Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2024, 2024:101

https://doi.org/10.28919/cmbn/8823

ISSN: 2052-2541

DEEP ENSEMBLE TRANSFER LEARNING FOR DETECTING BREAST CANCER IN HISTOPATHOLOGICAL IMAGES

INDO INTAN[1,*], BERTI JULIAN NELWAN[2], MATTHEW MARTIANUS HENRY[3], ANDREA STEVENS KARNYOTO[3], RIRIN ENDAH PUSPITASARI[2], BENS PARDAMEAN[3,4]

[1]Department of Informatics Engineering, Universitas Dipa Makassar, Makassar, 90245, Indonesia

[2]Department of Anatomical Pathology, Faculty of Medicine, Hasanuddin University, Indonesia

[3]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, 11480, Indonesia

[4]Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

Abstract. Among other cancer type, breast cancer is the leading cause of death worldwide. The traditional approach in detecting breast cancer malignancy relied on rigorous analysis, making the whole process prone to diagnostic error. This study proposes a deep learning solution to solve the problem using deep ensemble convolutional neural network (CNN) models constructed from single and smaller ensemble models. The utilized single models are ResNet50V2, InceptionResNet50V2, DenseNet201, EfficientNetB4, EfficientNetV2S, and Xception. Smaller ensemble combinations are also made from the single models. The deep ensemble models composed of ResNet50V2-EfficientNetV2S-DenseNet201, EfficientNetB4-EfficientNetV2S-Xception, EfficientNetB4-EfficientNetV2S, DenseNet201-EfficientNetB4, and ResNet50V2-DenseNet201. These models are trained using histopathological images acquired from Hasanuddin University Hospital and BreakHis with 400x magnification. Despite the data imbalance, the deep ensemble successfully obtained a 0.94 ROC-AUC score with a 0.97 average precision (AP) score, showing its capability to distinguish breast cancer malignancy from histopathological images. Further analysis revealed some distinctive patterns in the image that make the images easily

*Corresponding author

E-mail address: indo.intan@undipa.ac.id

Received August 09, 2024

classified by the deep ensemble model. This study has demonstrated that the deep ensemble CNN model constructed from smaller ensemble CNN models yields remarkable results in breast cancer detection.
Keywords: breast cancer; convolutional neural network; classification; deep learning; ensemble; histopathological images.
2020 AMS Subject Classification: 68T07, 92C55.

1. INTRODUCTION

Breast cancer is the leading cause of cancer-related deaths worldwide [1]. In 2020, there were a total of 2,261,419 cases of breast cancer, accounting for 11.7% of all cancer cases (Figure 1). The number is significantly increased, replacing lung cancer from their most prevalent disease title in years prior [2]. Additionally, the mortality rate for breast cancer remains alarmingly high, with 5,528,810 deaths reported in the same year. In 2023, there were 297,790 breast cancer incidences estimated in US alone, the highest among other cancer cases. In Southeast Asia, specifically in Indonesia, there were 47.1% female breast cancer cases and 62.1% female breast cancer mortality predicted in 2040 [3]. Recently, intronic MUTYH DNA glycosylase has been found to escalate the risk of breast cancer in Indonesian population [4]. This exceptional circumstance surrounding breast cancer is a topic of frequent discussion in health-related seminars held across the globe.

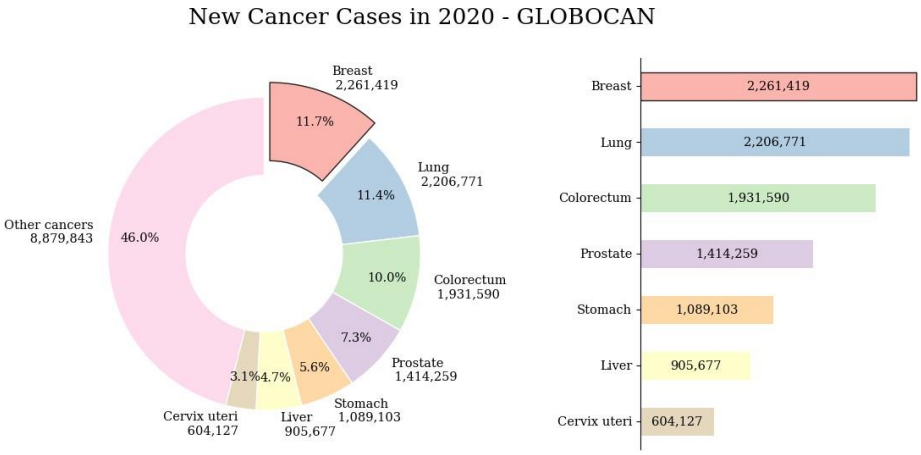


FIGURE 1. Cancer cases in 2020 according to GLOBOCAN

The pathologist approach in detecting breast cancer relied on examining microscope visualizations or the manual calculation of stained cell counts, proceeding by analyzing the data

