INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

Oral Lesion Image Classification Using Convolutional Neural Network (CovNets) Method Based on MobileNetV2 Architecture

Reza Alamsyah[1], Irwan Jani Tarigan[2]
STMIK Methodist Binjai, Teknik Informatika[1], STMIK Methodist Binjai, Sistem Informasi[2]

Article Info
Keywords:
Oral lesions,
Oral cancer,
Convolutional Neural Network (CNN)

ABSTRACT
Oral lesions, which can appear in various areas of the oral cavity, are often an early indication of oral cancer, one of the most common cancers worldwide and a leading cause of cancer death in South Asia, Southeast Asia, and the Western Pacific. Oral cancer can affect various parts of the mouth and throat, with contributing factors including genetics, smoking, and viral infections. Early detection is critical for effective management of oral cancer, allowing for early treatment that increases the chance of cure and reduces the risk of complications. This study used a Convolutional Neural Network (CNN) to detect images of oral lesions, including benign and malignant lesions, by utilizing the TensorFlow Object Detection API and data from the Oral Images Dataset. Testing with 40 images (20 benign and 20 malignant lesions) showed an accuracy of 92.5%, a precision of 95%, and a recall of 90%, demonstrating the potential of CNN in efficiently detecting oral lesions.

Corresponding Author:
Reza Alamsyah
STMIK Methodist Binjai, Teknik Informatika
reza@stmikmethodistbinjai.ac.id

INTRODUCTION

Oral lesions can appear in various areas of the mouth, such as the lips, tongue, base of the tongue, gums, roof of the mouth, back of the cheeks, and salivary glands. These lesions are often an early indication of oral cancer, which is one of the most common cancers in the world and is the leading cause of cancer death in men in many countries in South Asia, Southeast Asia, and the Western Pacific. Oral cancer can affect various parts of the mouth and throat, with causative factors including genetic factors, smoking, and viral infections (WHO, 2023).

When oral cancer spreads, symptoms can become more serious and difficult to treat. Therefore, it is very important to have regular check-ups and avoid risk factors that can trigger the growth of lesions in the oral cavity. Prevention and early detection play a crucial role in managing oral cancer, so that treatment can be started earlier, increasing the chances of cure, and reducing the risk of more severe complications.

According to Humaira et al. (2020), early detection is the main key in controlling cancer, especially oral cancer, effectively. Therefore, a routine screening process is needed, namely a procedure that aims to detect cancer or precancerous conditions in the oral cavity before clinical symptoms appear. This procedure is very important because it allows for more effective treatment and increases the chances of recovery. This is especially important in countries with limited resources, where early detection can significantly reduce the death rate from oral cancer. However, challenges such as limited resources, lack of trained medical personnel, and low public awareness can hinder the implementation of effective screening programs. Accurate diagnosis of this problem can be done through the biopsy specimen stage, also known as histopathological evaluation. Potentially malignant oral disorders are one of the important oral health problems.

In this study, the Convolutional Neural Network (CNN) method is applied to detect oral lesion images, which falls under the category of object detection and image classification. CNN is well suited for this task due to its ability to recognize complex visual patterns, such as skin features that indicate the presence of lesions. The use of TensorFlow Object Detection API in this study helps in developing and training the oral lesion image detection model more efficiently.

Literature review

a. Lesion
Lesion is a term used in medical media to describe any type of disorder or change in body tissue, especially those caused by disease or injury. Lesions can appear in various parts of the body and can be inflammation, wounds, or tumors, which can be benign or malignant (Esha et al., 2022).

b. Types of Lesions
There are various types of lesions that can occur in the human body, including skin lesions such as wounds and boils, lesions in internal organs such as tumors and abscesses. Lesions can also be found in the human nervous system such as brain tumors and lesions due to stroke, there are also lesions in other parts of the body such as cardiovascular, oral, digestive, respiratory, reproductive, and others. Such as: oral lesions.

