Proceedings of the 2023 International Conference on Mechatronics and Smart Systems
DOI: 10.54254/2755-2721/10/20230193

Image processing based on neural networks

Boming Yang

School of Beihang University, Beijing 100191, China

20373104@buaa.edu.cn

Abstract. This article integrates convolutional neural network (CNN) and graph convolutional network (GCN) techniques. Performance is improved by the suggested architecture's use of crucial methods such as dropout, batch normalization, and rank-based random pooling. The network was trained and tested on a sizable amount of breast lesion imaging data, and its precision was evaluated in comparison to other methods. The findings showed a significant increase in accuracy, resulting in high rates of malignant lesion diagnosis with few false positives. The effective fusion of CNN and GCN methods highlights the potential for improving the detection of malignant breast lesions and provides a viable path for further study in this area.

Keywords: convolutional neural network, breast lesions.

1. Introduction

CNNs are frequently used to handle data in a variety of disciplines, including audio, language, and images [1][2]. The advantage of CNNs is that because the networks in the convolution share parameters, there are fewer parameters needed compared with typical neural networks, and the model complexity is lower [2]. The convolution operation displays position invariance because it concentrates on the local information of the input data and extracts local features. As a result, although classic neural networks may have trouble with position changes in other data, CNNs can reliably classify features even if their position in the image changes.

The most common disease in women, breast cancer, can have serious effects if it is not detected and treated right away [3]. Rapid detection and diagnosis of medical pictures are made possible by machine learning technology, increasing the early detection rate and diagnostic precision for breast cancer. Moreover, machine learning helps doctors grasp and analyze enormous amounts of medical picture data, producing more accurate diagnostic results [4]. Throughout the last few decades, computer technology has been widely employed in image recognition and has come into touch with the medical business. Convolutional neural networks (CNNs) have made substantial advances in deep learning in recent years. Instead of merely performing tasks created by humans, modern deep learning models may interpret task content on their own and provide plausible replies based on training data [5][6]. In the past, to gather patient information and assess the severity of the condition, medical professionals would examine photographs of the malignant location. Deep learning models have made it possible for large data analytic approaches to assess and diagnose patients based on their symptoms, with some diseases already showing diagnostic performance that is on par with or better than that of humans. Although there are cancer treatments, the likelihood that a patient will survive the disease is considerably increased by early

© 2023 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/10/20230193

detection and rapid treatment [7]. Thus, it is still important to enhance the technology of picture recognition and machine learning.

This paper focuses on how to achieve better image classification results. Early detection is encouraged by this method's assistance in differentiating between healthy tissue and tissue vulnerable to malignant transformation. Many techniques have been created as image processing technology develops. By removing noise and increasing background, pre-processing methods such as thresholding, Z-score normalization, and median filtering can significantly increase the efficiency and classification accuracy of neural networks when processing picture data. Two networks that are often used for image processing are convolutional neural networks and graph convolutional networks. This study makes use of deep learning technology and experiments with different methods to decide which classification approach is most suitable for the selected dataset.

2. Problem formulation

Region-growing is a technique for identifying regions of interest (ROI) and eliminating the backdrop, illustrated by Algorithm 1. Here, R is the set of points in the area, and Q is the set of points to be processed (also known as the queue), are the two sets that the algorithm initializes. The seed point (x_0, y_0), which serves as the beginning point for the area expansion, is used to initialize both sets. When Q (the queue) is not empty, the algorithm enters a while loop that runs indefinitely. A point (x, y) is dequeued from Q (i.e., taken out of the queue and put back in) inside the loop. The procedure examines each nearby point (x', y') of the dequeued point (x, y) to see if the prerequisites for area expansion are satisfied. The nearby point (x', y') is added to the region R and the queue Q if the requirements are satisfied. Repeating the procedure until there are no more points in Q, enlarges the region R by analyzing the points' neighbors.

Algorithm 1: Region of Interest (ROI) Algorithm

```
Data: R ={(x_0, y_0)}
Data: Q = {(x_0, y_0)}
while Q is not empty do
    (x, y)+- dequeue(Q)
    for k ∈ {−1, 0, 1} do
        for l ∈ {−1,0,1} do
            if x + k ≥ 0 and x + k < M and y + l ≥ 0 and y + l < N then
                if (x + k, y + I) ∉ R and d(z + k, y + l) < T then
                    R ← R ∪ {(x + k, y + l)}
                    Q ← Q ∪ {(x + k, y + l)}
                end if
            end if
        end for
    end for
end while
```