using non-automated statistical methods. Among them, immunohistochemistry (IHC) [5] has emerged as the gold standard. This technique involves protein binding analysis in thin tissue samples obtained from surgery or biopsy to examine the structure and spread of cancer cells. However, the protocols to acquire the optimal result from IHC is strict [6]. For example, the freshly cut tissue need to be immediately examined, since the thick tissue produced background noise that hinder the IHC result. In endogenous enzyme blocking process, antigens like CD4 were destroyed by 3% diluted hydrogen peroxide (H2O2), so the hydrogen peroxide concentration need to be readjusted [6]. The analysis of IHC results–along with many other approaches alike–were also problematic due to several other factors: the scarcity of specialized physicians [7], the heavy workloads of doctors [8] that potentially leads to diagnostic errors, and reduced physician observational accuracy [9] that potentially leads to the necessity of multiple physicians repeating the examination to achieve reliable results.

The advances in deep learning method–especially in the development of transfer learning models–were helpful to thwart the previously described problems experienced by the pathologist. The ultimate goal of transfer learning is to improve the learning of target predictive function using the knowledge acquired from the source domain [10]. Transfer learning [11] has shown its capability to process numerous data, including unstructured input like text or images. The backbone of the image processing transfer learning models is convolutional neural net- work (CNN) [12]. The well-known CNN architecture are VGG [13], Inception [14], DenseNet [15], ResNet [16], MobileNets [17], and the most recent is EfficientNet [18] variants, ranging from EfficientNetB0 to EfficientNetB7. These models capabilities in processing images can be leveraged to detect cancer incidence [19, 20], classification of lung disease [21, 22], and classification of autism spectrum disorder [23]. The backbone models in [19] is EfficientNet that is combined with Squeeze and Excitation layer. In [20], the backbone model for learning the mammogram is CheXNet [24], which is a DenseNet [15] pre-trained on ChestX-ray14 dataset [25]. DenseNet is again utilized for chest radiograph classification task using COVID-19 radiographic images [22]. Different from [22], ResNet has been employed to classify whether the patient has pneumonia from lung chest x-ray images [21]. This proves the transfer learning

distinguished ability to outperform the conventional machine learning models, including tasks like breast cancer classification [26, 27]. However, CNN also suffers from bias and overfitting, since the parameter are complex and numerous. This is due to the data-hungry nature of CNN, signaling the large amount of data needed to train it [28]. CNN ensemble models, on the other hand, can reduce the bias and overfitting problem. The strategy to resist the bias and overfit- ting is to classify the class through the voting [29], stacking [30], or bagging mechanism. All mechanism yields satisfactory results [29, 30].

Nevertheless, the performance of CNN ensemble models can still be increased. In medical related fields, specifically in cancer detection, higher accuracy is preferable, since the fields leaves no room for mistake. Therefore, this study aims to build an ensemble model made not just from a single CNN as the baseline, but also included smaller ensemble CNNs. The ensemble model purpose is to classify cancer malignancy from histopathological images more accurately. The study contributes on the model ensemble novelty and the enriching population-based histopathological images used. The study materials and methods are better described in Section 2. The study result, along with the discussions is outlined in Section 3. The last section concludes the study.

2. MATERIAL AND METHODS

The study workflow comprises five stages, starting from data acquisition, proceeded by data preprocessing, data augmentation, model selection, and lastly, ensemble model building and evaluation. The workflow is summarized in Figure 2.

