumor classification, and avoid failure in generalizing new data, called overfitting, as well as evaluate the difference in accuracy generated from several variations of datasets and parameters. The method used in this study is MobileNet with hyperparameter tuning and fine-tuning, and it was compared with the MobileNet Baseline method. The dataset consists of 12,446 images classified as Normal, Cyst, Stone, and Tumor, collected from Kaggle. The findings of this study on the division of the 80:10:10 ratio of the proposed method image data resulted in 100% accuracy, 100% precision, 100% recall, and 100% F1-Score. This study is expected to produce architecture modifications that can classify kidney tumors with high accuracy, so that the hypothesis is achieved. In addition, various approaches in medical image analysis using deep learning have shown better results in identifying various tumors, especially in this research, in the classification and detection of kidney tumors.

Keywords— Kidney Tumor, Classification, MobileNet, Hyperparameter, Deep Learning

1. INTRODUCTION

The kidney is a vital organ that plays an important role in maintaining the body's homeostatic balance through blood filtration, fluid and electrolyte regulation, and excretion of metabolic waste [1]. One of the most important organs in the body, it is susceptible to acute damage in sepsis, known as Sepsis-Associated Acute Kidney Injury (SA-AKI), which significantly increases the risk of developing chronic kidney disease, cardiovascular disorders, and death [2]. 2020 marked a significant challenge in kidney transplantation, with a decrease in the overall number of transplants due to a decline in transplants from living donors, although transplants from deceased donors increased. The COVID-19 pandemic has had a significant impact on the transplant process, including fewer new candidates, increased deaths on the waiting list, and a high rate of unutilized organs. In addition, living donor transplantation remains low in both adult and pediatric populations, with marked disparities in access by race [3].

Tumors or neoplasms are one of the primary causes of morbidity (illness and health) and mortality (death from specific diseases) among kidney post-transplant patients, with the total risk of cancer nearly double that of the general population. This study found that Kaposi's sarcoma and hematological cancers are the most common tumor types, particularly in the first three years following transplantation, and that the majority of cases are linked to the effects of post-transplant immunosuppression. These findings highlight the significance of short and long-term oncologic monitoring in managing renal transplant patients [4]. Kidney tumors are one of the most frequent diseases worldwide, and early detection is key to treatment success [5], [6]. Malignant tumors can damage the kidneys and spread to other organs, resulting in lethal malignancies. TAMs are immune cells that play a dual role; [7], [8] they play a vital role in tumor microenvironmental

dynamics by encouraging angiogenesis and metastasis while also altering immune responses that promote tumor growth.

Recognizing chronic kidney disease (CKD) as a priority in national health policies is an important step toward improving renal healthcare globally[9], given that CKD remains a significant health burden, with a consistent trend of rising incidence, mortality, and disability from 1990 to 2021, particularly in countries with low socio-demographic indices[10]. There is a need to develop therapies to improve toxic protein clearance to maintain renal function by confirming that UMOD mutase-induced uromodulin aggregate formation causes endoplasmic reticulum stress and immunological reactions that directly contribute to progressive renal tubular damage in ADTKD. This method opens new avenues for targeted molecular therapies in hereditary kidney disease[11], [12]. [13] Fasler's study demonstrated that juvenile kidney tumor cells respond to diverse cytokines by activating the STAT and NF-kB pathways. This supports the in vitro model's potential for mapping cytokine signaling, which is pertinent to creating more effective immunotherapies. [14][15] TAMs are an important target in cancer immunotherapy due to their overwhelming polarization towards the M2 phenotype. Clinical efforts to restrict TAM recruitment or convert them to the M1 phenotype are ongoing. This method is very promising despite significant obstacles in terms of efficacy and widespread adoption in therapy.

AI-powered diagnostic devices are urgently required to detect early kidney illnesses such as stones, cysts, and malignancies. The AI model based on Vision Transformers and CNN is a possible technique for clinical integration to increase the accuracy of radiology diagnosis. It demonstrates accurate and consistent segmentation performance in detecting kidney and renal tumors on CT-Scan scans, particularly for tiny lesions [16]. Moreover, its performance remains consistent when evaluated using independent test data, indicating robust generalization. [17] According to Abdelrahman and Viriri's study, deep learning in kidney tumor segmentation, specifically convolutional neural networks, has significantly improved diagnostic accuracy and detection effectiveness and is now an essential part of modern radiology and treatment procedures. [18] Applying the lightweight, quick, and efficient MobileNet architecture in skin picture categorization produces excellent results, with a 90% accuracy on the HAM1000 dataset. The model can handle class imbalance and increase generalization by altering the architecture and employing data augmentation techniques. These findings suggest that the MobileNet model, which informed the decision-making process in learning algorithms, is suitable for early skin disease detection systems, particularly under low-power conditions.

The fundamental challenge is improving the model's performance through effective hyperparameter change. [19] André Luiz's statistical hyperparameter tuning research, using analysis of variance and Tukey's test, discovered the optimal and optimized CNN architecture configuration for visual categorization of construction machinery. This technique gives a more in-depth understanding of parameter interactions and allows for the most accurate model configuration selection. Consequently, the recommended combination of DenseNet and Adagrad has the highest average accuracy in binary testing (90.0%) and multiclass classification (77.8%). [20] Using modern image processing methods, transfer learning approaches, and ideally tuned hyperparameter optimization, this work accurately classifies renal CT images with high accuracy, allowing for better radiological diagnosis and early detection of kidney illnesses. [21] In Amini's research, the hyperparameter tuning strategy improves the model development component, resulting in prediction performance that is 10%-25% greater than that of the traditional univariate approach. The combination of hyperparameter optimization, ensemble approaches, and multivariate inputs shows that model parameter refinement is required to address the complexity of rainfall dynamics under changing climatic circumstances—the significance of autonomous hyperparameter tuning in increasing the accuracy of rainfall forecast models. Deep neural networks (DNNs) performed significantly better after utilizing methods such as random search and the Parzen tree structure estimator.

