*Computer Modeling in Engineering & Sciences*
DOI:10.32604/cmes.2020.011069
***Article***

Tech Science Press

# Least-Square Support Vector Machine and Wavelet Selection for Hearing Loss Identification

**Chaosheng Tang[1], Deepak Ranjan Nayak[2] and Shuihua Wang[1,3,4,*]**

[1]School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, 454000, China
[2]Department of Computer Science and Engineering, Malaviya National Institute of Technology, Jaipur, 302017, India
[3]School of Mathematics and Actuarial Science, University of Leicester, Leicester, LE1 7RH, UK
[4]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
*Corresponding Author: Shuihua Wang. Email: shuihuawang@ieee.org
Received: 17 April 2020; Accepted: 30 June 2020

**Abstract:** Hearing loss (HL) is a kind of common illness, which can significantly reduce the quality of life. For example, HL often results in mishearing, misunderstanding, and communication problems. Therefore, it is necessary to provide early diagnosis and timely treatment for HL. This study investigated the advantages and disadvantages of three classical machine learning methods: multilayer perceptron (MLP), support vector machine (SVM), and least-square support vector machine (LS-SVM) approach and made a further optimization of the LS-SVM model via wavelet entropy. The investigation illustrated that the multilayer perceptron is a shallow neural network, while the least square support vector machine uses hinge loss function and least-square optimization method. Besides, a wavelet selection method was proposed, and we found db4 can achieve the best results. The experiments showed that the LS-SVM method can identify the hearing loss disease with an overall accuracy of three classes as $84.89 \pm 1.77$, which is superior to SVM and MLP. The results show that the least-square support vector machine is effective in hearing loss identification.

**Keywords:** Hearing loss; wavelet entropy; multilayer perceptron; least square support vector machine

## 1 Introduction

Hearing loss (HL) is a kind of prevalent chronic condition affecting people. 46% of adults aged 48–87 years had a hearing loss [1]. Exchange of information with others, an important aspect of everyday life, can be seriously impaired in individuals with hearing loss. These difficulties with communication could lead to a perceived reduction in quality of life [2]. As life expectancy increases and older adults are living longer, an increasing number of individuals will be forced to endure hearing loss during their senior years. The influence of HL to the patients is various, such as speech development, language acquisition, etc. More importantly, it often results in physical and mental pain to humans, as well as serious harm and huge economic losses [3]. Statistical

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

results indicate that about $1.1 million social costs will be lost for every untreated HL person. On the contrary, if the HL can be detected in an early stage, these costs could be decreased by 75%.

Generally, HL can be attributed to many causes. For example, it can be caused by autosomal dominant, X-linked, and mitochondrial mutations, and can also be caused by simple Mendelian inheritance. Some studies proposed that HL is a result of genetics or the environment, such as a dangerous source of radiation, even resulted from many factors and their interaction [4].

HL has aroused extensive interest in the study. Some research suggests that HL derived from physical or emotional injure. The results of numerous analytical measurements demonstrated that most hearing loss is Sensorineural hearing loss (SNHL). SNHL not only triggers of HL, but also severely affecting brain functions, such as lower intelligence, motor proficiency and speech and language delay, etc.

Although HL is a serious health problem worldwide and its socio-economic consequences are globally significant, there are no effective measures to overcome it. However, with the development of medical science and advances in biotechnology, recent researches are focusing on the problem of prevention or early identification for the HL [5].

In the present studies, the researchers investigated the SNHL by sMRI, fMRI, and DTI. Profant et al. [6] studied the Age-related hearing loss problem by MR morphometry and diffusion tensor imaging (DTI) method, the result indicated that HL does not play a significant role in the auditory system in elderly subjects. Nayeem [7] used wavelet entropy (WE) and genetic algorithm (GA) to identify hearing loss patients. Gao et al. [8] employed cat swarm optimization (CSO) approach.

This paper aims to identify the hearing loss only from structure MRI images [9]. We did not utilize fMRI nor DTI. Our method still got promising performances, which validate that hearing loss may alter the brain structure. In addition, image processing [10–12], computer vision [13], and texture analysis [14–16] techniques were employed.

In this study, our contribution focus on the following three folds: (i) The db4 wavelet has proven to be the best results in three orientations and overall accuracy, compared with 4 different Wavelet Daubechies Function. (ii) Three different algorithms: MLP, SVM, and LS-SVM, has been analyzed by $10 \times 6$-fold cross validation. The results showed that the LS-SVM model we proposed has performed the best for identifying the hearing loss disease. (iii) Through comparing with state-of-the-art approaches, our LS-SVM algorithm performed better than WE + GA and CSO methods.

