EAI Endorsed Transactions
on AI and Robotics

Research Article EAI.EU

Hearing loss classification via AlexNet and Support Vector Machine

Jing Wang[1*]

[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, P R China

Abstract

This paper presents a new method for detecting hearing loss. Our approach is first to use AlexNet to extract the features. Then, we use the Support Vector Machine as a classifier to classify the images. 10-fold cross-validation results showed that the sensitivities of the healthy control group, the left-sided hearing loss group, and the right-sided hearing loss group in this method were 94.67%, 94.00%, and 95.17%, respectively, achieving a very good effect compared with other hearing loss detection methods. In conclusion, our method is effective for the identification of hearing loss.

Keywords: AlexNet, Support Vector Machine, Hearing loss.

Received on 07 March 2023, accepted on 21 March 2023, published on 21 April 2023

doi: 10.4108/airo.v2i1.3113

Corresponding author. Email: wangjing@home.hpu.edu.cn

1. Introduction

1.1. Background

Hearing loss is also known as deafness or hearing level. Hearing loss includes decreased hearing sensitivity, increased hearing threshold, impaired hearing function, and so on. With the development of medicine and technology, many diseases that cause hearing loss (such as otitis media, etc.) can be prevented by treatment [1, 2]. There are also measures to prevent hearing loss individually, such as practicing ear hygiene and staying away from loud noises.

On the one hand, the causative and preventive factors of hearing loss interact to determine the occurrence, nature, severity, and progression of hearing loss. There are many risk factors for hearing loss. Prenatal risk factors include genetic factors, intrauterine infections, hypoxia or birth asphyxia, hyperbilirubinemia, low birth weight, other prenatal infections, or ototoxic drugs. Risk factors for children and adolescents include otitis media, meningitis, and other infectious diseases. Risk factors for adults and the elderly include chronic diseases, smoking, otosclerosis, age-related hearing degradation, sudden hearing loss (sudden deafness), and non-modifiable risk factors (such as gender, ethnicity, etc.).

On the other hand, excessive earwax, ear or head trauma, noise (occupational, environmental, recreational), ototoxic drugs, ototoxic chemicals, nutritional deficiencies, viral infections, and other ear diseases can also cause hearing loss. According to the World Health Organization (WHO), an estimated 1.1 billion young people (ages 12 to 35) worldwide are at risk of hearing loss due to regular exposure to recreational noise [3].

According to the data released by the World Health Organization, there are about 1.1 billion young people (between 12 and 35 years old) in the world who are facing irreversible hearing loss [4, 5]. Without action, nearly 630 million people worldwide will have hearing loss by 2030; By 2050, that number will rise to more than 900 million (1 in 10). Hearing tests can be performed in hospitals or professional hearing aid fitting centers. Hearing loss can be accurately measured through hearing tests.

The International Health Organization (WHO-2021) classifies hearing loss as follows: average hearing loss of fewer than 20 decibels is normal; Mild hearing loss is defined as an average hearing loss between 20 and 35 decibels; Moderate hearing loss is defined as an average hearing loss of 35 to 50 decibels. Moderate to severe hearing loss was defined as an average hearing loss between 50 and 65 decibels [6, 7]; Severe hearing loss was defined as a mean hearing loss of 65 to 80 decibels; An average hearing loss of 80 to 95 decibels is considered a severe hearing loss. Total hearing loss/total deafness is a mean hearing loss of 95 decibels or more significant [8].

In addition, the average hearing loss was less than 20 decibels in normal ears and greater than or equal to 35 decibels in ears with hearing loss. Compared with the old classification standard, the new standard takes 15dB as the first level. The grading criteria for moderate and severe hearing loss, complete hearing loss, and unilateral deafness were improved. According to this classification, more than 20 percent (1.5 billion) of the world's population cannot hear.

1.2. The Latest Method

1.2.1. Related Work

With the rapid development of machine learning, more and more people focus on the combination of machine learning correlation algorithms and medical image processing [9]. Trying to identify hidden disease diagnosis automatically features from medical image big data through machine learning correlation algorithms such as linear regression, Support Vector Machine (SVM), nearest neighbor (KNN), logistic regression, dimension reduction, gradient enhancement, and so on.

The medical image reflects the internal structure of the human body, which is one of the main bases of modern medical diagnosis. It has the characteristics of availability, high quality, large volume, and unified standard. With the development of medical imaging and computer technology, medical image analysis has become an indispensable tool and technical means in medical research, diagnosis, and treatment of clinical diseases. In recent years, deep convolutional neural networks (CNNs) have become a research hotspot in medical image analysis [10].

More than half of all hearing loss is preventable. Therefore, early detection of hearing loss is very important for patients. It is particularly important to judge patients' pathology through medical images [11]. Through reading the recent papers published by scholars, it is found that scholars have proposed several feasible methods for image classification [12]. Tang, L. et al. (2018) [13] proposed the method of combining Hu moment invariant (HMI) and Support Vector Machine. Gao, R. et al. (2019) [14] used the hearing loss recognition method based on wavelet entropy and cat swarm optimization, Wang, L. et al. (2020) [15] suggested hearing loss recognition based on fractional Fourier entropy and direct acyclic graph Support Vector Machine. Yao, X. et al. (2020) [16], [17] proposed a hearing loss classification algorithm based on stationary wavelet entropy and genetic algorithm. Some methods have achieved certain effects, but there are still problems, such as weak generalization ability and unsatisfactory classification effect. Zhang, Y. (2017) [18] used stationary wavelet entropy (SWE) to detect unilateral hearing loss.

