Prostate Image Classification Using Pretrained Models: GoogLeNet and ResNet-50

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
muhammad.ahdan.2016@ft.umy.ac.id

3rd Feriandri Utomo
Faculty of Medicine and Health Sciences, Universitas Abdurrab
Pekanbaru, Indonesia

2021 15th International Conference on Signal Processing and Communication Systems (ICSPCS) | 978-1-6654-3699-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICSPCS53099.2021.9660334

***Abstract*—Anatomic pathology errors often occur around 1% to 43% in diagnosis (depending on the disease). This error resulted in delayed treatment and an incorrect diagnosis. This study aims to detect the prostate cancer cells by using pretrained models of deep learning based on the value of performance metrics using a confusion matrix. Image data was taken from a light microscope at the Universitas Indonesia (UI) Hospital. In this study, 10-fold cross-validation was used as a validation of performance metrics on the model. In addition to the accuracy assessment, there is an assessment of precision, recall (sensitivity), specificity and F-score, as well as running time in the process. In this study, the accuracy value in each iteration has a relationship with other performance metric values, if the accuracy value decreases, the other performance metric values will also decrease, and vice versa. In training, ResNet-50 has an average accuracy value of 99.83%, 0.2% higher than GoogLeNet. Testing performances show that ResNet-50 has an average accuracy value of 98.02%, slightly higher (0.28%) than GoogLeNet. In F-score as well, GoogLeNet has a slightly smaller 0.54% difference than ResNet-50. It can be concluded that the pretrained models has good performances in prostate images classification.**

Keywords—prostate cells, anatomic pathology, artificial intelligence, deep learning, classification.

I. INTRODUCTION

According to Cancer Statistics 2020, prostate cancer has been the most common cancer in men. In 2020, there was a 1% rise in cancer cases in the US (up 17,230 people) from the previous year. In the same year, 191,930 (21%) people possessed prostate cancer, and 701,730 (79%) for men had other cancers. The total number of deaths also rose by 1,710 people (31,620 people in 2019 and 33,330 in 2020). In 2020, this cancer remained in the second rank as the previous year (33,330 people (10%) out of 321,160 total cancer deaths)[1], [2].

Until now, the cause of this cancer remains unknown compared to other cancers. Factors influencing cancer include advanced age, ethnicity, genetic factors, and family history. In addition, other factors triggering cancer cover diet, obesity, lack of exercise, inflammation, hyperglycemia, infection, and exposure to chemicals or ionizing radiation [3].

A biopsy is one method used to detect cancer cells in the patient's organs [4]. This method takes organ tissue to examine whether it is cancerous or not [5]. In a study directed by Stephen S. Raab, MD, errors often occurred in diagnosing cancer. In diagnosis, all specimens had an anatomic pathology error of about 1% to 43% (depending on the disease). Accordingly, this error resulted in delayed treatment and incorrect diagnosis [6].

Artificial intelligence (AI) helps to detect cancer cells with more accurate results [7]. A team of researchers from Germany, the United States, and France studied AI to distinguish abnormally growing skin tissue in more than 100 thousand images. The system has been built based on CNN's deep learning designed to fight 58 dermatologists from 17 countries. The dermatologists consisted of 30 experts with more than five years of experience, 11 skilled with two to five years of experience, and seventeen beginners with less than two years of experience. In dermoscopic image reading only, AI gained 95% sensitivity, better than the average dermatologist, reaching 86.6% [8].

In addition to cancer, AI can also identify Covid-19 infections through coughing and vocal cords. The research has carried out by a team of researchers from MIT who developed AI. The study has utilized the extractor feature of the biomarker as a pre-screening for Covid-19. The research team took 5,320 respondents consisting of coughs and vocal cords from April to May 2020. The study obtained a sensitivity of 98.4% of respondents positive for Covid-19 in official tests [9].

