Al-Rafidain Journal of Engineering Sciences

Journal homepage https://rjes.iq/index.php/rjes

ISSN 3005-3153 (Online)

Using an Improved Visual Geometry Group Neural Network for Skin Cancer Classification

NOOR HAMID HAMEED[1] and WALEED ABDULLAH ARAHEEMAH[2]

[1,2]Middle Technical University,
[1]Dac2008@mtu.edu.iq, [2]Dr.waleed@mtu.edu.iq

ARTICLE INFO

Article history:
Received 23 April 2024,
Revised 25 April 2024,
Accepted 03 May 2024,
Available online 03 May 2024

Keywords:
Skin cancer
Visual Geometry Group
accuracy level
convolution neural network
image classification
supervised classification

ABSTRACT

Over time, deep learning has become more accurate and efficient at dealing with more complex risks. However, training deep neural networks remains a daunting problem and challenges remain, possibly due to several issues, including relevance, generalizability, and training time. Producing specific processors to manipulate artificial networks and reduce task-specific network learning time is natural, but the challenge of fit and generalizability remains undisputed. These networks have been used to detect and diagnose skin cancer, one of the most dangerous types of cancer caused by DNA damage. Automatic recognition of skin cancer is important to help doctors detect the disease in its early stages. In this research, the proposed vgg16 method was used to classify benign and malignant medical images. Image processing was used, where Gaussian noise and salt and pepper noise were added and filters for the arithmetic mean and median were added for each noise. The network was implemented using the Keras interface. The application of the suggested model is shown that the height accuracy 0.86 for train and accuracy 0.92 for test

1. 1. Introduction

An image is a projection of a 3D scene into a 2D plane of view where it represents a matrix of a 2D image using a limited number of cell elements, commonly referred to as pixels (image elements)[1].

The convolutional neural network medical image classifier used can suggest a dermatologist-level classifier in our phone or computer by taking a photo of the user's skin lesion and instantly getting the diagnosis. Noise and filters are added to enhance generalization and improve model performance by preventing excessive loss, improving image quality, and reducing computational costs.

Many researches have been conducted in this field, the most important of which are:

In 2021, Zaid G. Hadi and others used two image recognition methods. The first method was implemented using convolutional neural networks (CNN) using the residual network (ResNet-50) as a deep learning convolutional neural network (DLCNN) that maps images to class labels. It presents a classifier with a single CNN to automatically recognize benign and malignant skin images. As for the second method, we used a support vector machine. The

* Corresponding author E-mail address: Dac2008@mtu.edu.iq
https://doi.org/10.61268/xankz410

performance of the model was tested using non -coated images of the poster. This model determines the most common skin cancers and can be updated with an unlimited number of new images. DLCNN was trained by the Resnet-50 model and showed a good classification for the good and malicious leather categories. RESNET-50 achieved a significant definition rate of more than 97% in the test photos, and the SVM vector compiler accuracy (SVM) for the classification of skin cancer was accurately, for the step of extracting features, 86.9%. These results were obtained using Matlab[2].

In 2021, Md Shahin Ali and others a deep convolutional neural network (DCNN) model based on deep learning was proposed. The accurate classification method between benign and malignant skin lesions was compared with some transfer learning models such as AlexNet, ResNet, VGG-16, DenseNet, MobileNet, etc. The model was evaluated on the HAM10000 dataset and we finally obtained the highest training accuracy of 93.16% and testing accuracy of 91.93%, respectively.[3].

In 2022, Sara Medhat et al proposed three different methods for binary classification (melanoma and nevus) using smartphone images. Three CNN classifiers were compared: Alex-net, Mobilenet-V2, and Resnet-50. MobileNet_v2 achieved 99.36% accuracy, 96.364% sensitivity, 99.792% specificity, 99.167% accuracy, and 97.661% F1 score[4].

In 2023, Ashutosh Lembhe and others proposed technologies such as image processing and machine learning for an artificial skin cancer screening process. ISR techniques recreate a high-resolution image or series of visible LR images. A method that uses deep learning has been used on the superior accuracy of the image to enhance the accuracy of the convolution neural network model. This model was developed using the KERAS rear interface and the model was tested by adjusting the nerve network layers used for training. The model was built on a set of data with public sources from ISIC[5].

