Comparing the performance of linear regression versus deep learning on detecting melanoma skin cancer using apple core ML

Herry Sujaini, Enriko Yudhistira Ramadhan, Haried Novriando
Department of Informatics, Tanjungpura University, Indonesia

Article Info

ABSTRACT

Melanoma is a type of deadly skin cancer. The survival rate of the patients can fall as low as 15.7% if the cancer cell has reached its final stage. Delayed treatment of melanoma can be attributed to its likeness to that of common nevus (moles). Two machine learning models were developed, each with a different approach and algorithm, to detect the presence of melanoma. Image classification is using the regression algorithm, and object detection is using deep learning. The two models are then compared, and the best model is determined according to the achieved metrics. The testing was conducted using 120 testing data and is made up of 60 positive data and 60 negative data. The testing result shows that object detection achieved 70% accuracy than image classification's 68%. More importantly, linear regression's 43% false-negative rate is noticeably high compared to convolutional neural network's (CNN) 25%. A false-negative rate of 43% means almost half of sick patients tested using image classification will be diagnosed as healthy. This is dangerous as it can lead to delayed treatment and, ultimately, death. Thus it can be concluded that CNN is the best method in detecting the presence of melanoma.

Corresponding Author:

Enriko Yudhistira Ramadhan
Department of Informatics
Tanjungpura University
Prof. Dr. H. Hadari Nawawi Rd., Pontianak, West Borneo 78124, Indonesia
Email: enrikogreen@student.untan.ac.id

1. INTRODUCTION

Melanoma is a type of skin cancer that develops from melanocytes that are located in the human skin. According to a report issued by The World Health Organization, melanoma is the most lethal skin cancer and has accounted for 133.000 deaths globally every year [1]. According to a 2019 report issued by The Global Cancer Observatory, melanoma accounts for 1.392 deaths in Indonesia [2]. Melanoma is caused by an uncontrollable growth of melanocytes, a cell responsible for the production of melanin. Melanin is a pigment that gives the skin a dark tone. U.V. light from the sun can potentially damage the melanocyte and triggers uncontrollable cell growth. This is how melanoma formed. Early detection is critical because the survival rate is highly dependant on early treatment. The survival rate when the cancer has metastasized is only 15.7% [3]. It means 17 out of 20 patients will die when melanoma reaches its final stages. Meanwhile, when the cancer is detected early and still localized, the 5-year survival rate is 98.4% [4]. This is why it is crucial to detect melanoma when it is still in its early stage [5]. The problem is, melanoma is hard to detect in its early stages, and it looks very similar to common moles [6]. When it becomes a large lesion in the skin

and easy to detect, it is already too late, and the cancer cell may already be spreading to other organs. One technology that can be used to detect melanoma early is computer vision [7].

Until fairly recently, doctors had to rely on their eyes to diagnose melanoma [7]. This method of diagnosing melanoma visually is proven to be inaccurate [8]. This changes when imaging technology came to the scene. Reflectance confocal microscopy and optical coherence tomography are just some of the machines used to help doctors identify melanoma. Still, doctors are the ones who made the final decision. The breakthrough happened when A. I. was introduced in the decision-making. A.I. can potentially play a key role in diagnosing melanoma [7]. A study published by lancet oncology pits the best doctors in skin cancer, including 283 certified dermatologists, 118 dermatology residents, and 83 general practitioners against A.I. programs. The study concluded that doctors gave on average 18.78 correct diagnoses, and the A.I. on average gave 25.43 correct diagnoses. Even the three top-performing doctors couldn't compete against the A.I. [7]. This shows that the future of diagnosing melanoma lies in artificial intelligence and computer vision.

This research intends to pit two algorithms, namely linear regression and convolutional neural network (CNN), and deciding which one is the best at giving the correct diagnoses on melanoma cases. Both of these algorithms have their advantages and disadvantages. Image classification uses linear regression algorithm. Linear regression is a method to discover the relationship between one or more response variables and the predictors [9]. For example, does smoking causes lung cancer? Here, the response variable is "smoking," and the predictor is "lung cancer." A prediction model can then be made according to the data given. On the other hand, object detection uses a CNN, a deep neural network class. CNN is a specific type of neural network with a known grid-like topology for processing data. Examples include time-series data, which can be considered at normal time intervals as a 1-D grid collecting samples, and image data, which can be thought of as a 2-D pixel grid. Convolutional networks in functional applications are tremendously efficient. The term "CNN" means that a statistical operation called convolution is used for the network. A specialized type of linear action is convolution. Convolutional networks are neural networks that use convolution in place of general matrix multiplication in at least one of their layers. This model can both classify and pinpoint the position of the object being classified. However, this method requires giving a bounding box to the training images. The CNN also requires massive computational power and considerably longer training time [10] Both models will be trained using the Apple Core ML platform. Apple Core ML is proven to be good at detecting skin lesions [11].

This research will develop two diagnosis models, where one model uses linear regression algorithm and the other uses CNN algorithm. The model is then shown test pictures and asked to give diagnosis. The metrics achieved by both models are then compared and from there, the best algorithm could be determined.