Several kinds of research about computer-assisted prostate cancer have explored the methods used and the results. There is a study conducted by Geert Litjens et al. in detecting prostate cancer using 347 MRI images (165 cancer and 182 normal) with computer assistance. This study consisted of two stages, namely, initial candidate detection using prostate segmentation (multi-atlas based, voxel feature extraction, classification and detection of local maxima) and then segmenting the region so that the classification of each candidate was obtained. This system compares the prospective clinical performance of radiologists with sensitivities of 0.42, 0.075, and 0.89 at 0.1, 1, and 10 false positives per case [10].

In addition, there are other studies that use MRI images used by Shijun Wang et al. The techniques used are image segmentation, registration, feature extraction, and classification. The system compares 15 computer-aided-diagnosis (CADx)-based systems that have been carried out and published. As a result, the system noted that the AUC performance of the CADx system was still below 0.90. This shows that there are still improvements in this area in the future [11].

In addition to detecting the presence of prostate cancer, there are studies to provide the right dose of radiotherapy. A classification method for individual doses in three dimensions is developed. The method identified prostate cancer patients who are at risk for rectal bleeding. The results obtained 87 patients (in two years of treatment) who used radiotherapy obtained high sensitivity and specificity of performance [12].

In addition to using the semi-non-negative ICA algorithm, there is another classification using Deterministic Multi-directional Analysis (DMA). A total of 99 patients who were treated for prostate cancer were tested and analyzed for a model called CP-DMA using cross validation. The results of the study were then compared with the supervised (linear discriminant analysis, supporting vector machine, K-means, K-nearest neighbor), the unsupervised approach (the latest principal component analysis-based algorithms, and multidimensional classification methods) and the Normal Tissue Complication Probability (NTCP) model. The results show that this method shows as a good classification because it has good sensitivity and specificity values beating the classical approach [13].

ResNet-50 is a pretrained models which is used for several literature reviews. Research conducted by Qasem at all. using ResNet-50 in diagnosing breast cancer. The results obtained an accuracy of 99% compared to other models using the same dataset [14]. In the case of the same cancer, there are studies that use this model in its classification. The results obtained are very good results with an overall accuracy of 95.74% [15]. In addition to breast cancer, this model is also used in the classification of tissue types in colorectal cancer histology images. As a result of the study, this model achieved a classification accuracy of 94.4% for eight classes of CRC tissue separation [16]. In addition to the accuracy assessment, this model also obtained a good F-score of 94.11% for Yusuf at all. in the detection of invasive ductal carcinoma [17].

In previous studies, an assessment of prostate cell image classification has been carried out using AlexNet and GoogLeNet. Previous research concluded that GoogLeNet is superior in terms of performance metrics [18], so it is necessary to test using other pre-workout models. Another test using the pretrained model is to find out which model is the most suitable. Just like the tests that have been done, good performance metrics are used to measure the feasibility of the system [19]. This research is expected to assist in the selection of cancer detection models for researchers (pathologists, physician assistants, etc.).

II. Methods

The methods used in this study include pre-processing data, 10-fold cross-validation, training from pretrained models, prediction, and image classification. Pre-processing is required to prepare the data before processing. Image data is cropped, labeled, and reduced in size according to model requirements. The data were divided into five classes: normal, IIA, IIC, III, and IV. The pre-training model uses GoogLeNet and ResNet-50, which are trained to classify and predict images into five classes. The flow chart is shown in Figure 1.

A. Tools and Materials

The tools used encompassed software and hardware. The software was MATLAB R2019a with deep learning (DL) toolbox as a framework provider for implementing CNN [20]. In addition, the hardware used is exhibited in Table I.

