



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2022, 2022:67

<https://doi.org/10.28919/cmbn/7529>

ISSN: 2052-2541

## MULTICLASS CLASSIFICATION OF HISTOLOGY ON COLORECTAL CANCER USING DEEP LEARNING

AL MIRA KHONSA IZZATY<sup>1,\*</sup>, TJENG WAWAN CENGGORO<sup>2,3</sup>, GREGORIUS NATANAEL

ELWIREHARDJA<sup>3</sup>, BENS PARDAMEAN<sup>1,3</sup>

<sup>1</sup>Computer Science Department, BINUS Graduate Program – Master of Computer Science Program, Bina Nusantara University, Jakarta 11480, Indonesia

<sup>2</sup>Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta 11480, Indonesia

<sup>3</sup>Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia

Copyright © 2022 the author(s). This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract:** Colorectal cancer (CRC) is a type of cancer that occurs in the colon or rectum, caused by cells dividing uncontrollably. Deep learning has proven to perform image recognition accurately that rivals human capabilities. This method became popular and can handle various complex image data. This paper presents a multiclass classification of histology on colorectal cancer using a Convolutional Neural Network (CNN). We propose the usage of EfficientNet with transfer learning to create high-performance learners and combine the model with the attention Squeeze and Excitation layer (SE layer). In several studies, the SE layer can improve the model by extracting essential features of the images. We compare EfficientNet with other architectures such as ResNet-101, AlexNet, and VGG16. Our experiment result achieves 97% testing accuracy, whereas NN-Ensemble-CNNs as the baseline model achieves 96.16%. The combined EfficientNet model and SE layer performed better than regular EfficientNet and other models.

---

\*Corresponding author

E-mail address: [al.izzaty@binus.ac.id](mailto:al.izzaty@binus.ac.id)

Received May 28, 2022

**Keywords:** Convolutional Neural Network (CNN); EfficientNet; squeeze and excitation layer; multiclass classification, histology slides.

**2010 AMS Subject Classification:** 62M45, 68T45, 68T05, 68U10, 78M32, 82C32, 94A08.

## 1. INTRODUCTION

Cancer is a disease caused by cell growth and rapid uncontrollability [1]. The rapid spread of cells causes cancer to become one of the deadliest diseases in the world [2]. In 2020, World Health Organization (WHO) referred to the International Agency for Research of Cancer that colorectal cancer is one of the deadliest cancers in Indonesia [3], [4]. The cause of colorectal cancer is not only genetic factors but also lifestyle habits [5]–[7]. Colorectal cancer can be diagnosed using microsatellite instability, scan colonoscopy, and histology slides that contain quantitative information [8]. Histology slides have several stages of complete examination, starting from fixation, macroscopic cuts, and Hematoxylin Eosin (H&E) staining until they can be read in microscopic [9].

Machine learning methods can extract histological information better than manual [10]. In 2016, a study was conducted using histology slides on colorectal cancer and classified into eight classes. Best results were obtained with an accuracy of 87.4% using the Radial Basis Function SVM model [11]. Deep learning can learn more complex features from the image and obtain satisfactory results, such as in predicting risk CRC [4], [12] and indicating microsatellite instability [9], [13], [14]. CNN has become more popular and is generally used for computer vision. It can learn from previous evaluation errors and repeatedly train [15], [16]. CNN's success in giving the best performance in image captioning and object detection [17]–[18]. In [19], it compared some machine learning and deep learning methods to predict multiclass histology slides CRC using the Kather-CRC-2016 dataset [11]. They proposed NN-Ensemble CNNs, which combined four CNN pre-trained models such as ResNet-101, ResNet-50, Inception-v3, and DenseNet161. CNN models give the best performance with an accuracy of 96.16%.

EfficientNet is one of the most potent lightweight models. This model successfully handles complex features and provides accurate results in binary and multiclass classification[20]. Several

studies use visual attention to enhance CNN's performance in handling complex features. Visual attention uses to reduce parameters and extract better features for each layer. The use of soft attention, recurring attention, and attention map use to classify histology slides [21]. Soft attention is used to reduce poor segmentation and improve performance [22]. The contribution of this study is to design a model by combining pre-trained EfficientNet and attention SE Layer. This attention is used to recalibrate features, using global information to highlight informative features and selectively suppress less useful ones. We compared the new model with previous research and other architectures such as ResNet-101, AlexNet, and VGG19.

## **2. RELATED WORK**

Detection of colorectal cancer has become a concern in recent years. Many models have been proposed to cover this problem; one of them is an approach to computer vision. In 2019, machine learning was used to predict the spread of lymph nodes using histology slides T1 CRC (colorectal cancer) with a random forest method approach [10]. The results showed that the prediction of colorectal cancer using the random forest method obtained an AUROC value of 0.94, while traditional readings using microscopes only obtain a value of 0.84. The nearest mean classifier was used to predict the cancer stage to predict the-year survival of 87.6% [23]. Colorectal cancer research has progressed quite rapidly. In 2016 initial research for multiclass classification was proposed. This study used 5000 histology slides of colorectal cancer, consisting of eight tissue types with 625 for each class. Compare four machine learning methods, including Nearest Neighbor, Linear SVM, Radial Basic Function SVM, and Decision Tree. The result shows the radial basic function SVM as the best model with an accuracy of 87.4% [11].