A simple and popular strategy for reducing background (BG) or image noise during CNN data pre-processing is the thresholding technique [8]. Setting a pixel intensity threshold, designating pixels below the threshold as BG, and allocating pixels above the threshold as foreground are the essential principles of thresholding technology. These steps are demonstrated in Algorithm 2. The grayscale picture $I(i, j)$, where i and j are the row and column indices of the pixels, and the threshold value T are the two inputs that the algorithm requires. The process generates a binary picture called $I'(i, j)$ in which the row and column indices of the pixels are i and j, respectively. To determine if the input pixel value $I(i, j)$ is above or below the threshold T, an if-else structure is utilized. The output pixel value $I'(i, j)$ is set to zero if the input pixel value $I(i, j)$ is less than the threshold T. The output pixel value $I'(i, j)$ is set to 1 if the input pixel value $I(i, j)$ is equal to or greater than the threshold T.

DOI: 10.54254/2755-2721/10/20230193

Algorithm 2: Thresholding Algorithm

Data: $\boldsymbol{I(i,j)}$, T
Result: $I'(i,j)$
if $I'(i,j) < T$ **then**
    $I'(i,j) \leftarrow 0$
else
    $I'(i,j) \leftarrow 1$
end if

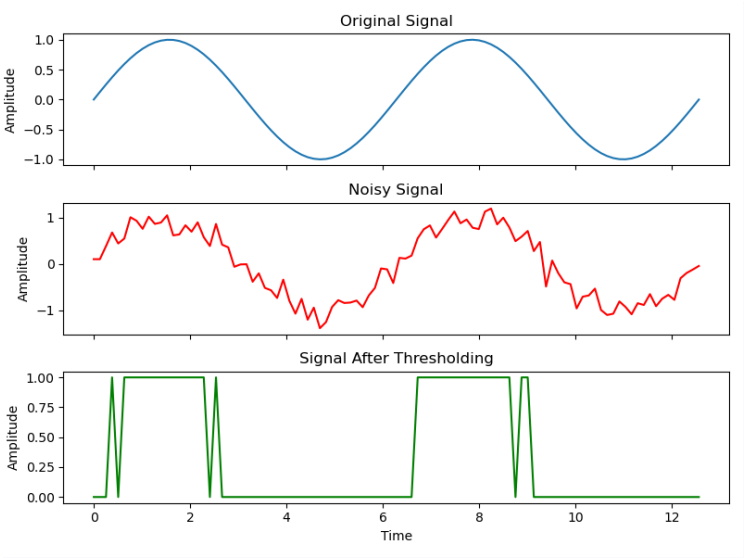

Figure1. Effects of threshold technology.

A common data pre-processing method used to transform data into a conventional normal distribution is called z-score normalization. The goal of Z-score normalization is to rescale the data so that the mean is set to 0 and the standard deviation is equal to 1. After subtracting the mean from each data point, this is done by dividing the result by the data's standard deviation. The formula for normalizing Z-scores is as follows

$$I' = (I - mean(I))/stddev(I) \tag{1}$$

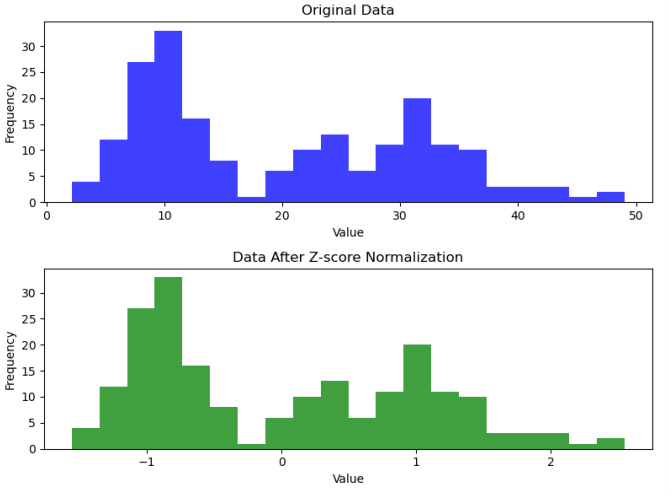

Figure2. Feature distribution of data before and after applying Z-score normalization.

DOI: 10.54254/2755-2721/10/20230193

Convolutional Neural Networks (CNNs) typically use the center crop pre-processing technique to resize images and center the object of interest. With this method, the longer side of the image is cropped to make a square, and the cropped image is then centered. According to the following principle, the top-left and bottom-right corners of the center crop region can be computed for an input picture I with dimensions and an anticipated output size of, where:

Calculate the rows' (height) and columns' (width) starting and ending indices:

$$x_{start} = (M - M)/2, y_{start} = (N - N)/2 \tag{2}$$

$$x_{end} = x_{start} + m, y_{end} = y_{start} + n \tag{3}$$

By extracting the center crop region $I' = I[x_{start}: x_{end}, y_{start}: y_{end}]$from the input image I, the center crop operation extracts a region of size m x n from the original image I, producing a new image I' with the desired size.