2. RESEARCH METHOD
2.1. Acquiring data

The model is trained with data acquired from the International skin imaging collaboration (ISIC). ISIC is an academic and industry partnership designed to facilitate digital skin imaging to help reduce melanoma mortality. ISIC's objective is to support efforts to reduce melanoma-related fatalities and excessive biopsies by enhancing the precision and reliability of early detection of melanoma. On that end, ISIC is establishing proposed guidelines for digital imaging and building a public database of clinical and dermoscopic skin lesion images [12]. Table 1 shows some samples of positive and negative images of melanoma obtained through ISIC's website.

These images show why a model is needed to differentiate a mole from melanoma. The two is extremely similar. Six hundred images will be used, split into training, validation, and testing data. The dataset consists of 300 images of melanoma and 300 images of common mole (nevus). The ratio used is 60:20:20, making 360 images for training, 120 images for validation, and another 120 images for testing. The images are chosen at random. Both models will be using the same images, ruling out any bias and making the algorithm of both models the only factors affecting the result. After the images are downloaded, it is then placed on a folder corresponding to the usage of the images (e.g., Testing images are put on a folder named "Testing"). Training, validation, and testing folders each contain two folders inside named "Positive" and "Negative." Figure 1 shows how the folder is structured in preparation for training the models.

2.2. Image pre-processing

Images retrieved from ISIC already meet DICOM standards [7]. Therefore, no further pre-processing and image alteration is needed. The only pre-processing work needed is giving a bounding box to the training and validation images as a CNN requires the training and validation data to be given a bounding box before training the model. IBM Annotation Cloud will be used to give the images labels and bounding boxes. Figure 2 shows the process of giving a label and bounding box to an image.
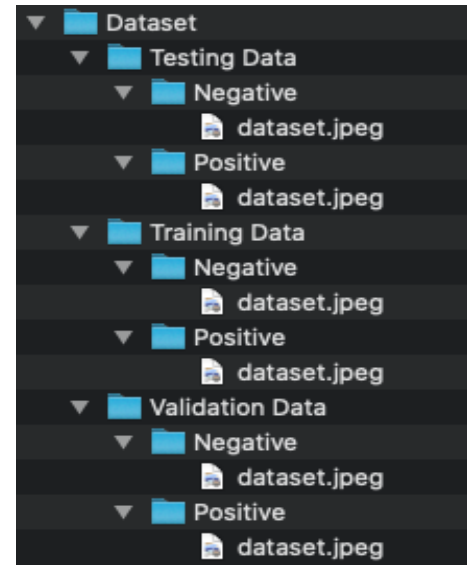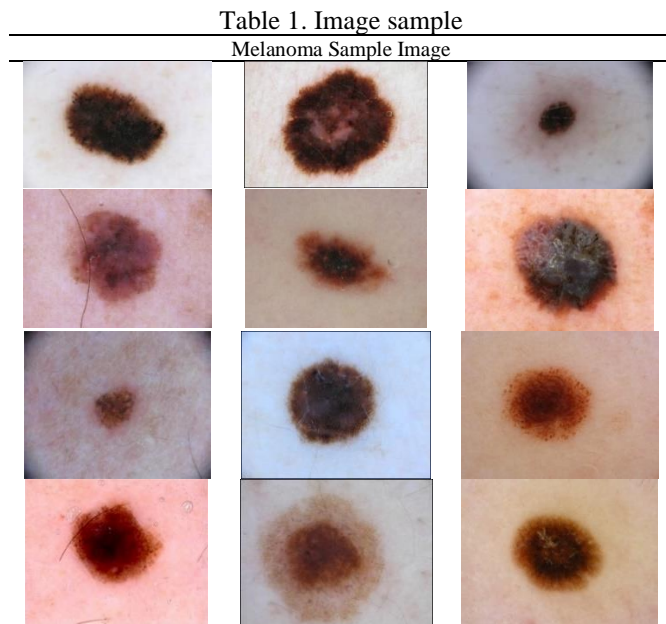
Table 1. Image sample

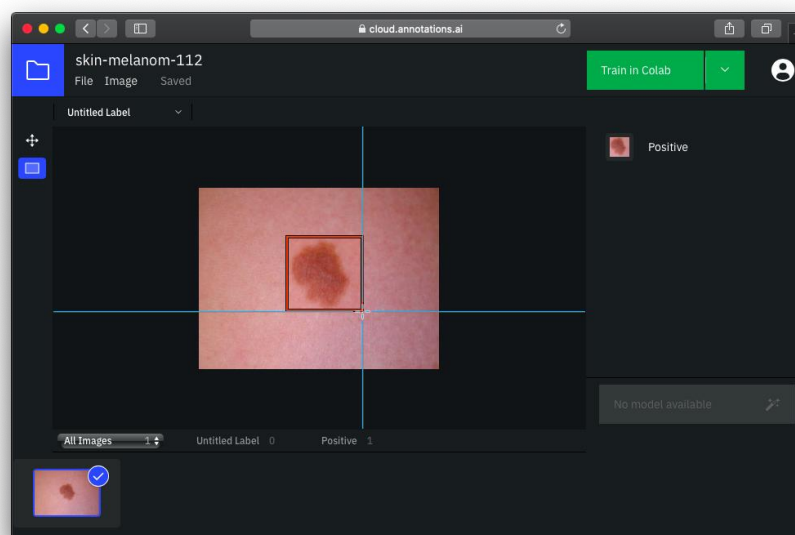| Melanoma Sample Image |
| --- |
| |

Figure 1. Folder structure

Figure 2. Drawing bounding box

2.3. Training process

Two models will be trained using Create ML, each with different algorithms. Create ML is an interface designed for training ML models. The output file is .mlmodel. This is a user interface that will communicate with the Core ML framework for ease of use. The following Figure 3 is the opening interface of Create ML.