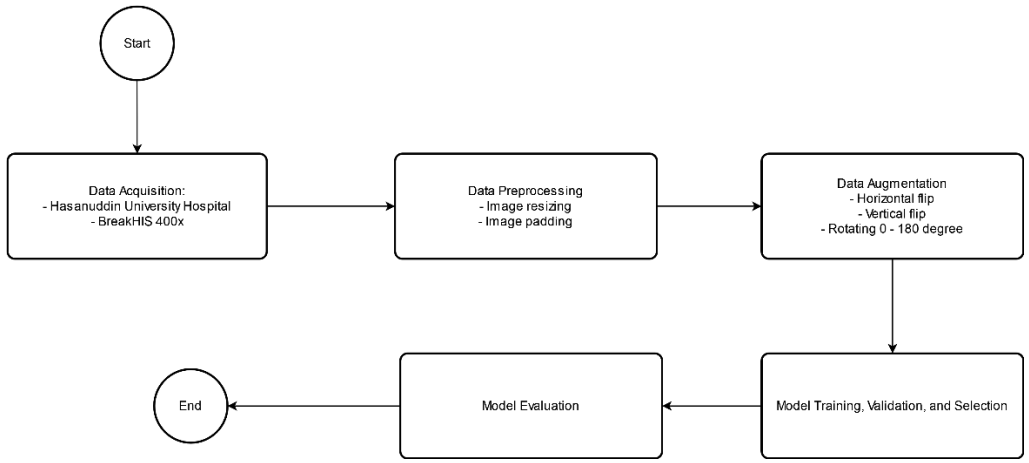
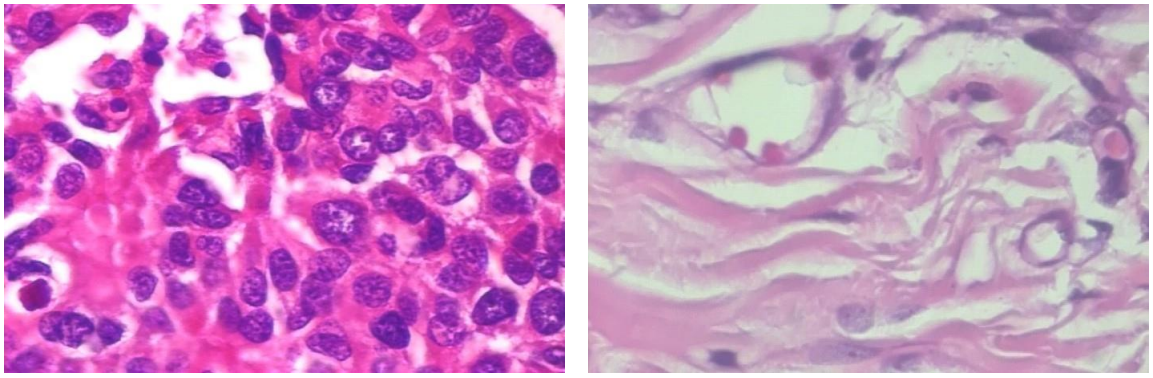


FIGURE 2. Flowchart of the proposed work

2.1. Data Acquisition

The histopathological image dataset used in this study is acquired from Hasanuddin University Hospital. The cancer malignancy from each image was classified manually by experts from Hasanuddin University Hospital. The resulting primary dataset consists 161 histopathological images split into 81 train images and 80 test images. The sample malignant and benign image is shown in Figure 3a and Figure 3b, respectively. Since the number of images is scarce, the histopathological images from Breast Cancer Histopathological Image Classification (BreakHis) [31] with 400X magnification is added into the primary dataset. The addition is based according to the expert clarification on the malignancy and the certainty of tumor, hence not all 400X images from BreakHis is attached. Only 1148 train and 545 test images from BreakHis 400x are selected. The final dataset consists of 1693 images, with 1145 benign im- ages and 548 malignant images. The dataset is further split into training, validation, and test set, consisting of 918, 230, and 545 images, respectively. The original images dimensions were then resized into 224x224 pixels resolution.



(a) (b)

FIGURE 3. The malignant (a) and benign (b) breast cancer histopathological images

2.2. Data Preprocessing and Augmentation

The original images dimensions were then resized into 224x224 pixels resolution. To ensure compatibility with the model, the padding technique was used to add black frames at the top and bottom of the images. These frames were added on a scale of 75.

The final dataset has imbalanced class distributions. To resolve the issue, augmentation techniques were employed. The original images were flipped vertically and horizontally to introduce variations. The original images were also rotated within a range of 0 - 180 degrees.

2.3. Baseline Model Training, Validation, and Selection

The preprocessed and augmented histopathological images were then fed into each CNN base models. The CNN models utilized were ResNet50V2 [32], DenseNet201 [15], InceptionResNetV2 [33], EfficientNetB4 [18], EffectiveNetV2S [34], and Xception [35]. The models are optimized with Adam [36] optimizer and a learning rate of 0.001 with weight decay. The dual or three model ensemble combinations from the single baseline were also incorporated, making the baseline of the ensemble model in this study composed of both single and ensemble models. Soft voting mechanism [10] was used in the baseline ensemble models. Each baseline was trained and validated for 35 epochs using the training and validation set, with 28 images in each batch and an early stopping applied to prevent overfitting. After training process, the baselines, be it a single CNN or ensemble CNN, were used to classify the cancer malignancy on the test set. The final classification result from each baseline were assessed using the ROC-AUC score [37]. The best five baseline models were chosen and ensembled.