Median filtering, a typical method of image pre-processing, replaces each pixel's value with the median value of its neighbors. By using median filtering, a filter or kernel K of size$(2k + 1) \times (2k + 1)$ cantered on a pixel $I(i, j)$ is slid over each pixel in an $M \times N$image matrix I, and the median value of the pixel values within the kernel region is calculated. In this instance, this median value will take the place of the existing pixel $I(i, j)$ value. The filtered picture might be shown as

$$F(i, j) = median\{I(_{\text{i-p}},\ j - p)|p \in \{-k, -(k - 1), \ldots, k - 1, k\}, q \in \{-k, -(k - 1), \ldots, k - 1, k\}\}. \tag{4}$$

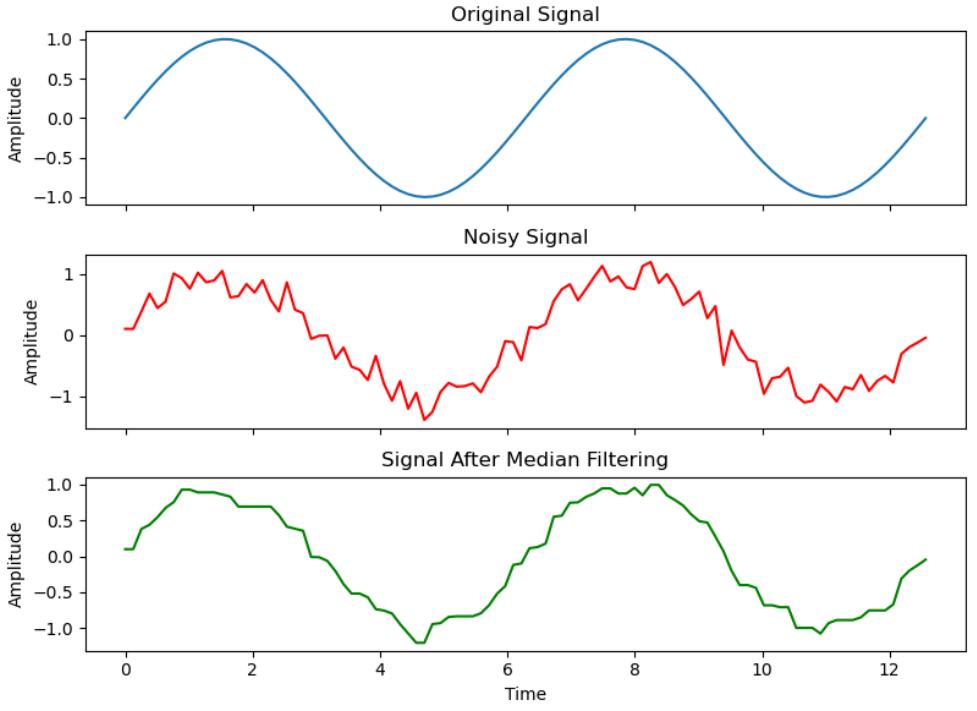

Figure3. The noise removal effect of median filtering.

For learning data that is graphically organized, a deep learning model called the Graph Convolutional Network (GCN) was developed. GCNs can more accurately reflect higher-order interactions and global context than CNNs, which typically concentrate on local patterns within a fixed receptive field, by taking advantage of the connectedness structure of the graph. GCNs can learn more adaptable and expressive feature representations by altering their receptive fields in accordance with the structure of the graph. Algorithm 3 explains the Graph Convolutional Network (GCN):

DOI: 10.54254/2755-2721/10/20230193

Algorithm 3: Basic Graph Convolution Network (GCN)

Require: Graph $G(V,E)$,Node features X, Number of layers L, Layer weights $W_1, \ldots, W_L$

$A \leftarrow adjacency\_matrix(G)$
$D \leftarrow degree\_matrix(A)$
$A' \leftarrow D^{-1/2} \leftarrow AD^{-1/2}$
function GRAPH_CONVOLUTION(A', X, W)
    return $ReLU(A'XW)$
end function
$H \leftarrow X$
for $l \leftarrow 1, \ldots, L$ **do**
    H$\leftarrow$*graph_ convolution (A', H, W_l)*
end for