TABLE I. Hardware Specifications

Hardware	Specification
Ram	16 Gb
Processor	Intel Core i5-9400 CPU @ 2.90 GHz
Graphics	GeForce GTX 970 4Gb

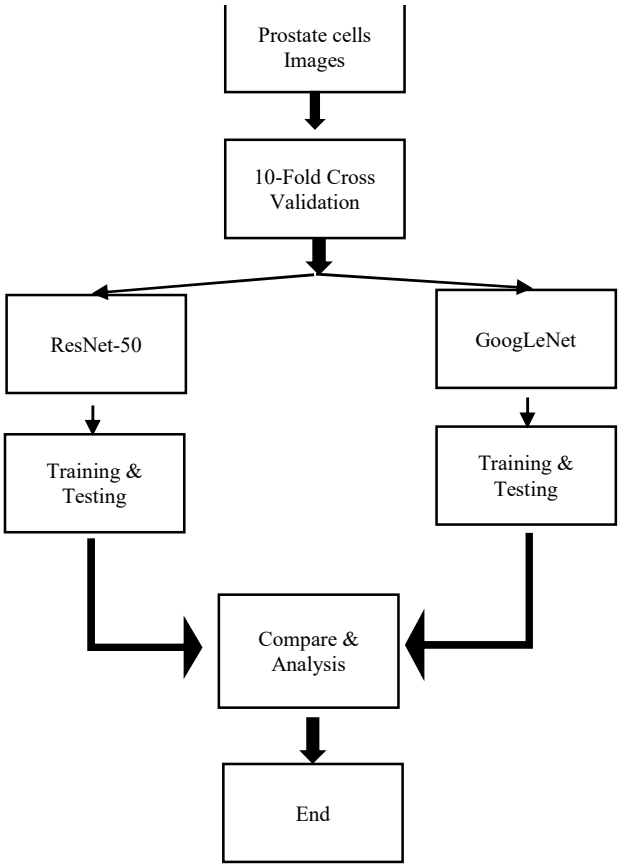


Fig. 1. The flowchart in Prostate Image Classification using Pretrained Model: GoogLeNet and ResNet-50

The images used are images that have been diagnosed and classified by the medical team to be used as research data. The image consisted of 57 prostate cells with five grades (normal, IIA, IIC, III, and IV). Images collected from a light microscope at the University of Indonesia (UI) Hospital, with copyright and code of ethics. The image is then divided into four parts to reproduce the sample data, the process of which can be seen in Figure 2.

From the cropping results, 268 images were produced consisting of five classification classes (one normal class and four cancer classes), of which the number of images is depicted in Table II.

For data input, all images must be reduced to 224x224 pixels. The reduced data was then processed using a 10-fold cross-validation technique. This technique was applied to assess and validate the accuracy of the model [21]. This technique divides the overall data into two parts, of which 90% is training data, and the remaining 10% is test data [22].

TABLE II. The Cropped Prostate Cell Images

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

B. *Training of Pretrained Model*

The pre-training model used is a model that has been trained for more than one million images and can classify 1000 object categories. The models used are GoogLeNet and ResNet-50. GoogLeNet is a model developed by Szegedy et al. [23], who won the 2014 ImageNet Large-Scale Visual Award (ILSVRC) [24]. The next model is ResNet-50 which is a model developed by He et al., which in 2015 was ranked in the top 5 with an error of 3.57% [25].

In this study, the same configuration was used so that it could be compared with previous studies of prostate cell classification [18]. The epoch used is 15, with a learning rate of 0.0001 (constant). Adam's optimization with a batch size of ten was implemented in the algorithm. The model uses fine-tuning—a concept that uses transfer learning, replacing the output layer of a pre-trained model with a prostate cell data set. Three layers were replaced consisting of fully connected output, softmax, and classification [26].