Deep learning models are generally better than conventional machine learning, including the detection of histology slides in cancer [24]. The working process of deep learning is described as how the human brain works in a neural network [25], [26]. Histology slides were also used in detecting microsatellite instability, as well as examining information related to genetic changes [9], [13], [14]. The deep learning approach is used in this study with an adversarial network. The data used were CGAN histology taken from two groups of patients; the first group data were 256

patients and the second group 1,457 patients. To improve the learning process in the model, they used 10,000 synthetic CGAN histology data. The best performance results were obtained with an AUROC value of 0.77 [13]. Another research used a deep learning approach in detecting tumors [27] and predicting the survival period of colorectal cancer [8], [28]. Based on the comparison of CNN, which includes VGG19, AlexNet, SqueezeNet version 1.1, GoogleNet, and ResNet-50, the best model performance is VGG19 architecture, with an accuracy of 98.7% [6].

Over the years, more research using CNN emerged in breast and colorectal cancer classification. This research compared several architectures, including EfficientNetB0 – EfficientNetB7, ExceptionNet, InceptionNetV3, VGG16, and ResNet152. The results show that the EfficientNet variant can outperform the performance of other architectures. The best variant was EfficientNetB0, with an accuracy of 97.85% [20]. Another research detected lymph node spread in breast cancer, and they proposed Boosted EfficientNet-B3 combined with Random Center Cropping, Downsampling, Feature Fusion, and attention mechanisms. The results show that Boosted EfficientNet-B3 can outperform other architectures, especially EfficientNet-B3 [29].

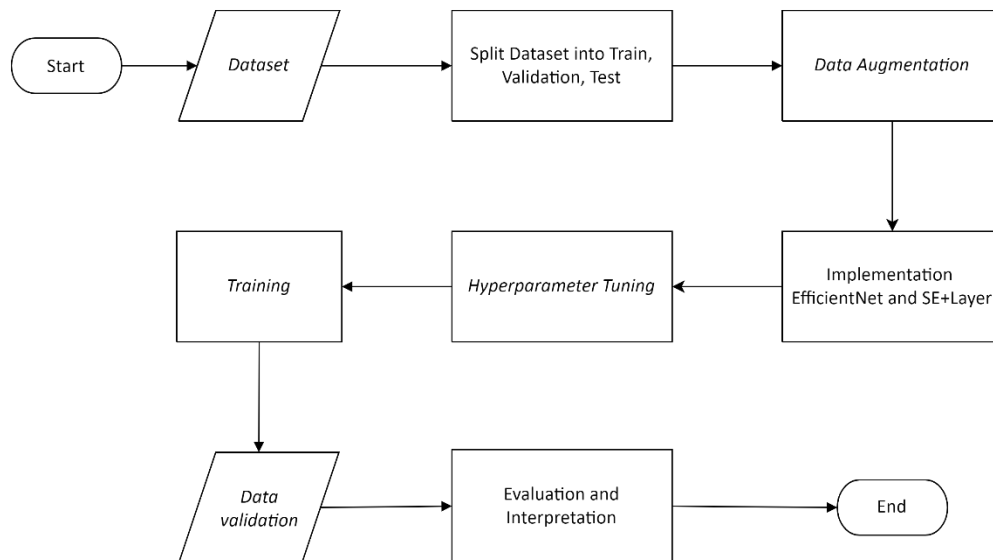
Visual attention is used to recognize and extract essential parts of the image. Visual attention was able to increase accuracy in the case of image captioning [30]. Comparison of soft attention such as pyramid features DenseNet, Squeeze and Excitation Layer (SE layer), and Residual Attention DenseNet. SE layer is designed to improve network representation by calibrating features and various experiments[31], [32]. ResNet with SE layer classified binary and multiclass classes, compared with general ResNet and proof that attention can help in increasing performance. The average accuracy results for binary class cases are between 98.87% to 99.34%, while for multiclass are 90.66% to 93.81%. The proposed method in this study is inspired by combining the pre-trained EfficientNet model [20] with SE Layer [25], [26].

### **3. METHOD**

A convolution neural network (CNN) is an architecture that is commonly used in computer vision. This research aims to build a multiclass classification model to classify histology slides of colorectal cancer. Previous studies detected colorectal cancer in a multiclass manner with a

performance of 96.16% [19]. SE Layer effectively improves network representation by extracting essential features [32]. This study proposes a novel approach to combining Efficient-Net architecture and visual attention SE-layer attention. EfficientNet with compound scaling and residual block can increase the model's efficiency. At the same time, the SE layer focus on improving the accuracy by extracting essential features in the image. The stages of this research are divided into three phases: planning, training, and model evaluation. The public dataset is downloaded and split into data train, validation, and data test at the planning stage. Data augmentation methods are used in this study, such as random resize crop, flip, and normalization.

The training stage compares the proposed method with four architectures, namely ResNet-101, AlexNet, VGG16, and EfficientNet-B0. All model adapts a pre-trained model on the image. The model with the best performance was selected as the primary model. Furthermore, hyperparameter tuning is carried out on the preceding model or setting parameters that affect optimization. The evaluation process is carried out by calculating the accuracy and F1 measure. The general flow of the system is shown in Figure 1.