Ensure: Node representations H

The method is fed a graph *G(V, E)*, V is the set of nodes and E is the set of edges, *X* is node characteristics, L is the number of layers and W_1,..., W_L are layer weights. The graph G's adjacency matrix A and degree matrix D are both constructed. The normalized adjacency matrix A' is obtained using the formula $A' = D^{-1/2}AD^{-1/2}$. The normalized adjacency matrix A', the node properties *X*, and the layer weights *W* are the three inputs for the function "graph_convolution" that is supplied. Rectified Linear Unit (ReLU) activation is used to compute and return the graph convolution. The node's *H* characteristics are started with *X*. A for loop iterates across the layers from one to *L*, one at a time. With *A'*, H, and W_l as inputs, the "graph_convolution" function performs the graph convolution at each layer, saving the outcome in *H*. The node representations the algorithm returned are represented by *H*.

3. Method

The mini-MIAS database of mammograms 2018 dataset is used in this study's analysis of breast cancer imaging data. It comes from the first MIAS dataset, which was released in 2018. The dataset is frequently used for the development and assessment of computer-assisted diagnostic (CAD) systems for cancer. It contains images of various resolutions and sizes from various sources. Basic factual information is labelled on these images, such as the presence of masses, calcifications, or healthy tissue. The photos from this dataset, which have a high level of trustworthiness and a variety of purposes, are also used in this study as training images. The dataset includes 322 photos (113 abnormal images and 209 normal images). To choose regions of interest (ROI), enhance, and normalize the ROIs, as well as to carry out a better feature extraction and classification technique later, pre-processing is necessary before classification.

3.1. Data pre-processing

Initially, noise is removed and picture features are improved by using a sliding window and a median filtering approach to replace each pixel value with the median of its nearby pixels. To ensure that characteristics have the same range of values and to improve algorithm performance, the images are then subjected to Z-score normalization. Thresholding is used in conjunction with the regional growth approach to choose Regions of Interest (ROI) to discern between the foreground and background of a picture. The target is in the middle of the square image after the longer side of the image is cropped using the center crop technique. Several methods are used to decrease noise, normalize pixel values, segment ROIs, and extract pertinent features during the preparation of the picture collection. Its

DOI: 10.54254/2755-2721/10/20230193

integration facilitates the preparation of picture data for additional analysis, improving the efficiency and precision of machine learning algorithms. The photos are subjected to Z-score normalization in the following step.

3.2. Network establishment:

First, to establish an 8-layer CNN backbone with 6 convolutional layers and 2 fully connected layers, this document will propose some improvements to the backbone. Assuming that the convolution kernel is represented by F_j, where all j in the range$[1, \ldots, J]$, $\otimes$ denotes the convolution operation, and X X represents the input matrix, the following operation is performed in the convolutional layer:

$$C(j) = X \otimes F_j, \forall j \in [1, \ldots, J] \tag{5}$$

Calculating the dot product of the convolution kernel and feature map overlap may alter the size of the output feature map since the convolution technique requires sliding the convolution kernel on the input data features. By adding zero values to the perimeter of the input feature map before convolving it, it is possible to keep the output map's size while successfully preventing information loss.

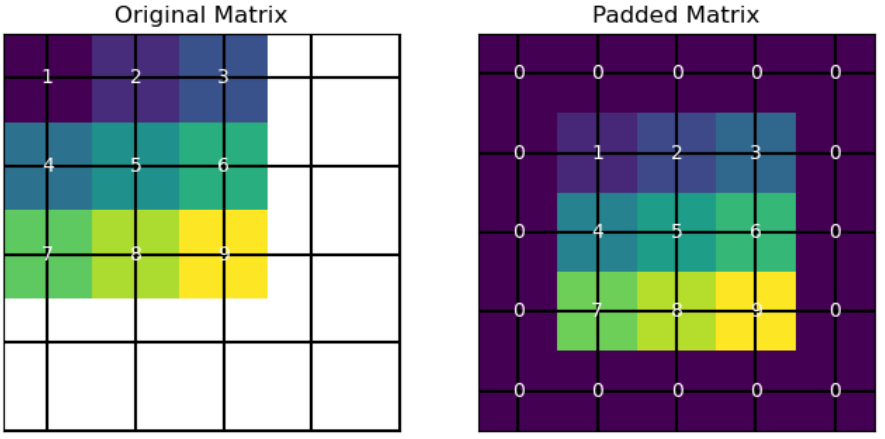

Figure 4. Feature diagram before and after the padding.