c. OpenCV
OpenCV is an open-source software library designed for image processing and computer vision. This library offers a variety of functions for image analysis, including object detection, facial recognition, motion tracking, and video processing, among others. OpenCV is written primarily in C++ and provides bindings for several programming languages, such as Python, Java, and MATLAB, making it very flexible and easy to integrate into various applications (Susim et al, 2021). OpenCV was first developed by Intel in 1999 as part of the company's efforts to increase computing power in the field of computer vision. The initial goal of developing OpenCV was to provide researchers and developers with a powerful tool for processing and analyzing images efficiently. In 2000, OpenCV was released as an open-source project under the BSD license, allowing users to freely access, modify, and distribute the source code. OpenCV offers a variety of features and functions that allow users to perform various tasks in image processing and computer vision. Some of these features include image processing, object detection and recognition, object tracking, facial recognition, video processing, and applications in machine learning and deep learning.

d. Deep Learning

INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

*Deep Learning*is a branch of Machine Learning that utilizes artificial neural networks inspired by the structure of neurons in the human brain to process data and make decisions. As a subfield of machine learning, deep learning focuses on the development and use of algorithms that can build models with a high level of abstraction in representing data. Unlike traditional machine learning approaches, deep learning uses artificial neural networks with many layers, known as deep neural networks. Deep learning models are designed to learn from large amounts of data, extract complex features, and make accurate predictions without the need for manual feature coding by humans (Ilahiyah & Nilogiri, 2018).

e. Image Processing

*Image processing*is a field of science in information technology that focuses on digital image processing. The main goal of image processing is to manipulate or analyze digital images so that the information contained in them can be extracted or enhanced. This process involves processing analog images or images to reduce data errors during transmission and detection of image signals, and to improve the quality of image display. The main goal is to facilitate the interpretation of images by the human visual system through image manipulation and analysis.

Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a type of deep neural network specifically designed to process two-dimensional data with a deep network structure. CNN is often used in image analysis, detection, and object recognition, due to its ability to handle high-dimensional vectors with many parameters, which is a characteristic of this network (Sriyati & Arief Setyanto, 2018). CNN is very effective in solving problems related to Machine Learning and Deep Learning, especially in applications involving two-dimensional image data, Computer Vision, and Natural Language Processing (Wulandari, 2020).

Convolution Layer

Convolutional *layer*is a common type of layer used to perform convolution operations, which transform input into feature maps by performing pointwise operations between the filter and the input matrix. The filter then computes the value at the current position, moves by shifting to the right, and computes again until all pixel positions have been computed. The result is a square kernel that is used as input to the next layer.

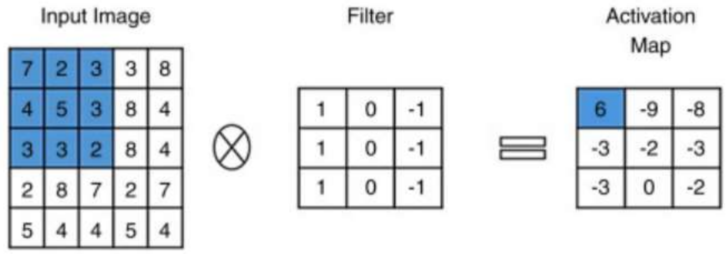


Figure 1. Convolution Operation (Mostafa & Wu, 2021)

The convolution calculation can be done using equation 1.1

$$h_x = f_x * g_x \quad (2.1)$$

Pooling Layer

Pooling Layeraims to reduce the number of parameters of the input tensor, which helps prevent overfitting, extract representative features from the input tensor, and reduce the computational burden, thereby improving efficiency. There are two types of pooling, namely Max Pooling and Average Pooling. In Max Pooling, as shown in Figure 2.4, a kernel of size n * n (e.g. 2x2) is shifted across the matrix, and the maximum value of each kernel position is taken and placed in the corresponding position in the output matrix. Whereas in Average Pooling, a kernel of size n * n is shifted across the matrix and the average of all values at each kernel position is taken and placed in the corresponding position in the output matrix. This process is carried out for each channel in the input tensor until the output tensor is obtained. It is important to note that pooling reduces the size of the image in terms of height and width, but the number of channels (depth) remains unchanged (Peryanto et al., 2020).