"Image Classifier" option uses linear regression model whereas "Object Detector" option uses CNN. After choosing the model, the training interface will be shown. For the regression model, the training set is set to default with 25 iterations. Setting the iteration beyond 30 did not make any difference to the model's accuracy because the training process converges at exactly 25 iterations. Set the training, validation, and testing data to its corresponding folders. The CNN model process is the same, except there is an extra. JSON file inside the folder. For a CNN, 3.000 iterations are used. This number is used because of the computational power's constraint. The model also achieved similar performance at 2000 iterations; thus, any iterations beyond 2000 is unnecessary. The following is Figure 4 that shows the Create ML interface.
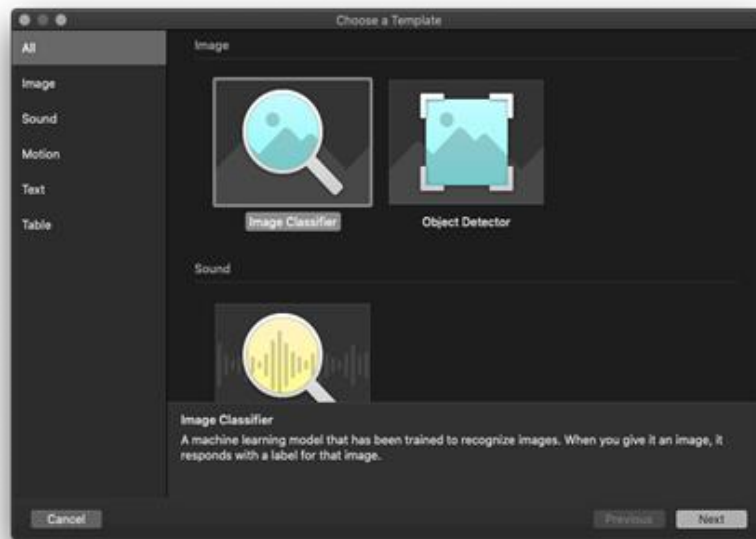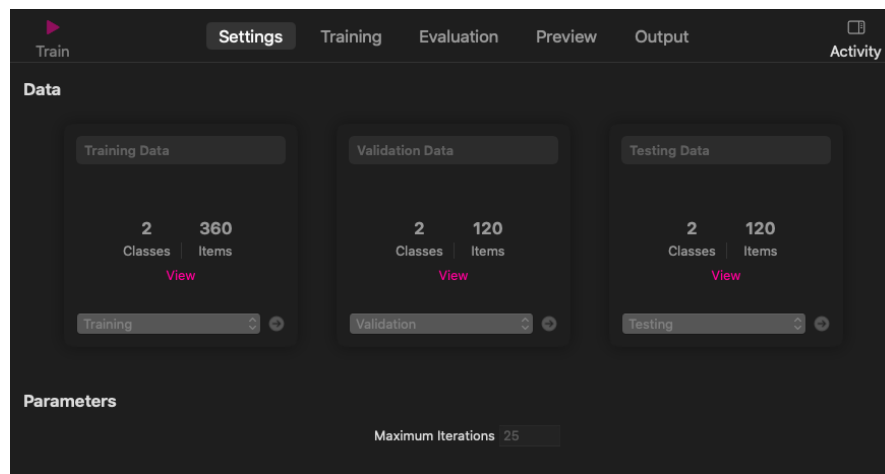
Figure 3. Create ML opening interface

Figure 4. Create ML interface

2.4. Linear regression

The first model uses linear regression algorithm. This algorithm is quite fast and lightweight but potentially less accurate. The model takes a labelled image and apply sharpening filters to accentuate the image's features. A grid is then superimposed on top of the image. This grid contains a value ranging from 0.00 to 1.00, where 0.00 is when the box is colored all white and 1.00 when the box is colored all black. This grid is the feature of the image, visualized by numerical values. The grid is transformed into an array. This array is then stored inside the model. This process is then repeated for each image. For each image trained, an array is created. These arrays are stored in two classes based on the labels given. These labelled arrays are what the models are made of. When identifying an image, the same process applies. Sharpening filters are applied, and a grid is made. The grid is transformed into an array. However, since this array has no label, the model then compares this unknown array to the model's databases of array from the aforementioned training process. A comparison is made, whether this new array is similar to "Positive" arrays or "Negative" arrays. A similarity percentage is then made based on the comparison. This is how linear regression model makes an identification. Figure 5 visualizes the process.
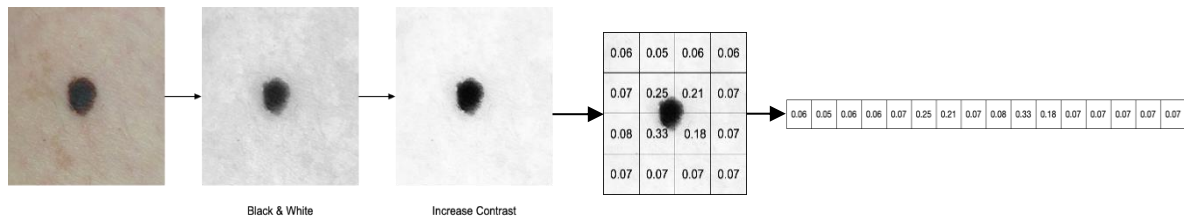
Figure 5. Learning process