1.2.2. Contributions

In this paper, AlexNet+SVM is used for feature extraction of image information. AlexNet extracts the features of the collected images, and SVM classifies the extracted feature information to achieve a higher classification accuracy. For SNHL classification tasks, I use the HLNet framework. In the HLNet framework, I use AlexNet model to extract more accurate and effective original information to train our model. Firstly, data enhancement was used to increase MRI image information to improve generalization ability. Then, CNN was used as the depth feature extractor [19, 20]. Finally, Support Vector Machine (SVM) was used to divide the extracted feature information into left-sided hearing loss SNHL (LSHL), right-sided hearing loss SNHL (RSHL), and a healthy control group (HC). After preprocessing, the model can accurately classify SNHL data sets, significantly improve the classification efficiency, and show good generalization ability. Experimental results show that this method has good performance.

The structure of this article is shown below. Section 2 describes the data set used, Section 3 analyzes the models and methods used, Section 4 gives the corresponding experimental results and comparative data with other methods, and Section 5 summarizes.

2. Dataset

Usually, when I extract data samples from image data sets, I will use new data sets to ensure the accuracy of data samples. First, I need to collect a large number of hearing loss images to get an excellent data set. The AlexNet model requires input of a normalized image with a size of 224pixel×224pixel. And then there's image classification. Before image classification, I should first divide the image information into the data set into two parts: training samples and test samples (a 10-fold cross-validation method is used to distinguish training samples and test samples).

The process of image classification mainly includes two steps: training and testing. The training step is to input the image information in the training sample into the Support Vector Machine, train the classifier of the Support Vector Machine, and get a good Support Vector Machine model. The test process is to input the image information in the test sample into the trained SVM model for image feature recognition, to achieve the correct classification of images. Different types of hearing loss image information are shown in Figure 1.

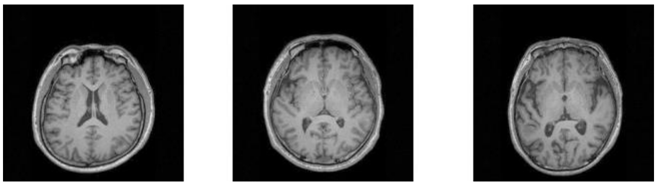


Figure 1. images of hearing loss

From left to right, they are healthy control (HC), left-sided hearing loss (LSHL), and right-sided hearing loss (RSHL).

3. Methodology

3.1. AlexNet

The current target recognition task is using the traditional machine learning method. Since there are thousands of mutable images in the real world but the current tagged data set is relatively small, it is important to use a very large training set to identify a large number of mutable images in the real world. The training set includes LabelMe, which contains tens of thousands of fully segmented images, and ImageNet, which contains 15 million high-resolution images with tags. To learn a large number of pictures, you need a model with a very large learning capacity. CNN(convolutional neural network) is proposed accordingly. The capacity of convolutional neural networks can be controlled by varying their width and depth, and they also make correct assumptions about the properties of the picture (the statistical stability and the dependence between local pixels).

Compared with other feedforward neural networks proposed before, CNN has fewer connections and parameters, so it is easier to train. But that's not enough, because using CNN is still extremely expensive compared to using larger, high-resolution images. AlexNet is a network model, and the name comes from Alex Krizhevsky, the first author of the paper [12]. In the past 20 years since LeNet's proposal, neural networks were once surpassed by other Machine learning methods, such as Support Vector Machine. However, AlexNet's proposal in 2012 proved for the first time that the features learned could surpass those designed by hand and break the current situation of computer vision research in one stroke.

Although, the structure of AlexNet's network model is similar to LeNet's on the whole in that it first convolves and then fully joins. However, the AlexNet and LeNet network models differ greatly in detail. AlexNet is more complex. AlexNet has 60 million parameters and 65,000 neurons, five layers of convolution, three layers of fully connected network, and the final output layer is the 1000-channel softmax. At the same time, AlexNet uses two GPUs for calculation, which greatly improves computing efficiency.

Moreover, in the ILSVRC-2012 competition, AlexNet obtained an error rate of 15.3% in the top-5 test, while the error rate of the method that won second place was 26.2%. The difference is large enough to illustrate the impact the network had on academia and industry at the time. AlexNet's innovations include 1. Use of ReLU activation function; 2. Put forward Dropout to prevent overfitting; 3. Use Data augmentation.

3.1.1. AlexNet Network Model Structure

The network structure of AlexNet model has 8 layers. 1 to 5 are convolution layers, and 6 to 8 are fully connected layers. The input parameters of each layer of the model are shown in Figure 2. Among the convolutional layers, it is noteworthy that the third and fourth convolutional layers are no Maxpooling [21, 22] the basic structure of the convolutional layer is Convolutional and ReLU, while the basic structure of the first, second, and fifth convolutional layers is Convolutional, ReLU, and Maxpooling. The full connection layers are named FC1, FC2, and FC3 respectively [23, 24]. The basic structure of the full connection layers FC1 and FC2 are Full connection, ReLU, and Dropout. Fully connection layer The basic structure of FC3 is Full connection and Softmax. The original image input 224 × 224 is randomly cropped, and the actual size of the image is 227 × 227.