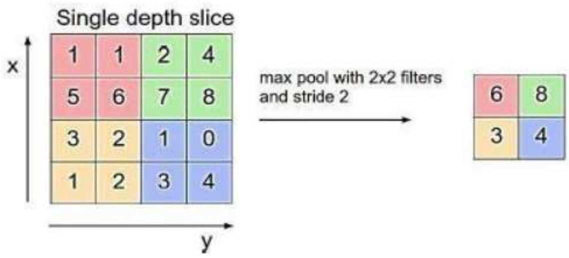


Figure 2. Max Pooling Operation (Peryanto et al., 2020)

ReLU Layer

ReLU (Rectified Linear Unit) activation is an activation function that changes pixel values to 0 if the value is less than 0, while pixel values greater than 0 remain unchanged. The purpose of this function is to reduce errors and improve quality by reducing the impact of negative pixel values, which can help speed up convergence and improve model performance in image processing. Here is the equation for the ReLU function:

$$f_x = \left(max\ (0, x)\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.2)$$

x is the output of the neuron, if the value of x is greater than 0 then the value is x itself, if the value is less than 1 or equal to 0, then the value is 0.

Fully Connected

Fully-connected layer is a layer where all activity neurons from the previous layer are connected to neurons in the next layer just like an artificial neural network can. Each activity from the previous layer needs to be converted into one-dimensional data before it can be connected to all neurons in the Fully-Connected layer.

Flatten

Flatten layer is a layer that converts multi-dimensional feature maps into one-dimensional feature maps. Flatten layer produces output in the form of one vector. The length obtained from this flatten layer will be used in the fully connected layer input.

Softmax

*Softmax*is another form of logistic regression algorithm that is used to classify more than two classes. Standard classification, which is generally performed by the logistic

INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

regression algorithm, is a binary classification task. Here is the formula for calculating the softmax value:

$$S_{O_i} = \frac{e^{o^i}}{\sum_{j=1}^{n} e^{oj}} \quad \text{...} (2.3)$$

MobileNetV2

*MobileNetV2*is a convolutional neural network architecture designed for optimal performance on mobile devices. This architecture uses an inverted residual structure with residual connections between bottleneck layers. Intermediate expansion layers utilize lightweight depthwise convolutions to process features and add non-linearity. Overall, MobileNetV2 starts with a full convolution layer that has 32 filters, and continues with 19 residual bottleneck layers (Mark et al, 2019).