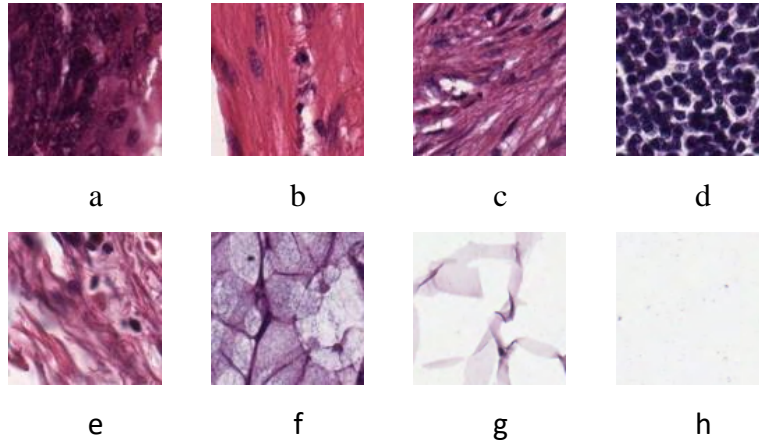


**Figure 1.** System workflow for multiclass classification on histology slide cancer colorectal.

Planning stage, Training, and Evaluation.

### 3.1 Histology Slides Cancer Colorectal

Datasets used are public datasets obtained from the Institute of Pathology, University Medical Center Mannheim, Heidelberg University, Mannheim, Germany. Datasets from Kather-CRC-2016 consisted of 5,000 histology data slides with eight classes [11]. All images are RGB 0.495m per pixel. The dataset was digitized with the Aperio ScanScope (Aperio/ Leica biosystem) at 20 times magnification. The following are examples of histology slides in cancer colorectal:



**Figure 2.** Dataset histology slides cancer colorectal [11]. The first images of all histology slides are shown. (a) tumor epithelium, (b) simple stroma, (c) complex stroma, (d) immune cell conglomerates, (e) debris, (f) normal mucosal glands, (g) adipose, (h) background

Normal tissue has a regular shape and a cluster of cells with boundaries. Whereas in Figure 2, the structure of abnormal cells is irregular, it is indicated by too many branches, and boundaries between cells are not visible. In abnormal tissue, the cell nucleus becomes more extensive and darker [33].

### 3.2 EfficientNet

CNN enhancements are used to achieve better performance. In [34] ResNet scaling increases from ResNet-18 to ResNet-200 using more layers. However, the CNN scaling process was never well understood. Several experiments were carried out to maximize model accuracy by scaling all layers with a constant ratio, formulated in Equation (1). Where  $\hat{F}_i$  are operators,  $\hat{H}_i \times \hat{W}_i$  are input resolution,  $\hat{C}_i$  is output channel, and  $\hat{L}_i$  is layer.

$$\max_{d,w,r} Accuracy(\mathcal{N}(d, w, r))$$

$$(\mathcal{N}(d, w, r) = \bigodot_{i=1..s} \hat{F}_i^{d, \hat{L}_i}(X_{(r, \hat{H}_i, r, \hat{W}_i, w, \hat{C}_i)}) \quad (1)$$

$$Memory(\mathcal{N}) \leq target\_memory$$

$$FLOPS(\mathcal{N}) \leq target\_flops$$

According to the paper [35], a balanced scaling process can improve performance. EfficientNet is a CNN architecture that scales using combined coefficients were scaling up is performed on Width (W), Height (H), and resolution, also known as compound scaling. The compound scaling is summarized in Equation (2) set, where  $\varphi$  is a global scaling factor that controls available resources.  $\alpha, \beta, \gamma$  to determine the allocation of resources to the network depth, width, and resolution. Once concluded, the definition of  $\alpha, \beta, \gamma, \varphi$  can be gradually increased to achieve better accuracy.

$$\begin{aligned} \text{depth} &= \alpha^\varphi \\ \text{width} &= \beta^\varphi \\ \text{resolution} &= \gamma^\varphi \\ \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\ \alpha \geq 1, \beta \geq 1, \gamma &\geq 1 \end{aligned} \quad (2)$$

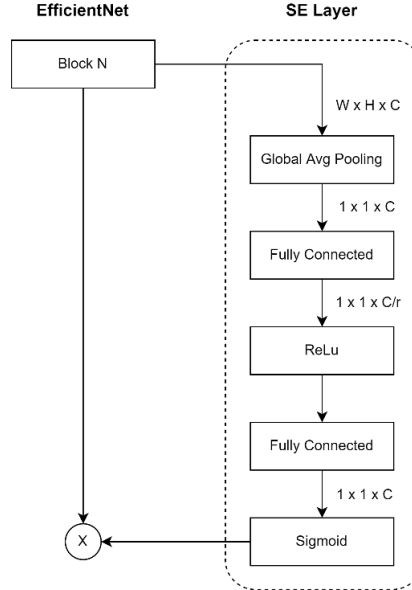
Several studies stated that EfficientNet could outperform other architectures that were considered state-of-the-art, both in terms of accuracy and efficiency [36]. For width scaling, a feature map is added for each layer. Depth scaling added layers to the network architecture. At the same time, resolution scaling increased the resolution of the input image [35].