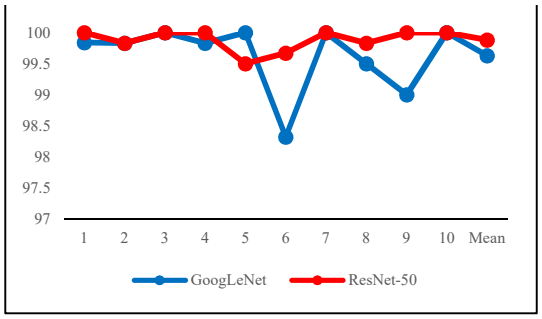
C. *Analysis*

After training, analysis was performed on the output of the model in the form of image classification for each class. In this process, the model's performance metrics could be determined, whether it was feasible or not. In assessing performance metrics, accuracy was considered insufficient; thus, more assessment using a confusion matrix was highly required. The confusion matrix has been known as the error matrix, containing a specific table about the visualization of the performance model usually used in supervised learning. Each row of the matrix represented an instance of the actual class, while each column represented an instance of the predicted class [27].

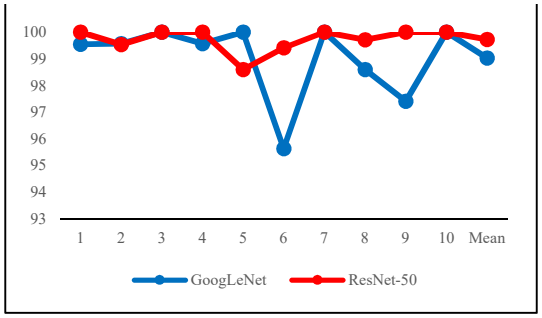
The confusion matrix contained the information required to analyze a model. The assessment in the confusion matrix did not only contain measurements of the accuracy value but also values of precision, sensitivity, specificity, and F-score [26]. There were terms in the confusion matrix representing the classification results: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In this study, we assessed the performance of the model using a confusion matrix of which the calculation formula is demonstrated in [28].

III. Results and Discussions

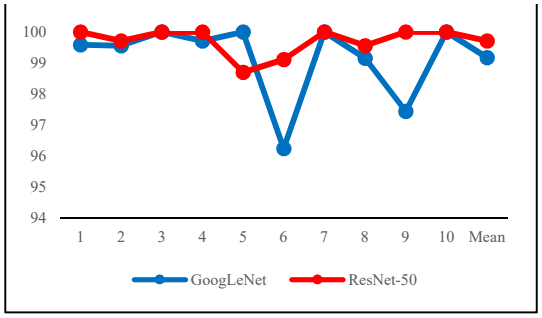
In this study, there are two assessments both on training and testing. In this assessment, a 10-fold cross-validation method was used to compare the classification performance of the previously trained models. Calculations for each model use standard performance metrics. The standard performance metrics used are accuracy, precision, sensitivity, specificity, and F-score.