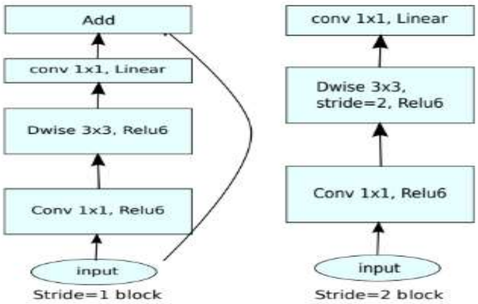


Figure 3. MobileNetV2 CNN Architecture Model (Mark et al, 2019

METHOD

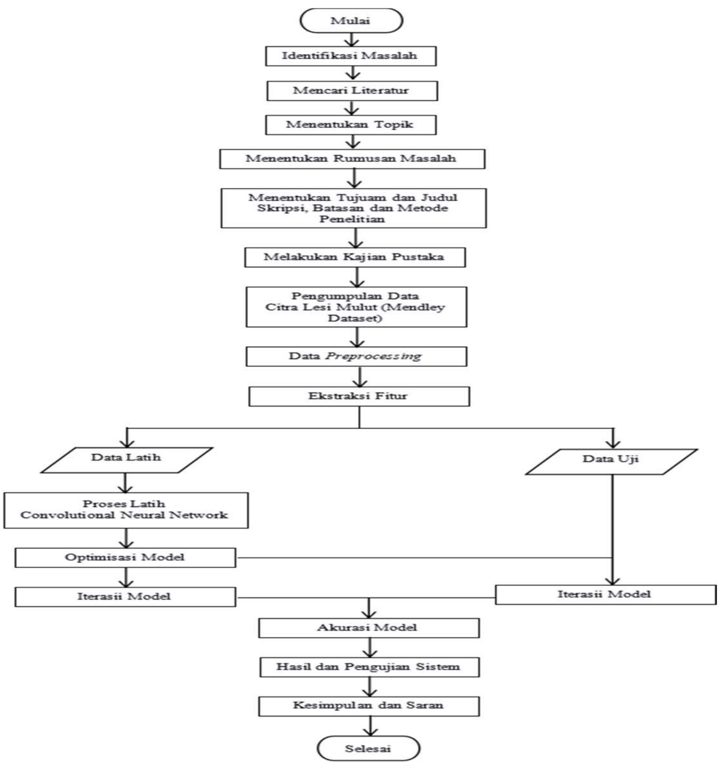


Figure 4. Research Design

INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

Data collection

Research Data

This study uses secondary data types in the form of oral lesion images from the Oral Images Dataset. This dataset can be accessed through the link provided: https://data.mendeley.com/datasets/mhjyrn35p4/2. The selection of this data supports an in-depth analysis of the characteristics of oral lesions, expanding the understanding of diagnosis and treatment. By utilizing this dataset, research can explore various visual features and utilize image analysis techniques to produce a deeper understanding in this field. The process of data analysis and implementation on oral lesions images from two types of oral lesion diseases, including: benign lesions and malignant lesions. The data used is 2230 images including training data and test data.

Literature Study Method

The process of collecting data from various sources such as journals, books related to the problems in the research objectives. Theories that are relevant to the problems studied as reference materials in discussing the results of the problems in the thesis research.

Data Analysis

Research Data Variables

The research data variables are as follows:

Table 3.1. Research Data Variables for Oral Lesions

Oral Lesions Image Data	Label	Label Name
	0	benign lesion
	1	malignant lesion

RGB image pixel values

In general, the input image is a 24-bit color image (true color) the pixel value consists of 3 bytes, namely the Red Green and Blue (RGB) values contained in binary bitmap data. To read the RGB value, it is done by reading the bitmap data which is 3 bytes long, each byte represents the R, G, and B components. Each byte of data represents 8 bits so in the color image there are 3 bytes x 8 bits = 24 bits of color content. The 24-bit color image can be seen as in Figure 3.2.

Figure 5. Image of Oral Lesions

Grayscale image pixel values

Grayscale images have color gradations ranging from white to black by averaging the intensity values of the RGB color image to obtain the intensity value of the gray image. The calculation process for grayscale images can be seen.

INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

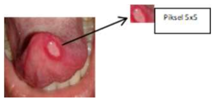


Figure 6. Example of RGB image extraction to Grayscale

From the image above, it is a representation of an RGB image with 5 x 5 pixels, so the color value intensity can be changed to grayscale as follows with equation (2.3), which can be seen in table 3.1:

Table 3.2 Example of 5x5 pixel RGB image

R	G	B	R	G	B	R	G	B	R	G	B	R	G	B
200	125	206	100	179	145	45	100	119	100	179	145	200	125	206
189	245	230	209	30	117	125	206	89	209	30	117	100	179	145
125	90	128	200	125	206	117	125	206	117	125	206	209	30	117
100	179	145	117	125	206	76	89	200	125	206	90	117	125	206
209	30	117	45	100	119	117	125	206	45	100	119	67	112	190