EfficientNet architecture consists of eight variants, namely EfficientNet-B0 to EfficientNet-B7. The basis of this architecture is EfficientNet-B0 which adapts the inverted bottleneck residual block MobileNetV2 or also called MBConv. This residual block is used to increase network efficiency [37].

### 3.3 Visual Attention

Visual attention identifies essential parts of an image to obtain a frame of reference for object recognition memory [38]. Visual attention includes identifying important features, increasing accuracy, feature binding, and object recognition[39]. Visual attention consists of two types, hard and soft attention. Hard attention is visual attention that focuses on an important part of an image

or the part with the highest value compared to others. Meanwhile, soft attention results from visual attention to all the features in the image [30]. The mechanism used by soft attention is by decreasing gradient descent. Some categories of soft attention include spatial attention, channel attention, mixed attention, and self-attention [40].



**Figure 3.** Squeeze and Excitation Layer

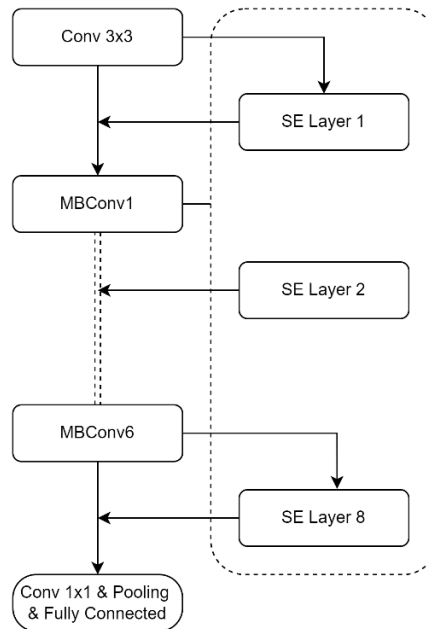
Squeeze and Excitation (SE) locate the original feature map areas by squeezing the global feature into one layer, then scaling back to their original size. This attention is used to recalibrate features, using global information to highlight informative features and selectively suppress less useful ones. To fulfill the objective of recalibration, the Sigmoid activation mechanism is used because it is flexible and can learn non-mutually exclusive relationships [31]. Mechanism with sigmoid activation formulated in the set of Equation (3), where  $\delta$  refers to ReLU function [41], and  $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$  and  $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ .

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

It works almost for all types of CNN architecture. For instance, it is combined with DenseNet and ResNet [32]. This study uses SE-Layer on the EfficientNet architecture to improve performance. SE-layer is considered easy to use in neural network architecture. EfficientNet



architecture combined with SE-Layer has an architectural description in Figure 4.



**Figure 4.** EfficientNet and SE Layer

#### 4. RESULT AND DISCUSSION

The training was carried out on 60% of the dataset or 3000 histology slides. This research was conducted using three scenarios, as shown in Table 1. This scenario determined the effect of hyperparameter tuning and visual attention on the performance of the classification model. There are three scenarios used to find the better model:

- First scenario: compare four pre-trained consists of ResNet-101, AlexNet, VGG, and EfficientNet, then choose the model with the highest accuracy as input for the second scenario
- Second scenario: fine-tune hyperparameter
- Third scenario: train proposed method with pre-trained model and hyperparameter tuning

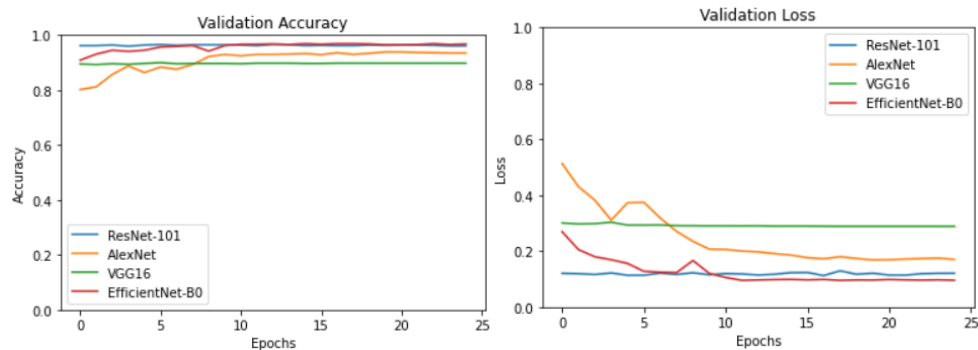
All architecture was initiated with the same parameter for the experimental setup, as shown in Table 1. Each model used a pre-trained model trained on 1000 ImageNet classes. The weights of each layer were unfrozen, so the model can continue to learn using the data train histology slides.

The learning rate used is the most used value in deep learning experiments. The decay step used is 10 for every 0.1 decreases in the learning rate.

**Table 1.** Initial Hyperparameter

Hyperparameter	Value
Batch	32
Learning rate	$10^{-3}$
Decay step	Dropped $10^{-1}$ every 10 steps
Optimizer	Adam
Epoch	25
Pre-trained	True

The result is shown in Figure 5 comparison of each architecture. EfficientNet-B0 can produce the highest accuracy values in train and validation.



**Figure 5.** Validation Accuracy and Validation Loss

The second scenario used EfficientNet to tune the hyperparameter, focusing on optimizing the model. Training used default epoch 30 and decay step 10. Trials were carried out for three different optimizer methods: SGD, Adam, and Adamax. The highest parameter in the tuning hyperparameter replaces the last parameter used in the proposed model.