[22] Parvathi and Jonnadula suggested a two-stage method for segmenting kidney tumors using a Custom U-Net and Mask R-CNN. This work offered accurate segmentation findings on

the TCGA-KIRC dataset, with an accuracy value of 91% and an IoU of 91%. The model can identify complex tumor shapes, fluctuations in image lighting, and overlapping objects. Although segmentation algorithms such as U-Net and Mask R-CNN produce good spatial detail results (91% accuracy), MobileNet-based classification approaches with hyperparameter-tuning can increase accuracy even further with limited computational resources.

Furthermore, the proposed CNN method for kidney tumor identification is versatile and straightforward. At epoch-19, the model in this investigation achieved 94.94% accuracy, 95.07% validation accuracy, and 0.1270 validation loss [23]. The key disadvantage of this model is that it does not employ pre-trained data, which puts the model's performance in danger of overfitting and can enhance accuracy. The technique or model can be developed by adjusting the architecture or hyperparameter-tuning to optimize the architecture further, resulting in higher accuracy for more accurate identification. Thus, the employment of approaches involving architecture change or hyperparameter tuning is urgently required to improve accuracy and efficiency in implementation.

In Addition [20] Pimpalkar's novel renal CT image categorization method employs transfer learning-based deep learning models such as VGG16, ResNet50, AlexNet, and InceptionV3, adjusted meticulously through a structured hyperparameter configuration process. This strategy combines image augmentation techniques, watershed segmentation, Otsu thresholding, and relief method-based feature selection to detect better kidney diseases, such as cysts, stones, tumors, and normal kidneys. The results showed that the suggested model attained the greatest classification accuracy of 99.96%, making it a high-performance solution in CT scan image-based automated diagnosis systems. It could help with early detection and more accurate clinical decision-making. Although this study obtained an accuracy of 99.96%, greater configuration and optimization of the CNN architecture are required to avoid model overfitting and underfitting.

Based on the aforementioned developments, findings show that deep learning-based methods in medical image analysis have demonstrated increased effectiveness in performing disease classification, including kidney tumors. One promising architecture is MobileNet, which provides lightweight and efficient computational performance. Classification requires a model that accurately identifies the presence of tumors and is lightweight enough to be widely implemented with hardware limitations. This research centers on developing a baseline MobileNet-based model into a proposed method for renal tumor classification with various dataset configurations and parameters, emphasizing achieving accurate results and strong generalization ability.

2. RESEARCH METHODS

This research begins with an initial data handling process used as the basis for developing deep learning models. After the data is acquired, a data preprocessing stage is performed, where the data is divided into two parts: training data, validation (initial testing), and testing data. The training process was conducted over 13 epochs, the length of time or number of iterations chosen to enable stable model learning while preventing excessive iterations that could cause overfitting in the deep learning architecture training workflow. The model is then built using the MobileNet CNN architecture as the baseline and compared with the optimized proposed method. In the final stage, the performance of each model is evaluated using the confusion matrix. The overall stages of this research can be seen in the methodology diagram presented in Figure 1.

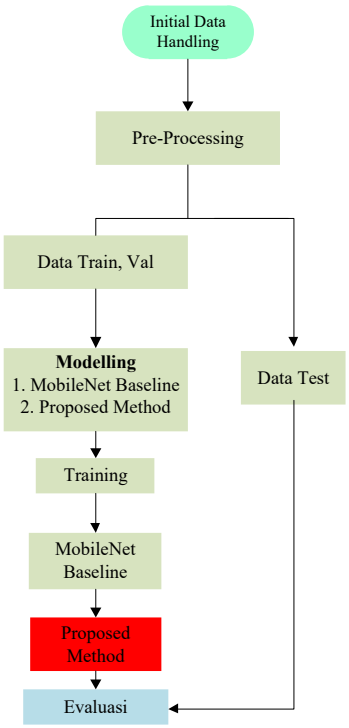
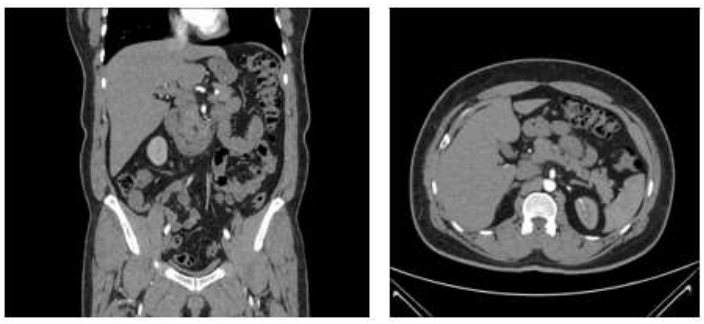


Figure 1. Research Stage

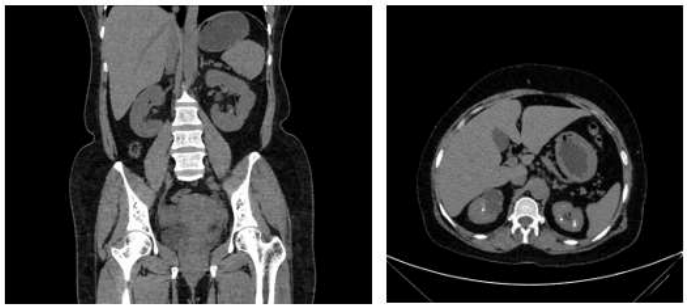
2.1. Initial Dataset Handling

The dataset used in this study was obtained from the Kaggle *CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone* [24] and consists of 12,446 wood images. The dataset is divided into four classes, namely: normal (5,077 images), cyst (3,709 images), stone (1,377 images), and tumor (2,283 images). The data is divided into three main parts: training, validation, and testing. In this study, the dataset was split into training, validation, and test sets using a standard split. Although the class composition was not perfectly balanced, random data splitting helped ensure that each dataset split had a relatively similar class distribution. In this way, the model still obtained a picture that represented the actual data conditions. Since there were no changes to the class composition, the evaluation results could be considered to reflect the model's ability to process the class distribution in the original data, so the interpretation of accuracy and other metrics still reflected the model's performance against natural class variations. This composition aims to evaluate the performance of the model on various proportions of training and test data. A representative sample of the dataset is illustrated in Figure 2.