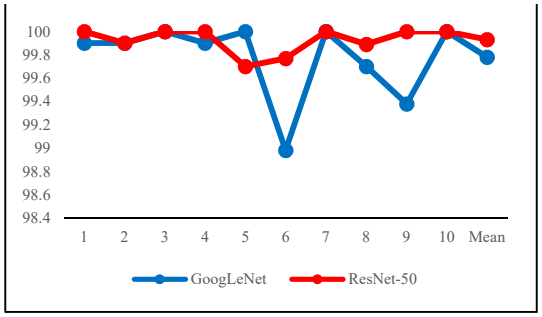
(a). Accuracy



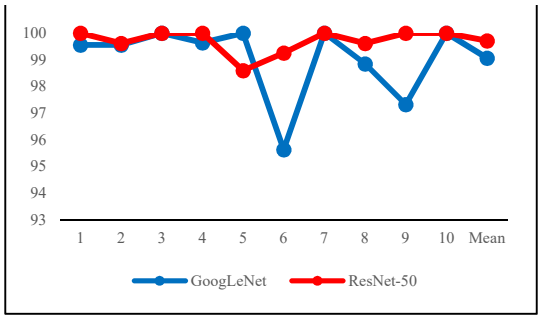
(b). Precision



(c). Sensitivity

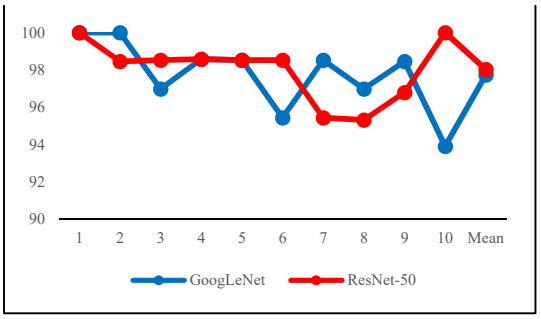


(d). Specificity

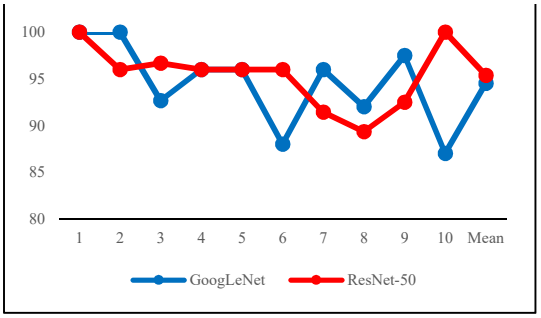


(e). F-score

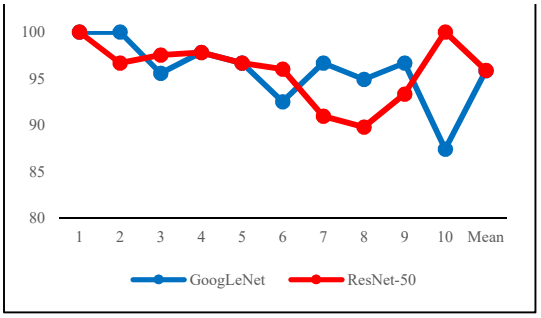
Fig. 2. Comparison of the performance of different classification methods at each fold in training



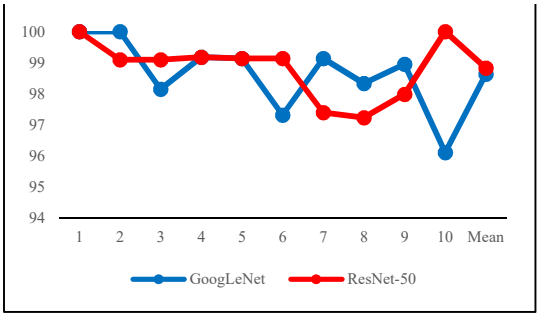
(a). Accuracy



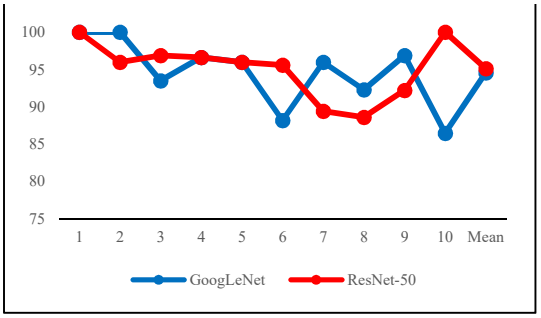
(b). Precision



(c). Sensitivity



(d). Specificity



(e). F-score

Fig. 3. Comparison of the performance of different classification methods at each fold in testing

The training performance of each model is shown in Figure 2. In Figure 2, it can be seen that each accuracy value in each iteration has a relationship with other performance metrics values. The better the accuracy value in the iteration, the better the other performance metrics will be. Figure 2(a) shows that GoogLeNet has decreased accuracy to 98.32% in the sixth iteration (k = 6) so that all performance metrics values also experience a decrease (you can see the same process of increasing and decreasing each performance metrics value in Figure 2(a) to 2(e)). Then, GoogLeNet increased sharply in the seventh iteration (k = 7) so that all performance metrics also increased. In Figure 2, the performance metrics of ResNet-50 are quite stable compared to GoogLeNet. The worst accuracy of ResNet-50 is in the fifth iteration (k = 5) which is 99.50% better than the worst accuracy of GoogLeNet in the sixth iteration (k = 6). Although GoogLeNet has a lower accuracy value, both pretrained models get 100% accuracy in the third, seventh and tenth iterations. In terms of accuracy, ResNet-50 has an average value of 99.83%, slightly higher by 0.2% than GoogLeNet. In addition to the accuracy assessment, GoogLeNet has lower performance metrics than ResNet-50. This can be seen from the average value of performance metrics for each training which can be seen in Figure 2.

