Comparison Performance of Prostate Cell Images Classification using Pretrained Convolutional Neural Network Models

1st Yessi Jusman
Department of Electrical Engineering
Faculty of Engineering, Universitas Muhamadiyah Yogyakarta
Yogyakarta, Indonesia
*Corresponding Email: yjusman@umy.ac.id

2nd Muhammad Ahdan Fawwaz Nurkholid
Department of Electrical Engineering
Faculty of Engineering, Universitas Muhamadiyah Yogyakarta
Yogyakarta, Indonesia

3rd Dhimas Arief Darmawan
Universitas Indonesia
Jakarta, Indonesia

4th Feriandri Utomo
Faculty of Medicine and Health Sciences
Universitas Abdurrab
Pekanbaru, Indonesia

2021 IEEE Region 10 Symposium (TENSYMP) | 978-1-6654-0026-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/TENSYMP52854.2021.9550865

Abstract— **Prostate cancer is the most common cancer in men in 2019. In that year, in the United States 174,650 men (20%) had prostate cancer and the remaining 696.32 men (80%) had other cancers (lung, bronchus) etc). In cancer diagnosis, there are several problems such as errors in reporting the diagnosis and the need for a long time. Artificial intelligence has long been known to facilitate the detection process, but a comparison analysis of the model is needed to get more optimal results. This study aims to compare the performance of two pretrained models (i.e. AlexNet and GoogLeNet). The data used is the image of prostate cells taken from a light microscope at the Universitas Indonesia (UI) Hospital. This study uses k-fold cross-validation to validate the accuracy of a model used. Performance evaluation of pretrained models is based on performance metrics: accuracy, precision, recall (sensitivity), specificity and f-score and running time in the testing process. The best accuracy is obtained by GoogLeNet with 99.63% and 97.74% and the lowest accuracy is obtained by AlexNet with 99.13% and 94.11%. During the training, AlexNet had a shorter time with 47 seconds than GoogLeNet with 112 seconds. In testing times, AlexNet was also faster with 0.307 seconds than GoogLeNet with 0.372. This research is expected to assist researchers (pathologists, physician assistants, etc.) in choosing the right architecture for the classification of prostate cancer images in terms of time and accuracy.**

Keywords—prostate cells, pathology images, deep learning, classification, analysis.

I. Introduction

According to Cancer Statistics, 2019, prostate cancer is the most common cancer in men in 2019. In that year, in the United States 174,650 men (20%) had prostate cancer and the remaining 696.32 men (80%) had other cancers. (lungs, bronchi and so on). This cancer accounts for the second largest mortality rate (31,620 people (10%)) after lung cancer and bronchi (76,650 people (24%)) of the 870,970 people died from cancer [1].

The cause of prostate cancer has yet to be determined, but genetic factors are a definite cause [2]. In research conducted by Prof. Jack Cuzick et al, there are various causes of prostate cancer. The cause is divided into three factors, namely factors that cannot be changed (age, race, geography, family and genetic factors), external factors (urinary tract infections, smoking, diet, weight, and physical activity) and blood and hormonal factors in the body (such as androgen hormone levels in male reproductive organs or IGF-1 that are not good caused by genetic and environmental factors) [3].

In cancer diagnosis, science is needed to diagnose diseases diagnosed by pathologists [4], [5]. But in the diagnosis of cancer, there is a number of problems as follows. First, as stated by Shah, et al that in 1,777 prostate biopsy specimens diagnosed by nine pathologists, there was a 45% difference in diagnostic reporting [6]. In addition, the diagnosis process by pathologists in general requires a relatively long time. The time period required is 24 hours (the fastest) and usually takes three days in a standard examination [7].

At present, there are many methods to solve problems in the diagnosis of various diseases, including prostate cancer. One of them is the application of machine learning which has great potential to improve performance in diagnosing cancer [8]. In 2016, a study was conducted by Xinggang Wang et al, with the aim to compare the use of deep learning versus non-deep learning in detecting prostate cancer. The results show that machine learning has better accuracy and reliability at 84% (deep learning), higher than 70% (non-deep learning) [9].