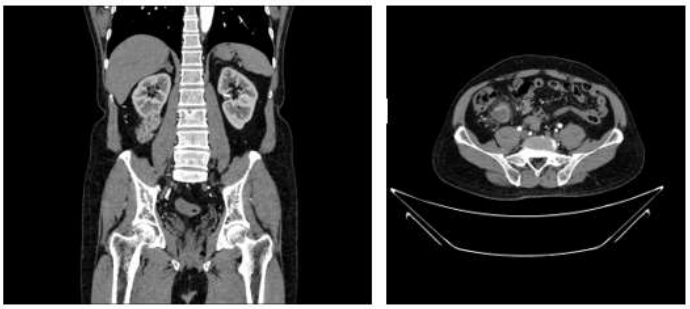


(a) Normal

(b) Cyst



(c) Stone



(d) Tumor

Figure 2. An illustrative subset of the dataset is shown to provide an overview: (a) Normal, (b) Cyst, (c) Stone, and (d) Tumor

The data samples illustrated in Figure 2, which are provided in JPG format, are used for model training, validation, and testing. The JPG format was chosen due to its storage efficiency and compatibility with various deep learning architectures.

2.2. Architecture Model

Model architecture in deep learning refers to a layered structure of computational units designed to learn a representation of data progressively. Each layer has a specific role in extracting and transforming features from the input, resulting in a goal-driven final output, such as classification, segmentation, or prediction. The suggested approach uses a hyperparameter-tuned version of the MobileNet architecture with an input image resolution of 224×224 pixels. In accordance with conventional techniques for training on large-scale datasets, a batch size of 32 was chosen. Figure 3 shows a structural comparison between the proposed technique and the baseline MobileNet model.

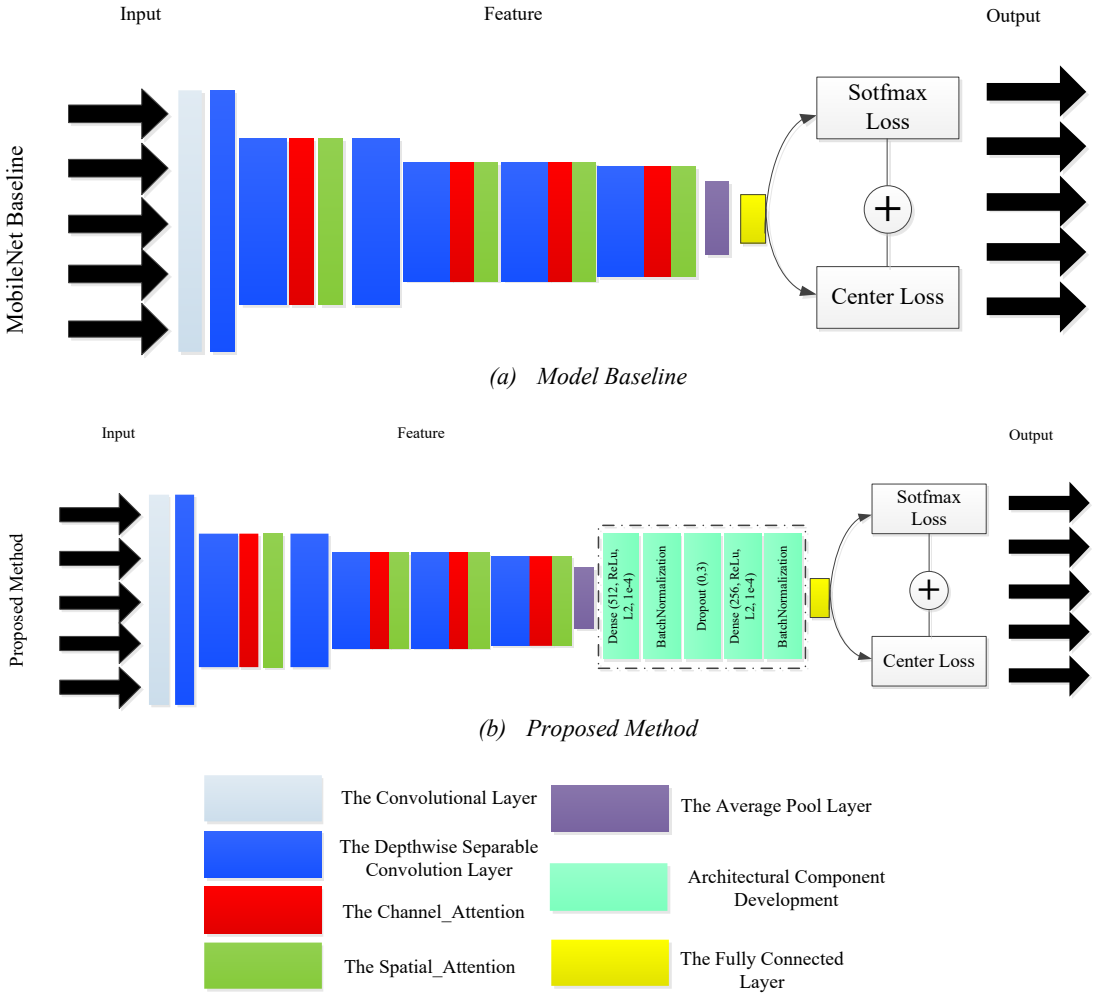


Figure 3. Model Architecture (a) MobileNet Baseline, (b) Proposed Method

Figure 3 presents the core components of the improved convolutional neural network architecture. The model comprises several essential building blocks, beginning with convolutional layers that capture low-level spatial features from the input image. To improve computational efficiency, the architecture adopts depthwise separable convolutions, which separate spatial filtering from channel-wise processing—significantly reducing the number of parameters while maintaining high performance. Furthermore, the model integrates channel attention and spatial attention mechanisms to enhance its ability to focus on the most relevant information across both feature dimensions. The extracted features are then condensed through average pooling and subsequently passed to the fully connected layers for final classification. The hyperparameters utilized in the model configuration are detailed in Table 1.

Table 1. Hyperparameter Tuning Values

Layer	Comparative Value	Activation	Regularization
Dense	512	ReLu	L2
BatchNormalization	-	-	-
Dropout	0,3	-	-
Dense	256	ReLu	L2
BatchNormalization	-	-	-

Table 1 above shows the hyperparameter tuning values for the architecture added to the baseline MobileNet model, called the proposed method.

