

Classification of cervical spine fractures using 8 variants EfficientNet with transfer learning

Adhitho Satyo Bayangkari Karno¹, Widi Hastomo², Tri Surawan³, Serlia Raflesia Lamandasa⁴,
Sudarto Usuli⁴, Holmes Rolandy Kapuy⁴, Aji Digdoyo³

¹Department of Information System, Faculty of Engineering, Gunadarma University, Depok, Indonesia

²Department of Information Technology, Ahmad Dahlan Institute of Technology and Business, Jakarta, Indonesia

³Department of Mechanical Engineering, Faculty of Technology Industry, Jayabaya University, Jakarta, Indonesia

⁴Department of Management, Faculty of Economics, Sintuwu Maroso University in Central Sulawesi, Indonesia

Article Info

Article history:

Received Oct 28, 2022

Revised Feb 4, 2023

Accepted Mar 9, 2023

Keywords:

Convolutional neural network

Deep learning

EfficientNet

Image classification

Spine fractures

ABSTRACT

A part of the nerves that govern the human body are found in the spinal cord, and a fracture of the upper cervical spine (segment C1) can cause major injury, paralysis, and even death. The early detection of a cervical spine fracture in segment C1 is critical to the patient's life. Imaging the spine using contemporary medical equipment, on the other hand, is time-consuming, costly, private, and often not available in mainstream medicine. To improve diagnosis speed, efficiency, and accuracy, a computer-assisted diagnostics system is necessary. A deep neural network (DNN) model was employed in this study to recognize and categorize pictures of cervical spine fractures in segment C1. We used EfficientNet from version B0 to B7 to detect the location of the fracture and assess whether a fracture in the C1 region of the cervical spine exists. The patient data group with over 350 picture slices developed the most accurate model utilizing the EfficientNet architecture version B6, according to the findings of this experiment. Validation accuracy is 99.4%, whereas training accuracy is 98.25%. In the testing method using test data, the accuracy value is 99.25%, the precision value is 94.3%, the recall value is 98%, and the F1-score value is 96%.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Widi Hastomo

Department of Information Technology, Ahmad Dahlan Institute of Technology and Business

Jakarta, Indonesia

Email: Widie.has@gmail.com

1. INTRODUCTION

The human spine is a critical body component that permits the body to stand upright. The spine is made up of segments that allow the spine to move freely and the body to conduct a range of movements. The spine is divided into segments, which include 7 segments (C1-C7) of the upper spine (cervical spine), 12 segments (T1-T12) of the chest, 5 waist segments (L1-L5), 5 sacral segments (S1-L5), and a caudal section [1] as shown in Figure 1. Muscles and tendons also exist in the spine and serve as connections between bone segments, nerves, and other essential tissues that link various organs in the brain and body. Bone is composed of two distinct tissues: thicker and more compact on the outside and a network of thin fibers on the interior [2].

A cervical spine fracture (CS-fx) can cause significant damage and a high likelihood of paralysis, and a delay in diagnosis can result in a long illness and a high risk of mortality. After one year of surgery, the risk of death remains considerable for people over the age of sixty [3], [4]. Several studies on the causes and hazards of spinal fractures have been conducted: Fredø *et al.* [5] determined that over 3,000 people met the

criterion for major cervical spine injury during a three-year period (2009-2012). Approximately 3,000 individuals had one or more CS-fx, with roughly 300 suffering significant non-fractured upper cervical spine damage. These patients are frequently above the age of fifty, and men suffer around 70% more than women. The most prevalent cause of accidents, according to Leucht *et al.* [6], is falling (39%), followed by driving accidents (26.5%). This incident caused a cervical spine fracture (65%) and extensive damage to several segments (80%). Watanabe and colleagues [7] suffering is most frequent in the elderly because to deterioration of strength and bone mass, and it spreads to the upper spine and neck. High-energy impacts generated by high levels of exercise result in more injuries among young people.

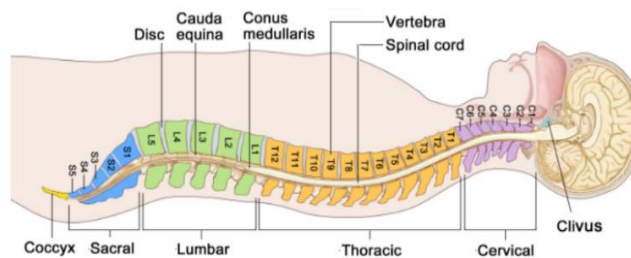


Figure 1. Human vertebral-column

Medical imaging technologies, such as computed tomography [8]–[12], have become widely used for image processing, allowing doctors to make a more detailed diagnosis. Magnetic resonance imaging [13]–[19] has been used to look at areas that are difficult to see because they are obscured by other organs. However, imaging the spine using modern medical technology is time-consuming, expensive, classified, and not frequently available in primary care. Computer-aided diagnostics systems are required to improve diagnostic speed, accuracy, efficiency, and accuracy.

Artificial neural networks (ANNs) with progressively thick layers have grown rapidly in recent years toward specialized computing technologies, particularly deep learning (DL). Because of their ability to examine a growing amount of data, DL techniques are gaining favor as computers develop. With the availability of additional devices capable of speeding up the computer process [20]. DL enhances medical image classification by using layers such as pooling, convolution, fully connected, activation, and a variety of other hyperparameter settings. This study used deep learning to classify upper cervical spine fractures (C1 segment). ImageNet is also used for transfer learning to improve training effectiveness and efficiency. In this investigation, we employed eight varieties from the EfficientNet family (variants B0 to B7).

2. RELATED WORK

Several research on the categorization of vertebral fractures using CNN have been reviewed briefly: Small and colleagues [21] CNN was evaluated for its ability to aid radiologists in the detection of cervical spine fractures. To identify cervical spine fractures, Salehinejad *et al.* [22] employed a long-short-term memory (LSTM) layer. Voter *et al.* [23] did research on the performance of artificial intelligence decision assistance systems, as well as an examination of failures and bad performance. Boonrod *et al.* [24] employed codeless DL as the foundation for their training and testing. Calculate the network model's level of diagnostic accuracy, sensitivity, and specificity. Merali *et al.* [25] employed the ResNet-50 convolutional neural network (CNN) architecture to assess the density of the cervical spinal cord in their study.