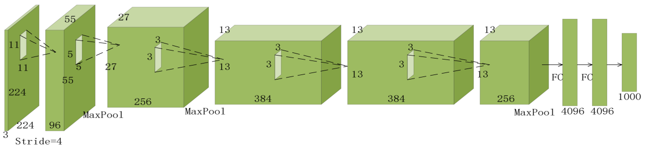


Figure 2. AlexNet network model structure

3.1.2. ReLU Nonlinearity

Under standard conditions, tanh or sigmoid is generally used as an activation function for neuron output. However, in terms of the training time of gradient descent, these saturated nonlinear functions are much slower than unsaturated nonlinear functions. Such nonlinear Units are referred to as Rectified Linear Units (ReLUs). The training speed of a deep convolutional neural network with ReLU is several times faster than that with tanh. Meanwhile, the value of ReLU gradient descent in AlexNet is always 1 to solve the problem of gradient disappearance. ReLU is shown in Figure 3. The formula of the ReLU activation function is:

$$f\left(x\right)=\max\left(0,x\right) \tag{1}$$

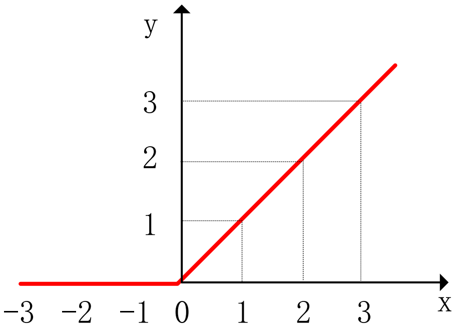


Figure 3. ReLU activation function

3.1.3. LRN

The concept of local response normalization (LRN) technique was first proposed in the AlexNet model. LRN technology is mainly a technical method to improve the accuracy of deep learning training, which is different from the activation function. LRN is generally an operation carried out after activation and pooling [25]. Why introduce an LRN layer? Firstly, a neurobiological concept should be introduced: lateral inhibition when an activated neuron suppresses an adjacent neuron. The purpose of normalization is "inhibition", and LRN uses this kind of lateral inhibition to achieve local inhibition [26]. Especially when I use ReLU, this "lateral inhibition" is very effective and can achieve better effects in AlexNet. The formula for local response normalization is as follows.

$$b^i_{x,y} = \frac{a^i_{x,y}}{\left(k+\alpha\sum_{j=\max\left(0,i-\frac{n}{2}\right)}^{\min\left(N-1.i+\frac{n}{2}\right)}(a^j_{x,y})^2\right)^\beta}, \qquad (2)$$

where a represents the output result after the convolution layer (including convolution operation and pooling operation).The structure of the output is a four-dimensional array $\left[batch, height, width, the\ channel\right]$ ($batch$: number of batches (each batch is one image); $height$: height of the image; $width$: width of the image; $the\,channel$: number of channels can be understood as the number of neurons output by an image in a batch after the convolution operation (or the depth of the image after processing)). $a^i_{x,y}$ represents a position in the output structure $\left[a,b,c,d\right]$, which can be understood as a point at a certain height and a certain width under a channel in a graph, that is, a point with a height of b and a width of c under the d channel in graph a. a , $n/2$, k , α , and β represent input, depth_ radius, bias, alpha, and beta respectively in the function, where $n/2$, k , α , and β are all customized. In particular, note that the direction of $\sum$ the stack is along the direction of the channel. That is, the sum of squares of the values of each point is along the direction of the 3rd channel in that is, the sum of squares of the points of the first $n/2$ channels (minimum is the 0th channel) and the second $n/2$ channels (maximum is the d-1 channel) in the same direction of a point (total n +1 points). In addition, it is also explained in the English annotations of the function that the input is regarded as d three-dimensional matrices, which means that the number of channels in the input is regarded as the number of three-dimensional matrices, and the direction of superposition is also in the direction of channels. i represents the output of the ith kernel at the position (x, y) after the activation function ReLU is applied, n is the number of adjacent kernal maps at the same position [27], and N is the total number of kernal. The parameters k , n , α , and β are all hyperparameters. Generally set $k=2$, $n=5$, $\alpha=0.0001, \beta=0.75$.

3.1.4. Dropout

Dropout is a relatively common method for suppressing overfitting. Two fully connected layers in AlexNet use Dropout [28], because the fully connected layer tends to overfit, whereas the convolutional layer does not. The exit in a neural network is realized by modifying the structure of the neural network itself [29]. For a layer of neurons, by defining the probability that the neuron is set to 0, the neuron will not participate in the forward and backward propagation as if it were deleted in the network [30]. Meanwhile, the number of neurons in the input and output layers remained the same. Then the parameters are updated according to the learning method of the neural network. In the next iteration, randomly delete some neurons again (set to 0) until the end of the training. The effect of Dropout is shown in

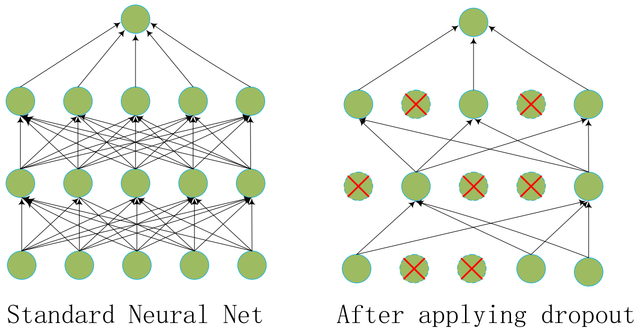


Figure **4**.