In 2024, Azadeh Noori Hoshyar and others performance comparison of five filters (average filter, adaptive median filter, median filter, Gaussian filter, adaptive Wiener filter) was performed to remove noise from Gaussian noise, salt and pepper noise, Poisson noise, and speckle noise. The best performance was achieved by the adaptive filters at different Gaussian and Poisson densities and low Speckle densities, the adaptive median filter in salt and pepper noise and the median filter at high Speckle densities.[6]

In this research, we used a developed neural network model based on classification of skin cancer medical images and compared its performance in three cases: without using noise, using noise, and using noise with filters.

2. Image classification

Image classification is a visual processing technique that distinguishes between several classes of subjects based on image attributes. In pattern recognition and computer vision[7]

An image is a projection of a 3D scene into a 2D plane of view where it represents a matrix of a 2D image using a limited number of cell elements, commonly referred to as pixels (image elements). Each pixel is represented by numerical values: for grayscale images, a single value representing the pixel intensity (usually in the range [0, 255]). For color images, three values are stored (representing the amount of red (R), green (G), and blue (B)). If an image contains only two intensities, the image is known as a binary image.[1]

2.1. Supervised image classification

Supervised machine learning assumes that the model has been trained on a dataset similar to the problem at hand, which consists of input data and corresponding output data. Once the model learns the relationship between inputs and outputs, it can classify new, unknown data sets and make predictions or decisions based on them. This type of learning is divided into two methods: classification and regression.[8]

2.2. Unsupervised image classification

Unsupervised machine learning differs from supervised machine learning in its use of unannotated data, which has not been previously labeled by humans or algorithms. The model learns from input data without expected values, and the available data set does

not provide answers to the given task. Instead of classifying or predicting outputs, this algorithm focuses on grouping data based on its characteristics. The goal is to teach the machine to detect patterns and aggregate data without a single correct answer. There are two types of unsupervised learning: clustering and association[8]

3. Artificial Neural Network

It is a computational model that has the ability to process information to handle tasks such as classification and regression. It is one of the components of artificial intelligence inspired by the human brain and nervous system. The neural network consists of neurons, which mainly consist of inputs, which are multiplied by weights. The weights assigned to each arrow represent the information flow. These weights are then calculated by a mathematical function that determines the activation of the neuron. Another function calculates the output of an artificial neuron (sometimes based on a certain threshold). The neurons in this network are just collecting their inputs. Since an input neuron has only one input, its output will be the input it received multiplied by the weight.[9][10]

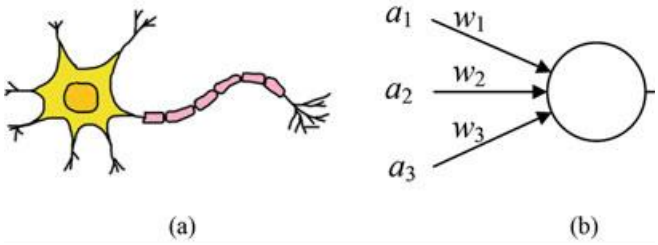


Figure (1): the ANN[9]

3.1. Convolution Neural Network

CNNs are a powerful type of neural network that is widely used in image recognition tasks. It consists of a series of convolutional and pooling layers that extract relevant features from the input image, followed by one or more fully connected layers that use these features to perform prediction. Includes CNN infrastructure

1. Convolutional Layer: The convolutional layer is the basic structure of a CNN. The convolutional operation is performed on the input image using a set of learnable filters. Filters are small arrays that pass over the input image.
2. Pooling layer: The pooling layer is used to reduce the spatial dimensions of the feature maps produced by the convolutional layer. It operates on each feature map independently and reduces its size it by taking the maximum or average value. Clustering helps reduce the computational complexity of the network.
3. Activation layer: The activation layer applies a non-linear activation function to the outputs of the previous layer. This introduces non-linearity into the network, allowing it to learn more complex features.
4. Batch Normalization Layer: The Batch Normalization layer normalizes the output of the previous layer by subtracting the mean and dividing by the standard deviation of the batch. This helps reduce internal variable shift and improves network convergence.
5. Dropout layer: The dropout layer randomly drops a proportion of the neurons in it.
6. Fully Connected Layer: The layer that connects each neuron in the previous layer to every other neuron.[11]

3.1.1. Visual Geometry Group (VGG16)

VGG is a convolutional architecture proposed by Karen Simonyan and Andrew Zisserman from the Visual Engineering Group at the University of Oxford. There are two types of VGG models based on the number of layers: VGG-16 and VGG-19. VGG-16 consists of thirteen convolutional layers, a max pooling layer, three fully connected layers, and an output layer.