3. METHOD

This study's technique consists of multiple steps, beginning with integrating several data files into one file and then filtering and sorting to extract the required data. Divide the data into training and validation sets before running the training process to determine the best model. Using the model to classify the validation data. Then, using the segmentation box, locate the fault. Figure 2 depicts the experimental flow.

3.1. Classification

The supervised classification method includes making one or more characteristics a target and categorizing the target, as well as other features serving as input data in the learning process. If the target class is not accessible, the procedure is known as unsupervised classification, and it is followed by the

clustering process. This study employs supervised classification for feature segment C1, which is separated into two classes, namely whether or not there is a break in segment C1, and the input for the data train is an image that has been translated into a specific value based on the pixel.

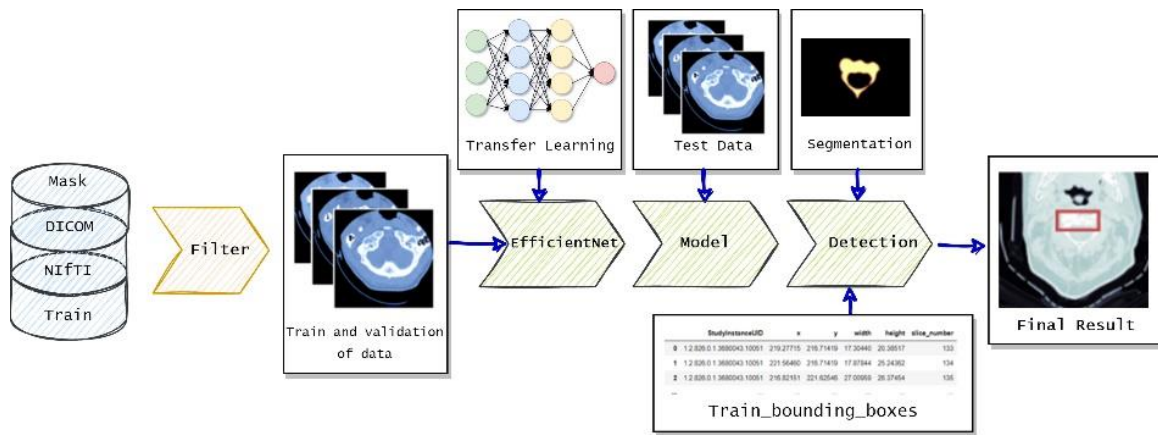


Figure 2. Chart of experimental flow

3.2. Convolutional neural network

Convolutional neural network (CNN) is a technique for assessing informative components of a picture that has been broken into smaller image sections (windows). The window will change based on the stride value to hunt for local characteristics that might give useful information. Then, convert the window into a numerical matrix (filter). In the filter matrix, many weight combinations can be used. Softmax and ReLU are two popular types of filters. To distinguish the picture bounds, a padding technique is typically utilized, which involves adding pixels to the image boundaries' edges as shown in Figure 3. A pooling strategy is used to summarize key information as shown in Figure 4. The window matrix is combined (pooled) into a vector using the pooling technique. Max pooling and average pooling are two forms of pooling that are often employed in neural networks. The fully connected approach is then utilized to aggregate all of the information collected for picture categorization.

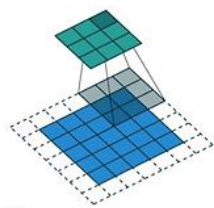


Figure 3. Illustration of padding and convolution

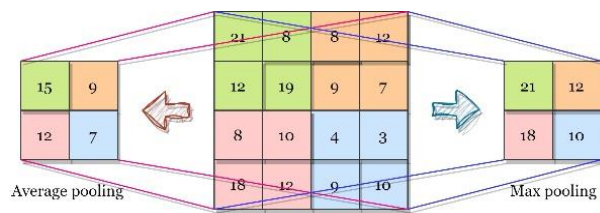


Figure 4. Example of maximum and average pooling

3.3. Dataset preprocessing and analysis

The dataset from www.kaggle.com [26] consists of two comma separated values (CSV) files: train.csv and train bounding boxes.csv, as well as two picture folders: train images and segmentation. The train images folder has a number of patient id subfolders (StudyInstanceUID), each of which contains a different number of images and has the digital imaging and communications in medicine (DCM or DICOM) extension. The segmentation folder comprises a series of patient id subfolders (StudyInstanceUID), and each subfolder contains mask pictures in neuroimaging informatics technology initiative (Nifti) format. The train.csv file contains 2,019 rows of patient data showing a fracture (number 1) or no fracture (number 0) for each segment (segments C1 to C7) and the entire cervical spine of each patient as shown in Figure 5. The train bounding boxes.csv file contains 7,217 patient rows with descriptions of the position and size of the box (mask) of the fault area as shown in Figure 6.

Because of restricted computer capabilities, only the required and relevant medical information is included in each DICOM picture file, namely:

- Obtaining data and generating a *PatientID* column from a *StudyInstanceUID* column
- Identifying patients with fracture data in their cervical vertebrae (patient overall=1)
- Counting the number of photographs associated with each *PatientID*
- Select data *TransferSyntaxIUD=Explicit VR Little Endian*
- Select *PhotometricInterpretationInfo=MONOCHROME2*
- Select *BitsStored=16*
- Select a picture with a resolution of 512×512 pixels.
- Retrieves *RowsColumns* and *PixelSpacing* information for each DICOM picture
- Retrieves *WindowCenter* and *WindowWidth* information for each DICOM image
- Selects each image with *PixelRepresentation=1*

Figure 7 shows some of the data resulting from the merger of the selected data.