3.2. Extract Features From AlexNet

AlexNet model mainly consists of a convolution layer, a fully connected layer, and a pooling layer. Among them, the function of the convolution layer is to extract the image feature method. The convolution kernel correlation will make the feature graph smaller and smaller. To prevent that, we are going to use padding before we do the convolution, which is called effective convolution. In this model, only the first convolution uses effective convolution. The feature mapping of each convolution layer is as follows:

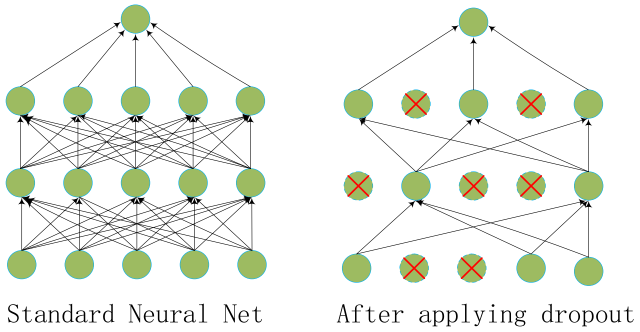


Figure 4. Role of Dropout

$$o^l_j = f\left(\sum_i o^{l-1}_j \times \omega^l_{ij} + b^l_j\right) \qquad (3)$$

where o^l_j represents the output of the Jth neuron after l layer convolution; ω^l_{ij} is the convolution kernel; b^l_j stands for bias. f is the activation function of neurons, and AlexNet uses the ReLU activation function.

The main function of the pooling layer is to reduce the model's size, improve the computing speed and improve the robustness of the model. The pooling layer can be divided into the maximum pooling layer and the average pooling layer. This model uses the maximum pooling layer. The size and stride of the pooling layer are shown in Figure 5.

3.3. Support Vector Machine

Support vector machine (SVM, also known as support vector network) is the algorithm that has received the most attention in machine learning. From the practical application perspective: SVM performs very well in various practical problems. It is widely used in handwritten recognition, digit recognition, and face recognition and plays an important role in classifying text and hypertext. At the same time, SVM is also used to perform image classification and image segmentation systems [31]. From an academic point of view: SVM is the closest machine learning algorithm to deep learning. SVM is a binary model. The basic idea is to solve the separation hyperplane that can correctly divide the training data set and has the largest geometric interval. Its learning strategy maximizes the interval and finally becomes the solution to a convex quadratic programming problem [32]. SVM can be divided into linear separable support vector machines, linear support vector machines, and nonlinear support vector machines. We often think of convolutional neural networks and deep learning for image classification. However, deep learning often requires much computational overhead, and code writing and debugging are relatively complex. For image classification of small and distinct data sets, we use the SVM algorithm to classify images.

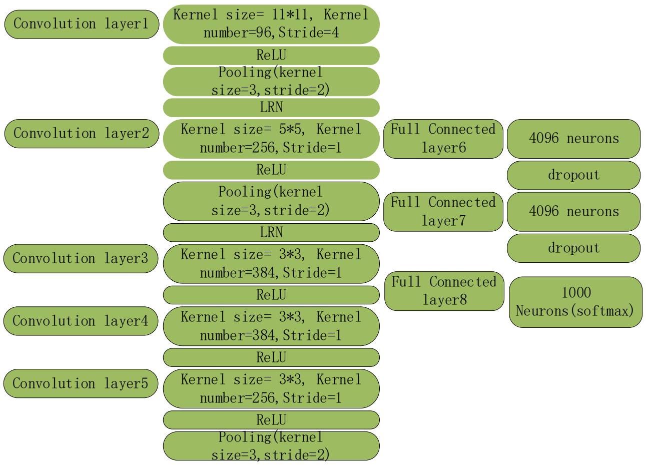

Convolution layer1
Kernel size= 11*11, Kernel number=96,Stride=4
ReLU
Pooling(kernel size=3,stride=2)
LRN
Convolution layer2
Kernel size= 5*5, Kernel number=256,Stride=1
ReLU
Pooling(kernel size=3,stride=2)
LRN
Convolution layer3
Kernel size= 3*3, Kernel number=384,Stride=1
ReLU
Convolution layer4
Kernel size= 3*3, Kernel number=384,Stride=1
ReLU
Convolution layer5
Kernel size= 3*3, Kernel number=256,Stride=1
ReLU
Pooling(kernel size=3,stride=2)
Full Connected layer6
4096 neurons
dropout
Full Connected layer7
4096 neurons
dropout
Full Connected layer8
1000 Neurons(softmax)

Figure 5. Layer structure schematic of AlexNet

Linear separable support vector machine: When the training samples are linearly separable, a linear separable support vector machine is learned by maximizing the hard interval.

If a linear function can separate the samples, the data samples are said to be linearly separable. A linear function is a straight line in two dimensions, a plane in three dimensions, and so on. If space dimension is not taken into account, such linear functions are collectively called hyperplanes. The hyperplane can be expressed as:

$$W^T x + b = 0, \tag{4}$$

where x is the input vector, that is, the vector in the sample set; W is an adjustable weight vector, each vector has an adjustable weight; T : transpose, the transpose of the vector; b : Bias, the deviation of the hyperplane from the origin.

Example of principle: W^T takes (w_1, w_2) and x takes (x_1, x_2) T , then the original formula is $w_1 x_1 + w_2 x_2 + b = 0$, which is the same as the traditional equation $Ax + By + C = 0$. The above formula is the general formula from two-dimensional and three-dimensional space to a higher-dimensional plane. W : is the normal vector of the plane, which determines the direction of the hyperplane. b : determines the distance between the hyperplane and the origin. The normal vector is the same as the number of sample attributes and the hyperspace dimension. The parameters we want in hyperspace are the values of W and b that determine the hyperplane.

The distance from any point x in hyperspace to the hyperplane is:

$$r = \frac{|w^T x + b|}{\|w\|}, \tag{5}$$

We can understand this formula from the special to the general, if in two-dimensional space, namely the plane, the distance from the point to the line is:

$$d = \left| \frac{Ax_0 + By_0 + C}{\sqrt{A^2 + B^2}} \right|, \tag{6}$$

where A, B, C are the parameters of the line and w, x_0, and y_0 are the coordinates of x. This distance is geometric distance, which is the distance that the human eye sees intuitively. Geometric distance has size but no direction. Now we artificially specify that the point in the sample data with a positive distance from the hyperplane is the sample point of class +1, which is this kind of point and label it +1, and the point with a negative distance from the hyperplane is labeled -1, why not +2, -2 actually can be used, 1 is for the convenience of subsequent calculation [33].