The paper is organized as follows: Section 2 will summarize an overview of the methods including wavelet entropy, multilayer perceptron, and least squares support vector machine. Section 3 shows the dataset and experiments. Section 4 is devoted to analyzing the experimental results. In Section 5, we conclude the paper with the direction of further research.

## 2 Methodology

We used wavelet entropy as feature extractors [17], and meanwhile, we investigated several artificial intelligence methods: wavelet transform, multilayer perceptron (MLP), support vector machine (SVM), and least square support vector machine (LS-SVM). Deep learning methods [18–21] and transfer learning [22–25] are not used in our method, due to its requirement of a relatively large dataset while our dataset is especially small.

### *2.1 Wavelet Daubechies Function*

Wavelet analysis is a special type of linear transformation of functions of a rather wide class. The basis of eigenfunctions, on which the decomposition is carried out, has many special properties. The correct application of these properties allows researchers to focus on those or other features of the analyzed process which cannot be identified by the traditionally used Fourier and Laplace transformations.

The basic idea of wavelet analysis is to represent or approximate a signal or function with a family of the function called wavelet function system, which is structured by the translation and telescopic of a basic wavelet function, and of which the conversion coefficient can be used to describe the original signal. The base of the wavelet transform is given by

$$\psi_{\alpha,\beta}(\xi) = |\alpha|^{-\frac{1}{2}}\psi((\xi-\beta)/\alpha)\alpha, \quad \beta \in P, \quad \alpha \neq 0 \tag{1}$$

where $\psi(x)$ is the basic wavelet function, $a$ and $b$ are the scaling and translation factors. The wavelet transform about the function $f(t)$ is defined as follows

$$W_f(a,b) = f^*\psi_{a,b}(x) = \frac{1}{\sqrt{a}}\int_R f(x)\,\overline{\psi}\left(\frac{x-b}{a}\right)dt \tag{2}$$

where $f(t) \in L^2(R)$ and * denotes convolution. It is observed from (2) that $W_f(a,b)$ have contributions from the scaling factor $a$ and the space position $b$ as a function of the wavelet transform. The wavelet transform can be used to analyze signal by the function $\psi(x)$ stretches in scale and translates in the spatial domain (time domain).

The name of the Daubechies family can be written as dbN, where N is the order (or associated filter) and db is the short name of the wavelet. For $N \in \mathbb{N}$, the Daubechies wavelet transform of class $D-2N$ is a function $\psi = N\psi(x) \in L^2(\mathbb{R})$,

$$\psi(x) := \sqrt{2}\sum_{k=0}^{2N-1}(-1)^k h_{2N-1-k}\varphi(2x-k) \tag{3}$$

where $h_0, \ldots, h_{2N-1} \in \mathbb{R}$ and those are the filter constant coefficients that satisfy the conditions below:

$$\sum_{k=0}^{N-1} h_{2k} = \frac{1}{\sqrt{2}} = \sum_{k=0}^{N-1} h_{2k+1} \tag{4}$$

where $\varphi = N\varphi\colon \mathbb{R} \to \mathbb{R}$ is the scaling (Daubechies) function given by the recursion [26]:

$$\varphi(x)\colon = \sqrt{2}\sum_{k=0}^{2N-1} h_k\varphi(2x-k) \tag{5}$$

The dbN function is compactly supported standard orthogonal wavelet, it appears makes the discrete wavelet analysis becomes possible. In this paper, we use Daubechies 4 wavelet (db4).

### 2.2 Entropy

The coefficients of wavelet decomposition subbands are further extracted by the entropy method. Wavelet Entropy (WE) is a powerful tool to study the transient behavior of the waves. In the beginning, it was proposed to analyze the wavelet sub-bands distribution, and then extends to wide scientific areas.

Usually, the entropy is assumed to be a time series system, and provides a crucial idea to settle the question that for the data contain complex features. Considering a random variable $x \in R^n$, the entropy can be defined as in Eq. (1)

$$S = \int_0^{\infty} -(p)\, log\,(p)\, dx \tag{6}$$

where $p(.)$ refers to the probability density function (PDF) [27].

### 2.3 Multilayer Perceptron

Multi-layer neural networks [28] comprise an input layer, a hidden layer, and an output layer. It can have several output units, where the output units of the hidden layer function as the input units for the next layer. However, multiple layers of linear units still only produce linear functions [29–31].