2.4. Final Ensemble Model

The final ensembled CNN models were then used together to classify the histopathological images in the test set. The final decision also based on soft voting, as has been described earlier. Following that the false positives in breast cancer detection is fatal, added with the data imbalance, AP score [38] is employed to assist the ROC-AUC in further analysis. AP score is calculated through the balance of precision and recall represented by the area under the precision-recall (PR) curve, where more area under the curve indicates a good balance between the two metrics.

3. Results and Discussions

3.1. Baseline Model Training and Evaluation

From Figure 4, the lowest training loss is achieved by InceptionResNetV2 after 35 epochs

elapsed. The other CNN models were stopped earlier with training loss above the InceptionResNetV2. Among other models with early stopping applied, DenseNet-201 achieved a relatively lower loss than the other early-stopped CNN models. This suggests that DenseNet-201 can distinguish the cancer malignancy faster from the image, with loss similar to InceptionResNet50V2 and better than other CNN models. Analyzed from the losses, all the single CNN models achieve superior performance on the training set and have the potential to be a better classifier when ensembled.

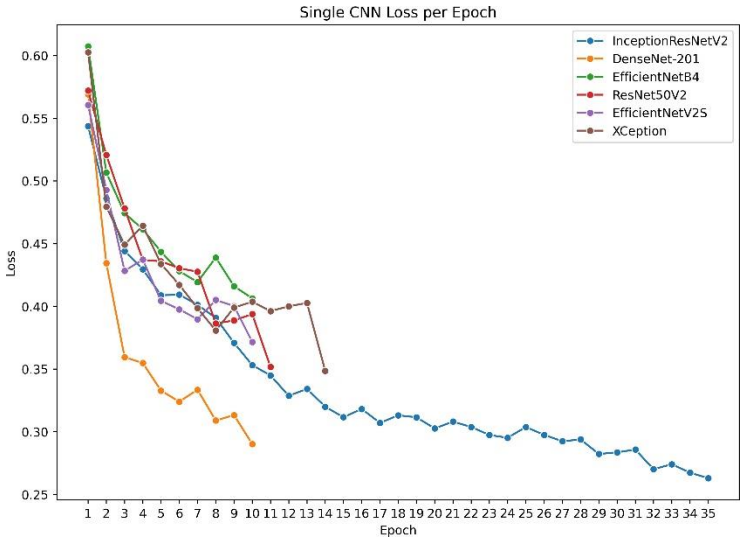


FIGURE 4. Loss from Single CNN Models

However, as examined in Table 1, InceptionResNet50V2 only achieved an ROC-AUC score of 0.5026. This score shows that InceptionResNet50V2 has overfitted the training set. DenseNet201, on the other hand, serves as the single CNN model with the highest ROC-AUC scor during vali- dation, followed by the Xception and ResNet50V2. The ROC-AUC score from the three models only differs slightly, with DenseNet201 showing its superiority by converging with less training epoch. Both EfficientNetB4 and EfficientNetV2S variants were also trained with the same number of epochs as DenseNet201, even though their ROC-AUC score is below the three best single models.

The performance of some best-performing dual or three ensemble models is shown in Table 1.

TABLE 1. Model performance on the validation set

Model Type	Model(s)	Epoch Elapsed	ROC-AUC
	ResNet50V2	10	0.9067107
	InceptionResNetV2	35	0.5026
DenseNet201		10	0.9388382
Single	EfficientNetB4	10	0.8715361
	EfficientNetV2S	10	0.8973904
	Xception	14	0.9126721
	ResNet50V2-DenseNet201	11	**0.9566492**
Dual-	DenseNet201-EfficientNetB4	9	**0.9679375**
Ensemble	EfficientNetB4-EfficientNetV2S	10	**0.9700793**
	EfficientNetV2S-Xception	13	0.9525312
Multi-	EfficientNetB4-EfficientNetV2S-Xception	13	**0.9708004**
Ensemble	ResNet50V2-EfficientNetV2S-DenseNet201	9	**0.9747339**

In the dual ensemble model, the EfficientNetB4 and EfficientNetV2S ensemble achieve the best ROC-AUC score, while in the three-model ensemble, the highest ROC-AUC score is achieved by the combination of ResNet50V2, EfficientNetV2S, and DenseNet201. The dual or multi-ensemble model omits InceptionResNetV2 since the single model performance is poor in the training set (Table 1). The best ROC-AUC is obtained by five ensemble models: ResNet50V2-EfficientNetV2S-DenseNet201, EfficientNetB4-EfficientNetV2S-Xception, EfficientNetB4-EfficientNetV2S, DenseNet201-EfficientNetB4, and ResNet50V2-DenseNet201. These five ensemble models are combined to create the final ensemble classifier.