Now assume that this hyperplane can classify samples correctly, but we do not know the values of w and b of this hyperplane, which is exactly what we require, but this plane must exist, there are:

$$y_i = +1 \rightarrow w^T x_i + b > 0 \tag{7}$$

$$y_i = -1 \rightarrow w^T x_i + b < 0 \tag{8}$$

After removing the absolute value of the numerator and denominator in equation r of geometric distance (because they are both positive), the remaining $w^T + b$ can determine whether the sample is in the category of +1 or -1, and the functional distance is defined as:

$$y_i \left(w^T x_i + b \right). \tag{9}$$

The distance is the sample class times w^T plus b . Because the positive sample category is +1 and $w^T + b$ is also positive, The negative sample class is -1 and $w^T + b$ is negative. So the function distance only has magnitude and no direction. Functional distance is equivalent to the molecular part of the geometric distance [34]. In all samples, each point has a functional distance and a geometric distance. The geometric distance is the direct observable distance, and the functional distance has the following properties: The functional distance from a point to the hyperplane depends on the normal vector and b values of the hyperplane. The same hyperplane can have multiple sets of w and b values, but each set of values is proportional. The distance between the points varies depending on the values of w and b. In general, to solve the SVM algorithm, meet the constraint condition of $y_i\left(w^T x_i + b\right) \geq 1$, under the premise of solving $\| w \|^2$ minimum value.

3.4. Implementation

The concrete implementation process of this paper is shown in Figure 6. The image data is trained in the AlexNet model and then classified using support vector machines. The 10-fold cross-validation calculation was performed on the processed data, and the classification results were compared in the next step.

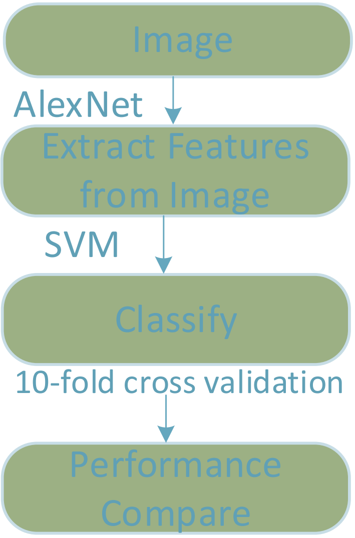


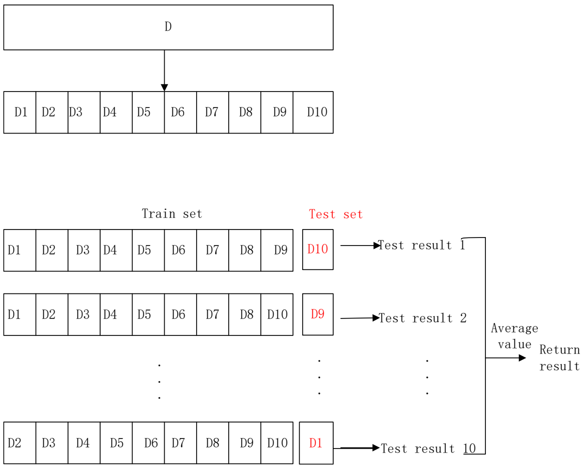
Image
AlexNet
Extract Features from Image
SVM
Classify
10-fold cross validation
Performance Compare

Figure 6. Implementation process

3.5. Measure

3.5.1. 10-Fold Cross-Validation

The 10-fold cross-validation method divides the collected image information into ten parts. Take one part as the test sample and the rest as the training sample. This process was repeated ten times in different training and tests. This process is shown in Figure 7.


D
D1 D2 D3 D4 D5 D6 D7 D8 D9 D10
Train set
Test set
D1 D2 D3 D4 D5 D6 D7 D8 D9 D10
Test result 1
D1 D2 D3 D4 D5 D6 D7 D8 D10 D9
Test result 2
Average value
Return result
D2 D3 D4 D5 D6 D7 D8 D9 D10 D1
Test result 10

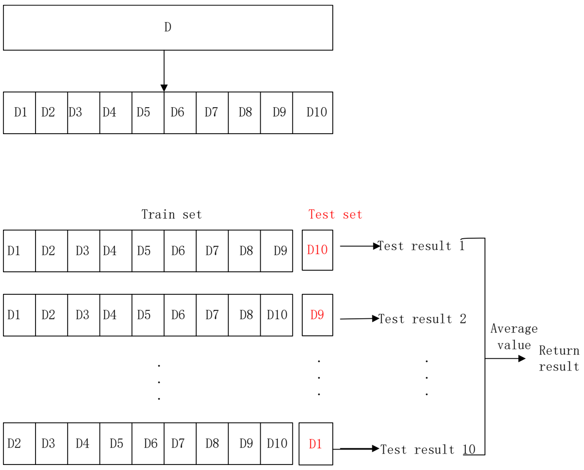

D
D1 D2 D3 D4 D5 D6 D7 D8 D9 D10
Train set
Test set
D1 D2 D3 D4 D5 D6 D7 D8 D9 D10
Test result 1
D1 D2 D3 D4 D5 D6 D7 D8 D10 D9
Test result 2
Average value
Return result
D2 D3 D4 D5 D6 D7 D8 D9 D10 D1
Test result 10

Figure 7. 10-fold cross-validation

The purpose of cross-validation is to reduce the randomness caused by a single division of train set and test set, make full use of existing data sets of multiple divisions, and avoid the selection of hyperparameters and models without generalization ability due to special division. Cross-validation is used to reduce overfitting probability and improve generalization ability.