2.5. Convolutional neural network

The second model uses an algorithm called "CNN". This algorithm mirrors how the human's brain work. First, the training image is divided into multiple parts. Each of those parts are then given multiple filters. These parts with filters are then associated with the labels and forms the "neurons" of the model. With multiple images trained and each of those images containing multiple parts and those parts again contains multiple filters, one model could potentially have more than 100.000 neurons inside them. These neurons are the knowledge base of the model, similar to the arrays in regression model. The main difference is that one image only results in one array (or one "knowledge") in regression model. So, if 60 images are used in the training phase, the model only has 60 arrays to form the knowledge base. In CNN, 60 images will result in more than 60 neurons forming. CNN has the benefit of having more knowledge to work with when identifying an image. For the identifying process, the same process is done to the image.

However, since these neurons have no label, the neurons are then compared to the thousands of already labelled neurons inside the model's knowledge base. If these neurons connect more to the neurons labelled "positive", then this image is "positive" and vice-versa. The following Figure 6 shows the process. Since the model is learning an image labelled "positive", all the neurons will associate it with a "positive" neuron.
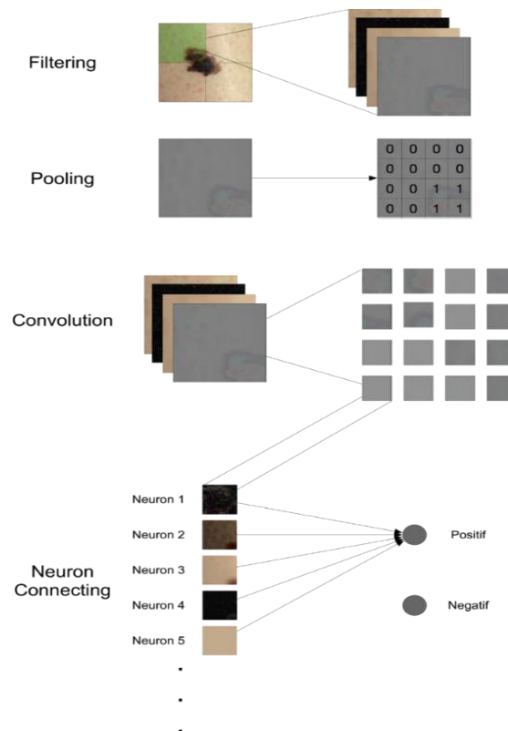

Figure 6. Learning process

2.5. Testing process

After the training process is done, the testing process is conducted. Testing data consisted of 120 images, in which 60 images are labelled "positive-1-60", and another 60 are labelled "negative-1-60". Each

model is fed with testing images, and the classification of each model is noted. The testing is only done once because the model is consistent at classifying each image, and there are no variations on multiple testing (i.e., Picture 1 is always identified as "negative" by the model no matter how much repetition is done on testing). The following Figure 7 shows how the testing is conducted.
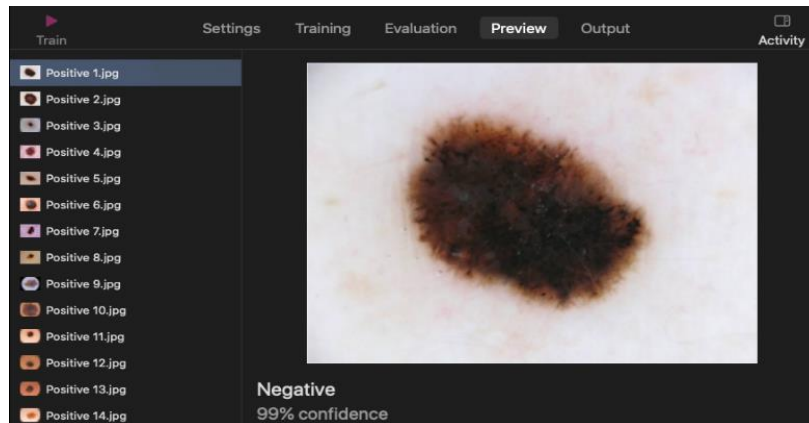