The VGG16 architecture consists of:

1. Convolutional layer with 64 filters of size 3 3, followed by a maximum 2 2 pooling layer.
2. Convolutional layer with 128 filters of size 3 3, followed by a 2 max pooling layer.
3. Convolutional layer with 256 filters of size 3 3, followed by a 2 max pooling layer.
4. Convolutional layer with 512 filters of size 3 3, followed by 2 max pooling layers.
5. Convolutional layer with 512 filters of size 3 3, followed by a 2 max pooling layer.
6. Fully connected layer containing 4096 units and ReLU activation function, followed by dropout Organization.

7. Fully connected layer with 4096 units and ReLU activation function, followed by dropout Organization.[12,13]

Figure (2): the VGG16 Model[12]

3.2. Improved Visual Geometry Group (VGG16)

The last layer was deleted and the following layers were added

1. Flatten layer
2. Fully connected layer (dense) containing 4096 units and ReLU activation function
3. Dropout layer: The dropout layer randomly drops a proportion of the neurons in it by (0.5).
4. Fully connected layer (dense) containing 4096 units and ReLU activation function
5. Dropout layer: The dropout layer randomly drops a proportion of the neurons in it by (0.3).
6. Fully connected layer (dense) containing 4096 units and ReLU activation function.
7. Dropout layer: The dropout layer randomly drops a proportion of the neurons in it by (0.4).
8. Fully connected layer (dense) containing 2 units and Softmax activation function.

Model: "sequential"

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 7, 7, 512)	14714688
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 4096)	102764544
dropout (Dropout)	(None, 4096)	0
dense_1 (Dense)	(None, 4096)	16781312
dropout_1 (Dropout)	(None, 4096)	0
dense_2 (Dense)	(None, 4096)	16781312
dropout_2 (Dropout)	(None, 4096)	0
dense_3 (Dense)	(None, 2)	8194

Total params: 151050050 (576.21 MB)
Trainable params: 136335362 (520.08 MB)
Non-trainable params: 14714688 (56.13 MB)

4. Image Noise Model

Image processing is the process of converting an image into a digital form and performing certain operations to obtain some useful information from it. The image processing system treats images as two-dimensional

signals when processing methods are applied[14]

Noise It is a random difference in brightness or color information caused by unsuitable conditions such as low light and temperature or due to dust particles.

4.1. Gaussian Noise

It is a statistical noise with a probability density function equal to the natural distribution. The Gaussian noise is added to the pixels of the image that contains noise.

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Where x represent the grey level, μ the mean and σ its standard deviation

4.2. Salt and pepper Noise

They are black and white pixels that appear at random intervals, and occur as a result of data transfer in pepper noise a and b. The two values are different and salt and each of them has a probability of less than 0.1 on average. The damaged pixels are assigned alternately to the upper and lower value, which gives the image the appearance of salt and pepper. [6, 13]

5. Image Filtering

Image filtering is a technique used in image processing to perform various operations on an image to improve its quality, enhance its features, or extract useful information. Filtering involves the application of a filter, which can be a mathematical operation or convolution, to an image. Different types of filters serve different purposes, such as smoothing, sharpening, edge detection, noise reduction.

Here is some image filtering

5.1. Mean Filter

It is a type of linear filter used in computer vision that replaces the value of each pixel in an image with the average value of neighboring pixels, it is used to reduce noise in the image also used for image smoothing and blurring

The (3X3) filter will be

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

5.2. Median Filter

It is a type of nonlinear filter used in computer vision that replaces the value of each pixel in an image with the average value of neighboring pixels. It is a simple and effective technique to reduce noise in images and can also be used to smooth the image and preserve its edges.[15]