3.5.2. Micro Average F1 Score

F1 Score (F1-score, F1-measure) is a Measure of a classification problem used to weigh Precision and Recall and is defined as the harmonic average of precision rate and recall rate.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{10}$$

precision: How much of a supposedly positive sample is actually positive.

$$Precision = \frac{TP}{TP + FP}, \tag{11}$$

Recall: How many samples have been recovered from a positive sample.

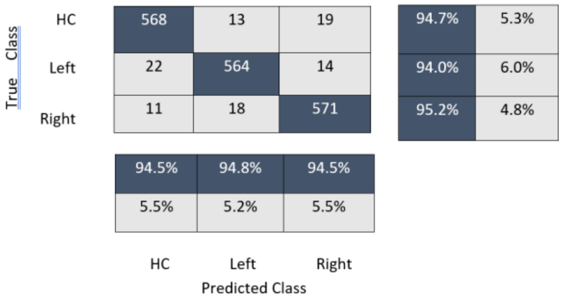
$$Recall = \frac{TP}{TP + FN}, \tag{12}$$

where TP(True Positive): True positive indicates that it is predicted to be positive but is positive. FP(False Positive): False positive indicates that the prediction is positive, but the actual is negative. TN(True Negative): The prediction is negative, but the result is negative. FN (False Negative): False negative indicates that the prediction is negative, but the result is positive.

4. Experiment Results and Discussion

4.1. Confusion Matrix of the Proposed method

Confusion matrix, also known as the possibility table or error matrix.


True Class
Predicted Class

	HC	Left	Right		
HC	568	13	19	94.7%	5.3%
Left	22	564	14	94.0%	6.0%
Right	11	18	571	95.2%	4.8%
	94.5%	94.8%	94.5%		
	5.5%	5.2%	5.5%		

Figure 8. Confusion matrix

It is a specific matrix used to visualize the performance of an algorithm, usually supervised learning (non-supervised learning, usually with a matching matrix). Each column represents the predicted value, and each row represents the actual category. The object matrix is a summary of an algorithm's test results for further performance analysis.

See Figure 8, the number of samples with hearing loss on the left-sided was 600. The number of correctly predicted samples was 564, the number of healthy controls was 22, and the number of right-sided hearing loss was 14. The number 11 in the first column of the third row indicates that 11 samples with hearing loss on the right-sided were incorrectly predicted to be healthy controls; row3, column2, indicates that 18 samples with right-sided hearing loss were incorrectly predicted to have left-sided hearing loss, while row3, column3, indicates that 571 samples with actual right-sided hearing loss were correctly predicted to have right-sided hearing loss. From the above data, it can be seen that the accuracy rate and recall rate of the healthy control group is about 94.5% and 94.7%, respectively. The accuracy rate of hearing loss in the left-sided was 94.8 percent and the recall rate was 94 percent. The accuracy rate of hearing loss in the right-sided was about 94.5%, and the recall rate was 95.2%. Among them, the right-sided hearing loss class had the highest recall rate, while the left-sided hearing loss class had the highest accuracy rate.

4.2. Statistical Results

Figure 8 shows our confusion matrix. As can be seen from Figure 8, 568 samples of the healthy control group were predicted to be healthy controls, 13 samples of the healthy control group were predicted to have hearing loss in the left-sided, and 19 samples of the healthy control group were predicted to have hearing loss in the right-sided. We collected a total of 1800 subjects with hearing loss image information. There were 600 SNHL subjects with left-sided hearing loss, 600 SNHL subjects with right-sided hearing loss, and 600 healthy control subjects (HC). Before model training, we first need to obtain relevant image data information from subjects, process the image information, adjust the image scale, and send the image to the HLNet model. Then AlexNet model is used to complete the extraction of image feature information, and the extracted feature information is handed to Support Vector Machine (SVM) for classification. In the classification process, a 10-fold cross-validation method is used to obtain the classification result of image information. The specific results of model-related performance are shown in Table 1.

Table 1. Model Performance

Model Performance			
Type	Precision	Accuracy	F1 Score
Healthy Control Group (HC)	94.5	94.7	94.6
Left-sided Hearing loss (LSHL)	94.8	94.0	94.4
Right-sided Hearing loss (RSHL)	94.5	95.2	94.7
Micro-Average	94.6	94.6	94.6

As can be seen from Table 1, the accuracy rate of the healthy control group was the same as that of the right-sided hearing loss group, both being 94.5%. The left-sided hearing loss group had the highest accuracy of 94.8 percent. The accuracy and F1 Score of right-sided hearing loss were 95.2% and 94.7%, respectively. The accuracy and F1 Score of left-sided hearing loss are 94.0% and 94.4%, respectively.