An activation function is needed to transform the nonlinear data after the convolution procedure is finished. The Rectified Linear Unit (ReLU) function, which works on the basis that $f(x) = max(0, x)$, outputs the highest value between 0 and x, is an efficient activation function. The ReLU function addresses the problem of vanishing gradients while assuring quick computation speed when compared to other activation functions like the sigmoid and hyperbolic tangent.

During the neural system drive process, batch standardization and abandonment techniques are used to avoid overflow[9][10]. In batch normalization, the mean and variance of each layer's input are first calculated, as follows:

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i, \sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2 \tag{6}$$

where m is the number of samples in a batch and x_i represents the input of the ith sample in the layer. Then, each input sample is normalized by

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \tag{7}$$

where ε is a small constant added to avoid division by zero. The normalized input is then subjected to a linear transformation, where γ and β are learnable parameters that modify the distribution and offset

DOI: 10.54254/2755-2721/10/20230193

of each layer. To increase the network's speed and convergence and to make it more generic, this method is applied to each batch of data.

Dropout Is an effective way to prevent network overfitting, by randomly setting some neuron output to zero during training, reducing neuron dependence. The following equation can be used to model the dropout process: $r_i \sim Bernoulli(p), \hat{h}_i = r_i \cdot h_i$, where r_i is a random variable, $\widehat{h_i}$ is the output of the ith neuron, and p is the pre-defined dropout probability. After applying dropout, the dependencies between neurons are reduced, which enhances the robustness of the model.

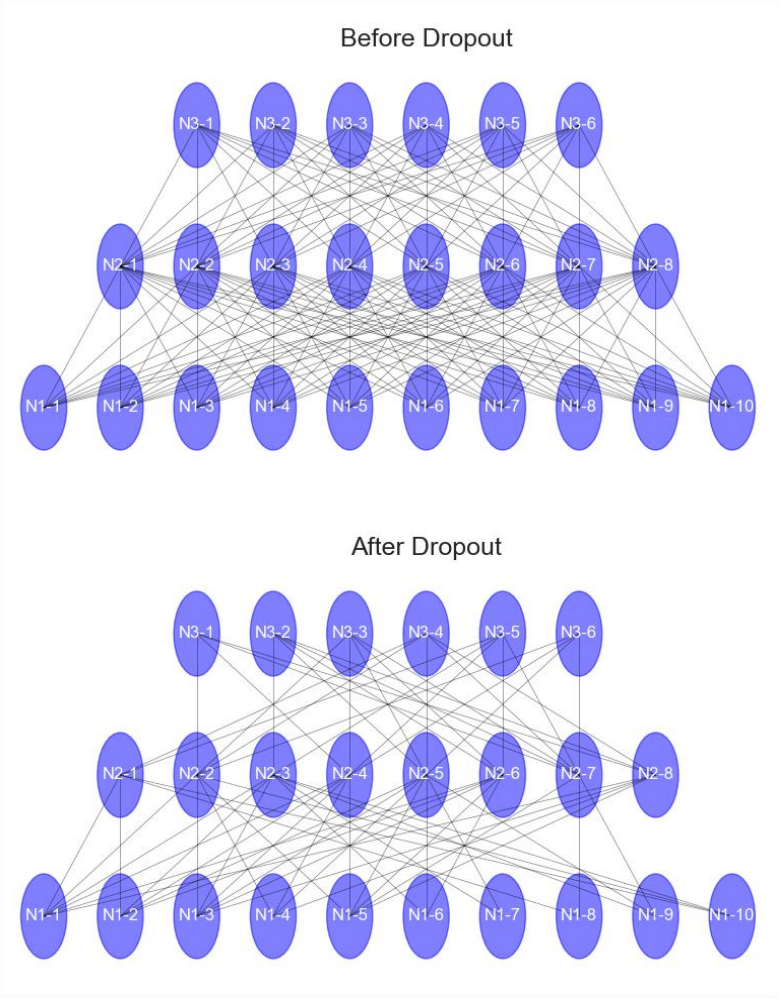

Figure 5. Comparison diagram of the network before and after using dropout (p=0.5).

To add some unpredictability to the pooling process and avoid overfitting, rank-based stochastic pooling is employed as an alternative to max pooling. Rank-based stochastic pooling randomly chooses activation values from a group of nearby neurons based on their rank proportionally to their probability distribution, as opposed to the maximum or average activation value utilized in max or average pooling. The probability distribution is based on the rank of each neuron within the group [11]. Max pooling can be expressed by $Y_{i,j,c} = max_{p=0}^{k_h-1} max_{q=0}^{k_w-1} X_{i\times s_h+p, j\times s_w+q, c}$, where X is an input tensor with dimensions (H, W, C), and the sliding window size is (k_h, k_w) with stride (s_h, s_w). Rank-based stochastic pooling can be expressed using the following equation: where y_i is the pooled output of the ith feature, $x_{i,j}$ is the activation value of the jth feature k is the number of pooled features, and $p_{i,j}$ is a binary random variable indicating whether or not to select the jth feature for pooling.