2.3. Model Evaluation

In this evaluation stage, the trained MobileNet model was tested using separate datasets consisting of three kidney tumour disease data classes and one class of normal kidney. The testing process involves making predictions and evaluating them using a confusion matrix to determine the actual classification performance of the model. The model's performance was assessed using four primary metrics: accuracy, precision, recall, and F1 Score. Accuracy measures a model's classification skill by comparing the proportion of correctly predicted examples to the total number of samples. Precision is the number of correct positives divided by the total number of anticipated positives. Recall (also known as sensitivity) refers to the proportion of actual positive cases properly detected by the model. Finally, the F1 score indicates the harmonic mean of precision and recall, making it a useful metric when dealing with unbalanced data, particularly when false positives and false negatives are equally relevant.

3. RESULT AND DISCUSSION

The dataset used in this study was collected from Kaggle (CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone) [24][25] and included 12,446 photos of kidney tumours and normals. The collection is organized into four categories: normal (5,077 photos), cyst (3,709 images), stone (1,377 images), and tumour (2,283 images). The data was separated into three main sections: training, validation, and testing. Table 2 illustrates the data-splitting strategy. The image to be trained has numerous size variations, including 705x569 pixels, 512x512 pixels, 804x651 pixels, 652x528 pixels, and 798x646 pixels. Model comparison is configured with Adam's optimizer, input size 224, epoch 13, and batch size 32. Table 2 displays the data framework that will be used to train the baseline MobileNet model and the proposed method.

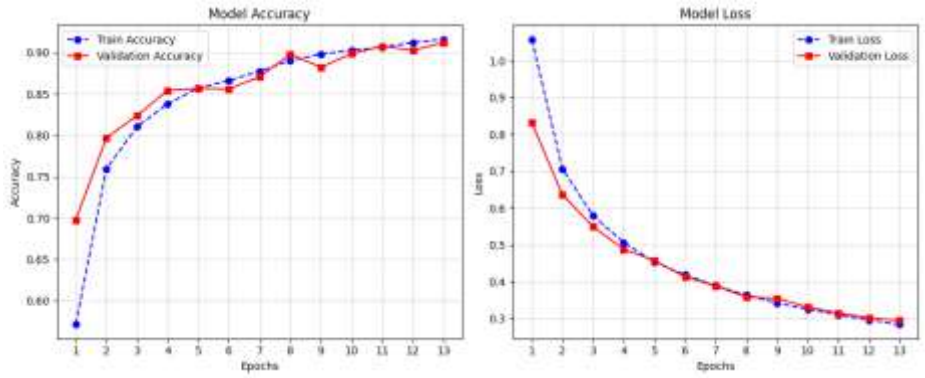
Table 2. Dataset Splitting Framework

Splitting Data	Class Name	Training Data	Validation Data	Testing Data
80-10-10	Normal Cyst Stone Tumor	9955	1245	1246
70-15-15	Normal Cyst Stone Tumor	8710	1867	1869
60-20-20	Normal Cyst Stone Tumor	7466	2489	2491

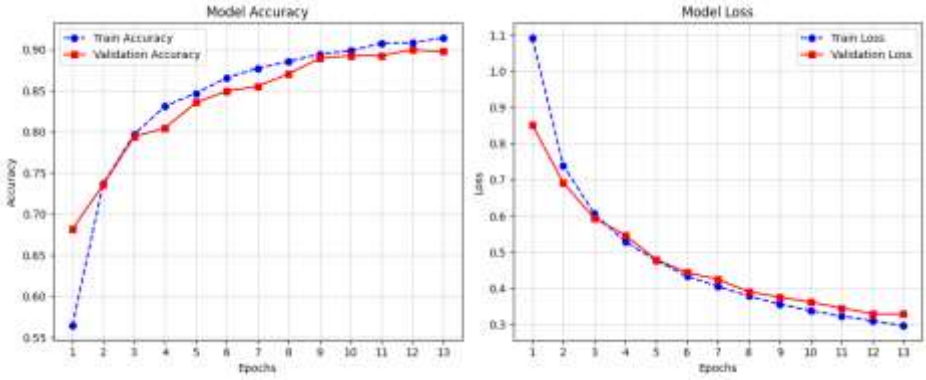
Table 2 provides information on training the baseline MobileNet model and the proposed model using datasets divided 80-10-10 (80% train data, 10% valid data, and 10% test data), 70-15-15, and 60-20-20.

3.1. MobileNet Baseline

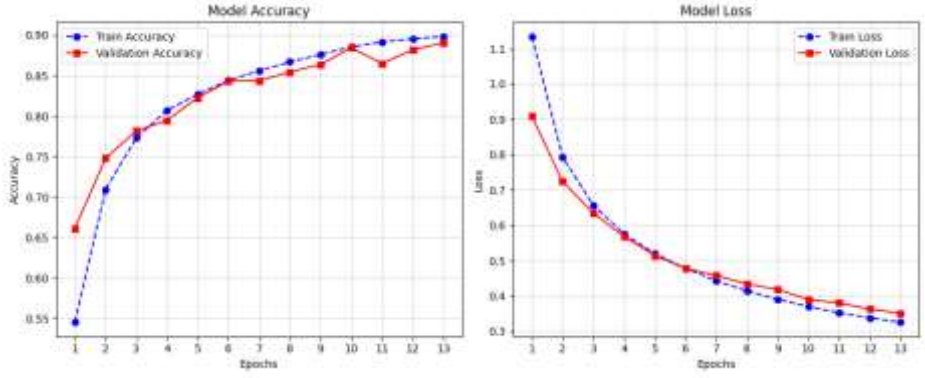
This architecture utilizes depthwise separable convolution to significantly reduce computational complexity compared to standard convolution, while maintaining competitive accuracy on benchmark datasets such as ImageNet.This baseline MobileNet model was trained using dataset splits of 80-10-10 (80% train data, 10% valid data, and 10% test data), 70-15-15, and 60-20-20.