	StudyInstanceUID	patient_overall	C1	C2	C3	C4	C5	C6	C7
0	1.2.826.0.1.3680043.6200	1	1	1	0	0	0	0	0
1	1.2.826.0.1.3680043.27262	1	0	1	0	0	0	0	0
2	1.2.826.0.1.3680043.21561	1	0	1	0	0	0	0	0
...
2016	1.2.826.0.1.3680043.14341	0	0	0	0	0	0	0	0
2017	1.2.826.0.1.3680043.12053	0	0	0	0	0	0	0	0
2018	1.2.826.0.1.3680043.18786	1	0	0	0	0	0	0	1

2019 rows × 9 columns

	StudyInstanceUID	x	y	width	height	slice_number
0	1.2.826.0.1.3680043.10051	219.27715	216.71419	17.30440	20.38517	133
1	1.2.826.0.1.3680043.10051	221.56460	216.71419	17.87844	25.24362	134
2	1.2.826.0.1.3680043.10051	216.82151	221.62546	27.00959	26.37454	135
...
7214	1.2.826.0.1.3680043.9940	298.00000	119.00000	87.00000	58.00000	142
7215	1.2.826.0.1.3680043.9940	299.00000	120.00000	89.00000	56.00000	143
7216	1.2.826.0.1.3680043.9940	299.04425	121.88319	89.71326	53.46549	144

7217 rows × 6 columns

Figure 5. Train.csv file

Figure 6. Train bounding boxes.csv file

	StudyInstanceUID	PatientID	patient_overall	Jum_dcm	TransferSyntaxIUD	PhotometricInterpretation	BitsStored	RowsColumns	PixelSpacing	WindowCenter	WindowWidth	PixelRepresentation
0	1.2.826.0.1.3680043.10016	10016	1	645	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.275391, 0.275391]	300.0	2000.0	1
1	1.2.826.0.1.3680043.10051	10051	1	272	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.253906, 0.253906]	400.0	2500.0	1
2	1.2.826.0.1.3680043.10204	10204	1	331	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.416016, 0.416016]	500.0	2000.0	1
3	1.2.826.0.1.3680043.10261	10261	1	308	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.544921994209, 0.544921994209]	752.0	2706.0	1
4	1.2.826.0.1.3680043.10400	10400	1	270	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.332031, 0.332031]	500.0	2000.0	1
...
437	1.2.826.0.1.3680043.9796	9796	1	310	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.302734, 0.302734]	250.0	2500.0	1
438	1.2.826.0.1.3680043.9886	9886	1	329	Explicit VR Little Endian	MONOCHROME2	16	512x512	[0.332031, 0.332031]	500.0	2000.0	1

Figure 7. The results of merging with the selected data

The selection of data is done based on the number of image slices from each patient so that the training data is separated into two parts. The first data set is patient data with the number of slices ranging from 0 to 350 images, while the second data set consists of the number of slices of more than 350 images. In preparation for the training procedure, the data were further separated into three parts: training data (90%), validation data (5%), and test data (5%).

3.4. Pre-trained EfficientNet architecture

Because of its ability to handle vast volumes of input, CNN is a strong and widely used multi-layer neural network [27]. In the past, most computer vision researchers extracted features manually in order to gain better classification results. CNN currently conducts feature extraction automatically during the training phase by utilizing the pooling layer and convolution layer [28]. In general, increasing the available resources allows CNN to improve its accuracy. Increasing the layer depth [29] or breadth [30] is a typical method. A less common, but more popular method is to raise the image resolution size [31]. In comparison to other designs, the EfficientNet family has a balanced layer thickness, layer breadth, and picture resolution as shown in Figure 8(a)-(e). Scaling the model using transfer learning datasets (ImageNet, CIFAR-10, CIFAR-100, CIFAR-101, and Flower), the EfficientNet family significantly outperforms other architectures with high effectiveness, efficiency, fewer parameters, and faster computation [32].

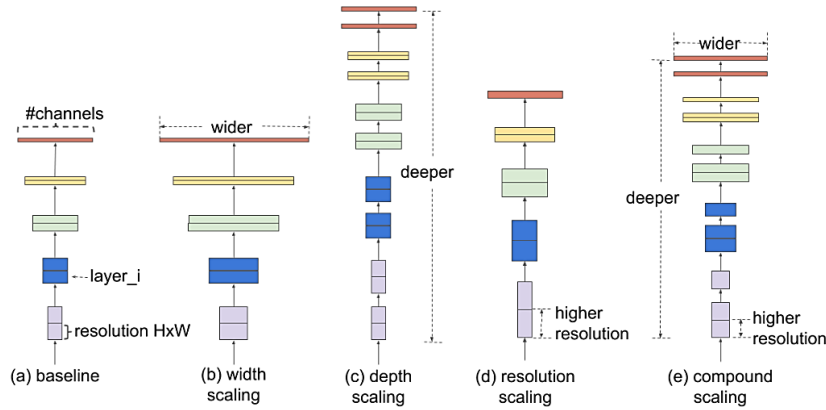


Figure 8. Model sizing EfficientNet layer (a) illustrates a rudimentary network, (b)-(d) are traditional scaling methods that enhance only one network dimension (width, depth, or resolution), and (e) is our suggested integrated scaling approach that evenly scales each of the three aspects at a constant ratio

We present a framework in this research that makes use of eight architectural model versions from the EfficientNet family: B0 through B7. These eight variants are utilized as a comparison to acquiring a decent accuracy value; the training procedure with these eight variants is carried out independently. The mobile subblock array (MBCConv) introduced by Tan *et al.* [33], [34] is the foundation of EfficientNet. The greater the variance, the greater the number of channels as shown in Table 1.

Because of the restricted quantity of data, transfer learning (ImageNet) is utilized to achieve model convergence rapidly. Pooling is needed to turn the feature map into a number because the categorization is binary. The sigmoid activation function is then used to connect this layer to the dense layer. Adamax optimization, loss calculation using categorical cross-entropy, metric accuracy, learning rate 0.001, and epoch 50 are used to compile the model [35].