The graph convolution network (GCN) is given by

DOI: 10.54254/2755-2721/10/20230193

$$Y = Pool(X). \tag{8}$$

The input feature matrix X is subjected to graph-based pooling operations by the pool function to condense its spatial size and extract the most crucial information. GCN effectively makes use of graph convolutional processes and graph-based pooling layers, making it better equipped to handle non-Euclidean input and scale to massive graph datasets.

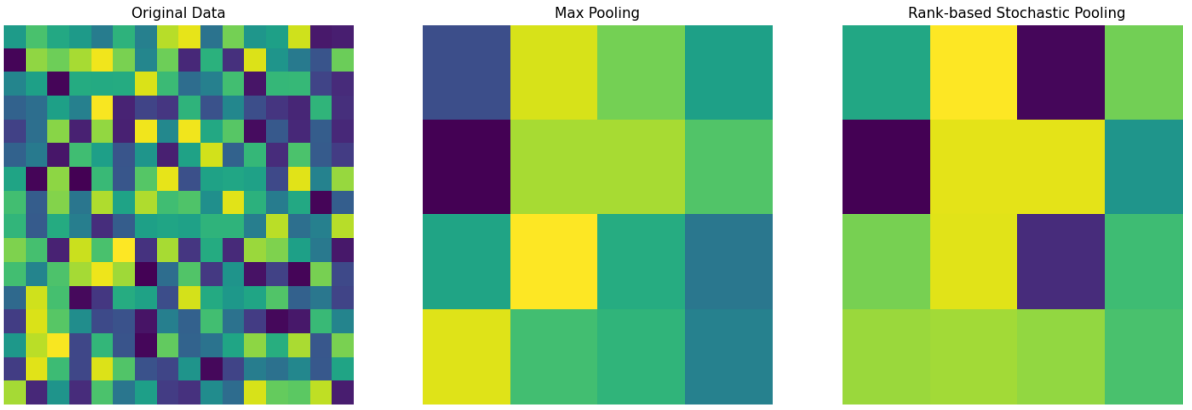

Figure 6. The contrast between maximum pooling and rank-based stochastic pooling.

4. Result

Net 1 was the network with batch normalization and dropout operation, Net 0 was the base network, Net 2 was the network with maximum pooling in Net 1 that was changed to rank-based stochastic pooling, and Net 0-Net 5 was added to the GCN. The base network was given the name Net 0. The results for the different networks are shown in the table below:

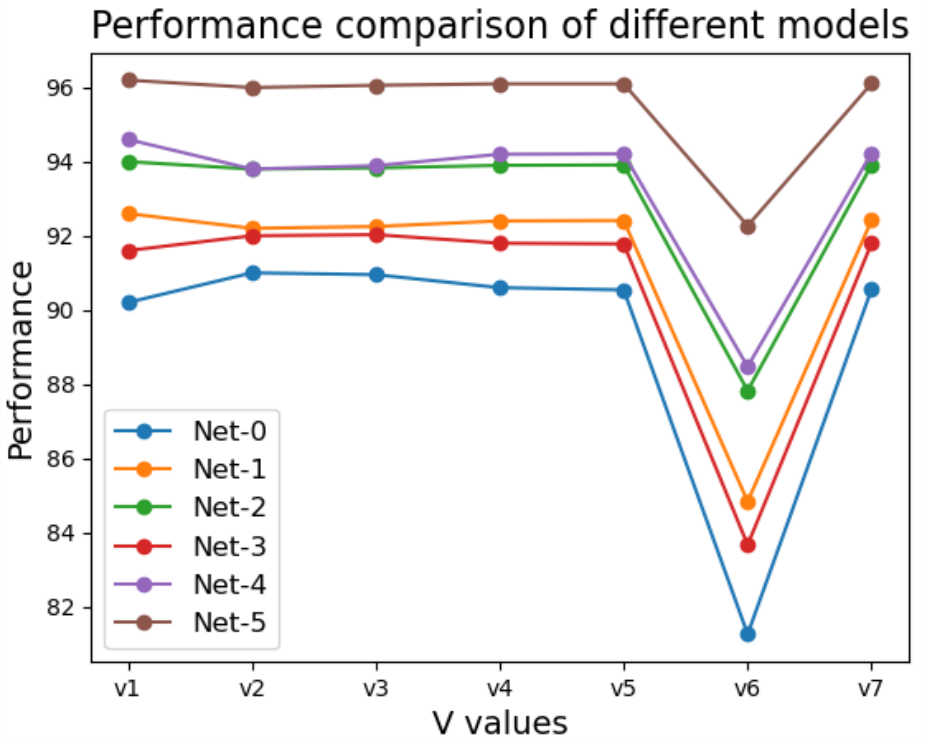

Figure 7. Results of different networks [12].