From the table above, the grayscale image value can be calculated using equation (2.3) as follows:

$$I_{bw}(x,y) = \frac{I_R(x,y) + I_G(x,y) + I_B(x,y)}{3}$$

$$f(1,1) = \frac{(200 + 125 + 206)}{3} = 177$$

$$f(1,2) = \frac{(189 + 245 + 230)}{3} = 221$$

$$f(1,3) = \frac{(125 + 90 + 128)}{3} = 114$$

$$f(1,4) = \frac{(100 + 179 + 145)}{3} = 141$$

$$f(1,5) = \frac{(209 + 30 + 117)}{3} = 119$$

So, if all the image data in the example of the 5 x 5 pixel RGB image above is changed to grayscale or gray values as in the table below.

Table 3.3 Results of RGB Image Value Calculation *Grayscale*

177	141	88	141	177
221	119	140	119	141
114	177	149	149	119
141	149	122	140	149
119	88	149	88	123

RESULTS AND DISCUSSION

Convolution Process

0	0	0	0	0	0	0
0	50	10	21	12	23	0
0	23	13	60	67	11	0
0	65	43	78	16	11	0
0	66	21	10	45	67	0
0	78	60	21	11	15	0
0	0	0	0	0	0	0

0	0	0	0	0	0	0
0	56	45	67	23	15	0
0	55	14	70	44	24	0
0	54	56	10	13	15	0
0	22	32	10	43	77	0
0	17	20	56	26	27	0
0	0	0	0	0	0	0

0	0	0	0	0	0	0
0	43	43	66	80	27	0
0	45	15	12	56	27	0
0	90	88	54	55	26	0
0	43	73	12	21	32	0
0	67	10	22	43	31	0
0	0	0	0	0	0	0

INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

Figure 7. RGB Image Intensity Values

Red, Green and Blue Convolution

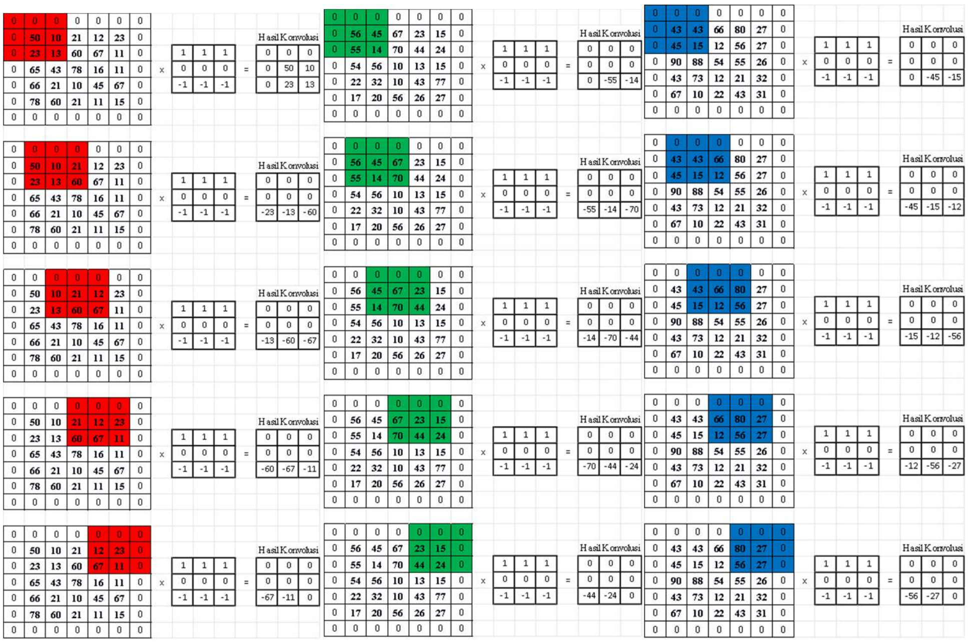


Figure 8. Red, Green and Blue Convolution

So from the convolution results in Figure 8 above, the results of the addition of the Red, Green and Blue convolutions and the ReLU process can be obtained.