Table 1. Number of channels per stage from 8 variants of EfficientNet architecture

Subblock	B0	B1	B2	B3	B4	B5	B6	B7
Conv3×3	32	32	32	40	48	48	56	64
MBCConv1, k3×3	16	16	16	24	24	24	32	32
MBCConv1, k3×3	24	24	24	32	32	40	40	48
MBCConv1, k3×3	40	40	48	48	56	64	72	80
MBCConv1, k3×3	80	80	88	96	112	128	144	160
MBCConv1, k3×3	112	112	120	136	160	176	200	224
MBCConv1, k3×3	192	192	208	232	272	304	344	384
MBCConv1, k3×3	320	320	352	384	448	512	576	640
Conv1×1+Pooling+FC	1,280	1,280	1,408	1,536	1,792	2,048	2,304	2,560

4. EXPERIMENTAL RESULTS AND DISCUSSION

The procedure was carried out in two steps in this investigation. That example, the first step runs the classification model on the training data, while the second stage uses bounding boxes and segmented image data to locate fracture locations on the test data. The input data is split into two categories. The first group is patient data, which consists of picture fragments ranging from 0 to 350. This first group has a total of 18,288 photos. This input data (18,288 photos) is split into three categories: 90% for training data (16,459 images), 5% for validation data (915 images), and 5% for test data (915 images). The second group consists of patient data, which includes over 350 picture fragments. This second group has a total of 10,614 images. This input data (10,614 images) is split into three categories: 90% for training data (9,552 images), 5% for validation data (531 images), and 5% for test data (531 images).

Figure 9 shows the training results in the form of loss and accuracy graphs for the patient data group that has a number of slices between 0-350 images. The graphs of training loss and validation of loss appear to coincide with loss values between 3.6 and 4.1. In general, the graphs for accuracy training and validation appear to be spread out, except for versions B2 and B3 which look rather tight.

Figure 10 is the result of training for the patient data group which has more than 350 image slices. The training loss and validation loss graphs seem to coincide, but still have a loss value between 2.7 and 3.2. The graphs of accuracy training and accuracy validation look better and coincide, although B4 and B7 seem to spread at the beginning of the epoch but then narrow at the end of the epoch.

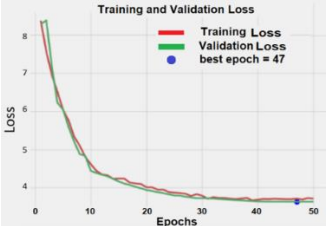













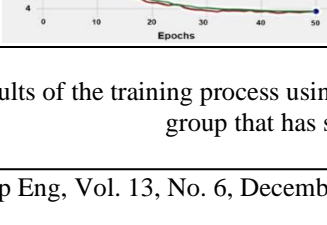
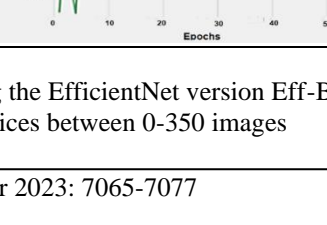
Ver	Graph of loss and accuracy		Best weight			
B0			Train_loss=3.7146	Train_acc=0.7650	Val_loss=3.6356	Val_acc=0.7823
B1			Train_loss=3.7635	Train_acc=0.7525	Val_loss=3.7850	Val_acc=0.7101
B2			Train_loss=3.8196	Train_acc=0.7850	Val_loss=3.8150	Val_acc=0.7954
B3			Train_loss=3.7246	Train_acc=0.8200	Val_loss=3.8762	Val_acc=0.7549
B4			Train_loss=4.0363	Train_acc=0.8175	Val_loss=4.1676	Val_acc=0.6455
B5			Train_loss=3.8555	Train_acc=0.8250	Val_loss=3.9375	Val_acc=0.7505
B6			Train_loss=3.6374	Train_acc=0.8550	Val_loss=3.7517	Val_acc=0.8074
B7			Train_loss=3.8687	Train_acc=0.7825	Val_loss=3.8874	Val_acc=0.7801

Figure 9. The results of the training process using the EfficientNet version Eff-B0-Eff-B7 for the patient data group that has slices between 0-350 images

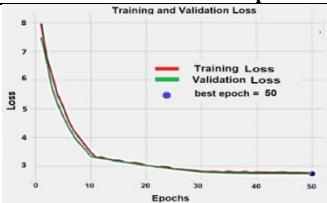

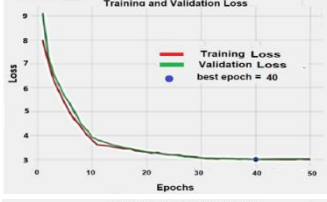

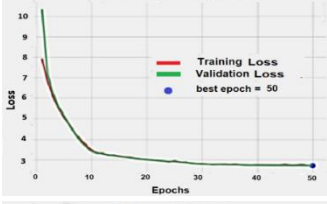

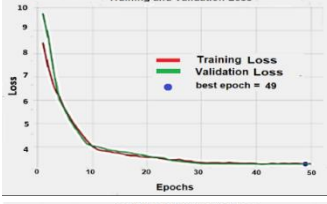

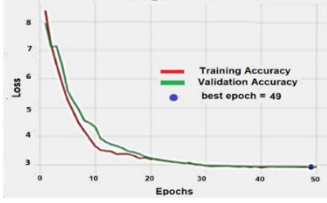
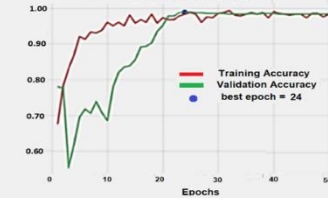






Ver	Graph of loss and accuracy		Best weight			
B0			Train_loss =2.7584	Train_acc=0.9800	Val_loss=2.7401	Val_acc=0.9736
B1			Train_loss=2.9985	Train_acc=0.9725	Val_loss=3.0127	Val_acc=0.9755
B2			Train_loss=2.7296	Train_acc=0.9825	Val_loss=2.7164	Val_acc=0.9887
B3			Train_loss=3.2519	Train_acc=0.9775	Val_loss=3.2390	Val_acc=0.9623
B4			Train_loss =2.9239	Train_acc=0.9750	Val_loss=2.9313	Val_acc=0.9831
B5			Train_loss=3.0613	Train_acc=0.9700	Val_loss=2.9986	Val_acc=0.9906
B6			Train_loss=2.8169	Train_acc=0.9825	Val_loss=2.7675	Val_acc=0.9944
B7			Train_loss=3.1506	Train_acc=0.9820	Val_loss=3.1485	Val_acc=0.9605

Figure 10. Results of the training process using EfficientNet versions Eff-B0-Eff-B7 for patient data groups that have slices of more than 350 images

The training process takes a significant period of time due to limited infrastructure. As a result, it was concluded in this study that the training process was carried out with epoch 50, with the purpose of achieving a reasonably excellent accuracy value. Table 2 displays the processing time for each version, and it can be observed that the time has increased owing to the thicker number of layers. Between versions B4 and B5, there was a significant increase in time.