As in training, in testing the accuracy value of each iteration is closely related to the value of other performance metrics. It can be seen in Figure 3(a), GoogLeNet in the tenth iteration (k = 10) decreased accuracy to 93.9%, so that the entire value of performance metrics decreased. In testing, ResNet-50 also got an average accuracy value of 98.02%, greater than GoogLeNet which only got 97.74%. In Figure 3, it can be concluded that the better the accuracy value for each iteration, the better the other performance metrics.

Although the assessment of each iteration has been carried out, it is also necessary to assess the classification of each class. The classification of assessments has been carried out in Table III. This assessment consists of an assessment of each class along with its average. This assessment uses a confusion matrix in each class. It can be seen that ResNet-50 has an average accuracy value of 0.9802, slightly higher than 0.0028 from GoogLeNet. In addition to accuracy, precision and recall (also known as sensitivity) can be used to measure the accuracy and quality of the model. As with accuracy, ResNet-50 also has slightly higher average precision and sensitivity values. ResNet-50 got a precision and sensitivity of 0.9539 and 0.9655 slightly higher than GoogLeNet which got 0.9452 and 0.9601, respectively. High precision values indicate that the model can return more related than unrelated results, while recall focuses only on true positive sensitivity. In this study, assessment of performance metrics is also required to calculate the true negative rate. Calculations use ideal specificity to calculate negative case accuracy. In calculating the average specificity, ResNet-50 also has a slight difference with GoogLeNet, which is only 0.0019. Furthermore, to measure the level of model performance, the F-score is also used. The F-score uses precision and sensitivity values for each class with the best value of 1 and the worst of 0. It can be seen in Table III that, GoogLeNet has a difference of 0.0054 which is slightly smaller than ResNet-50. Finally, it is seen that GoogLeNet has a very small difference in performance metric values with ResNet-50, so the performance of the two is not much different.

TABLE III. COMPARISON OF TESTING MODELS TAKEN THROUGH AVERAGE OF EACH PERFORMANCE METRIC

Performance Metrics	Classes	Test	
		GoogLeNet	*ResNet-50*
Accuracy	Normal	1.0000	0.9963
	Stadium IIA	0.9627	0.9695
	Stadium IIC	0.9702	0.9736
	Stadium III	0.9731	0.9806
	Stadium IV	0.9809	0.9807
	Average	0.9774	0.9802
Precision	Normal	1.0000	1.0000
	Stadium IIA	0.9000	0.9150
	Stadium IIC	0.9625	0.9446
	Stadium III	0.9233	0.9500
	Stadium IV	0.9400	0.9600
	Average	0.9452	0.9539
Sensitivity	Normal	1.0000	0.9800
	Stadium IIA	0.9292	0.9348
	Stadium IIC	0.9413	0.9653
	Stadium III	0.9490	0.9667
	Stadium IV	0.9809	0.9807
	Average	0.9601	0.9655
Specificity	Normal	1.0000	1.0000
	Stadium IIA	0.9782	0.9822
	Stadium IIC	0.9847	0.9814
	Stadium III	0.9818	0.9864
	Stadium IV	0.9865	0.9911
	Average	0.9863	0.9882
F-score	Normal	1.0000	0.9889
	Stadium IIA	0.9004	0.9177
	Stadium IIC	0.9491	0.9476
	Stadium III	0.9291	0.9527
	Stadium IV	0.9500	0.9485
	Average	0.9457	0.9511