Hasil Akhir Konvolusi Red				
96	-96	-140	-138	-78
-48	-105	-94	-49	8
-51	-1	64	15	-34
-30	27	45	58	1
87	97	76	122	112

Proses ReLU				
96	0	0	0	0
0	0	0	0	8
0	0	64	15	0
0	27	45	58	1
87	97	76	122	112

Hasil Akhir Konvolusi Green				
-69	-139	-128	-138	-68
-9	48	56	67	10
15	75	43	8	-52
73	27	-23	-71	-25
54	64	85	85	120

Proses ReLU				
0	0	0	0	0
0	48	56	67	10
15	75	43	8	0
73	27	0	0	0
54	64	85	85	120

Hasil Akhir Konvolusi Blue				
-60	-72	-83	-95	-83
-92	-80	-8	38	26
-56	-56	-23	30	30
101	133	122	39	7
116	128	106	106	53

Proses ReLU				
0	0	0	0	0
0	0	0	38	26
0	0	0	30	30
101	133	122	39	7
116	128	106	106	53

Figure 9. Final Result of RGB Convolution Sum and ReLU Process

So in Figure 9 above, the results of the pooling layer and the sum of the RGB pooling layers in the ReLU process can be obtained.

INFOKUM
Volume 12, Number 04, 2024, DOI 10.58471/infokum.v12i04
ESSN 2722-4635 (Online)
https://infor.seaninstitute.org/index.php/infokum

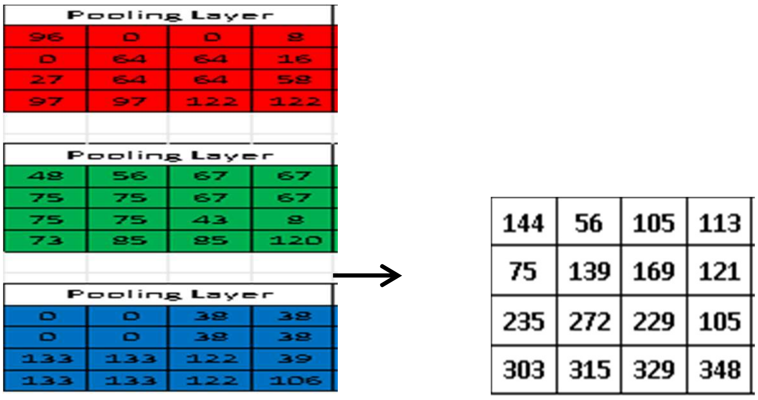


Figure 10. Pooling layer summation results

Flatten

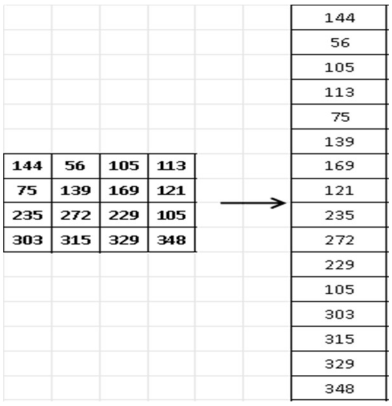
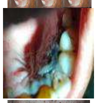


Figure 11. Flatten Results on Pooling Layer

Confusion Matrix Test Results

Table 3.4. Malignant Lesion Image Prediction Results

No Testing	Input Image	Classification Results	Information
1		Malignant Lesions	Correct
2		Malignant Lesions	Correct
3		Malignant Lesions	Correct
4		Malignant Lesions	Correct
5		Benign Lesions	Wrong
6		Malignant Lesions	Correct