Figure 7. Testing menu

For example on the Figure 7, the testing picture has the label "Positive 1", meaning it shows an image of a positive melanoma. The classification made by the model is 99% negative. However, the true classification of the image is positive. Hence, this model failed to identify the testing image numbered 1. The test continues to image labelled "Positive 2" and so on until the test image "Positive 60". The same process applies to images labelled "Negative 1" to "Negative 60". The long and the short of it is that when the classification made by the model is different from the label, then the model made a mistake in identification. These correct and false identifications are then tallied to form the confusion matrix for each of the models. From the confusion matrix, various metrics can be calculated to determine which algorithm is the best at diagnosing melanoma. The first basic metrics are accuracy and miss. Accuracy is a metric that measures the percentage of data correctly classified by the test under evaluation [13]. In (1) shows the formula for calculating accuracy.

$$Accuracy(\%) = \frac{true\ positive + true\ negative}{total\ data} \times 100 \tag{1}$$

Miss is a metric that measures how many data are incorrectly classified by the test. Miss is always the complement of accuracy. In (2) shows the formula for calculating miss.

$$Miss(\%) = \frac{false\ negative + false\ positive}{total\ data} \times 100 = 100 - Accuracy \tag{2}$$

The next metrics used are sensitivity and specificity. Sensitivity is a measure of how accurate the model is at identifying a true positive or a sick (positive) person [14]. The higher the sensitivity, the better the model at detecting a sick individual. In (3) shows the formula for calculating sensitivity.

$$Sensitivity(\%) = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100 \tag{3}$$

On the other hand, specificity measures how accurate the model is at identifying a true negative or a healthy person [14]. A diagnostic model with high specificity is better at detecting a healthy individual. In (4) shows the formula for calculating specificity.

$$Specificity(\%) = \frac{True\ Negative}{True\ Negative + False\ Positive} \times 100 \tag{4}$$

The last metrics are false-positive rate (FPR) and false-negative rate (FNR). FPR and FNR is a measure of how many errors a diagnostic tool makes in identifying a disease [15]. When a model has a high

FPR, it means that the model made too many false-positive diagnoses, i.e., healthy individuals identified as sick. In (5) shows the formula for calculating FPR.

$$False\ Positive\ Rate(\%) = \frac{False\ Positive}{False\ Positive + True\ Negative} \times 100 \qquad (5)$$

On the other hand, the model with high FNR tends to make many false-negative diagnoses, i.e., sick individuals identified as healthy. In (6) shows the formula for calculating FNR.

$$False\ Negative\ Rate(\%) = \frac{False\ Negative}{False\ Negative + True\ Positive} \times 100 \qquad (6)$$

3.   RESULTS AND DISCUSSION

The following Table 2 shows the metrics achieved by the regression model along with the duration it took to complete the training process. Linear regression's training duration is fairly fast, taking only 1 minute and 31 seconds. As can be seen in Figure 8 below, linear regression only took ten iterations to complete the training process.

As the training process converges to 100% at ten iterations, more iterations are not needed. The training process is fast, but there are potential consequences on the accuracy of the model. The following Table 3 shows the metrics achieved by the convoluted neural network model along with the duration it took to complete the training process.

Convoluted neural network took an astonishingly long 71 hours and 35 minutes to complete the training process. That is almost three days. Convolution process and the forming neuron is responsible for the long duration. The following Figure 9 shows the training performance of convoluted neural network. The graph in Figure 9 shows the loss in features when the model is being trained. To minimize this loss, the model ran multiple iterations and used the previous iteration to complete the features. As more iterations pass, the model will be more accurate. At 3.000 iterations, the model only lost 0.6 part of the feature.

Table 2. Linear regression metrics

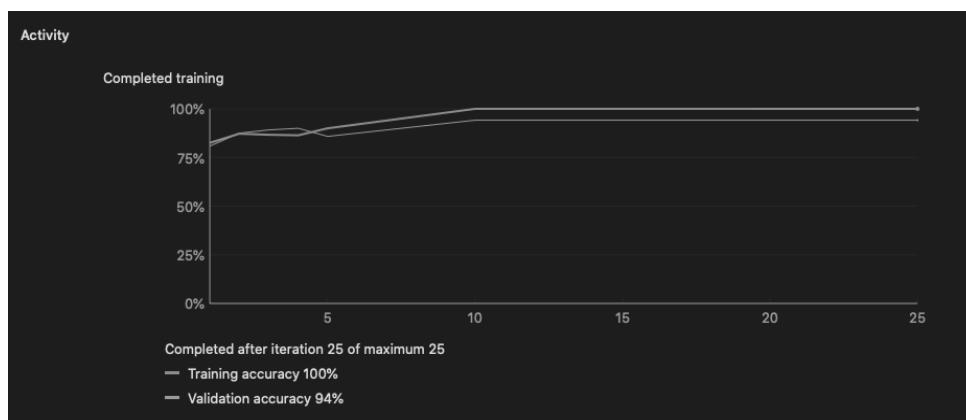| Linear Regression	Training Duration 1 minute 31 seconds	
Confusion Matrix		
True Prediction	Positive	Negative
Positive	34	12
Negative	26	48
Accuracy		Miss
68%		32%
Sensitivity		Specificity
56%		80%
False Positive Rate		False Negative Rate
20%		43%