3.2. Model Evaluation

The baseline performance on the test set is shown in Table 2. On the test set, the baseline models from single to multi-ensemble CNN shows linear result, with best performing model that attained an excellent performance score during validation also attaining good performance scores

during testing. Not all five of the best models during training successfully achieved the best scores in the test set, as dual ensemble EfficientNetB4-EfficientNetV2S performance is lower than EfficientNetV2S-Xception. However, the difference is slight, suggesting that incorporating EfficientNetV2S-Xception in the final ensemble model instead of EfficientNetB4-EfficientNetV2S makes no significant difference.

TABLE 2. Model performance on the test set

Model Type	**Model(s)**	**ROC-AUC**
Single	ResNet50V2	0.86339
	InceptionResNetV2	0.5
	DenseNet201	0.92343
	EfficientNetB4	0.86386
	EfficientNetV2S	0.81967
	Xception	0.85892
Dual-Ensemble	ResNet50V2-DenseNet201	0.92054
	DenseNet201-EfficientNetB4	0.94036
	EfficientNetB4-EfficientNetV2S	0.91122
	EfficientNetV2S-Xception	0.91299
Multi-Ensemble	ResNet50V2-DenseNet201-EfficientNetB4	0.93106
	EfficientNetB4-EfficientNetV2S-Xception	0.91567
Proposed Model (Ensemble of ResNet50V2-EfficientNetV2S-DenseNet201, EfficientNetB4-EfficientNetV2S-Xception, EfficientNetB4-EfficientNetV2S, DenseNet201-EfficientNetB4, and ResNet50V2-DenseNet201)		**0.94736**

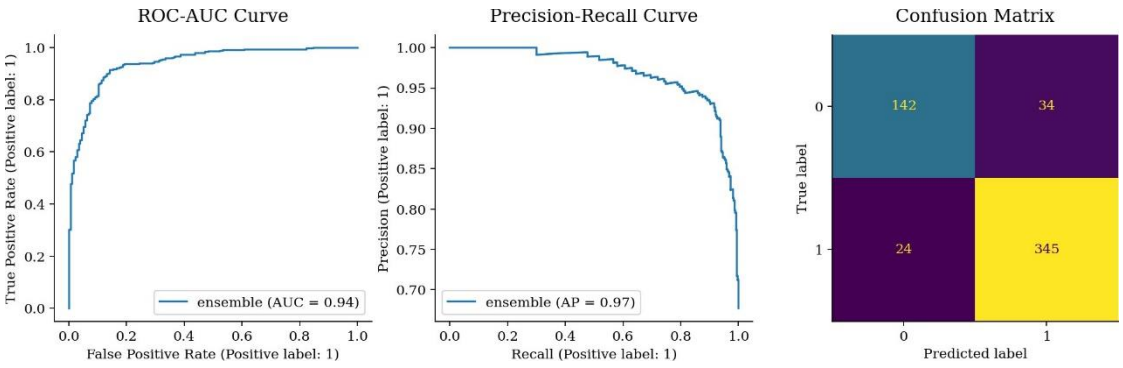


FIGURE 5. ROC-AUC and PR curve from the final ensemble model classification

The final ensemble model achieved best 0.94736 ROC-AUC score, with 0.97195 AP score, as shown in Figure 5. The ROC-AUC metric shows that cancer malignancy is easily distinguished from the images by the ensemble model. This finding is supported by the AP curve, where the balanced between precision and recall is shown, even though the class distribution is imbalanced. The final ensemble model classification samples are also shown in Figure 6.