6. Skin Cancer

Melanoma is a type of skin cancer that develops when melanocytes begin to grow out of control due to DNA damage in the outer layer of skin. Although skin cancer is the least common type of cancer, it is the most dangerous. In addition, there is a greater chance of spreading to other parts of the body unless it is diagnosed and treated early. A benign skin tumor is called a mole (nevus) that develops from melanocytes. Almost all moles (birthmarks) are harmless, but some types can increase your risk of melanoma. Malignant tumors develop when skin cells multiply excessively due to these mutations. Exposure to the sun's ultraviolet rays and UVB rays from tanning beds is major risk factors for malignant tumors and skin cancer.[16]

7. Datasets

A database of skin cancer images available on the Kaggle platform is used, which contains 1,800 healthy images and 1,500 affected images size 224x224.

https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign/data

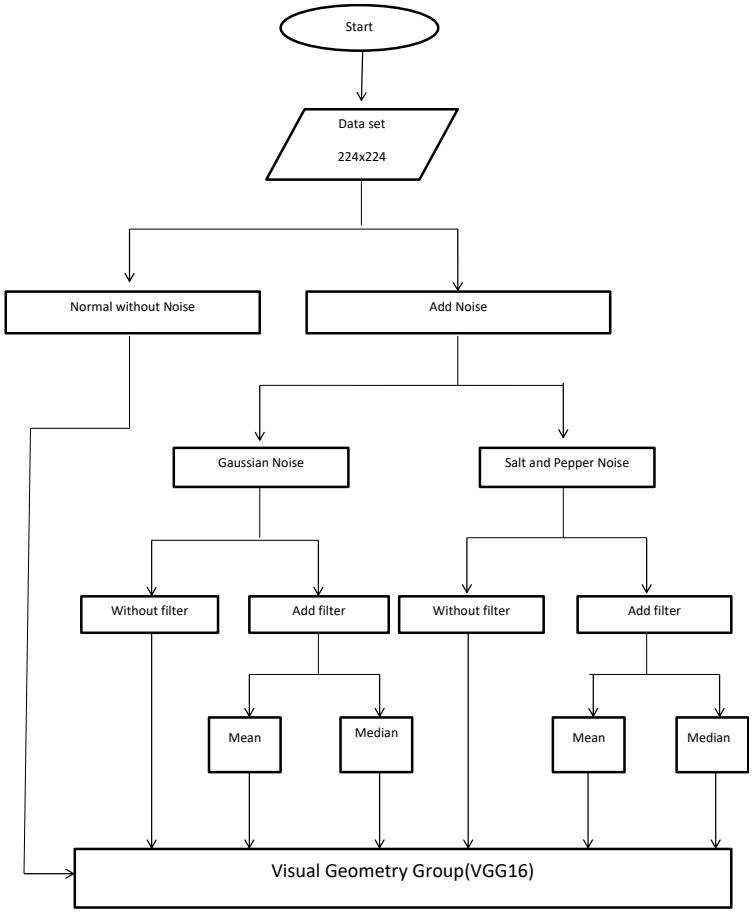
A. Benign B. Malignant

Figure (3) a sample of data

8. Result

Each image with color (R,G,B) bands and size (224x224) pixels,

The suggest model can be shown by the following Block Diagram



Data were entered into the vgg16 network with a batch size of 128 for 10 epochs (round-trip network execution for the training and testing phase) and an early stop command was given (technique to stop The training process when the results do not improve after a specified period of time in order to avoid over-training the network and overfitting). After Appling of suggested model on the image data for the following results was shown. It turns out that diagnosing skin cancer through processing its images according to the modified VGG16 method is affected by the state of the data, whether it is normal, noisy, or filtered with noise. It is also affected by the size of the batch, and that the best classification obtained in the normal case in the training stage is 0.86 and better. Classification in the testing phase was 0.91 at a batch size of 1280

Table (1): Normal stage

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	1.3969	0.4874	0.7864	0.5391
256	0.6993	0.5638	0.6595	0.5547
384	0.5951	0.6812	0.5870	0.6250
512	0.5125	0.7466	0.5529	0.6914
640	0.4861	0.7534	0.3423	0.8320
768	0.4164	0.8171	0.3069	0.9023
896	0.3662	0.8448	0.2738	0.9023
1024	0.3396	0.8565	0.2570	0.8984
1152	0.3299	0.8641	0.2439	0.9062
1280	0.3261	0.8582	0.2181	0.9180