(a) Model Baseline Dataset 80-10-10



(b) Model Baseline Dataset 70-15-15



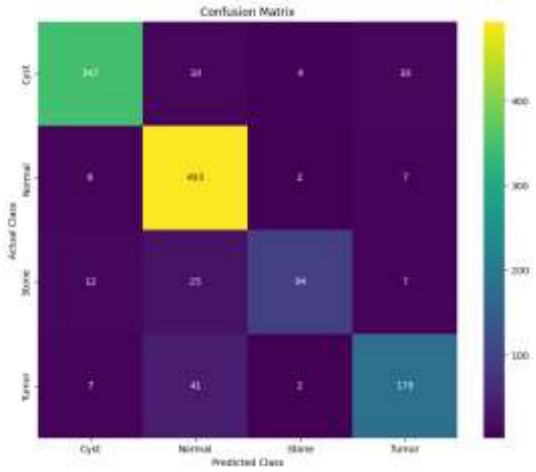
(c) Model Baseline Dataset 60-20-20

Figure 4. The training process shows the performance of the model (a) Dataset 80-10-10, (b) Dataset 70-15-15, and (c) Dataset 60-20-20

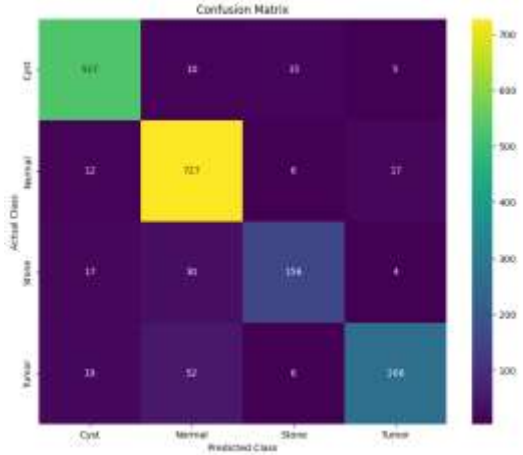
As depicted in Figure 4, the training process was closely monitored by observing the accuracy and loss graphs. The figure presents the outcomes of the MobileNet Baseline without fine-tuning and hyperparameter tuning. The resulting training curve, shown in Figure 4 (a), illustrates the model's performance over time. The outcomes indicate a training accuracy of 89.33% and a corresponding loss value of 0.3120. Figure 4 (b) The outcomes indicate a training

accuracy of 89.67% and a corresponding loss value of 0.3210. Figure 4 (c) The outcomes indicate a training accuracy of 88.36% and a corresponding loss value of 0.3573.

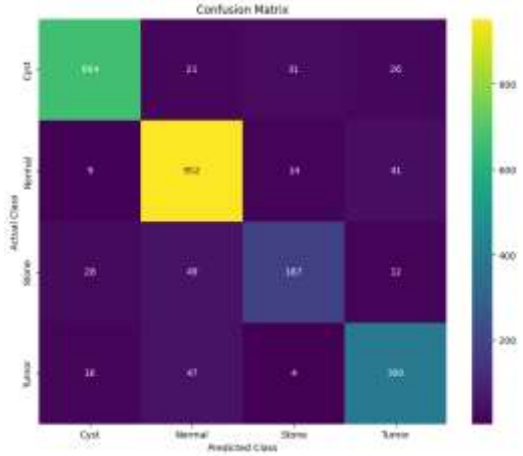
Figure 5 shows a detailed performance evaluation of the model using a confusion matrix. This matrix provides granular insight into categorization results by showing the distribution of true positives, false positives, true negatives, and false negatives. It identifies distinct misclassification patterns and illustrates the model's ability to discriminate across classes. The results support the model's reliability in practical classification tasks, particularly when assessed alongside precision, recall, and F1-score metrics.



(a) Metrics Dataset 80-10-10



(b) Metrics Dataset 70-15-15



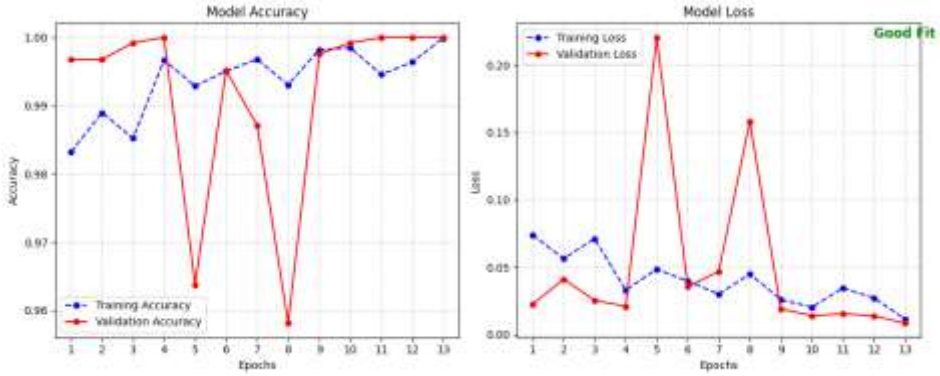
(c) Metrics Dataset 60-20-20

Figure 5. Confusion Metrics Derived from Dataset Evaluation (a) Dataset 80-10-10, (b) Dataset 70-15-15, and (c) Dataset 60-20-20