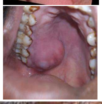
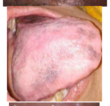
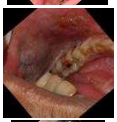
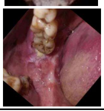
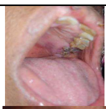
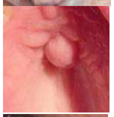
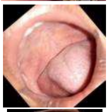
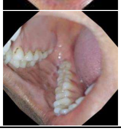
No Testing	Input Image	Classification Results	Information
7		Malignant Lesions	Correct
8		Malignant Lesions	Correct
9		Malignant Lesions	Correct
10		Malignant Lesions	Correct
11		Benign Lesions	Wrong
12		Malignant Lesions	Correct
13		Malignant Lesions	Correct
14		Malignant Lesions	Correct
15		Malignant Lesions	Correct
16		Malignant Lesions	Correct
17		Malignant Lesions	Correct
18		Malignant Lesions	Correct
19		Malignant Lesions	Correct
20		Malignant Lesions	Correct

Table 3.5. Benign Lesion Image Prediction Results

No Testing	Input Image	Classification Results	Information
1		Benign Lesions	Correct

No Testing	Input Image	Classification Results	Information
2		Benign Lesions	Correct
3		Benign Lesions	Correct
4		Malignant Lesions	Wrong
5		Benign Lesions	Correct
6		Benign Lesions	Correct
7		Benign Lesions	Correct
8		Benign Lesions	Correct
9		Benign Lesions	Correct
10		Benign Lesions	Correct
11		Benign Lesions	Correct
12		Benign Lesions	Correct
13		Benign Lesions	Correct
14		Benign Lesions	Correct
15		Benign Lesions	Correct
16		Benign Lesions	Correct

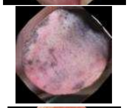
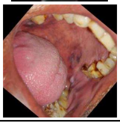
No Testing	Input Image	Classification Results	Information
17		Benign Lesions	Correct
18		Benign Lesions	Correct
19		Benign Lesions	Correct
20		Benign Lesions	Correct

Table 4.3 Confusion Matrix Results of Lesion Images

	Malignant Lesions	Benign Lesions	Total Test Image	Accuracy
Malignant Lesions	18	2	20	90%
Benign Lesions	1	19	20	95%

Accuracy Level

Based on the results of the confusion matrix in table 4.3, the average accuracy value can be calculated using equation 2.3, as follows:

Accuracy =(TP + TN) / (TP + TN + FP + FN) * 100 %
= (18 + 19) / (18 + 19 + 1 + 2) * 100%
= (37 / 40) * 100 %
= 0.925 * 100 %
= 92.5%

Recall

From the confusion matrix results, the recall results can be known, where the recall value is used to determine how precise the model is when it is matched again using different images on the input through the system. The following are the results of the recall value calculation for oral lesion image data:

a. Benign Lesions

Recall = (TP) / (TP + FN) * 100%
= (19) / (19 + 2) * 100%
= 19 / 21 * 100%
= 0.90 * 100%
= 90%

b. Malignant Lesions

Recall = (TP) / (TP + FN) * 100%
= (18) / (18 + 2) * 100%
= 18 / 20 * 100%

= 0.90 * 100%
= 90%

Precision

In the confusion matrix, precision results are also obtained, where the precision value results function to determine how precise the CNN model is when tested with input images. The following are the results of calculating the precision value for oral lesion image data:

a. Benign Lesions

Precision = (TP) / (TP + FP) * 100%
= (19) / (19 + 1) * 100%
= 19 / 20 * 100%
= 0.95 * 100%
= 95%

b. Malignant Lesions

Precision = (TP) / (TP + FP) * 100%
= (18) / (18 + 1) * 100%
= 18 / 19 * 100%
= 0.95 * 100%
= 95%

REFERENCES

WHO, International Agency for Research on Cancer, 2023. *Launch of IARC Handbooks of Cancer Prevention Volume 19: Oral Cancer Prevention.*

Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov and Liang-Chieh Chen. 2019. *MobileNetV2: Inverted Residuals and Linear Bottlenecks.*

Sayuti Rahman, Arnes Sembiring, Dodi Siregar, Husnul Khair, I Gusti Prahmana, Ratih Puspadini and Muhammad Zen. 2023. PYTHON : DASAR DAN PEMROGRAMAN BERORIENTASI OBJEK.