Table 2. The length of time required for the training process for each version of EfficienNet

Slice	B0	B1	B2	B3	B4	B5	B6	B7
0-350	1:27:36	1:18:22	1:29:47	1:30:15	1:51:9	3:29:52	3:16:49	4:30:28
>350	0:38:31	0:47:25	0:55:21	1:26:25	1:20:36	5:40:29	5:44:43	5:47:55

After the model is obtained using 8 versions of the EfficienNet architecture and two different input datasets. Furthermore, test data will be fed to the model to measure the performance capabilities of the model. This measurement uses a confusion matrix consisting of 5 variables, namely precision, recall, F1-score, support, and accuracy as shown in Figure 11. Figure 12 is the confusion matrix, the test results of the model for a dataset group of 0-350 images and a dataset group of more than 350 images. Measurement variables from training and testing results are numerically collected in one Table 3.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{F1 score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{Accuracy} = \frac{TN+TP}{TN+FP+TP+FN} \quad (4)$$

		Predicted Label	
		Negative	Positive
True Label	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Figure 11. Calculation accuracy with the confusion matrix

Table 3 shows that the accuracy values for the patient group with the number of slices 0-350 in the “TRAINING” column vary from 0.65 to 0.85, while the loss values range from 3.5 to 4.1. The “TESTING” column displays accuracy (0.65 to 0.78), precision (0.5), and recall (0.5 to 0.7). The F1-score for the 0-350 group is roughly 50%, showing that there is some overfitting since the data used during training gives strong predictions but produces poorer predictions during testing. The model for this group does not generalize well, thus when tests are performed using other data, the accuracy is reduced, or the results are not as predicted.

The accuracy values for ‘Train’ and ‘Validation’ in the set of patients with more than 350 slices in the “TRAINING” column vary from 0.97 to 0.99, while the loss values range from 2.7 to 3.2. The “TESTING” column displays accuracy scores ranging from 0.96 to 0.99, precision values ranging from 0.6 to

0.95, and recall levels ranging from 0.75 to 1.0. The F1-score is approximately 90%, and the training and testing accuracy appears to be extremely good. The loss amount must still be increased by increasing the percentage from 5% to 10% for each validation and testing data set. EfficientNet B6 achieved the maximum accuracy value of 0.9925 for patient data groups with slices of more than 350 pictures. Select the EfficientNet B6 kernel model at the end of the procedure to be utilized in deciding the classification and segmentation of the test data as shown in Figures 10 and 11. An Intel(R) Core i5-10400F CPU operating at 2.90 GHz, 8 GB of RAM, Windows 10 (64-bit) OS, and NVIDIA GeForce GT 710 graphics powered this experiment. The training method in this study was carried out on a Kaggle notebook, and we employed CPU and GPU accelerators (T4x2) on Kaggle concurrently to boost the notebook environment's power and shorten training time.