Figure 8. Linear regression performance

Table 3. Convolutional neural network metrics

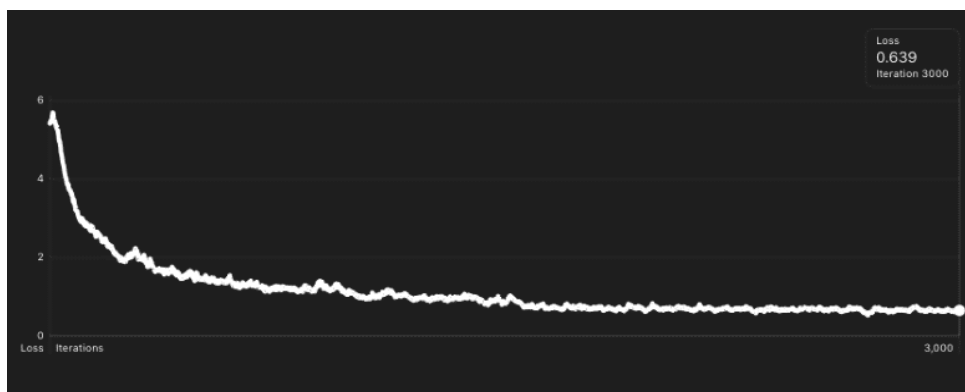| Convoluted Neural Network	Training Duration 71 hours 35 minutes	
Confusion Matrix		
True Prediction	Positive	Negative
Positive	45	18
Negative	15	39
Accuracy		Miss
70%		30%
Sensitivity		Specificity
75%		68%
False Positive Rate		False Negative Rate
30%		25%

Figure 9. Convoluted neural network performance

3.1. Metric analysis

At first glance, the two models appear to achieve similar results. Table 2 and Table 3 shows that the accuracy is only at a 2% difference, with CNN holding the lead. It is important to note that CNN algorithm failed to identify three images, giving neither positive nor negative diagnosis. These three images will not be included in the calculation. This is why accuracy is not the be-all and end-all of machine learning's metrics. Other metrics can hold valuable information regarding which method is the best-for example, the sensitivity and specificity metric. Sensitivity measures how accurate a diagnostic model can identify a sick individual, whereas specificity measures how accurate a diagnostic model can identify a healthy individual. CNN's model has a higher sensitivity at 75% than 56% achieved by the regression model. However, the regression model has a much higher specificity at 80% than CNN's 68%. This means that CNN is better at identifying an individual with melanoma, but regression is better at identifying a healthy individual without melanoma. This is a delicate balance because we cannot maximize both metrics at the same time. Other metrics that can be analyzed are the FPR and FNR. CNN has a lower FPR but a lower FNR. These metrics can be analyzed to determine which method is the best at detecting melanoma.

For accuracy and miss, the analysis is pretty simple. The higher the accuracy, the better, and the lower the miss, the better. Table 3 shows CNN achieved the best result at 70% accuracy and 30% miss. However, the difference with the regression model shown in Table 2 is a mere 2%. This is too small of a difference to call which method is the best. We need to analyze other metrics to see which method is the best. For sensitivity metric, CNN has a higher sensitivity than linear regression, as shown in Table 3; therefore, the CNN model is more accurate at identifying the presence of melanoma. Meanwhile, the regression model is more accurate at identifying a healthy person. However, in a predictive diagnostic test, identifying a sick individual is more of a priority than identifying a healthy one [16]. Therefore, CNN achieved the best metric in this category.

The next metrics are FPR and FNR. This is an important metric to analyze because FPR and FNR can affect how the model worked in real life. Table 2 shows that the regression model has a lower FPR at 20% from the test conducted compared to CNN's 30%, shown in Table 3. This means the regression model is less prone to false-positive error than the CNN model. However, the regression model has an alarmingly high FNR of 43%. This means that if 100 sick individuals are using this model as a diagnostic tool, 43 will be identified as healthy. A diagnostic test with a high false-negative rate is hazardous because sick individuals with melanoma can potentially be identified as healthy, thus missing a much needed early treatment. A false-negative result can lead to death in some diseases [17], including diseases that need early treatment, like

melanoma [8]. Therefore, CNN achieved an optimum metric in this category. The following Table 4 shows a comparison of metrics achieved by both regression and the CNN model.

Table 4. Comparisons of both methods

Linear Regression	Metrics	Convolutional Neural
34	True Positive	45
48	True Negative	39
12	False Positive	18
26	False Negative	15
68%	Accuracy	70%
32%	Miss	30%
56%	Sensitivity	75%
80%	Specificity	68%
20%	False Positive Rate	30%
43%	False Negative Rate	25%
4	Best Metrics	6

2.7. Comparison with other diagnosis methods

The following Table 5 shows how the model made in this research compares with other conventional diagnosis methods and other computer-vision based diagnoses based on accuracy. While deep residual network and optical coherence tomography in Table 5 have higher accuracy than CNN, there are caveats to both methods. Deep residual network uses a custom-developed model by IBM and requires higher computational power than CNN. The custom model also will not run on a mobile device. Meanwhile, everyone with a Macbook can develop a CNN model, and it can run on a laptop or an iPhone making it a more practical diagnostic method. Optical coherence tomography requires a specialized device to take the image. CNN can use images taken with a cellphone camera or any .jpeg or .png file for that matter. While CNN has a lower accuracy than biopsy methods, the former only needs a photo of the mole in question while the latter requires some body operations to be conducted.