**Table 2.** Hyperparameter Space

Optimizer	Decay Step	Epoch	Batch	Learning Rate	Validation	
					Accuracy	Loss
SGD	10	30	[32, 64]	$[10^{-3}, 10^{-4}]$	94.30	0.187
Adam	10	30	[32, 64]	$[10^{-3}, 10^{-4}]$	97.40	0.090
Adamax	10	30	[32, 64]	$[10^{-3}, 10^{-4}]$	97.00	0.095

The proposed model uses a hyperparameter with the highest accuracy in hyperparameter tuning. The validation results of three optimizer experiments show that a learning rate of 0.001 can produce optimal performance. Experiments with the SGD optimizer have the highest validation and loss values using batch 32, while Adam and Adamax obtain the highest performance using batch 64. Based on Table 2, batch size 64 and Adam optimizer was chosen as a hyperparameter to provide good performance for both training and validation.

Transfer learning and hyperparameter tuning are used in EfficientNet-B0 and SE Layer. Attention SE Layer is implemented in each MBConv layer, an inverted residual block on EfficientNet. The attention results in global feature maps multiplied by MBConv in element-wise multiplication. Table 3 shows validation accuracy and loss for each architecture.

**Table 3.** Comparison Models

Model	Validation Accuracy	Validation Loss
ResNet-101	96.90	0.1051
AlexNet	91.80	0.2726
VGG19	79.50	0.5558
EfficientNet-B0	97.50	0.0965
<b>EfficientNet-B0 + SE Layer</b>	<b>98.10</b>	<b>0.0698</b>

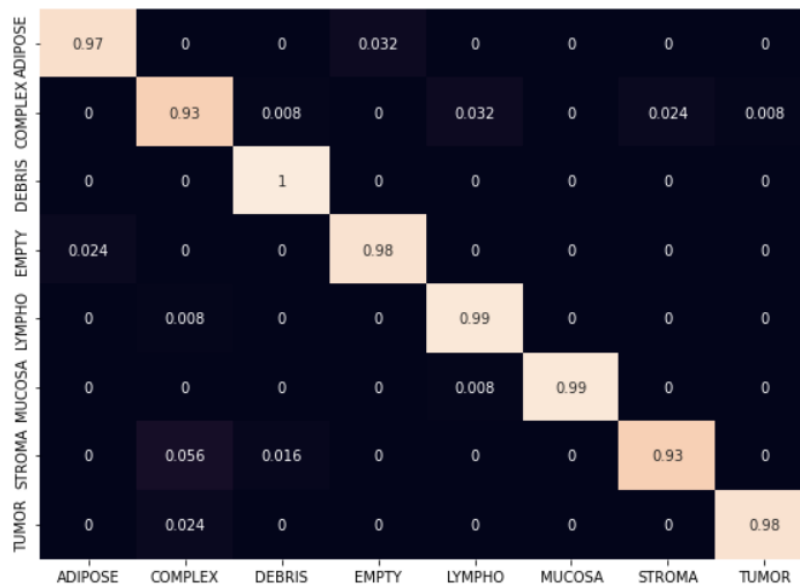
The validation results for all architectures were compared to determine the highest validation value. Table 3 showed that the EfficientNet-B0 + SE Layer has better results than another method. The highest validation accuracy of 98.10%, and the lowest validation loss with a value of 0.0698.

The testing process uses 20% dataset consisting of 1,000 histology slides. The test results showed the performances of each architecture, as shown in Table 4. Based on the accuracy, precision, recall, and F1-score values, EfficientNet-B0 has the highest performance in all aspects compared to other architectures. The experimental results showed that usage of SE Layer could support the performance of EfficientNet-B0 even when compared to previous studies, there is an increase in accuracy of 0.84%.

**Table 4.** Performance Models

Model	Testing (confusion matrix)			
	Accuracy	Precision	Recall	F1-Score
<b>ResNet-101</b>	95.91	97.61	96.61	98.28
<b>AlexNet</b>	88.42	92.80	92.50	96.11
<b>VGG19</b>	78.14	87.38	85.07	91.93
<b>EfficientNet-B0</b>	96.31	98.23	96.41	98.17
<b>EfficientNet-B0 + SE Layer</b>	<b>97.00</b>	<b>98.21</b>	<b>97.43</b>	<b>98.70</b>

The confusion matrix from EfficientNet-B0 + SE Layer compares predictions with the original target. Figure 6 shows the accuracy value for each class. Binary calculations determine the total true negative, true positive, false positive, and false negative.