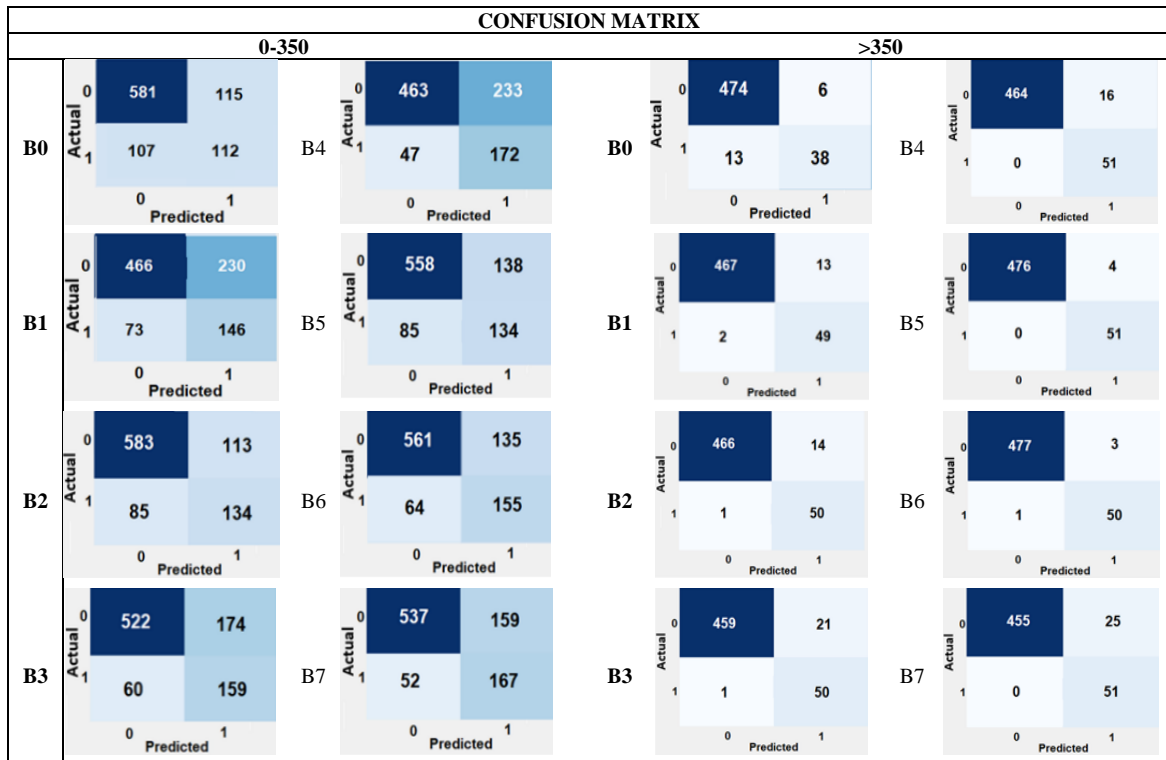


Figure 12. Confusion matrix: test findings and data

Table 3. Variables for measuring the results of training and testing

Slice	TRAINING				TESTING				
	Train		Validation		Classification Report				
	acc	loss	acc	Loss	acc	precision	recall	F1-score	
0-350	B0	0.7650	3.7146	0.7823	3.6356	0.7574	0.4934	0.5114	0.5022
	B1	0.7525	3.7635	0.7101	3.7850	0.6689	0.3883	0.6667	0.4908
	B2	0.7850	3.8196	0.7954	3.8150	0.7836	0.5425	0.6119	0.5751
	B3	0.8200	3.7246	0.7549	3.8762	0.7443	0.4775	0.7260	0.5761
	B4	0.8175	4.0363	0.6455	4.1676	0.6940	0.4247	0.7854	0.5513
	B5	0.8250	3.8555	0.7505	3.9375	0.7563	0.4926	0.6119	0.5458
	B6	0.8550	3.6374	0.8074	3.7517	0.7825	0.5345	0.7178	0.6090
B7	0.7825	3.8687	0.7801	3.8874	0.7694	0.5123	0.7626	0.6128	
>350	B0	0.9800	2.7584	0.9736	2.7401	0.9642	0.8636	0.7451	0.8000
	B1	0.9725	2.9985	0.9755	3.0127	0.9718	0.7903	0.9608	0.8673
	B2	0.9825	2.7296	0.9887	2.7164	0.9718	0.7812	0.9804	0.8696
	B3	0.9775	3.2519	0.9623	3.2390	0.9586	0.7042	0.9804	0.8197
	B4	0.9750	2.9239	0.9831	2.9313	0.9699	0.7612	1.000	0.8644
	B5	0.9700	3.0613	0.9906	2.9986	0.9925	0.9273	1.000	0.9623
	B6	0.9825	2.8169	0.9944	2.7675	0.9925	0.9434	0.9804	0.9615
B7	0.9820	3.1506	0.9605	3.1485	0.9529	0.6711	1.000	0.8031	

This report only presents one patient who was diagnosed as having a fracture in the C1 segment to avoid using too many pages to graphically demonstrate the results of classification and segmentation. Figure 13 depicts an extraction picture from a patient with the ID: 1.2.826.0.1.3680043.12281, and Figure 14 depicts a visual segmentation result from the same patient, which clearly indicates an upper neck fracture (C1) in slices 126 and 127.

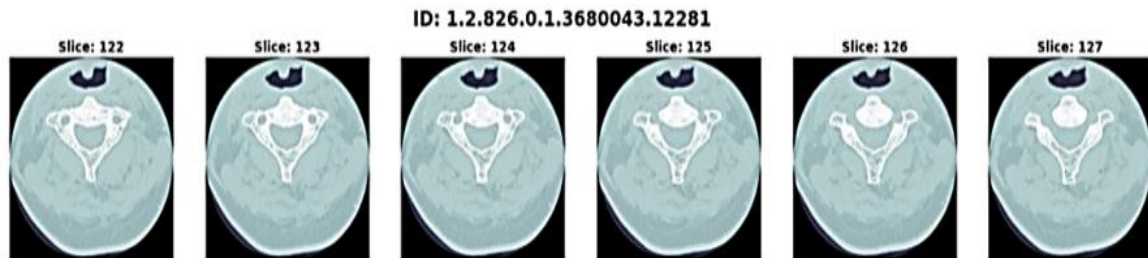


Figure 13. Extraction images of test data

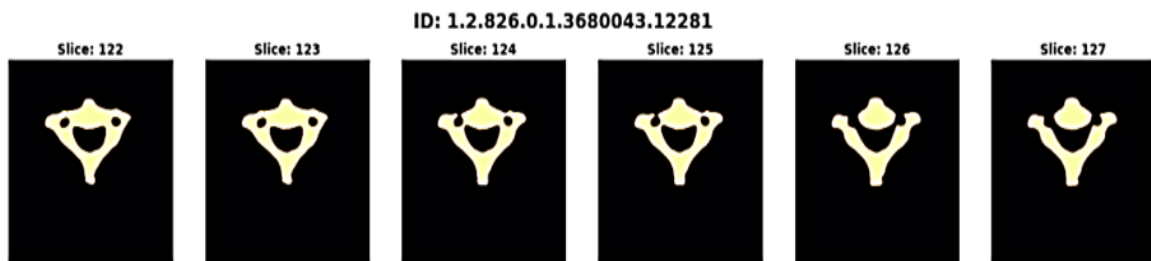


Figure 14. The results of the segmentation of the test data

5. CONCLUSION

Steps in identifying C1 segment cervical spine fractures include data preparation (combining, choosing, and sorting), training the dataset to construct the model, calculating accuracy and loss values, predicting test data, and creating fracture localization boxes. Even though the training and validation accuracy is around 80%, the training and validation loss values are still around 30%, and the F1-score is still about 50% for the patient group with a number of slices ranging from 0 to 350 images. This model cannot be utilized to effectively generalize test data since it produces unexpected predictions. Adding patient data with a number of slices of 0-350 images can enhance results by balancing the data.

Meanwhile, version B6 earned the greatest training accuracy score (98.25%), training validation value of 99.4%, testing accuracy of 99.25%, precision, recall, and F1-score extremely excellent in the patient data group with more than 350 slices. Despite the fact that the training and validation loss values still need to be improved by raising the original percentages for validation and testing data by 5% to 10%. Because it produces predicted outcomes, the model for this data group may be used to generalize to additional test data.