In 2016, Google collaborated with Moorfields Eye Hospital in London, England, in developing artificial intelligence (AI) to detect eye disease. Examination of more than 50 eye diseases has an accuracy of 94% [10]. In addition, Google has also used machine learning and augmented reality to create a microscope that can detect cancer, called the Augmented Reality Microscope (ARM). The device was designed to be able to detect prostate cancer with an accuracy rate of 96% (0.92-0.99) [11].

Artificial intelligence has long been known to facilitate the process of detecting prostate cancer, but a comparative analysis of the model is needed in order to obtain more optimal results. Input from artificial intelligence is prostate cancer image which is classified using pretrained models in classifying the image. Pretrained models used include, AlexNet and GoogLeNet. This research is expected to assist researchers (pathologists, physician assistants, etc.) in choosing the right model for the detection and classification of prostate cancer.

II. METHODOLOGY

This research includes pre-processing step, dataset management using 10-fold cross validation, training process of pretrained models, and image classification. We use five classes of prostate cells images that are normal, IIA, IIC, III, and IV stages. Pre-processing is used to prepare prostate cells image data so that it can be used appropriately. This process includes image cropping, image labeling, and image resizing. We use two pretrained models (i.e. AlexNet and GoogleNet). The pretrained models training process aims to train pretrained models in order to classify images into the 5 classes. The flow chart can be shown as Figure 1.

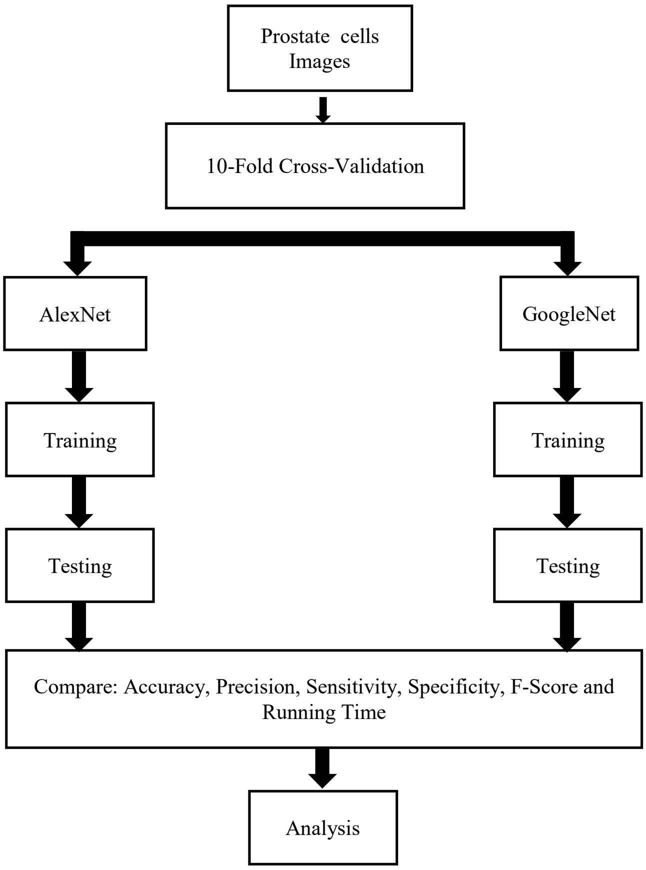

Fig. 1. Flow chart in comparison of pretrained convolutional neural network models

A. Tools and Materials

In this study, there are tools used. The tools used consist of software and hardware. The hardware used can be seen in Table 1, while the software used is MATLAB R2019a with Deep Learning (DL) Toolbox, as a framework provider for designing and implementing CNN.

TABLE I. HARDWARE SPECIFICATIONS

Hadware Memory	Characteristics
Processor	16 Gb
Graphics	Intel Core i5-9400 CPU @ 2.90 GHz
Hadware	GeForce GTX 970 4Gb

In this study using 57 prostate cell images taken from the Universitas Indonesia (UI) Hospital which had been taken care of by the code of ethics at the Institute for the code of ethics at the Hospital. Then, the image is cropped into four parts with the aim of multiplying the dataset. The total number of images after cropping is 268 images. Table 2 is the amount of the dataset after it has been cropped. The image that has been cropped is then divided into 4 staged cancer classes and one normal class.