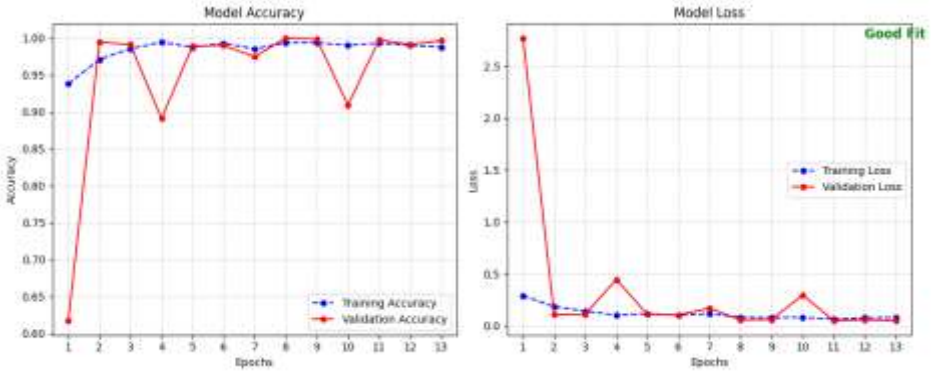
Figure 5 MobileNet Baseline Confusion Matrix (a) illustrates the classification outcomes across four classes. Within the Cyst category, the model accurately identified 347 instances, while 24 were misclassified. For the Normal class, a total of 493 images were correctly predicted, and 15 were incorrectly classified. In the case of the Stone category, the model achieved 94 correct predictions and 44 misclassifications. Lastly, for the Tumor class, 179 instances were correctly classified, whereas 50 samples were incorrectly labeled. For the confusion matrix with dataset (b) 70-15-15, there are 527 correct cyst images and 30 incorrect images, 727 correct normal images and 35 incorrect images, 156 correct stone images and 51 incorrect images, and 286 correct tumor images and 77 incorrect images. Dataset (3) 60-20-20, there are 664 correct cyst images and 78 incorrect images, 952 correct normal images and 64 incorrect images, 187 correct stone images and 81 incorrect images, and 390 correct tumor images and 67 incorrect images.

3.2. Proposed Method

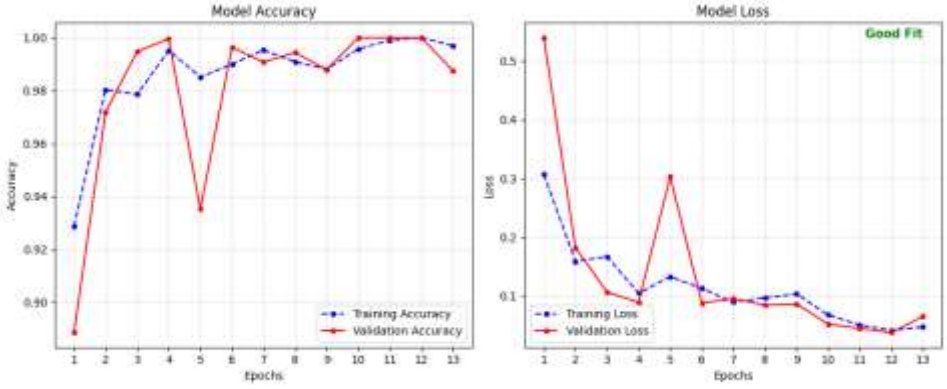
This design uses depthwise separable convolution to drastically reduce computational costs compared to traditional convolution, while maintaining comparable accuracy on test datasets such as ImageNet. The output of the proposed approach is illustrated through modifications to the CNN architecture, specifically by replacing the MobileNet Baseline, and adjusting several hyperparameters such as dropout application, regularization techniques, batch normalization, and other configurations. The base MobileNet model was trained on datasets split into 80-10-10, 70-15-15, and 60-20-20. The outcomes of the training phase, Figure 6 below:



(a) Model with Dataset 80-10-10



(b) Model with 70-15-15

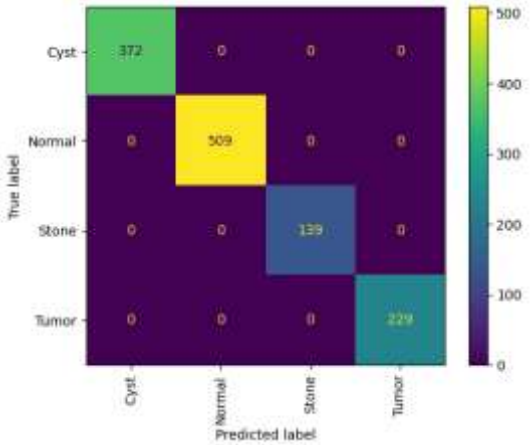


(c) Model with Dataset 60-20-20

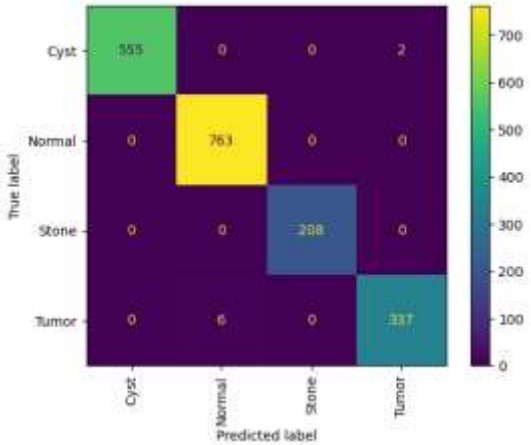
Figure 6. The Training Process Shows the Performance of the Proposed Model (a) Dataset 80-10-10, (b) Dataset 70-15-15, and (c) Dataset 60-20-20

As shown in Figure 6, the training process is carried out systematically, emphasizing accuracy and loss graphs. This training provides the results of the proposed method. The resulting training curve, shown in Figure 6 (a), illustrates the model's performance over time. The outcomes indicate a training accuracy of 100% and a corresponding loss value of 0.0083. (b) The outcomes indicate a training accuracy of 99,57% and a corresponding loss value of 0.0482. (c) The outcomes indicate a training accuracy of 99,04% and a corresponding loss value of 0.0595.