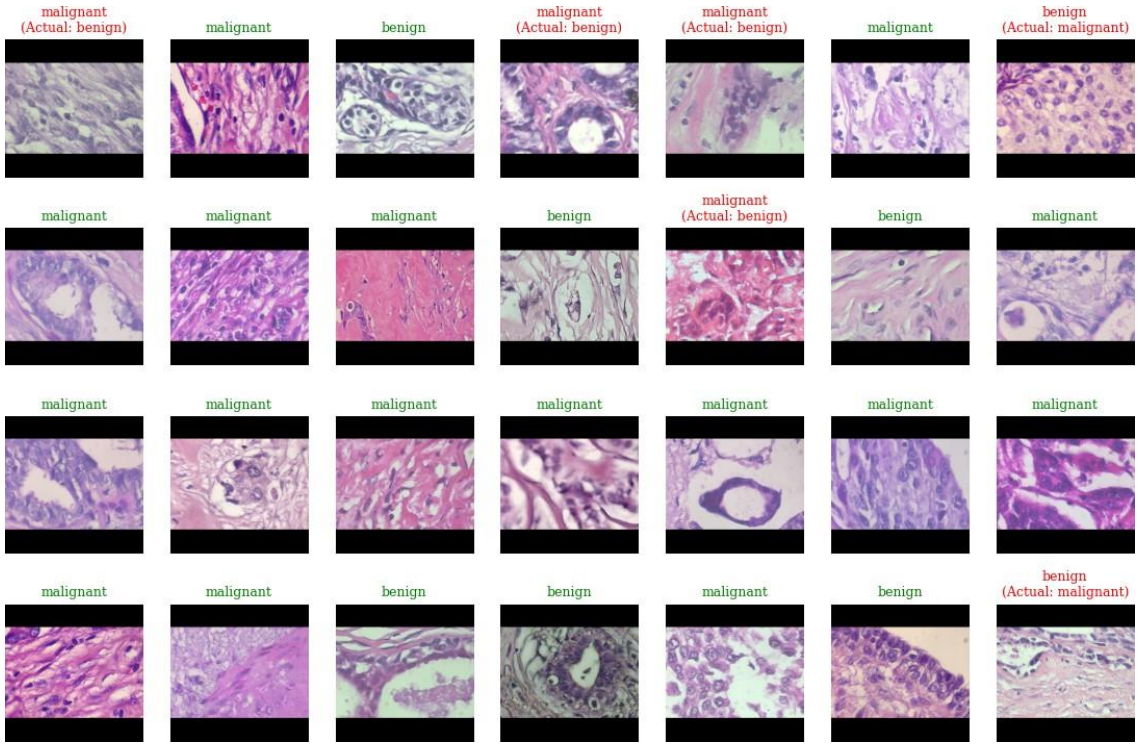


FIGURE 6. Classification samples from the final ensemble model

3.3. Performance Analysis of Single and Ensemble Model

Table 1 and Table 2 show that ensemble classifiers, whether dual or multi-ensemble, always obtain higher ROC-AUC scores than the single CNN models. In the dual ensemble model, the ensemble of both EfficientNet (EfficietNetB4 and EfficientNetV2S) gained remarkable performance on the validation set despite their single model performance. This result suggests that both single EfficientNet models capture different patterns in the image that complement each other when combined. However, relative to another dual ensemble model, the ROC-AUC score of EfficientNetB4-EfficientNetV2S decreased by 6% at the test set, while the ensemble that contains DenseNet-201 decreased by only 2% - 3% (Table 2). This proves the DenseNet model's capability in preventing overfitting [15], which is shown not only by the less epoch

elapsed (Figure 4) but also when ensembled with other single CNN models (Table 1 and Table 2).

The ensemble of ResNet50V2-DenseNet201-EfficientNetB4 attained a higher ROC-AUC score in the training set, exhibiting its prowess in distinguishing cancer malignancies almost perfectly. Nevertheless, the performance on the set dropped, with the dual ensemble of DenseNet201- EfficientNetB4 even achieving higher scores. This condition suggests that when smaller ensemble models are combined, the models are more prone to overfit even with small epochs; hence, early stopping must be applied with stricter patience than the single models.

Despite the slight difference, the final ensemble model comprising the smaller ensemble models achieved an even higher ROC-AUC score than the multi-ensemble models in both the training and test sets. This performance suggests that combining smaller ensemble models leads to better performance. With increasing dataset size, the deep ensemble model demonstrated a prime choice among the smaller ensembles and even single CNN models since a higher number of models serves as an expert better than a single model.

3.4. Identification of Misclassification Causes

Regardless of the class distribution imbalance, the malignancy is easily distinguished, even with single CNN models. It showcases that the images in the dataset have distinctive patterns that differ between benign and malignant tumors. A pattern observed from malignant images is the irregular spreads of dark purple dots (Figure 6, first row, rightmost column) or an empty white circle (Figure 6, second row, leftmost column). If this pattern somehow exists in the benign image, the image is mistakenly classified by the model as malignant (Figure 6, first row, fourth column from the left). Different from the malignant images, in benign images, the black dots are usually neatly arranged either inside the circle cell (Figure 6, first row, third column from the left), circulating the white cell (Figure 6, last row, fourth column from the left), or localized in a particular area (Figure 6, last row, third column from the left). Similarly, in malignant images, if the distinctive pattern of benign images exists, the malignant images are mistakenly classified (Figure 6, last row, rightmost column). This distinctive pattern makes the images easily classified despite the unbalanced class distribution. Still, some images possessed a distinctive pattern from other classes, suggesting that further analysis is needed to clarify the

final ensemble model result.