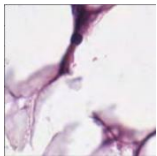
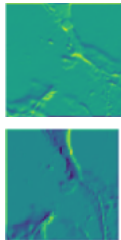
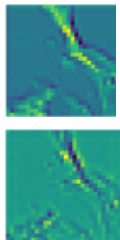
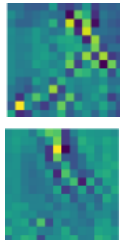
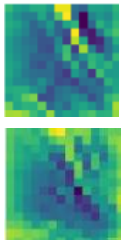
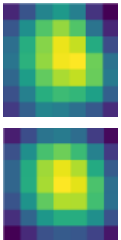
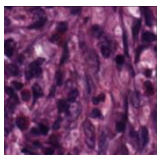
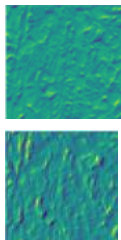
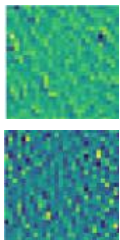
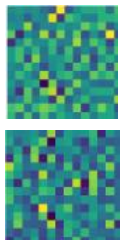
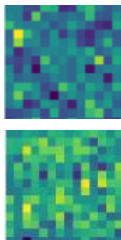
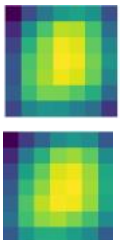
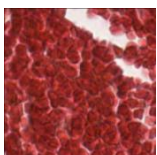
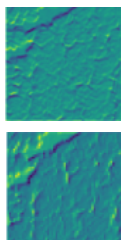
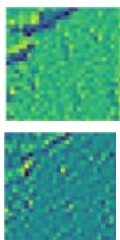
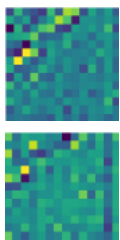
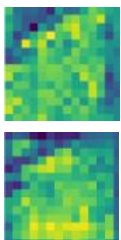



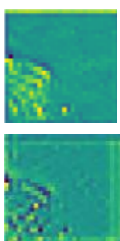
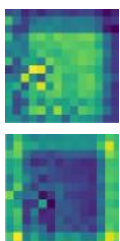
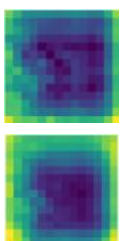
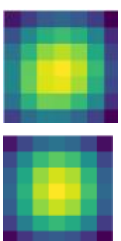


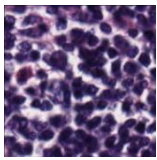
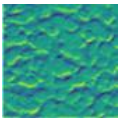
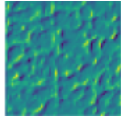
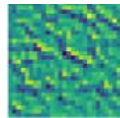
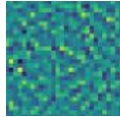
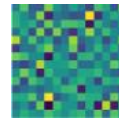
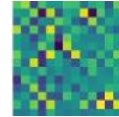
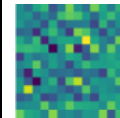
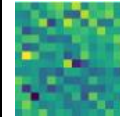
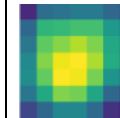

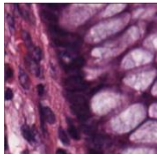
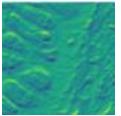
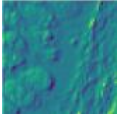
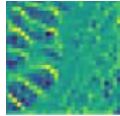
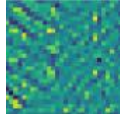
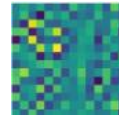
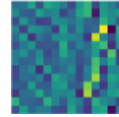
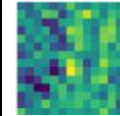
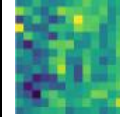
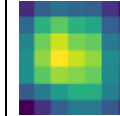

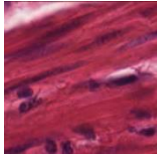
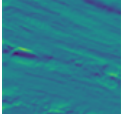
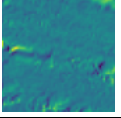
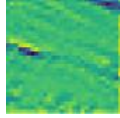
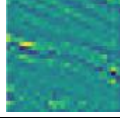
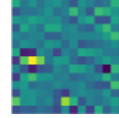
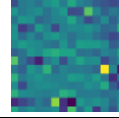
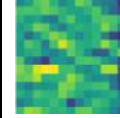
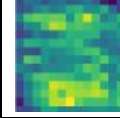

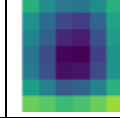
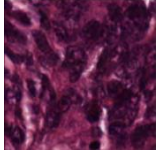
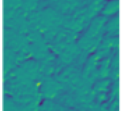
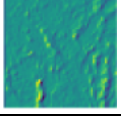
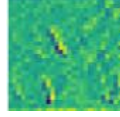
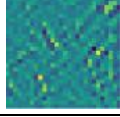
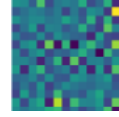
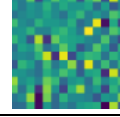
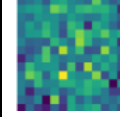
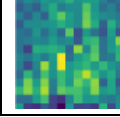
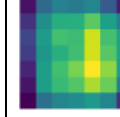
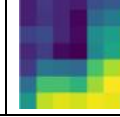
**Figure 6.** Confusion Matrix. Accuracy for each class.

## CLASSIFICATION COLORECTAL CANCER USING DEEP LEARNING

Table 5 is the result of the feature map for each block using the SE Layer. Attention helps define the essential features of each histology slide for both normal and abnormal tissues. The attention process in the first block brings up detailed features that still describe the original image. Block 2 identifies the more essential parts. In blocks 4 to 5, more detailed attention sharpens important features on the histology slides.