Table 5. Comparisons to other methods

Accuracy	Diagnosis Method
93%	Full Biopsy [18]
82%	Cell Biopsy [18]
80%	Dermatologist with 10 years of experience [19]
77.6%	Logistic regression using initial variables and Product Units [20]
77%	Punch Biopsy [18]
76%	Deep Residual Network (IBM) [21]
73%	Optical Coherence Tomography [22]
70%	Convolutional Neural Network
68%	Linear Regression
67%	Shave Biopsy [18]
64%	Visual Biopsy [23]
62%	Dermatologist with 3-5 years of experience [19]
59%	Dermatologist with 1-2 years of experience [19]

It is important to highlight that CNN is still more accurate compared to the two biopsies method, shave biopsy and visual biopsy. From this research, it can be determined that diagnosis with CNN, while not that accurate compared to a full-body biopsy, is accurate enough for a quick and fast assessment of a highly suspicious mole. It does not need an operation to be conducted, and the result comes within 5 seconds. The diagnosis can be done with a MacBook or an iPhone and does not need highly specialized equipment. Again, while the accuracy is not that high, immediate treatment is crucial when it comes to melanoma cases, and rapid analysis with a cellphone is one of the methods that can be used [24]. With this quick method, a diagnosis can be made at home, and when the result shows >50% positive, the person can get to a hospital and receive a more accurate biopsy and other treatments. As melanoma cases continue to rise every year [25] and skin disease being the most common form of disease globally [26], it is necessary to provide everyone with immediate treatments. Easy excision is usually curative when melanoma is detected when it's only isolated to the skin's outer layers, and the 5-year relative survival rate is around 90%. The need to enhance the effectiveness, efficacy, and consistency of the diagnosis of melanoma is apparent. The emotional and financial consequences of getting a diagnostic error is a prevalent, harmful, and costly phenomenon [27]. This method can prevent the aforementioned consequences with ease.

4. CONCLUSION

This research was made to develop two machine learning models that can detect melanoma when it is still in its early stage, compare both models, see which model is the best, and decide if it can diagnose melanoma skin cancer. This research proved that machines could diagnose melanoma better than human eyes, beating two biopsy methods, and beating dermatologists with 1-5 years of experience. It is also concluded that using a deep learning CNN is the best approach to detecting melanoma compared to linear regression. With higher accuracy and a lower false-negative rate, it is the best method to use. It is essential to highlight that while CNN has achieved a higher metric, it is by no means can be used as a one-and-only diagnostic tool. An accuracy of 70%, while seemingly high, is just too low to be reliable. 70% accuracy still means that for every 100 individuals using the model, 30 of them will have the wrong diagnosis. Of course, the perfect model is the one that can detect every individual, healthy or not, with 100% accuracy. However, there is currently no diagnostic tool that can detect melanoma with that kind of accuracy. Rather, this model can be used to do a quick test before the full biopsy is done. This model can be integrated into a smartphone application. It then uses a smartphone camera to capture skin imagery, giving predictions in real-time. This system can then be deployed anywhere, from an online doctor consultation app to staff in the hospital. Thus, this model can potentially be used as a secondary diagnostic tool, giving every doctor and physician the ability to detect melanoma with reasonably high accuracy, removing the need for unnecessary biopsies.