Humaira Nazir and Wakambum Monalisa, 2020. *Study of Precancerous Lesions and Conditions by Clinical Examination, Chemiluminescence, and Toluidine Blue as Early Detection Tool- Retrospective Study.* https://aimdrjournal.com/wp-content/uploads/2021/06/DE1_OA_Humaira-edit.pdf

Aarushi Shah, Manan Shah, Aum Pandya, Rajat Sushra, Ratnam Sushra, Manya Mehta, Keyur Patel, and Kaushal Patel. 2023. *A comprehensive study on skin cancer detection using artificial neural network (ANN) and convolutional neural network* (CNN).

Rohim, A., Sari, Y. A., & Tibyani. (2019). Convolution neural network (cnn) untuk pengklasifikasian citra makanan tradisional. Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 3(7), 7038–7042.

Swasono, D. I., Wijaya, M. A. R., & Hidayat, M. A. (2023). Klasifikasi Penyakit pada Citra Buah Jeruk Menggunakan Convolutional Neural Networks (CNN) dengan Arsitektur Alexnet. INFORMAL: Informatics Journal, 8(1), 68.

Tsany, A., & Dzaky, R. (2021). Deteksi Penyakit Tanaman Cabai Menggunakan Metode Convolutional Neural Network. E-Proceeding of Engineering, 8(2), 3039–3055.

Ulya, H., Darmanti, S., & Ferniah, R. S. (2020). Pertumbuhan Daun Tanaman Cabai (Capsicum annuum L.) yang Diinfeksi Fusarium oxysporum pada Umur Tanaman yang Berbeda. Jurnal Akademika Biologi, 9(1), 1–6.

Wulandari, I., Yasin, H., & Widiharih, T. (2020). Klasifikasi Citra Digital Bumbu Dan Rempah Dengan Algoritma Convolutional Neural Network (Cnn). Jurnal Gaussian, 9(3), 273–282. https://doi.org/10.14710/j.gauss.v9i3.27416

Yuliany, S., Aradea, & Andi Nur Rachman. (2022). Implementasi Deep Learning pada Sistem Klasifikasi Hama Tanaman Padi Menggunakan Metode Convolutional Neural Network (CNN). Jurnal Buana Informatika, 13(1), 54–65.

Ezar Al Rivan, M., & Giovri Riyadi, A. (2021). Perbandingan Arsitektur LeNet dan AlexNet Pada Metode Convolutional Neural Network Untuk Pengenalan American Sign Language. Jurnal Komputer Terapan, 7(Vol. 7 No. 1 (2021)), 53–61.

Hayati, E., Mahmud, T., & Fazil, R. (2012). PENGARUH JENIS PUPUK ORGANIK DAN VARIETAS TERHADAP PERTUMBUHAN DAN HASIL TANAMANCABAI (Capsicum annum L.). Urnal Floratek, 7(2), 173–181.

Hidayat, D. (2022). Klasifikasi Jenis Mangga Berdasarkan Bentuk Dan Tekstur Daun Menggunakan Metode *Convolutio Nalneural Network(Cnn) Classification of Types of Mango Based on Leave Shape and Texture Using Convolutio Nalneural Network(Cnn) Method*. Journal of Information Technology and Computer Science (INTECOMS), 5(1), 98–103.

Ilahiyah, S., & Nilogiri, A. (2018). Implementasi Deep Learning Pada Identifikasi Jenis Tumbuhan Berdasarkan Citra Daun Menggunakan Convolutional Neural Network. JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia), 3(2), 49–56.