**Table 5.** Result Visual Attention

Class	Original Image	Blok 1 Attention	Blok 2 Attention	Blok 3 Attention	Blok 4 Attention	Blok 5 Attention
Adipose						
Complex						
Debris						
Empty						

Lympho		 	 	 	 	 
Mucosa		 	 	 	 	 
Stroma		 	 	 	 	 
Tumor		 	 	 	 	 

In the SE layer, sigmoid activation function is used to distribute the important features of the histology slides. The deeper the EfficientNet-B0 block, the SE layer extracts more relevant features to speed up model learning time. Compared with the regular EfficientNet-B0, there is a significant difference in execution time, 8 minutes 47 seconds. While from the accuracy, there is a difference of 0.39%. The results show that attention can identify essential features on histology slides of colorectal cancer, thereby helping to improve the performance and learning time of the model.

## 5. CONCLUSION

We present a systematic evaluation of different methods for multiclass classification of colorectal cancer histological data. In this study, the usage of the EfficientNet-B0 model with SE Layer attention was proposed to capture important features in the image to obtain better accuracy values. Compared to the basic architecture of EfficientNet-B0, the proposed model outperforms the accuracy value of 0.39% and the F1-score of 0.53%. The attention SE Layer can produce better performance than the model without attention. Meanwhile, compared with previous studies, the difference in accuracy is 0.84% , and the difference in F1 score is quite far, namely 11.45%. So, by using the attention model, EfficientNet-B0 can work better.

Suggestions for further research are implementing other soft attention on Efficient-Net and conducting experiments on the use of augmented data to improve the training process. This approach can be used by medical experts in diagnosing colorectal cancer and can be implemented for histology slides of other cancers.

## CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interest.

## REFERENCES

- [1] G. Mathur, S. Nain, P.K. Sharma, Cancer an Overview, Acad. J. Cancer Res. 8 (2015), 1-9.
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: A Cancer Journal for Clinicians. 68 (2018), 394–424. <https://doi.org/10.3322/caac.21492>.
- [3] R. N. Latifah, Karakteristik Klinis Penderita Kanker Kolorektal di Indonesia, Thesis, Universitas Hasanuddin 2020. <http://repository.unhas.ac.id/id/eprint/2090>.
- [4] I. Yusuf, B. Pardamean, J.W. Baurley, et al. Genetic risk factors for colorectal cancer in multiethnic Indonesians, Sci Rep. 11 (2021), 9988. <https://doi.org/10.1038/s41598-021-88805-4>.
- [5] T.W. Cenggoro, B. Mahesworo, A. Budiarto, et al. Features importance in classification models for colorectal

- cancer cases phenotype in Indonesia, *Procedia Computer Sci.* 157 (2019), 313–320.  
<https://doi.org/10.1016/j.procs.2019.08.172>.
- [6] B. Mahesworo, A. Budiarto, B. Pardamean, Systematic Evaluation of Cross Population Polygenic Risk Score on Colorectal Cancer, *Procedia Computer Sci.* 179 (2021), 344–351. <https://doi.org/10.1016/j.procs.2021.01.015>.
- [7] B. Pardamean, J.W. Baurley, C.I. Pardamean, J.C. Figueiredo, Changing colorectal cancer trends in Asians, *Int. J. Colorectal Dis.* 31 (2016), 1537–1538. <https://doi.org/10.1007/s00384-016-2564-z>.
- [8] J.N. Kather, J. Krisam, P. Charoentong, et al. Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study, *PLoS Med.* 16 (2019), e1002730.  
<https://doi.org/10.1371/journal.pmed.1002730>.
- [9] J.N. Kather, A.T. Pearson, N. Halama, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer, *Nat. Med.* 25 (2019), 1054–1056. <https://doi.org/10.1038/s41591-019-0462-y>.
- [10] M. Takamatsu, N. Yamamoto, H. Kawachi, et al. Prediction of early colorectal cancer metastasis by machine learning using digital slide images, *Computer Methods Programs in Biomed.* 178 (2019) 155–161.  
<https://doi.org/10.1016/j.cmpb.2019.06.022>.
- [11] J.N. Kather, C.-A. Weis, F. Bianconi, et al. Multi-class texture analysis in colorectal cancer histology, *Sci. Rep.* 6 (2016), 27988. <https://doi.org/10.1038/srep27988>.
- [12] S. Amadeus, T.W. Cenggoro, A. Budiarto, et al. A design of polygenic risk model with deep learning for colorectal cancer in multiethnic Indonesians, *Procedia Computer Sci.* 179 (2021) 632–639.  
<https://doi.org/10.1016/j.procs.2021.01.049>.
- [13] J. Krause, H.I. Grabsch, M. Kloor, et al. Deep learning detects genetic alterations in cancer histology generated by adversarial networks, *J. Pathol.* 254 (2021), 70–79. <https://doi.org/10.1002/path.5638>.
- [14] A. Echle, H.I. Grabsch, P. Quirke, et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning, *Gastroenterology.* 159 (2020), 1406–1416.e11. <https://doi.org/10.1053/j.gastro.2020.06.021>.
- [15] Y. LeCun, B. Boser, J.S. Denker, et al. Backpropagation applied to handwritten zip code recognition, *Neural Comput.* 1 (1989), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>.
- [16] B. Pardamean, H.H. Muljo, T.W. Cenggoro, et al. Using transfer learning for smart building management system,