REFERENCES

[1] World Health Organization, "World Cancer Report," IARC Press, Lyon, 2003.
[2] Global Cancer Observatory, "Indonesia Cancer Report," World Health Organization, 2020.
[3] L. A. Pollack *et al.*, "Melanoma survival in the United States, 1992 to 2005," *Journal of the American Academy of Dermatology*, vol. 65, no. 5, pp. S78.e1-S78.e10, 2011, doi: 10.1016/j.jaad.2011.05.030.
[4] Melanoma Research Alliance, "Melanoma Survival Rates," MRA, Washington DC, 2018.
[5] N. Codella *et al.*, "Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted By The International Skin Imaging Collaboration (ISIC)," *Arxiv.Org*, pp. 1-12, 2021.
[6] D. Schadendorf, "Melanoma," *Nature Reviews Disease Primers*, no. 15003 pp. 1-8, 2015.
[7] M. Kenneth, "A Sharper Eye: How Advanced Imaging Technologies and Artificial Intelligence Are Transforming Melanoma Diagnosis," *The Melanoma Letter*, vol. 38, no. 1, p. 1 2020.
[8] J. Dinnes *et al.*, "Visual inspection for diagnosing cutaneous melanoma in adults," *Cochrane Database Systematic Reviewer*, 2018.
[9] X. Yan and X. Su, "Linear Regression Analysis, Singapore: World Scientific," 2009.
[10] I. Goodfellow, Y. Bengio and A. Courville, "Deep Learning, Massachusetts," The MIT Press, 2016.
[11] B. Schnalzer and B. Alcalde, "Skin Lesion Boundary Detection with Neural Networks on iOS Devices," *Proceedings of the third International Conference on Medical and Health Informatics*, 2019, pp. 161-166, doi: 10.1145/3340037.3340057.
[12] M. A. Marchetti, "Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images," *Journal of the American Academy of Dermatology*, vol. 78, no. 2, pp. 270-277, 2018, doi: 10.1016/j.jaad.2017.08.016.
[13] A. S. Glas, J. G. Lijmer, M. H. Prins, G. J. Bonsel and P. M. M. Bossuyt, "The diagnostic odds ratio: a single indicator of test performance," *Journal of clinical epidemiology*, vol. 56, no. 11, pp. 1129-1135, 2003, doi: 10.1016/S0895-4356(03)00177-X.
[14] A. G. Lalkhen, M. C. FRCA, A. McCluskey, B. M. C. FRCA, "Clinical tests: sensitivity and specificity," *Continuing Education in Anaesthesia Critical Care & Pain*, vol. 8, no. 6, pp. 221-223, 2008, doi: 10.1093/bjaceaccp/mkn041.
[15] Ross Roeser, "Audiology," Thieme, pp. 1-14, 2000.
[16] A. M. Simundic, "Measures of Diagnostic Accuracy: Basic Definitions," *EJIFCC*, vol. 19, no. 4, pp. 203-211, 2009.
[17] M. P. Petticrew, A. J. Sowden, D. Lister-Sharp and K. Wright, "False-negative results in screening programs: systematic review of impact and implications," *Health Technol Assess*, vol. 4, no. 5, pp. 1-120, 2000.
[18] T. Hieken, H. Irizarry, J. Boll and J. Coleman, "Accuracy of Diagnostic Biopsy for Cutaneous Melanoma: Implications for Surgical Oncologists," *International Journal of Surgical Oncology*, vol. 2013, ID. 196493, pp. 1-8, 2013, doi: 10.1155/2013/196493.
[19] M. Ca and M. Rm, "Clinical accuracy of the diagnosis of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 138, no. 2, pp. 283-287, 1998, doi: 10.1046/j.1365-2133.1998.02075.x.

[20] A. Sáez, J. S. Monedero, P. A. Gutiérrez and C. Hervás-Martínez, "Machine Learning Methods for Binary and Multiclass Classification of Melanoma Thickness From Dermoscopic Images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1036-1045, April 2016, doi: 10.1109/TMI.2015.2506270.

[21] N. C. F. Codella *et al*., "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 5:1-5:15, 1 July-Sept. 2017, doi: 10.1147/JRD.2017.2708299.

[22] T. M. Jørgensen, A. Tycho, M. Morgensen, P. Bjerring and G. B. E. Jemec, "Machine-learning classification of non-melanoma skin cancers from image features obtained by optical coherence tomography," *Skin Research and Technology*, vol. 14, no. 3, pp. 364-369, 2008, doi: 10.1111/j.1600-0846.2008.00304.x.

[23] A. W. Kopf, M. Mintzis and R. S. Bart, "Diagnostic accuracy in malignant melanoma," *Arch Dermatol*, vol. 111, no. 10, pp. 1291-1292, 1975, doi: 10.1001/archderm.1975.01630220055001.

[24] A. Afdhal, Z. Zulfikar, Z. Zulhelmi and T. Arif, "Image investigation using higher moment statistics and edge detection for recognizing abnormal skin," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 1, pp. 109-120, 2020, doi: 10.11591/eei.v9i1.1287.

[25] D. S. Rigel, "Epidemiology of melanoma," *Seminars in cutaneous medicine and surgery*, vol. 29. no. 4. pp. 205-209, 2010, doi: 10.1016/j.sder.2010.10.005.

[26] S. Hameed, A. Haddad, M. H. Habaebi and A. Nirabi, "Dermatological diagnosis by mobile application," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 3, pp. 847-854, 2019, doi: 10.11591/eei.v8i3.1502.

[27] A. P. J. Olson, M. L. Graber and H. Singh, "Tracking Progress in Improving Diagnosis: A Framework for Defining Undesirable Diagnostic Events," *Journal of General Internal Medicine*, vol. 33, pp. 1187-1191, 2018.

BIOGRAPHIES OF AUTHORS

Herry Sujaini is a lecturer of the Informatics Faculty at Tanjungpura University. He got his bachelor's degree at Tanjungpura University. He continued his study at Bandung Institute of Technology where he got his master's degree. He then continued his study again at Bandung Institute of Technology and got his Doctorate Degree. He currently lectures at Tanjungpura University, teaching various subjects such as machine learning, algorithm, programming, and natural language processing.

Enriko Yudhistira Ramadhan is a bachelor student of the Informatics Faculty at Tanjungpura University. He got his bachelor's degree in 2021 and currently in the process of getting master's degree at Bandung Institute of Technology.

Haried Novriando is a lecturer at Tanjungpura University. He got his bachelor's degree at Indonesia Islamic University in Yogyakarta. He then studied at Gadjahmada University and got his master's degree. He currently lectures at Tanjungpura University, teaching automata theory.