4.3. Comparison to State-of-the-art Approaches

We compared the AlexNet+SVM method used in this hearing loss model with the advanced HMI+SVM [13], WE+CSO [14], FRFE+DAG-SVM [15], SWE+GA [16], SWE+CSO [17] methods and obtained the data shown in Table 2. The data comparison of each model is also shown in detail in Figure 9. In HMI + SVM [13], the accuracy of healthy control group (HC), left-sided SNHL(LSHL) and right-sided SNHL(RSHL) was 76.83%, 75.00%, and 76.33%, respectively. The accuracy of SNHL(LSHL), SNHL(RSHL), and healthy control group (HC) by WE+CSO [14] method was 84.33%, 83.33%, and 84.67%, respectively. In FRFE+DAG-SVM [15] method, SNHL in the left-sided (LSHL) was 93.83%, SNHL in the right-sided (RSHL) was 94.17%, and that in healthy control group (HC) was 94.17%. In the method of SWE+GA [16], SNHL of left-sided (LSHL) was 91.00%, SNHL of right-sided (RSHL) was 89.00%, and that of healthy control group (HC) was 89.67%. In the SWE+CSO [17] method, the healthy control group (HC), left-sided SNHL(LSHL) and right-sided SNHL(RSHL) was

91.00%, 89.00%, 90.67%, and 90.22%, respectively. The MAF1 of HMI + SVM [13], WE+CSO [14], FRFE + DAG-SVM [15], SWE+ GA [16], and SWE+CSO [17] were 76.06%, 84.11%, 94.06%, 89.89% and 90.22%, respectively. In our AlexNet+SVM method, the left SNHL(LSHL), right SNHL(RSHL), healthy control group (HC) and MAF1 were 94.00%, 95.17%, 94.67% and 94.61%, respectively. The data presentation is shown in Figure 10.

Combined with the data in Table 2, it can be seen that the "AlexNet+SVM" method has the highest accuracy in classifying the highlighted messages of healthy control group, left-sided SNHL, and right-sided SNHL respectively (94.67%, 94.00%, 95.17).HMI+SVM [13] had the lowest accuracy for image classification (76.83%, 75.00%, 76.33%). WE+CSO [14] uses genetic algorithms, a method of finding optimal solutions by simulating natural evolutionary processes. HMI+SVM [13] uses 7 Hu moment invariants to extract features and uses Support Vector Machines as classifiers. SWE+GA [16] adopts the BBO algorithm for optimization to avoid local optimal dilemmas. Experimental results show that our method has good performance in image feature extraction and classification. At the same time, through the experiment, our method obtained the highest score.

5. Conclusions

This paper proposes a classification method of hearing loss based on AlexNet and Support Vector Machine (SVM). AlexNet neural network is used in this method. AlexNet uses a variety of technologies such as Dropout, ReLU, and Data Augmentation to address the overfitting of deep neural networks, allowing them to converge well with a large number of image parameters. At the same time, it also provides a better method for the development of image classification. Experimental results show that this method has high recognition accuracy, robustness, and generalization ability.

Table 2. Model effect comparison

Model effect comparison				
Model	HC	LSHL	RSHL	MAF1
HMI+SVM [13]	76.83	75.00	76.33	76.06
WE+CSO [14]	84.67	84.33	83.33	84.11
FRFE+DAG-SVM [15]	94.17	93.83	94.17	94.06
SWE+GA [16]	89.67	91.00	89.00	89.89
SWE+CSO [17]	91.00	89.00	90.67	90.22
AlexNet-SVM(Ours)	94.67	94.00	95.17	94.61

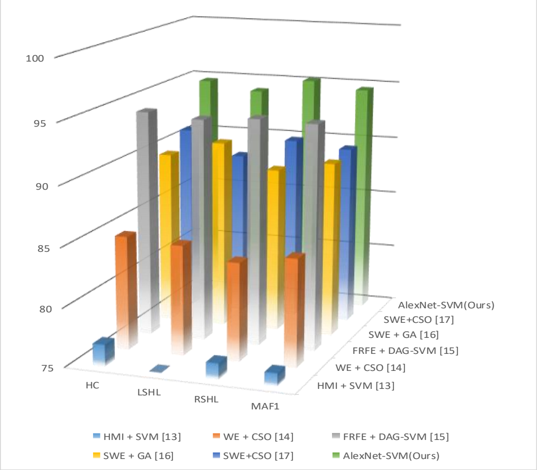


Figure 9. Columnar comparison of data from different models

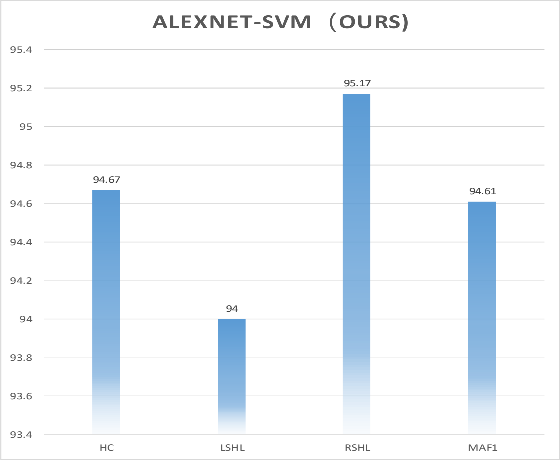


Figure 10. AlexNet-SVM(Ours)

In the future, I will collect more hearing data from hearing-impaired people and test their performance using deep learning methods. Such as classification deep neural networks and transfer learning. In the following work, we will also try better hearing loss methods to improve classification recognition accuracy and deep neural networks' performance.