4. Conclusion and Future Works

The conventional approach done by pathologists for breast cancer detection involves rigorous analysis, which is problematic due to the physician's heavy workload, which can lead to diagnostic error. In this study, a deep ensemble CNN model is constructed to thwart the problem. ResNet50V2, InceptionResNet50V2, DenseNet201, EfficientNetB4, EfficientNetV2S, Xception, and small ensembles made from a combination of these single models are candidates for the deep ensemble model baselines. The models are trained using histopathological images from Hasanuddin University Hospital and BreakHis histopathological images with 400x magnification. The final ensemble comprised of ResNet50V2-EfficientNetV2S-DenseNet201, EfficientNetB4-EfficientNetV2S-Xception, EfficientNetB4-EfficientNetV2S, DenseNet201-EfficientNetB4, and ResNet50V2-DenseNet201 ensembles. The ensembles achieved a 0.94 ROC-AUC score with a 0.97 AP score, suggesting that the model easily distinguished benign or malignant cancer from the given image despite the class imbalance. It is observed that the utilized histopathological images have a distinctive pattern that differs between the classes. The ROC-AUC score from the deep ensemble model is higher than the single or smaller CNN models. The deployment of the final ensemble model in an application is the direction for future study.

Acknowledgement

Special thanks to Directorate General of Higher Education, Research, and Technology, Indonesia for support in funding of this research. This research was funded by grant number 0557/E5.5/AL.04/2023.

Conflict of Interests

The authors declare that there is no conflict of interests.

REFERENCES

[1] M. Arnold, E. Morgan, H. Rumgay, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040, The Breast 66 (2022), 15–23. https://doi.org/10.1016/j.breast.2022.08.010.

[2] H. Sung, J. Ferlay, R.L. Siegel, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: Cancer J. Clin. 71 (2021), 209–249. https://doi.org/10.3322/caac.21660.

[3] B. Ng, H. Puspitaningtyas, J.A. Wiranata, et al. Breast cancer incidence in Yogyakarta, Indonesia from 2008–2019: A cross-sectional study using trend analysis and geographical information system, PLoS ONE 18 (2023), e0288073. https://doi.org/10.1371/journal.pone.0288073.

[4] S. Panigoro, E. Listiyaningsih, I. Nurlaila, et al. Intronic variant of MUTYH gene exhibits a strong association with early onset of breast cancer susceptibility in Indonesian women population, Asian Pac. J. Cancer Prev. 22 (2021), 3985–3991. https://doi.org/10.31557/apjcp.2021.22.12.3985.

[5] S. Magaki, S.A. Hojat, B. Wei, et al. An introduction to the performance of immunohistochemistry, in: W.H. Yong (Ed.), Biobanking, Springer, New York, 2019: pp. 289–298. https://doi.org/10.1007/978-1-4939-8935-5_25.

[6] S.W. Kim, J. Roh, C.S. Park, Immunohistochemistry for pathologists: Protocols, pitfalls, and tips, J. Pathol. Transl. Med. 50 (2016), 411–418. https://doi.org/10.4132/jptm.2016.08.08.

[7] X. Zhang, D. Lin, H. Pforsich, et al. Physician workforce in the United States of America: Forecasting nationwide shortages, Human Resour. Health 18 (2020), 8. https://doi.org/10.1186/s12960-020-0448-3.

[8] S.A. Al-Dabbagh, H.M. Sulaiman, N.A. Abdulkarim, Workload assessment of medical doctors at primary health care centers in the Duhok governorate, Human Resour. Health 19 (2022), 117. https://doi.org/10.1186/s12960-021-00664-2.

[9] J.G. Elmore, R.L. Barnhill, D.E. Elder, et al. Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy and reproducibility study, BMJ 357 (2017), j2813. https://doi.org/10.1136/bmj.j2813.

[10] Y. Zheng, C. Li, X. Zhou, et al. Application of transfer learning and ensemble learning in image-level classification for breast histopathology, Intell. Med. 3 (2023), 115–128. https://doi.org/10.1016/j.imed.2022.05.004.

[11] F. Zhuang, Z. Qi, K. Duan, et al. A comprehensive survey on transfer learning, Proc. IEEE 109 (2021), 43–76. https://doi.org/10.1109/jproc.2020.3004555.

[12] K. O'Shea, R. Nash, An introduction to convolutional neural networks, preprint, (2015). https://doi.org/10.48550/ARXIV.1511.08458.

[13] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, (2015). https://doi.org/10.48550/ARXIV.1409.1556.

[14] C. Szegedy, W. Liu, Y. Jia, et al. Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Boston, MA, USA, 2015: pp. 1–9. https://doi.org/10.1109/CVPR.2015.7298594.

[15] G. Huang, Z. Liu, L. Van Der Maaten, et al. Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017: pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243.