TABLE II. THE CROPPED PROSTATE CELLS IMAGES

Classes	Number of Images
Normal	44
IIA	48
IIC	76
III	52
IV	48

All images used in the study were reduced to 227×227 pixels for AlexNet, and 224×224 pixels for GoogLeNet. In this study, we use cross-validation techniques to manage the datasets. Cross-validation is useful for assessing and validating the accuracy of a model used [12], [13]. Finally, the data is divided into two sets using the 10-fold cross-validation method, where 90% of the data in the training set and the remaining 10% in the testing set.

In this stage using pretrained models from ImageNet which have trained more than one million images and are able to classify 1000 categories of objects. This process uses fifteen epochs with a learning rate of 0.0001 (constant). Optimization algorithm used is Adam optimization with a batch size of ten. In this study using the concept of fine-tuning. Fine-tuning is a transfer learning concept, which consists of replacing the pretrained output layer with a layer that contains the number of prostate cell dataset classes. The layers that are replaced are the last three layers: a fully connected layer, a softmax layer, and a classification output layer.

B. AlexNet Pretrained Model

AlexNet is one of the pretrained models developed by Alex Krizhevsky et al. [14]. AlexNet has trained more than one million images and is able to classify images into 1000 object categories. AlexNet is considered a good feature extraction because it has studied various feature representations in various images. In 2012, AlexNet won the most difficult challenge organized by ImageNet Large Scale Visual Recognition (ILSVRC) [15]. The competition aims to evaluate the algorithm in detecting objects and image classification on a large scale.

C. GoogleNet Pretrained Model

GoogLeNet is one of the pretrained models developed by Szegedy et al. [16]. GoogLeNet is the winner of the 2014 ImageNet Large Scale Visual Recognition (ILSVRC) [17]. GoogLeNet achieved a relatively lower error rate (6.67%) compared to VGGNet and AlexNet and was included in the top-5 [18]. GoogLeNet is also like AlexNet, which has trained more than one million images and is able to classify 1000 object categories. Broadly speaking, Inception is similar to GoogLeNet but has more layers [16]. The purpose of the

Inception layer is to increase the width of the training parameters in order to get a better classification. The increase in the number of parameters results in the training progress time required for Inception being longer than GoogLeNet. GoogLeNet then introduced the Inception module to reduce parameters very much because it uses a small convolution layer.

D. *Analysis*

This output layer contains the number of prostate cell dataset classes. Each output has a probability to classify images because each model has an automatic ability to study datasets in the training data process; then the model chooses the highest probability as a class prediction [14]. This output layer uses fine-tuning where the last three layers are replaced by the number of prostate cell dataset classes.

Assessment based on accuracy is usually not enough to judge the performance of a model. There is a comparison of the time needed to assess the efficiency and other performance metrics such as precision, sensitivity, specificity, f-score, accuracy using confusion matrix [19]. Confusion matrix contains a summary of the performance of the model used. Confusion matrix itself can provide an assessment of a model through accuracy.

In a confusion matrix, each column is a predicted class and each row is an actual class representing each class. The confusion matrix, which consists of many classes, has different False Positives (FP) and False Negatives (FN), so it is necessary to know the positive and negative definitions of each class. The formula for calculating performance metrics for one class can be seen in [20].

III. Results and Discussions

In this study, the classification of prostate cells has been completed, but performance evaluation is required in training and testing. Its function is to compare the performance of previously trained models. In comparing the performance of the pre-trained models, the comparison of performance metrics used is accuracy, precision, sensitivity, specificity, and f-score. In addition to comparisons of performance metrics, time comparisons are also needed to assess performance results. The higher the performance metric value and the faster run time, the better and more efficient the performance of the previously trained model.

All models show the average performance results using the performance metrics shown in Tables 3, and 4. Starting from the F-score, AlexNet got the lowest average result with 97.94%, then GoogLeNet with 99.06% as the best F-score. Besides, on the accuracy metric, AlexNet has the lowest with 99.19%, followed the best result is obtained by GoogLeNet with 99.63%. In the calculation of precision, AlexNet gets the lowest result with 98.02% and the highest result is GoogLeNet with 99.03%. Finally, on the measurement of sensitivity and specificity, AlexNet also got the lowest results with 98.30% and 99.49%, in contrast to the GoogLeNet results which had the highest results with 99.17% and 99.78%.