Using a higher version of the design (deeper layers) does not always result in better accuracy, according to our observations. By selecting the appropriate data set, high accuracy may be reached. Based on the processing time of each version, the training duration for the EfficientNet version with a thicker layer tends to be longer when utilizing the same data. Several parameters, such as the number of epochs, learning rate, number of layers, and input pixel size utilized during training, are chosen to attain a high accuracy value while taking into account the computer infrastructure employed.

REFERENCES




- [1] K. S. Saladin, "Anatomy and physiology," *SEER Training Modules*, National Cancer Institute, 2012. <https://training.seer.cancer.gov/anatomy/> (accessed Aug. 20, 2022).
- [2] A. Lichtenegger, "Modeling and simulation of the cervical spine: mechanical stress in injuries," Diploma Thesis, reposiTUM, 2015. doi: 10.34726/hss.2015.24612.
- [3] T. Delcourt, T. Bégué, G. Saintyves, N. Mebtouche, and P. Cottin, "Management of upper cervical spine fractures in elderly patients: current trends and outcomes," *Injury*, vol. 46, pp. 24–27, Jan. 2015, doi: 10.1016/S0020-1383(15)70007-0.

- [4] M. B. Harris *et al.*, "Mortality in elderly patients after cervical spine fractures," *The Journal of Bone and Joint Surgery-American Volume*, vol. 92, no. 3, pp. 567–574, Mar. 2010, doi: 10.2106/JBJS.I.00003.
- [5] H. L. Fredø, I. J. Bakken, B. Lied, P. Rønning, and E. Helseth, "Incidence of traumatic cervical spine fractures in the Norwegian population: a national registry study," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 22, no. 1, Dec. 2014, doi: 10.1186/s13049-014-0078-7.
- [6] P. Leucht, K. Fischer, G. Muhr, and E. J. Mueller, "Epidemiology of traumatic spine fractures," *Injury*, vol. 40, no. 2, pp. 166–172, Feb. 2009, doi: 10.1016/j.injury.2008.06.040.
- [7] M. Watanabe, D. Sakai, Y. Yamamoto, M. Sato, and J. Mochida, "Upper cervical spine injuries: age-specific clinical features," *Journal of Orthopaedic Science*, vol. 15, no. 4, pp. 485–492, Jul. 2010, doi: 10.1007/s00776-010-1493-x.
- [8] L. Tanzi, E. Vezzetti, R. Moreno, A. Aprato, A. Audisio, and A. Massè, "Hierarchical fracture classification of proximal femur X-Ray images using a multistage deep learning approach," *European Journal of Radiology*, vol. 133, Dec. 2020, doi: 10.1016/j.ejrad.2020.109373.
- [9] F. Yang, G. Wei, H. Cao, M. Xing, S. Liu, and J. Liu, "Computer-assisted bone fractures detection based on depth feature," *IOP Conference Series: Materials Science and Engineering*, vol. 782, no. 2, Mar. 2020, doi: 10.1088/1757-899X/782/2/022114.
- [10] P. A. Grützner and N. Suhm, "Computer aided long bone fracture treatment," *Injury*, vol. 35, no. 1, pp. 57–64, Jun. 2004, doi: 10.1016/j.injury.2004.05.011.
- [11] D.-Y. Gu, K.-R. Dai, S.-T. Ai, and Y.-Z. Chen, "Computer-aided fracture diagnosis and classification package embedded in the integrated electronic patient record system," in *IFMBE Proceedings*, Springer Berlin Heidelberg, 2009, pp. 1–4.
- [12] L. Nascimento and M. G. Ruano, "Computer-aided bone fracture identification based on ultrasound images," in *2015 IEEE 4th Portuguese Meeting on Bioengineering (ENBENG)*, Feb. 2015, pp. 1–6, doi: 10.1109/ENBENG.2015.7088892.
- [13] M. P. Koivikko and S. K. Koskinen, "MRI of cervical spine injuries complicating ankylosing spondylitis," *Skeletal Radiology*, vol. 37, no. 9, pp. 813–819, Sep. 2008, doi: 10.1007/s00256-008-0484-x.
- [14] N. D. Tomycz *et al.*, "MRI is unnecessary to clear the cervical spine in obtunded/comatose trauma patients: the four-year experience of a level I trauma center," *Journal of Trauma: Injury, Infection and Critical Care*, vol. 64, no. 5, pp. 1258–1263, May 2008, doi: 10.1097/TA.0b013e318166d2bd.
- [15] T. E. Darsaut *et al.*, "A pilot study of magnetic resonance imaging-guided closed reduction of cervical spine fractures," *Spine*, vol. 31, no. 18, pp. 2085–2090, Aug. 2006, doi: 10.1097/01.brs.0000232166.63025.68.
- [16] A. R. Vaccaro, K. O. Kreidl, W. Pan, J. M. Cotler, and M. E. Schweitzer, "Usefulness of MRI in isolated upper cervical spine fractures in adults," *Journal of Spinal Disorders*, vol. 11, no. 4, Aug. 1998, doi: 10.1097/00002517-199808000-00003.
- [17] W. Yuan *et al.*, "Establishment of intervertebral disc degeneration model induced by ischemic sub-implant in rat tail," *The Spine Journal*, vol. 15, no. 5, pp. 1050–1059, May 2015, doi: 10.1016/j.spinee.2015.01.026.
- [18] Y. Kumar and D. Hayashi, "Role of magnetic resonance imaging in acute spinal trauma: a pictorial review," *BMC Musculoskeletal Disorders*, vol. 17, no. 1, Dec. 2016, doi: 10.1186/s12891-016-1169-6.
- [19] M. Utz, S. Khan, D. O'Connor, and S. Meyers, "MDCCT and MRI evaluation of cervical spine trauma," *Insights into Imaging*, vol. 5, no. 1, pp. 67–75, Feb. 2014, doi: 10.1007/s13244-013-0304-2.
- [20] M. Pandey *et al.*, "The transformational role of GPU computing and deep learning in drug discovery," *Nature Machine Intelligence*, vol. 4, no. 3, pp. 211–221, Mar. 2022, doi: 10.1038/s42256-022-00463-x.
- [21] J. E. Small, P. Osler, A. B. Paul, and M. Kunst, "CT cervical spine fracture detection using a convolutional neural network," *American Journal of Neuroradiology*, vol. 42, no. 7, pp. 1341–1347, Jul. 2021, doi: 10.3174/ajnr.A7094.
- [22] H. Salehinejad *et al.*, "Deep sequential learning for cervical spine fracture detection on computed tomography imaging," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2021, pp. 1911–1914, doi: 10.1109/ISBI48211.2021.9434126.
- [23] A. F. Voter, M. E. Larson, J. W. Garrett, and J.-P. J. Yu, "Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of cervical spine fractures," *American Journal of Neuroradiology*, vol. 42, no. 8, pp. 1550–1556, Aug. 2021, doi: 10.3174/ajnr.A7179.
- [24] A. Boonrod, A. Boonrod, A. Meethawolgul, and P. Twinprai, "Diagnostic accuracy of deep learning for evaluation of C-spine injury from lateral neck radiographs," *Heliyon*, vol. 8, no. 8, Aug. 2022, doi: 10.1016/j.heliyon.2022.e10372.
- [25] Z. Merali, J. Z. Wang, J. H. Badhiwala, C. D. Witiw, J. R. Wilson, and M. G. Fehlings, "A deep learning model for detection of cervical spinal cord compression in MRI scans," *Scientific Reports*, vol. 11, no. 1, May 2021, doi: 10.1038/s41598-021-89848-3.
- [26] A. Flanders *et al.*, "RSNA 2022 cervical spine fracture detection," *Kaggle*, <https://www.kaggle.com/competitions/rsna-2022-cervical-spine-fracture-detection/overview> (accessed Jul. 29, 2022).
- [27] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.
- [28] N. Remzan, K. Tahiry, and A. Farchi, "Brain tumor classification in magnetic resonance imaging images using convolutional neural network," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 6, pp. 6664–6674, Dec. 2022, doi: 10.11591/ijece.v12i6.pp6664-6674.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [30] B. Chang, L. Meng, E. Haber, F. Tung, and D. Begert, "Multi-level residual networks from dynamical systems view," *arXiv preprint arXiv:1710.10348*, 2017.
- [31] R. Luo, F. Tian, T. Qin, E. Chen, and T.-Y. Liu, "Neural architecture optimization," *Advances in neural information processing systems*, vol. 31, 2018.
- [32] M. Tan and Q. V. Le, "EfficientNet: rethinking model scaling for convolutional neural networks," in *36th International Conference on Machine Learning*, 2019, pp. 10691–10700.
- [33] M. Tan *et al.*, "MnasNet: platform-aware neural architecture search for mobile," *arXiv preprint arXiv:1807.11626*, Jul. 2018.
- [34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," *arXiv preprint arXiv:1801.04381*, Jan. 2018.
- [35] W. Hastomo, A. S. B. Karno, N. Kalbuana, A. Meiriki, and Sutarno, "Characteristic parameters of epoch deep learning to predict Covid-19 data in Indonesia," *Journal of Physics: Conference Series*, vol. 1933, no. 1, Jun. 2021, doi: 10.1088/1742-6596/1933/1/012050.