IV. CONCLUSIONS

This study uses pretrained models using fine-tuning as a classification of prostate cancer. This study aims to assess the performance of the pretrained model based on the value of performance metrics using a confusion matrix. In this study, the accuracy value in each iteration has a relationship with other performance metric values, if the accuracy value decreases, the other performance metric values will also decrease, and vice versa. In training, ResNet-50 has an average accuracy value of 99.83% (0.2% higher than GoogLeNet). In addition to training, there is an assessment of performance metrics in testing. Tests show that ResNet-50 has an average accuracy value of 98.02%, slightly higher by 0.28% than GoogLeNet. In F-score as well, GoogLeNet has a slightly smaller 0.54% difference than ResNet-50. Higher accuracy values cause ResNet-50 to excel in all performance metrics values. Finally, it can be concluded that GoogLeNet has a very small difference in the value of performance metrics with ResNet-50, so the performance of the two is not much different.

ACKNOWLEDGMENT

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

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," CA: A Cancer Journal for Clinicians, vol. 70, no. 1, pp. 7–30, 2020, doi: 10.3322/caac.21590.

[2] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," CA: A Cancer Journal for Clinicians, vol. 69, no. 1, pp. 7–34, 2019, doi: 10.3322/caac.21551.

[3] L. A. Mucci, K. M. Wilson, and E. L. Giovannucci, "Epidemiology of prostate cancer," Pathology and Epidemiology of Cancer, vol. 10, no. 2, pp. 107–125, 2016, doi: 10.1007/978-3-319-35153-7_9.

[4] V. Kasivisvanathan et al., "Magnetic Resonance Imaging-targeted Biopsy Versus Systematic Biopsy in the Detection of Prostate Cancer: A Systematic Review and Meta-analysis(Figure presented.)," European Urology, vol. 76, no. 3, pp. 284–303, 2019, doi: 10.1016/j.eururo.2019.04.043.

[5] L. Fenzl, K. Bubel, M. Mehrmann, and G. Schneider, "Bildgebung und Biopsie von Weichteiltumoren," Radiologe, vol. 58, no. 1, pp. 79–92, 2018, doi: 10.1007/s00117-017-0331-y.

[6] M. J. Fromer, "Study Pathology Errors Can Have Serious Effect on Cancer Diagnosis & Treatment," Oncology Times, vol. 27, no. 22, pp. 25–26, Nov. 2005, doi: 10.1097/01.COT.0000291164.08133.9a.

[7] S. Huang, J. Yang, S. Fong, and Q. Zhao, "Artificial intelligence in cancer diagnosis and prognosis: Opportunities and challenges," Cancer Letters, vol. 471, pp. 61–71, Feb. 2020, doi: 10.1016/j.canlet.2019.12.007.

[8] H. A. Haenssle et al., "Man against Machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," Annals of Oncology, vol. 29, no. 8, pp. 1836–1842, 2018, doi: 10.1093/annonc/mdy166.

[9] J. Laguarta, F. Hueto, and B. Subirana, "COVID-19 Artificial Intelligence Diagnosis Using Only Cough Recordings," IEEE Open Journal of Engineering in Medicine and Biology, vol. 1, pp. 275–281, 2020, doi: 10.1109/ojemb.2020.3026928.

[10] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman, "Computer-aided detection of prostate cancer in MRI," IEEE Transactions on Medical Imaging, vol. 33, no. 5, pp. 1083–1092, 2014, doi: 10.1109/TMI.2014.2303821.

[11] P. Choyke, B. Turkbey, K. Burtt, R. M. Summers, and S. Wang, "Computer Aided-Diagnosis of Prostate Cancer on Multiparametric MRI: A Technical Review of Current Research," BioMed Research International, vol. 2014, pp. 1–11, 2014, doi: 10.1155/2014/789561.