[16] K. He, X. Zhang, S. Ren, et al. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016: pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

[17] A.G. Howard, M. Zhu, B. Chen, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications, preprint, (2017). https://doi.org/10.48550/ARXIV.1704.04861.

[18] M. Tan, Q.V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, preprint, (2019). https://doi.org/10.48550/ARXIV.1905.11946.

[19] A.M.K. Izzaty, T.W. Cenggoro, G.N. Elwirehardja, et al. Multiclass classification of histology on colorectal cancer using deep learning, Commun. Math. Biol. Neurosci. 2022 (2022), 67. https://doi.org/10.28919/cmbn/7529.

[20] B. Pardamean, T.W. Cenggoro, R. Rahutomo, et al. Transfer learning from chest X-ray pre-trained convolutional neural network for learning mammogram data, Procedia Computer Sci. 135 (2018), 400–407. https://doi.org/10.1016/j.procs.2018.08.190.

[21] E. Selvano, A.Y. Paulindino, G.N. Elwirehardja, et al. Evaluating self-supervised pre-trained vision transformer on imbalanced data for lung disease classification, ICIC Express Lett. Part B: Appl. 15 (2024), 83–89. https://doi.org/10.24507/icicelb.15.01.83.

[22] H.H. Muljo, B. Pardamean, K. Purwandari, et al. Improving lung disease detection by joint learning with COVID-19 radiography database, Commun. Math. Biol. Neurosci. 2022 (2022), 1. https://doi.org/10.28919/cmbn/6838.

[23] N. Dominic, Daniel, T.W. Cenggoro, et al. Transfer learning using inception-ResNet-v2 model to the augmented neuroimages data for autism spectrum disorder classification, Commun. Math. Biol. Neurosci. 2021 (2021), 39. https://doi.org/10.28919/cmbn/5565.

[24] P. Rajpurkar, J. Irvin, K. Zhu, et al. CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning, preprint, (2017). https://doi.org/10.48550/ARXIV.1711.05225.

[25] X. Wang, Y. Peng, L. Lu, et al. ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017: pp. 3462–3471. https://doi.org/10.1109/CVPR.2017.369.

[26] S. Boumaraf, X. Liu, Y. Wan, et al. Conventional machine learning versus deep learning for magnification dependent histopathological breast cancer image classification: A comparative study with visual explanation, Diagnostics 11 (2021), 528. https://doi.org/10.3390/diagnostics11030528.

[27] S. Sharma, R. Mehra, Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight, J. Digit. Imaging. 33 (2020), 632–654. https://doi.org/10.1007/s10278-019-00307-y.

[28] L. Alzubaidi, J. Zhang, A.J. Humaidi, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data 8 (2021), 53. https://doi.org/10.1186/s40537-021-00444-8.

[29] O. Lantang, G. Terdik, A. Hajdú, et al. Comparison of single and ensemble-based convolutional neural networks for cancerous image classification. Ann. Math. Inf. 54 (2021), 45–56. https://doi.org/10.33039/ami.2021.03.013.

[30] L.F. de J. Silva, O.A.C. Cortes, J.O.B. Diniz, A novel ensemble CNN model for COVID-19 classification in computerized tomography scans, Results Control Optim. 11 (2023), 100215. https://doi.org/10.1016/j.rico.2023.100215.

[31] F.A. Spanhol, L.S. Oliveira, C. Petitjean, et al. A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (2016), 1455–1462. https://doi.org/10.1109/tbme.2015.2496264.

[32] B. Leibe, J. Matas, N. Sebe, M. Welling, eds., Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV, Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-46493-0.

[33] C. Szegedy, S. Ioffe, V. Vanhoucke, et al. Inception-v4, inception-ResNet and the impact of residual connections on learning, Proc. AAAI Conf. Artif. Intell. 31 (2017), 4278-4284. https://doi.org/10.1609/aaai.v31i1.11231.

[34] M. Tan, Q.V. Le, EfficientNetV2: Smaller models and faster training, preprint, (2021). https://doi.org/10.48550/ARXIV.2104.00298.

[35] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, 2017: pp. 1800–1807. https://doi.org/10.1109/CVPR.2017.195.

[36] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, preprint, (2014). https://doi.org/10.48550/ARXIV.1412.6980.

[37] T. Fawcett, An introduction to ROC analysis, Pattern Recogn. Lett. 27 (2006), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

[38] P. Henderson, V. Ferrari, End-to-end training of object class detectors for mean average precision, in: S.-H. Lai, V. Lepetit, K. Nishino, Y. Sato (Eds.), Computer Vision – ACCV 2016, Springer International Publishing, Cham, 2017: pp. 198–213. https://doi.org/10.1007/978-3-319-54193-8_13.