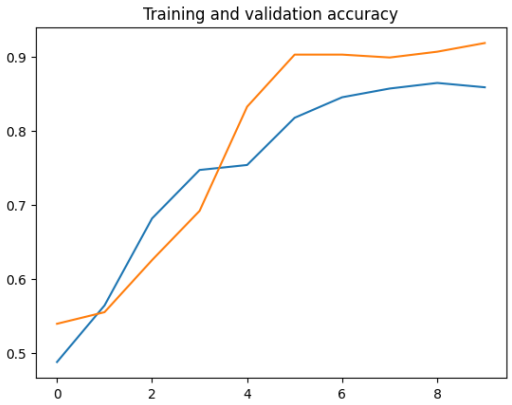


Figure (4): accuracy Normal stage

We also note that the classification accuracy was not greatly affected when adding noise of the Gaussian type, as an accuracy of 0.85 was obtained in the training phase and 0.91 in the testing phase at a batch size of 1024, where the batch size was less than the batch size in The natural state in this type of noise.

Table (2): Gaussian **Noise stage**

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	1.1090	0.5310	0.5598	0.8008
256	0.6565	0.6535	0.5070	0.7227
384	0.5049	0.7424	0.4043	0.7852
512	0.4442	0.7987	0.3699	0.8750
640	0.4106	0.8431	0.3127	0.8594
768	0.3677	0.8289	0.2490	0.8984
896	0.3526	0.8523	0.2598	0.8984
1024	0.3351	0.8507	0.2593	0.9102

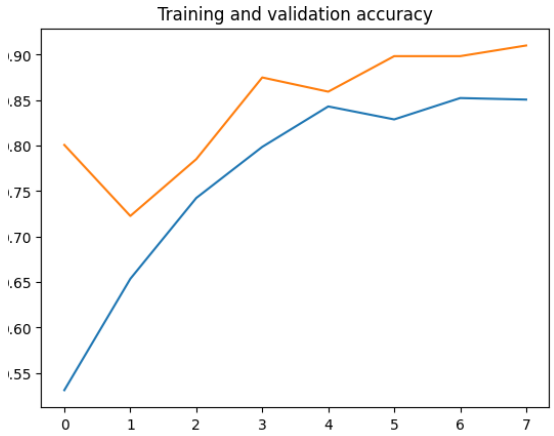


Figure (5): Gaussian Noise stage

But in the case of salt and pepper noise, the results appeared almost similar to the effect of Gaussian noise in the training and testing stages, and in the batch size.

Table (3): Salt and pepper Noise stage

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	1.4727	0.4832	0.9479	0.5312
256	0.7275	0.5671	0.6623	0.5312
384	0.6363	0.6435	0.5191	0.7227
512	0.5240	0.7357	0.3968	0.8516
640	0.4291	0.8138	0.3227	0.8828
768	0.3892	0.8347	0.2625	0.8867
896	0.3510	0.8508	0.2839	0.9062
1024	0.3444	0.8456	0.2745	0.9062

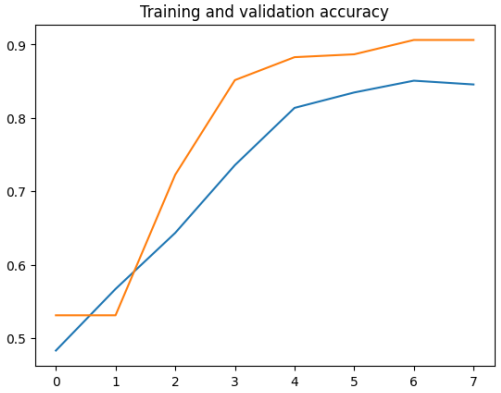


Figure (6): accuracy Salt and pepper Noise stage

As for processing with filters, we note that the batch size was 1152 when adding Gaussian type noise and an arithmetic mean-type filter, and it was accurate. The training accuracy was 0.87 and the test accuracy was 0.89. We note that the training accuracy was higher than the training accuracy and lower than the test accuracy when adding only two caustic noises, which is an unreasonable case, as the arithmetic mean filter, although it led to an increase in diagnosis in the training stage, but it reduced the diagnosis rates in the testing stage. This is because the correction carried out in a specific place in the image may lead to blurring in another, less important place, and this leads to the modified VGG16 method not having the ability to fully diagnose the injury, as the arithmetic mean correctly diagnoses the injury or not.