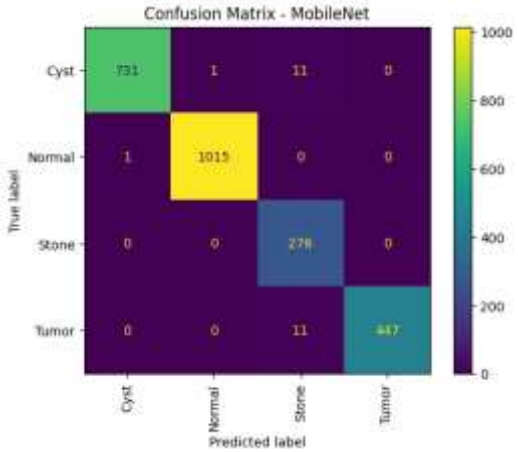
Figure 7 depicts the performance evaluation of the model using a confusion matrix. Can be seen in Figure 7 below:



(a) Proposed Method Dataset 80-10-10



(b) Proposed Method Dataset Variation 70-15-15



(c) Proposed Method Dataset Variation 60-20-20

Figure 7. Confusion Metrics Derived from Dataset Evaluation (a) Dataset 80-10-10, (b) Dataset 70-15-15, and (c) Dataset 60-20-20

Figure 7 presents the confusion matrix generated from the proposed method, offering a comprehensive view of the classification results across four distinct categories. Within the Cyst category, the model accurately identified 272 instances, while 0 were misclassified. For the Normal class, a total of 509 images were correctly predicted, and 0 were incorrectly classified. In the case of the Stone category, the model achieved 139 correct predictions and 0 misclassifications. Lastly, for the Tumor class, 229 instances were correctly classified, whereas 0 samples were incorrectly labeled. For the confusion matrix with dataset (b) 70-15-15, there are 555 correct cyst images and 2 incorrect images, 763 correct normal images and 0 incorrect images, 208 correct stone images and 0 incorrect images, and 337 correct tumor images and 6 incorrect images. Confusion matrix with dataset (3) 60-20-20, there are 731 correct cyst images and 12 incorrect images, 1015 correct normal images and 1 incorrect image, 276 correct stone images and 0 incorrect images, and 447 correct tumor images and 11 incorrect images.

3.3. Empirical Evaluation Results

The results confirm the capability of the MobileNet-based approach in managing intricate datasets while effectively learning meaningful feature representations. Using depthwise separable convolutions, MobileNet substantially minimizes computational complexity and parameter size while maintaining a high level of accuracy. MobileNet outperforms other picture classification algorithms, despite its small size. The evaluation, as shown in Table 3, comprises performance measures such as precision, recall, F1-score, and total test correctness.

Table 3. Dataset Splitting Framework

Model	Variation Dataset	Precision	Recall	F1-Score	Accuracy
MobileNet Baseline	Set 80-10-10	90%	84,25%	86,50%	89,33%
	Set 70-15-15	89,25%	85,75%	87,25%	89,67%
	Set 60-20-20	86%	84%	84,75%	88,04%
Proposed Method	**Set 80-10-10**	**100%**	**100%**	**100%**	**100%**
	Set 70-15-15	99,50%	99,50%	99,75%	99,57%
	Set 60-20-20	98,25%	99%	98,50%	99,04%

Table 3 shows, based on the classification results, MobileNet Baseline follows with the accuracy result on the 80-10-10 dataset reaching 89,67%, the 70-15-15 dataset reaching 89,33%, and the accuracy of the dataset reaching 88,04. Meanwhile, the proposed method achieved the

highest performance, reaching 100% in terms of precision, recall, F1-Score, and accuracy beyond all other models in accuracy identifying samples on the 80-10-10 dataset. These findings collectively highlight the superiority of the proposed method in image classification tasks, confirming the effectiveness of the study. The following Figure 8 comparison of classification results can be seen as a comparison chart.

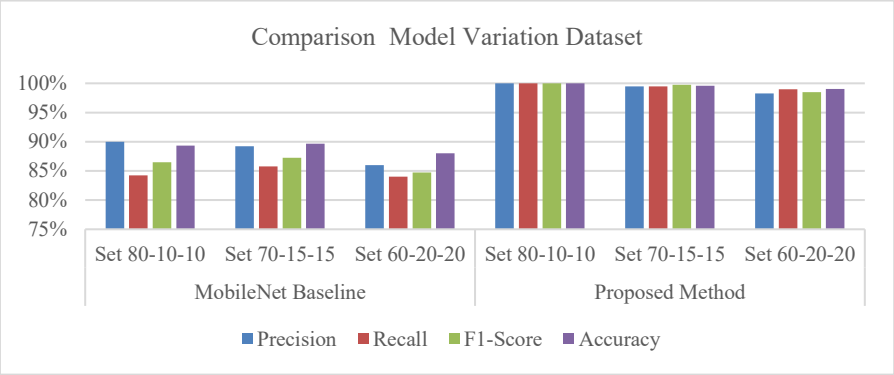


Figure 8. Chart Dataset Evaluation (a) Dataset 80-10-10, (b) Dataset 70-15-15, and (c) Dataset 60-20-20

4. CONCLUSION

This study emphasizes the significance of proper hyperparameter tuning in a convolutional neural network (CNN) architecture, particularly in diagnosing kidney tumor disorders. A comparison was conducted between the MobileNet Baseline model and the improved Proposed Method using dataset splits of 80-10-10, 70-15-15, and 60-20-20. The Proposed Method outperformed the others with the 80-10-10 dataset, achieving 100% precision, recall, F1-score, and accuracy, thereby demonstrating high reliability and accuracy in managing the challenges inherent to the classification process. Experimental results reveal that the baseline MobileNet architecture provides efficient and computationally lightweight classification performance but still falls short of obtaining ideal accuracy on complex datasets. In contrast, the proposed strategy, which applies hyperparameter tuning to MobileNet, successfully enhances the model's overall performance. This gain shows that parameter optimization is crucial for improving the model's generalization capacity in image classification tasks. This paper makes a substantial contribution to medical imaging by offering a robust and high-performance CNN-based classification framework specifically customized to effectively identify and categorize the visual characteristics of kidney cancer. Subsequent studies may examine the model's applicability across different imaging modalities and associated diagnostic problems and incorporate interpretability elements to further its clinical integration potential.

5. ACKNOWLEDGMENTS

Thank you very much for the knowledge and guidance provided by the STIKOM Tunas Bangsa Informatics lecturers in completing this paper.

REFERENCES

[1] P. E. Stevens *et al.*, "KDIGO 2024 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease," *Kidney Int.*, vol. 105, no. 4, pp. S117–S314, 2024, doi: 10.1016/j.kint.2023.10.018.

[2] A. Zarbock *et al.*, "Sepsis-associated acute kidney injury: consensus report of the 28th Acute Disease Quality Initiative workgroup," *Nat. Rev. Nephrol.*, vol. 19, no. 6, pp. 401–417, 2023, doi: 10.1038/s41581-023-00683-3.