[12] J. Coloigner et al., "A novel classification method for prediction of rectal bleeding in prostate cancer radiotherapy based on a semi-nonnegative ICA of 3D planned dose distributions," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 3, pp. 1168–1177, 2015, doi: 10.1109/JBHI.2014.2328315.

[13] A. Fargeas et al., "On feature extraction and classification in prostate cancer radiotherapy using tensor decompositions," Medical Engineering and Physics, vol. 37, no. 1, pp. 126–131, 2015, doi: 10.1016/j.medengphy.2014.08.009.

[14] Q. A. Al-Haija and A. Adebanjo, "Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network," IEMTRONICS 2020 - International IOT, Electronics and Mechatronics Conference, Proceedings, vol. 50, 2020, doi: 10.1109/IEMTRONICS51293.2020.9216455.

[15] X. Yu, C. Kang, D. S. Guttery, S. Kadry, Y. Chen, and Y. D. Zhang, "ResNet-SCDA-50 for Breast Abnormality Classification," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 18, no. 1, pp. 94–102, 2021, doi: 10.1109/TCBB.2020.2986544.

[16] J. N. Kather et al., "Multi-class texture analysis in colorectal cancer histology," Scientific Reports, vol. 6, pp. 2019–2020, 2016, doi: 10.1038/srep27988.

[17] Y. Celik, M. Talo, O. Yildirim, M. Karabatak, and U. R. Acharya, "Automated invasive ductal carcinoma detection based using deep

transfer learning with whole-slide images," Pattern Recognition Letters, vol. 133, pp. 232–239, 2020, doi: 10.1016/j.patrec.2020.03.011.

[18] Y. Jusman, M. Ahdan Fawwaz Nurkholid, D. Arief Darmawan, and F. Utomo, "Comparison Performance of Prostate Cell Images Classification using Pretrained Convolutional Neural Network Models," in 2021 IEEE Region 10 Symposium (TENSYMP), Aug. 2021, pp. 1–4. doi: 10.1109/TENSYMP52854.2021.9550865.

[19] Q.-Q. Zhou et al., "Automatic Detection and Classification of Rib Fractures on Thoracic CT Using Convolutional Neural Network: Accuracy and Feasibility," Korean Journal of Radiology, vol. 21, no. 7, p. 869, 2020, doi: 10.3348/kjr.2019.0651.

[20] Mathworks Inc., "Matlab Statistics and Machine Learning Toolbox," 2015, 2019.

[21] J. G. Moreno-Torres, J. A. Saez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k-fold cross-validation," IEEE Transactions on Neural Networks and Learning Systems, vol. 23, no. 8, pp. 1304–1312, 2012, doi: 10.1109/TNNLS.2012.2199516.

[22] D. Berrar, "Cross-validation," Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics, vol. 1–3, no. January 2018, pp. 542–545, 2018, doi: 10.1016/B978-0-12-809633-8.20349-X.

[23] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015, vol. 07-12-June-2015. doi: 10.1109/CVPR.2015.7298594.

[24] P. Ballester and R. M. Araujo, "On the performance of googlenet and alexnet applied to sketches," 30th AAAI Conference on Artificial Intelligence, AAAI 2016, pp. 1124–1128, 2016.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[26] 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, p. 1830, Nov. 2017, doi: 10.3390/en10111830.

[27] S. Haghighi, M. Jasemi, S. Hessabi, and A. Zolanvari, "PyCM: Multiclass confusion matrix library in Python," Journal of Open Source Software, vol. 3, no. 25, p. 729, 2018, doi: 10.21105/joss.00729.

[28] D. Bowes, T. Hall, and D. Gray, "Comparing the performance of fault prediction models which report multiple performance measures: Recomputing the confusion matrix," ACM International Conference Proceeding Series, pp. 109–118, 2012, doi: 10.1145/2365324.2365338.