Based on performance metrics calculations, GoogLeNet has obtained better performances than AlexNet in term of accuracy, precision, sensitivity and specificity, however AlexNet requires less time in training with an average of 47.5 seconds.

TABLE III. AlexNet Performance Metrics Measurement in Training Classification

Run	**Acc (%)**	**Prec (%)**	**Sens (%)**	**Spec (%)**	**F-Score (%)**	**Time (sec)**
1	100	100	100	100	100	52
2	99.16	98	97.92	99.47	97.86	48
3	99.50	98.60	98.80	99.70	98.66	48
4	99.15	97.74	98.04	99.49	97.84	45
5	99.67	99.41	99.11	99.77	99.25	47
6	98.82	97.07	97.26	99.27	97.04	47
7	99.33	98.49	98.19	99.57	98.31	47
8	99.84	99.71	99.58	99.89	99.64	47
9	97.96	95.12	94.73	98.71	94.74	47
10	98.47	96	96.37	99.04	96.05	47
Average	99.19	98.02	98	99.49	97.94	47.5

TABLE IV. GoogLeNEt Performance Metrics Measurement in Training Classification

Run	**Acc (%)**	**Prec (%)**	**Sens (%)**	**Spec (%)**	**F-Score (%)**	**Time (sec)**
1	99.84	99.55	99.58	99.90	99.56	115
2	99.83	99.57	99.55	99.90	99.56	113
3	100	100	100	100	100	113
4	99.83	99.57	99.71	99.90	99.64	112
5	100	100	100	100	100	112
6	98.32	95.63	96.25	98.98	95.63	112
7	100	100	100	100	100	112
8	99.50	98.60	99.15	99.70	98.85	112
9	99	97.41	97.44	99.38	97.32	112
10	100	100	100	100	100	112
Average	99.63	99.03	99.17	99.78	99.06	112.5

Based on performance metrics calculations, GoogLeNet obtain the highest results. As in training, AlexNet takes less time than others and this is an advantage of AlexNet. The average time needed for AlexNet in testing is 0.307 seconds, faster than GoogLeNet with 0.372 seconds. In testing, the comparison of performance models is also presented on the bar graph which is shown in Figure 2.

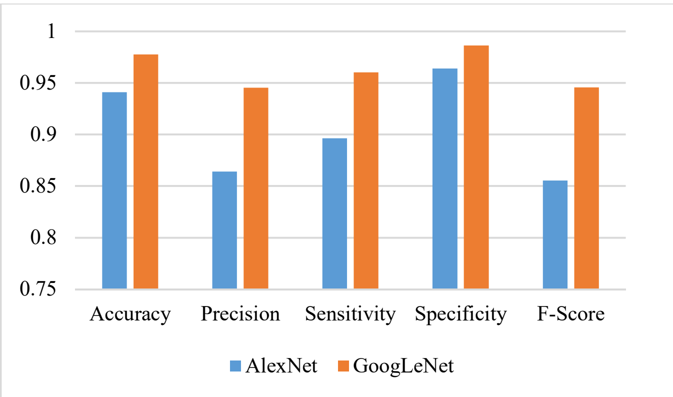

Fig. 2. Graph comparison of testing performance metrics

IV. CONCLUSIONS

This study proposes a deep learning method with pretrained convolutional neural network models and fine-tuning for prostate cancer classification. The purpose of this study is to analyze and compare the results of the performance of pretrained models based on the value of performance metrics and running time needed to do the classification. The results show that GoogleNet has the highest accuracy in training and testing with 99.63% and 97.74% and the lowest accuracy is obtained by AlexNet with 99.13% and 94.11%. During the training, AlexNet had a shorter time with 47 seconds than GoogLeNet with 112 seconds. In testing times, AlexNet was also faster with 0.307 seconds than GoogLeNet with 0.372 seconds. The time required is proportional to the accuracy obtained, the more time it takes, the better the accuracy will be obtained in the classification. This research is expected to assist researchers (pathologists, physician assistants, etc.) in choosing the right architecture for the classification of prostate cancer images in terms of time and accuracy. For future research, we suggest other pretrained models to train and test the prostate images classification.