BIOGRAPHIES OF AUTHORS

Adhitio Satyo Bayangkari Karno    obtained a Bachelor's degree (S-1) majoring in Mathematics and Natural Sciences in 1992, and a Master's degree (S-2) from the Faculty of Computer Science, Master of Information Technology in 2010 from the Universitas Indonesia (UI), Indonesia. His research interests include artificial intelligence, deep learning, and machine learning. His occupation until now is as a lecturer at several universities in Indonesia. He can be contacted at email: Adh1t10.2@gmail.com.






Widi Hastomo    received Bachelor of Computer Science and Master in information technology degree from STMIK Jakarta. His research interests include artificial intelligence and deep learning. His work has been documented in more than 25 papers. He can be contacted at email: Widie.has@gmail.com.






Tri Surawan    obtained a Bachelor's degree (S-1) in 1992 and Master's degree (S-2) in 2005 majoring in Mathematics and Natural Sciences, from the Universitas Indonesia (UI), Indonesia. His research interests include materials and artificial intelligence. His occupation until now is as a lecturer at several universities in Indonesia. He can be contacted at email: tri.surawan@gmail.com.






Serlia Rafflesia Lamandasa    Department of Management, Faculty of Economics, University of Sam Ratulangi Manado in 1988 (S1) and Department of Management of Development Resources, University of Sam Ratulangi Manado in 2002 (S2). Permanent Lecturer at the Faculty of Economics, UNSIMAR since January 1, 1989-now. Lecturer in human resources management, production operational management, operational research, HR planning and control, performance assessment. She can be contacted at email: serlia@unsimar.ac.id.






Sudarto Usuli    Bachelor's degree (S-1) in University of Sintuwu Maroso, and Master's degree (S-2) in University of Muhammadiyah Makassar. Currently, the focus is on research in the field of operational management and public financial management. He can be contacted at email: sudarto@unsimar.ac.id.



Holmes Rolandy Kapuy    holds a Doctor in Economy majoring Management from Faculty of Economics and Business Airlangga University, and a Magister of Management from Tadulako University. Currently the focus of research is on marketing strategy and management information systems. He can be contacted at email: rolandykapuy@gmail.com.



Aji Digdoyo    obtained of a Bachelor's Mechanical Engineer Degree (S-1) in 1988 and Master's degree (S-2) in 1998 majoring in Environmental sciences, from the "University of Indonesia (UI)," Indonesia. His research interests include renewable energy and artificial intelligence. His Occupation until now is a lecturer Faculty of Technology Industry, University Jayabaya in Department of Mechanical Engineer, Indonesia. He can contact at email: digdoyoaji@gmail.com.