[3] K. L. Lentine *et al.*, "OPTN/SRTR 2020 Annual Data Report: Kidney," *Am. J. Transplant.*, vol. 22, no. S2, pp. 21–136, 2022, doi: 10.1111/ajt.16982.

[4] G. Ietto, M. Gritti, G. Pettinato, G. Carcano, and D. D. Gasperina, "Tumors after kidney transplantation: a population study," *World J. Surg. Oncol.*, vol. 21, no. 1, 2023, doi: 10.1186/s12957-023-02892-3.

[5] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography," *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-15634-4.

[6] D. Alzu'Bi *et al.*, "Kidney Tumor Detection and Classification Based on Deep Learning Approaches: A New Dataset in CT Scans," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/3861161.

[7] M. Li, L. He, J. Zhu, P. Zhang, and S. Liang, "Targeting tumor-associated macrophages for cancer treatment," *Cell Biosci.*, vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1186/s13578-022-00823-5.

[8] Z. Z. Ji *et al.*, "Tumour-associated macrophages: versatile players in the tumour microenvironment," *Front. Cell Dev. Biol.*, vol. 11, no. October, pp. 1–18, 2023, doi: 10.3389/fcell.2023.1261749.

[9] B. Zhou *et al.*, "The Global, Regional, and National Burden of Pancreatitis From 1990 to 2021: A Systematic Analysis for the Global Burden of Disease Study 2021," *J. Gastroenterol. Hepatol.*, vol. 40, no. 5, pp. 1297–1306, 2025, doi: 10.1111/jgh.16906.

[10] B. L. Neuen *et al.*, "National health policies and strategies for addressing chronic kidney disease: Data from the International Society of Nephrology Global Kidney Health Atlas," *PLOS Glob. Public Heal.*, vol. 3, no. 2, pp. 1–13, 2023, doi: 10.1371/journal.pgph.0001467.

[11] G. Schiano *et al.*, "Allelic effects on uromodulin aggregates drive autosomal dominant tubulointerstitial kidney disease," *EMBO Mol. Med.*, vol. 15, no. 12, pp. 1–21, 2023, doi: 10.15252/emmm.202318242.

[12] H. Mabillard, J. A. Sayer, and E. Olinger, "Clinical and genetic spectra of autosomal dominant tubulointerstitial kidney disease," *Nephrol. Dial. Transplant.*, vol. 38, no. 2, pp. 271–282, 2023, doi: 10.1093/ndt/gfab268.

[13] E. Fasler-Kan *et al.*, "Cytokine Signaling in Pediatric Kidney Tumor Cell Lines WT-CLS1, WT-3ab and G-401," *Int. J. Mol. Sci.*, vol. 25, no. 4, 2024, doi: 10.3390/ijms25042281.

[14] J. Peng *et al.*, "Dual-targeting of artesunate and chloroquine to tumor cells and tumor-associated macrophages by a biomimetic PLGA nanoparticle for colorectal cancer treatment," *Int. J. Biol. Macromol.*, vol. 244, 2023, doi: 10.1016/j.ijbiomac.2023.125163.

[15] B. Toledo, L. Zhu Chen, M. Paniagua-Sancho, J. A. Marchal, M. Perán, and E. Giovannetti, "Deciphering the performance of macrophages in tumour microenvironment: a call for precision immunotherapy," *J. Hematol. Oncol.*, vol. 17, no. 1, pp. 1–34, 2024,

doi: 10.1186/s13045-024-01559-0.

[16] N. Karunanayake *et al.*, "Dual-Stage AI Model for Enhanced CT Imaging: Precision Segmentation of Kidney and Tumors," *Tomography*, vol. 11, no. 1, pp. 1–15, 2025, doi: 10.3390/tomography11010003.

[17] A. Abdelrahman and S. Viriri, "Kidney Tumor Semantic Segmentation Using Deep Learning: A Survey of State-of-the-Art," *J. Imaging*, vol. 8, no. 3, 2022, doi: 10.3390/jimaging8030055.

[18] V. Anand, S. Gupta, and D. Koundal, "Detection and Classification of Skin Disease Using Modified Mobilenet Architecture," *ECS Trans.*, vol. 107, no. 1, pp. 5059–5067, Apr. 2022, doi: 10.1149/10701.5059ecst.

[19] A. L. C. Ottoni, M. S. Novo, and M. S. Oliveira, "A Statistical Approach to Hyperparameter Tuning of Deep Learning for Construction Machine Classification," *Arab. J. Sci. Eng.*, vol. 49, no. 4, pp. 5117–5128, 2024, doi: 10.1007/s13369-023-08330-6.

[20] 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," *Sci. Rep.*, vol. 15, no. 1, pp. 1–20, 2025, doi: 10.1038/s41598-025-94905-2.

[21] A. Amini, M. Dolatshahi, and R. Kerachian, "Effects of Automatic Hyperparameter Tuning on the Performance of Multi-Variate Deep Learning-Based Rainfall Nowcasting," *Water Resour. Res.*, vol. 59, no. 1, 2023, doi: 10.1029/2022WR032789.

[22] S. S. L. Parvathi and H. Jonnadula, "An Efficient and Optimal Deep Learning Architecture using Custom U-Net and Mask R-CNN Models for Kidney Tumor Semantic Segmentation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, pp. 314–320, 2022, doi: 10.14569/IJACSA.2022.0130639.

[23] D. Lubis *et al.*, "Deteksi Ginjal Berdasarkan Citra CT Scan Dengan Algoritma Convolution Neural Network Kidney CYST Detection Based On CT Scan Image With Convolution Neural Network Algorithm," *J. Pendidik. dan Teknol. Indones.*, vol. 4, no. 11, pp. 369–375, 2024.

[24] M. N. Islam and M. H. K. Mehedi, "CT KIDNEY DATASET: Normal-Cyst-Tumor and Stone," 2022. https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone.

[25] M. N. Islam, M. Hasan, M. K. Hossain, M. G. R. Alam, M. Z. Uddin, and A. Soylu, "Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography," *Sci. Rep.*, vol. 12, no. 1, pp. 1–14, 2022, doi: 10.1038/s41598-022-15634-4.