References

[1]. Wang, S., Detection of Left-Sided and Right-Sided Hearing Loss via Fractional Fourier Transform. Entropy, 2016. 18(5): Article ID. 194

[2]. Jeong, J., Neutrophil-to-lymphocyte ratio as a prognostic inflammatory factor in sudden sensorineural hearing loss. European Journal of Inflammation, 2022. 20: Article ID. 1721727x221144452

[3]. Muus, J.S., et al., Hearing loss in children with growth hormone deficiency. International journal of pediatric otorhinolaryngology, 2017. 100: p. 107

[4]. Wang, S., et al., Wavelet entropy and directed acyclic graph support vector machine for detection of patients with unilateral hearing loss in MRI scanning. Frontiers in Computational Neuroscience, 2016. 10: Article ID. 160

[5]. Jaradeh, K., et al., Hearing Loss Screening, Diagnosis, and Treatment for Refugees and Asylees in an Urban Clinic, 2014-2017. Oto Open, 2022. 6(4): Article ID. 2473974x221132509

[6]. Wang, S., Texture Analysis Method Based on Fractional Fourier Entropy and Fitness-scaling Adaptive Genetic Algorithm for Detecting Left-sided and Right-sided Sensorineural Hearing Loss. Fundamenta Informaticae, 2017. 151(1-4): p. 505-521

[7]. Loughrey, D.G., Is age-related hearing loss a potentially modifiable risk factor for dementia? Lancet Healthy Longevity, 2022. 3(12): p. E805-E806

[8]. Tripathi, P., et al., Sudden Sensorineural Hearing Loss: A Review. Cureus, 2022. 14(9): p. e29458

[9]. Zhou, T., et al., GAN review: Models and medical image fusion applications. Information Fusion, 2023. 91: p. 134-148

[10]. van der Velden, B.H.M., et al., Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal, 2022. 79: p. 102470

[11]. Alzubaidi, L., et al., Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data, 2021. 8(1): p. 53

[12]. Krizhevsky, A., et al., ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017. 60(6): p. 84-90

[13]. Tang, L., et al. Hu Moment Invariant: A New Method for Hearing Loss Detection. in International Conference Advanced Engineering & Technology Research. 2018.

[14]. Gao, R., et al. Hearing loss identification by wavelet entropy and cat swarm optimization. in ADVANCES IN MATERIALS, MACHINERY, ELECTRONICS III: 3rd International Conference on Advances in Materials, Machinery, Electronics (AMME 2019). 2019.

[15]. Wang, L., et al. Hearing Loss Identification via Fractional Fourier Entropy and Direct Acyclic Graph Support Vector Machine. in International Conference on Multimedia Technology and Enhanced Learning. 2020.

[16]. Yao, X., et al. Hearing loss classification via stationary wavelet entropy and genetic algorithm. in 2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC). 2020.

[17]. Yao, C., Hearing loss classification via stationary wavelet entropy and cat swarm optimization. Cognitive Systems and Signal Processing in Image Processing, 2022: p. 203-221

[18]. Zhang, Y., Detection of unilateral hearing loss by Stationary Wavelet Entropy. CNS & Neurological Disorders - Drug Targets, 2017. 16(2): p. 15-24

[19]. Zainal, A.G., et al., Recognition of Copy Move Forgeries in Digital Images using Hybrid Optimization and Convolutional Neural Network Algorithm. International Journal of Advanced Computer Science and Applications, 2022. 13(12): p. 301-311

[20]. Le Gratiet, B., et al., Deployment of convolutional neural network solutions for image computing in semiconductor manufacturing environment. Journal of Micro-Nanopatterning Materials and Metrology-Jm3, 2022. 21(4)

[21]. Alsharabi, N., et al., Implementing Magnetic Resonance Imaging Brain Disorder Classification via AlexNet-Quantum Learning. Mathematics, 2023. 11(2): Article ID. 376

[22]. Guo, Y., et al., Radar Moving Target Detection Method Based on SET2 and AlexNet. Mathematical Problems in Engineering, 2022. 2022: Article ID. 3359871

[23]. Zhang, Y.-D., High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. Journal of Medical Imaging and Health Informatics, 2019. 9(9): p. 2012-2021

[24]. Sethy, P.K., et al., Lung cancer histopathological image classification using wavelets and AlexNet. Journal of X-Ray Science and Technology, 2023. 31(1): p. 211-221

[25]. Yee, N.L., et al., Apex Frame Spotting Using Attention Networks for Micro-Expression Recognition System. Cmc-Computers Materials & Continua, 2022. 73(3): p. 5331-5348

[26]. Kuo, C.F.J., et al., Automatic detection, classification and localization of defects in large photovoltaic plants using unmanned aerial vehicles (UAV) based infrared (IR) and RGB imaging. Energy Conversion and Management, 2023. 276: Article ID. 116495

[27]. Tripathi, S., et al., Denoising of magnetic resonance images using discriminative learning-based deep convolutional neural network. Technology and Health Care, 2022. 30(1): p. 145-160

[28]. Halle, S.D., et al., Bayesian dropout approximation in deep learning neural networks: analysis of self-aligned quadruple patterning. Journal of Micro-Nanopatterning Materials and Metrology-Jm3, 2022. 21(4): Article ID. 041604

[29]. Zhang, Y.-D., Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU. Journal of Computational Science, 2018. 28: p. 1-10

[30]. Saha, R.K., et al., Automated quantification of meibomian gland dropout in infrared meibography using deep learning. Ocular Surface, 2022. 26: p. 283-294

[31]. Gharekhani, M., et al., Quantifying the groundwater total contamination risk using an inclusive multi-level modelling strategy. Journal of Environmental Management, 2023. 332: Article ID. 117287

[32]. Fouad, I.A., A robust and efficient EEG-based drowsiness detection system using different machine learning algorithms. Ain Shams Engineering Journal, 2023. 14(3): Article ID. 101895

[33]. Behera, J., et al., Prediction based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets. Engineering Applications of Artificial Intelligence, 2023. 120: Article ID. 105843

[34]. Yan, Y., et al., Alcoholism via wavelet energy entropy and support vector machine, in Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing Companion. 2022, Association for Computing Machinery: Leicester, United Kingdom. p. Article 3.