Table (4): Gaussian Noise and Mean Filter

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	1.2903	0.5268	0.7535	0.4570
256	0.7301	0.5805	0.5438	0.8008
384	0.5316	0.7349	0.4284	0.8320
512	0.4510	0.8029	0.3342	0.8828
640	0.3995	0.8305	0.3313	0.8789
768	0.3793	0.8289	0.2732	0.8906
896	0.3421	0.8599	0.2497	0.8906
1024	0.3456	0.8473	0.2636	0.8945
1152	0.3085	0.8674	0.2899	0.8867

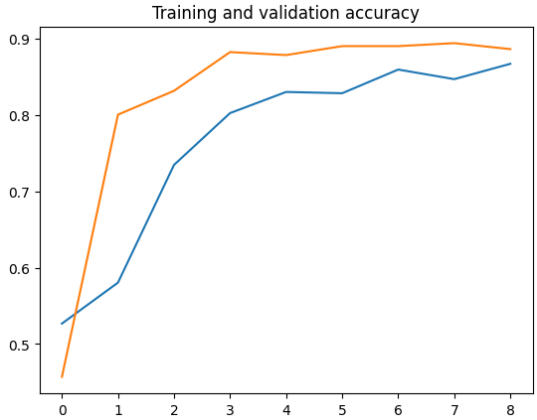


Figure (7): accuracy Gaussian Noise and Mean Filter

But in the case of adding Gaussian noise with a median filter. Accuracy was obtained at a smaller batch size, where the training accuracy was 0.83, the test accuracy was 0.86, and the batch size was 768.

Table (5): Gaussian Noise and Median Filter

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	0.8521	0.5545	0.5310	0.6914
256	0.5782	0.7122	0.3570	0.8750
384	0.4701	0.7685	0.2968	0.9023
512	0.4141	0.8255	0.2782	0.9219
640	0.3841	0.8305	0.3558	0.8789
768	0.3969	0.8328	0.3051	0.8555

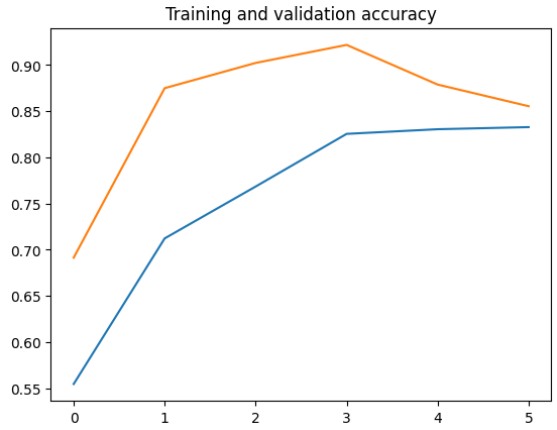


Figure (8): accuracy Gaussian Noise and Median Filter

If salt and pepper noise was added with an arithmetic mean filter, the training accuracy was 0.84 and the test accuracy was 0.88 at a batch size of 768.

Table (6): Salt and Pepper Noise and Mean Filter

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	1.4429	0.5243	0.6177	0.7383
256	0.6957	0.5696	0.5469	0.7500
384	0.5484	0.7273	0.4244	0.8047
512	0.4416	0.7919	0.2790	0.9102
640	0.4164	0.8188	0.2933	0.8945
768	0.3567	0.8398	0.2913	0.8789

Figure (9): accuracy Salt and Pepper Noise and Mean Filter

If salt and pepper type noise was added with a mediator type filter, the training accuracy was 0.85 and the test accuracy was 0.87 at a batch size of 1280.

Table (7): Salt and Pepper Noise and Median Filter

Batch Size	Train		Test	
	Loss	Accuracy	Loss	Accuracy
128	1.2909	0.5109	0.6193	0.6016
256	0.6417	0.6342	0.4999	0.7383
384	0.5179	0.7500	0.4104	0.8633
512	0.4329	0.8037	0.3163	0.8867
640	0.3972	0.8305	0.2830	0.8828
768	0.4112	0.8227	0.3863	0.8672
896	0.3714	0.8398	0.2664	0.8906
1024	0.3213	0.8708	0.2775	0.9023
1152	0.3393	0.8515	0.2404	0.9062
1280	0.3366	0.8490	0.3548	0.8672