- J. Big Data. 6 (2019), 110. <https://doi.org/10.1186/s40537-019-0272-6>.
- [17] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, 2012. <http://code.google.com/p/cuda-convnet/>
- [18] 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>.
- [19] E. Paladini, E. Vantaggiato, F. Bougourzi, et al. Two ensemble-CNN approaches for colorectal cancer tissue type classification, *J. Imaging.* 7 (2021), 51. <https://doi.org/10.3390/jimaging7030051>.
- [20] A. Kallipolitis, K. Revelos, I. Maglogiannis, Ensembling EfficientNets for the classification and interpretation of histopathology images, *Algorithms.* 14 (2021), 278. <https://doi.org/10.3390/a14100278>.
- [21] N. Tomita, B. Abdollahi, J. Wei, et al. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides, *JAMA Netw. Open.* 2 (2019), e1914645. <https://doi.org/10.1001/jamanetworkopen.2019.14645>.
- [22] K.H. Tam, M.F. Soares, J. Kers, et al. Predicting clinical endpoints and visual changes with quality-weighted tissue-based renal histological features, *medRxiv* 2022.03.30.22269826. <https://doi.org/10.1101/2022.03.30.22269826>.
- [23] R. Salazar, P. Roepman, G. Capella, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer, *J. Clinic. Oncol.* 29 (2011), 17–24. <https://doi.org/10.1200/jco.2010.30.1077>.
- [24] L.D. Tamang, B.W. Kim, Deep learning approaches to colorectal cancer diagnosis: a review, *Appl. Sci.* 11 (2021), 10982. <https://doi.org/10.3390/app112210982>.
- [25] 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), Article ID 1. <https://doi.org/10.28919/cmbn/6838>
- [26] Jimmy, T.W. Cenggoro, B. Pardamean, Systematic Literature Review: An Intelligent Pulmonary TB Detection from Chest X-Rays, in: 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), IEEE, Jakarta, Indonesia, 2021: pp. 136–141. <https://doi.org/10.1109/ICCSAI53272.2021.9609717>.
- [27] H. Yoon, J. Lee, J.E. Oh, et al. Tumor identification in colorectal histology images using a convolutional neural

- network, *J. Digit. Imaging.* 32 (2019), 131–140. <https://doi.org/10.1007/s10278-018-0112-9>.
- [28] L. Bottaci, P.J. Drew, J.E. Hartley, et al. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions, *The Lancet.* 350 (1997), 469–472. [https://doi.org/10.1016/s0140-6736\(96\)11196-x](https://doi.org/10.1016/s0140-6736(96)11196-x).
- [29] J. Wang, Q. Liu, H. Xie, et al. Boosted EfficientNet: detection of lymph node metastases in breast cancer using convolutional neural networks, *Cancers.* 13 (2021), 661. <https://doi.org/10.3390/cancers13040661>.
- [30] K. Xu, J. Ba, R. Kiros, et al. Show, attend and tell: neural image caption generation with visual attention, (2016). arXiv:1502.03044. <https://doi.org/10.48550/arXiv.1502.03044>.
- [31] J. Hu, L. Shen, S. Albanie, et al. Squeeze-and-excitation networks, 2017, arXiv:1709.01507. <https://doi.org/10.48550/arXiv.1709.01507>.
- [32] Y. Jiang, L. Chen, H. Zhang, et al. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module, *PLoS ONE.* 14 (2019) e0214587. <https://doi.org/10.1371/journal.pone.0214587>.
- [33] M.J. Munro, S.K. Wickremesekera, L. Peng, et al. Cancer stem cells in colorectal cancer: a review, *J. Clinic. Pathol.* 71 (2017), 110–116. <https://doi.org/10.1136/jclinpath-2017-204739>.
- [34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, (2015), arXiv:1512.03385. <https://doi.org/10.48550/arXiv.1512.03385>.
- [35] M. Tan, Q.V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, arXiv:1905.11946. <https://doi.org/10.48550/arXiv.1905.11946>.
- [36] M. Afif, R. Ayachi, Y. Said, et al. Deep learning based application for indoor scene recognition, *Neural Process Lett.* 51 (2020), 2827–2837. <https://doi.org/10.1007/s11063-020-10231-w>.
- [37] M. Sandler, A. Howard, M. Zhu, et al. MobileNetV2: Inverted residuals and linear bottlenecks, arXiv:1801.04381. <https://doi.org/10.48550/arXiv.1801.04381>.
- [38] C.H. Anderson, D.C. Van Essen, B.A. Olshausen, Directed visual attention and the dynamic control of information flow, *Neurobiol. Attention.* (2005), 11–17. <https://doi.org/10.1016/b978-012375731-9/50007-0>.
- [39] K.K. Evans, T.S. Horowitz, P. Howe, et al. Visual attention, *WIREs Cogn. Sci.* 2 (2011), 503–514. <https://doi.org/10.1002/wcs.127>.

- [40] X. Yang, An overview of the attention mechanisms in computer vision, *J. Phys.: Conf. Ser.* 1693 (2020), 012173.  
<https://doi.org/10.1088/1742-6596/1693/1/012173>.
- [41] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: *Proceedings of the 27<sup>th</sup> International Conference on Machine Learning*, Haifa, Israel, 2010.