DOI: 10.54254/2755-2721/10/20230193

It can be shown that adding rank-based stochastic pooling, dropout operation, batch normalization, and the GCN network increases the accuracy of the network, with Net 5 doing the best. The comparison shows that the approach proposed in this study has high accuracy when compared to other methods. The proposed network makes use of dropout, batch normalization, and rank-based random pooling, among other techniques, to enhance performance. The experimental results demonstrate that the suggested network beats cutting-edge techniques in terms of accuracy. To confirm its efficacy and evaluate it against other strategies, the performance of the proposed network must be measured through testing. The trials' findings demonstrate that the suggested network provides greater accuracy, showing its applicability in real-world scenarios. Due to dataset restrictions, it is yet unknown whether the strategy described in this work is effective in real-world applications, despite its encouraging results in identifying malignant breast cancer lesions. Clinical trials still need to be conducted.

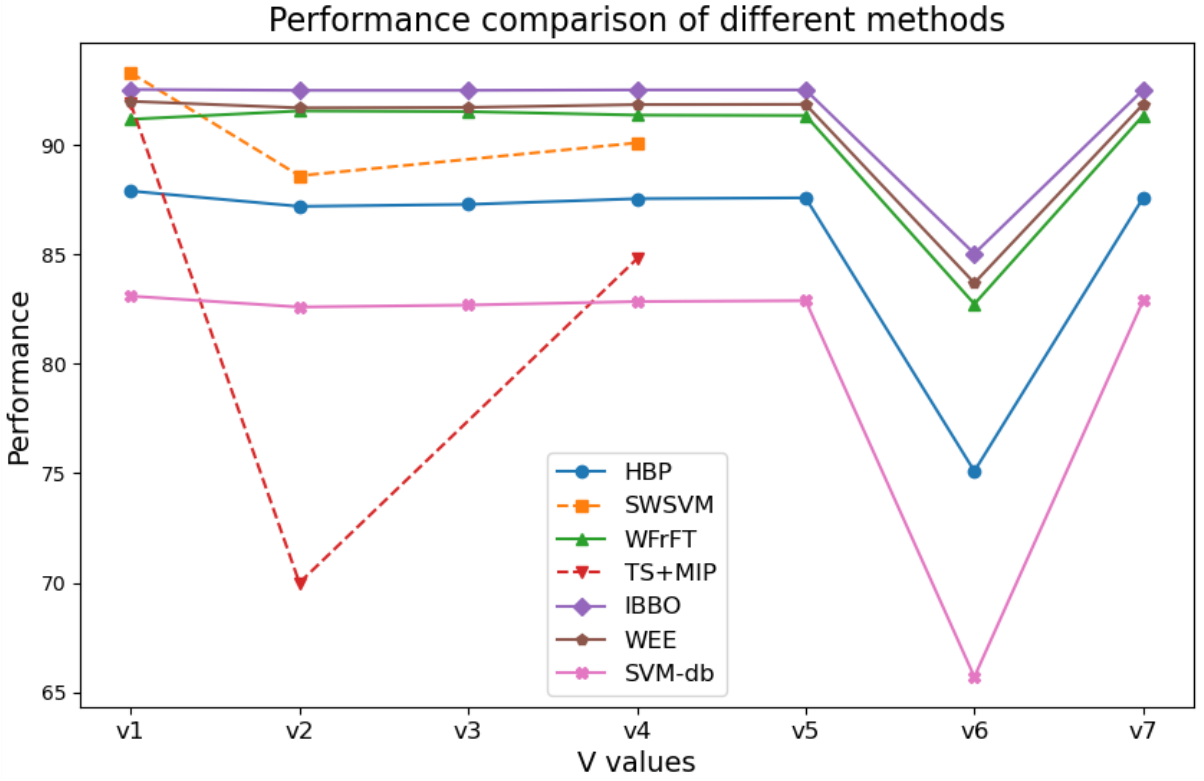

Figure 8. The effectiveness of existing methods [13-20].