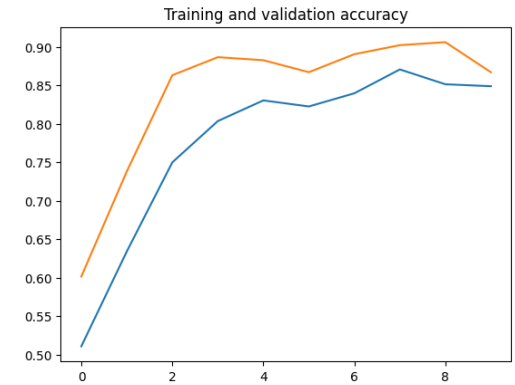


Figure (10): accuracy Salt and Pepper Noise and Median Filter

9. Conclusions and Suggestions

The results showed the method's ability to classify skin cancer incidence under noise conditions with different models according to the applied filters. The proposed method can be applied to classifying brain cancer images as well as lung cancer classification

References

[1] V. Tyagi, "Understanding Digital Image Processing," no. September, 2018, doi: 10.1201/9781315123905.

[2] Z. G. Hadi, A. R. Ajel, and A. Q. Al-Dujaili, "Comparison Between Convolutional Neural Network CNN and SVM in Skin Cancer Images Recognition," *J. Tech.*, vol. 3, no. 4, pp. 15–22, 2021, doi: 10.51173/jt.v3i4.390.

[3] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Mach. Learn. with Appl.*, vol. 5, no. February, p. 100036, 2021, doi: 10.1016/j.mlwa.2021.100036.

[4] S. Medhat, H. Abdel-Galil, A. E. Aboutabl, and H. Saleh, "Skin cancer diagnosis using convolutional neural networks for smartphone images: A comparative study," *J. Radiat. Res. Appl. Sci.*, vol. 15, no. 1, pp. 262–267, 2022, doi: 10.1016/j.jrras.2022.03.008.

[5] A. Lembhe, P. Motarwar, R. Patil, and S. Elias, "Enhancement in Skin Cancer Detection using Image Super Resolution and Convolutional Neural Network," *Procedia Comput. Sci.*, vol. 218, pp. 164–173, 2022, doi: 10.1016/j.procs.2022.12.412.

[6] A. Noori, A. Al-jumaily, and A. Noori, "Comparing the Performance of Various Filters on Skin Cancer Images," *Procedia - Procedia Comput. Sci.*, vol. 42, no. 02, pp. 32–37, 2014, doi: 10.1016/j.procs.2014.11.030.

[7] D. P. Mishra, S. Mishra, S. Jena, and S. R. Salkuti, "Image classification using machine learning," vol. 31, no. 3, pp. 1551–1558, 2023, doi: 10.11591/ijeecs.v31.i3.pp1551-1558.

[8] S. Kocot *et al.*, "What Is Machine Learning , Artificial Neural Networks and Deep Learning ?— Examples of Practical Applications in Medicine," no. Ml, 2023.

[9] Y. Tian, "Artificial Neural Network," no. January, 2022, doi: 10.1007/978-3-030-26050-7.

[10] G. Asadollahfardi, "Artificial Neural Network," vol. 3, no. 1, pp. 77–91, 2015, doi: 10.1007/978-3-662-44725-3_5.

[11] M. Krichen, "Convolutional Neural Networks : A Survey," pp. 1–41, 2023.

[12] S. Rajarajeswari, J. Prassanna, M. Abdul Quadir, J. Christy Jackson, S. Sharma, and B. Rajesh, "Skin Cancer Detection using Deep Learning," *Res. J. Pharm. Technol.*, vol. 15, no. 10, pp. 4519–4525, 2022, doi: 10.52711/0974-360X.2022.00758.

[13] "0b89da8e6a68bd0216f7025c8f7bdbf0310983a7 @ datagen.tech." [Online]. Available: https://datagen.tech/guides/computer-vision/vgg16/

[14] "image-processing-article @ www.simplilearn.com." [Online]. Available: https://www.simplilearn.com/image-processing-article

[15] "noise-filtering-mean-median-mid-point-filter-72ab3be76da2 @ medium.com." [Online]. Available: https://medium.com/@sarves021999/noise-filtering-mean-median-mid-point-filter-72ab3be76da2

[16] S. Nazari, "Automatic Skin Cancer Detection Using Clinical Images : A Comprehensive Review," 2023.