ACKNOWLEDGMENT

This research is supported by Universitas Muhammadiyah Yogyakarta and research project grant from Ministry of Research and Technology of the Republic of Indonesia.

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019 (US statistics)," CA. Cancer J. Clin., vol. 69, no. 1, pp. 7–34, 2019, doi: 10.3322/caac.21551.

[2] S. W. Leslie, T. L. Soon-Sutton, H. Sajjad, and L. E. Siref, Prostate Cancer. 2019.

[3] J. Cuzick et al., "Prevention and early detection of prostate cancer," Lancet Oncol., vol. 15, no. 11, pp. e484–e492, 2014, doi: 10.1016/S1470-2045(14)70211-6.

[4] G. P. Pena and J. S. De Andrade-Filho, "How does a pathologist make a diagnosis?," Arch. Pathol. Lab. Med., vol. 133, no. 1, pp. 124–132, 2009, doi: 10.1016/s1077-9108(09)79304-1.

[5] R. J. Horner, "Cancer Treatment Drugs," Cancer Concepts A Guideb. Non-Oncologist, pp. 0–5, 2015, doi: 10.7191/cancer.

[6] M. D. Shah, A. V. Parwani, and D. L. Zynger, "Impact of the pathologist on prostate biopsy diagnosis and immunohistochemical stain usage within a single institution," Am. J. Clin. Pathol., vol. 148, no. 6, pp. 494–501, 2017, doi: 10.1093/AJCP/AQX103.

[7] A. A. Ahmed and M. Abedalthagafi, "Cancer diagnostics: The journey from histomorphology to molecular profiling," Oncotarget, vol. 7, no. 36, pp. 58696–58708, 2016, doi: 10.18632/oncotarget.11061.

[8] R. Cuocolo et al., "Machine learning applications in prostate cancer magnetic resonance imaging," Eur. Radiol. Exp., vol. 3, no. 1, 2019, doi: 10.1186/s41747-019-0109-2.

[9] X. Wang et al., "Searching for prostate cancer by fully automated magnetic resonance imaging classification: Deep learning versus non-deep learning," Sci. Rep., vol. 7, no. 1, pp. 1–8, 2017, doi: 10.1038/s41598-017-15720-y.

[10] J. De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease," Nat. Med., vol. 24, no. 9, pp. 1342–1350, 2018, doi: 10.1038/s41591-018-0107-6.

[11] K. Gadepalli, R. Macdonald, T. Kohlberger, J. Hipp, and M. Stumpe, "An Augmented Reality Microscope for Real-time Automated Detection of Cancer," Google, 2018, [Online]. Available: https://research.googleblog.com/2018/04/an-augmented-reality-microscope.html.

[12] D. M. Hawkins, S. C. Basak, and D. Mills, "Assessing model fit by cross-validation," J. Chem. Inf. Comput. Sci., vol. 43, no. 2, pp. 579–586, 2003, doi: 10.1021/ci025626i.

[13] G. J. McLachlan, K.-A. Do, and C. Ambroise, Analyzing microarray gene expression data. Wiley, 2004.

[14] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[15] O. Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

[16] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[17] P. Ballester and R. M. Araujo, "On the performance of googlenet and alexnet applied to sketches," 30th AAAI Conf. Artif. Intell. AAAI 2016, pp. 1124–1128, 2016.

[18] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "Deepfruits: A fruit detection system using deep neural networks," Sensors (Switzerland), vol. 16, no. 8, 2016, doi: 10.3390/s16081222.

[19] D. L. MacLellan and S. B. Bauer, "Physiology of the lower urinary tract," Pediatr. Neurogenic Bl. Dysfunct. Diagnosis, Treat. Long-Term Follow., pp. 13–20, 2006, doi: 10.1007/3-540-30867-9_2.

[20] M. Ali, D. H. Son, S. H. Kang, and S. R. Nam, "An accurate CT saturation classification using a deep learning approach based on unsupervised feature extraction and supervised fine-tuning strategy," Energies, vol. 10, no. 11, 2017, doi: 10.3390/en10111830