5. Conclusion

There are additional faults in the proposed network that will need to be addressed in the future. To begin, the network was tested using a tiny dataset that may not entirely represent all scenarios. Second, because it was designed specifically for radiographic breast cancer pictures, the recommended network may not be applicable for other forms of medical imaging. When it comes to image processing, modern computers struggle to use an immense amount of visual data for computations. There are additional difficulties in the data collection and annotation procedures. Inadequate annotations may diminish the content learned by the network from the data, demanding even more data, but a lack of variety in the gathered data may result in a trained model that does not apply to all individuals. Human annotation, on the other hand, can be costly. These concerns may have an influence on the cost and efficiency of network training, making it more difficult to deploy deep learning on a large scale in the medical business. Despite these limitations, this study provides useful information for developing a network to detect malignant breast cancers. The experiment can be improved by: (1) obtaining a larger dataset to ensure diversity, which is required for training a model that is applicable to all situations; (2) exploring more neural network combinations; applying various techniques to the network for comparison and

DOI: 10.54254/2755-2721/10/20230193

testing to find better approaches; (3) enhancing the current network by investigating the effects of networks with different levels of image analysis depth.

Reference

[1] Esteva A ,Chou K and Yeung S 2021 Deep learning-enabled medical computer vision npj *Digital Medicine* 4(1)p.5

[2] Yue L, Tian D and Chen W 2020 Deep learning for heterogeneous medical data analysis. *World Wide Web* 23 (Austin: The University of Texas at Austin) (5) p.2715-2737

[3] Conti A, Duggento A and Indovina I 2021 Radiomics in breast cancer classification and prediction *Seminars in cancer biology* ed T Vincent 72 p.238-250

[4] Brenner D ,Weir H and Demers A 2020 Projected estimates of cancer in Canada in 2020 *Canadian Medical Association Journal* (Canada) 192(9)p.199-E205

[5] Chen J, Xie Z and Dames P 2022 The semantic PHD filter for multi-class target tracking: from theory to practice *Robotics and Autonomous Systems* 149 103947

[6] Chen J and Dames P 2022 Multi-class target tracking using the semantic phd filter In Robotics Research: The 19th International Symposium ISRR p 526-541

[7] Lu L, Dercle L and Zhao B 2021 Deep learning for the prediction of early on-treatment response in metastatic colorectal cancer from serial medical imaging *Nat Commun*. 12(1) p. 6654

[8] Rodriguez-Esparza E, Zanella-Calzada L and Oliva D 2020 An efficient Harris hawks-inspired image segmentation method. *Expert Systems with Applications* p.155

[9] Srivastava N, Hinton G and Krizhevsky A 2014 Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15 p 1929-1958

[10] Li W and Yan N 2018 Adaptive Batch Normalization for practical domain adaptation. *Pattern Recognition* 80 p. 109-117

[11] Zhang Y, Satapathy S and Zhu L 2022 A seven-layer convolutional neural network for chest CT-based COVID-19 diagnosis using stochastic pooling *IEEE Sensors Journal* 22(18): p.17573-17582

[12] Zhang Y and Dsg D 2021 Improved Breast Cancer Classification Through Combining Graph Convolutional Network and Convolutional Neural Network. *Information Processing & Management* 58(2)

[13] Milosevic M, Jankovic D and Peulic A 2015 Comparative analysis of litection in mammograms and thermograms. *Biomed Tech (Berl)* 60(1) p. 49-56

[14] Liu F, Nakamura K and Payne R 2017 Abnormal breast detection via combination of particle swarm optimization and biogeography-based optimization

[15] Gorgel P, Sertbas A and Ucan N 2015 Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines. *Expert Systems* 32: p. 155 - 164

[16] Zhang Y, Wang S and Yang J 2016 Computer-aided diagnosis of abnormal breasts in mammogram images by weighted-type fractional Fourier transform *Advances in Mechanical Engineering* 8(2)

[17] Huang C, Liao H and Shiang Y 2009 Synthesis of wavelength-tunable luminescent gold and gold/silver nanodots *J. Mater. Chem* 19(6) p. 755-759

[18] Wu F, Halatek J and Reiter M 2016 Multistability and dynamic transitions of intracellular Min protein patterns *Mol Syst Biol* 12(6) p. 873

[19] Chen X, Yun Z and Pan Y 2016 Wavelet energy entropy and linear regression classifier for detecting abnormal breasts. *Multimedia Tools and Applications* 77(3) p. 3813-3832

[20] Liu F and Brown M 2019 Breast cancer recognition by support vector machine combined with daubechies wavelet transform and principal component analysis *Breast Detection* Cham: Springer International Publishing 32