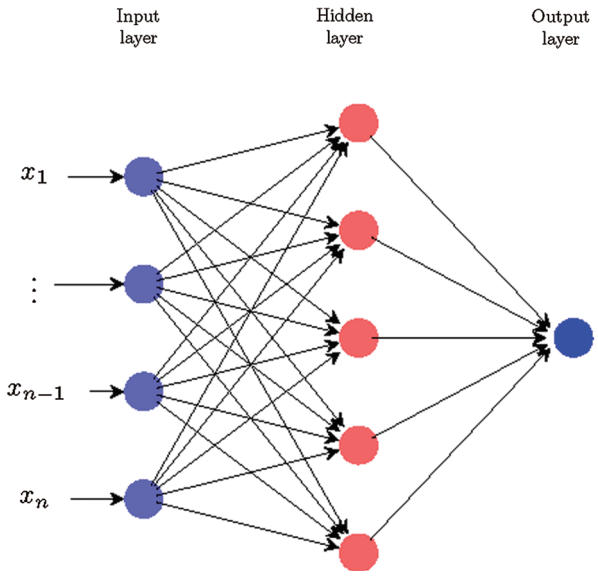


**Figure 1:** Structure of multilayer perceptron

Using a step function in a perceptron is another choice, but it is not differentiable, and thus it is not suitable for gradient descent search. The solution is a sigmoid function, which is a non-linear, differentiable threshold function. Fig. 1 shows the structure of the multilayer perceptron. It comprises an input layer, a hidden node, and an output layer. It can have several output units, where the output units of the hidden layer function as the input units for the next layer. However, multiple layers of linear units still only produce linear functions [32]. Using a step function in a perceptron is another choice, but it is not differentiable, and thus it is not suitable for gradient descent search. The solution is a sigmoid function, which is a nonlinear, differentiable threshold

function. There are many methods to train the MLP, e.g., the cat swarm optimization [33], particle swarm optimization, Jaya algorithm [34], artificial bee colony [35], biogeography-based optimization, bat algorithm [36], etc. In this study, we used plain back-propagation [37] for simple.

When we train deep models, it must be computationally tractable. To achieve this goal, it is extremely important to introduce a back-propagation algorithm. It can update the network parameters during training the neural network model. And more notably, there are a lot of hyper-parameters in the neural network model. We must make full use of professional knowledge and trial and error to find the appropriate super parameters. It is an ideal method to automate the tuning process. In this paper, we use a grid search to find an optimal number of hidden neurons.

### *2.4 Support Vector Machine*

In this paper, except using a multilayer perceptron, we also test the least squares support vector machine (LS-SVM). The LS-SVM is a variant of standard SVM. Support vector machine (SVM) is a machine learning tool that uses statistical learning theory to solve multi-dimensional functions. It is based on structural risk minimization principles, which overcomes the extra-learning problem of neural networks. Traditional SVM gets the solutions with optimal quadratic functions. In the process of the optimal solution, the dimension of the matrix is directly related to the number of training samples, and it is feasible to use inner products to solve the medium-scale optimal solution.

Given a set of training data like $\{(x_1, y_1), \ldots, (x_l, y_l)\}$ with $x_l \in R$ and $y_l \in [-1, 1]$. Nonlinear SVM classifiers use the kernel trick to produce nonlinear boundaries. The decision function given by an SVM is

$$y = \sum_{l=1}^{m} \phi_l g_l(x) + b \tag{7}$$

With $\phi \in R^N$, $g(\cdot) \in R^N \to R^M$, $M \to \infty$, $b \in R$, where $b$ is the weight vector and b is the bias.

In the N-dimension space, we can introduce slack variables $\lambda$, $\lambda_i'$ $(i = 1, \ldots, n)$ and formulate the optimization problem of SVM as below:

$$\min \frac{1}{2} \|\phi\|^2 + C \sum_{i=1}^{n} (\lambda_i + \lambda_i')$$

$$s.t. \begin{cases} y_i - f(\boldsymbol{x}_i, \phi) \le \delta + \lambda_i' \\ f(\boldsymbol{x}_i, \phi) - y_i \le \delta + \lambda_i \\ \lambda_i, \lambda_i' \ge 0, \quad i = 1, \ldots, n \end{cases} \tag{8}$$

### *2.5 LS-SVM*

For estimating least square support vector machine (LS-SVM) [38], we have comprehensively considered the complexity of function and fitting error, and express the constrained optimization

problem according to the structural risk minimization principle:

$$\sum_{w,b,e} \min J(w,e) = \frac{1}{2}\left(w^T w + \gamma \sum_{i=1}^{l} e_i^2\right) \tag{9}$$

$$s.t., \quad y_i = W^T \varphi(x_i) + b + e_i \quad i = 1,2,\ldots,l \tag{10}$$

where $\gamma$ is the margin parameter; and $e_i$ is the slack variable for $x_i$.

In order to solve the optimization problems in Eqs. (10) and (11), by changing the constrained problem into an unconstrained problem and introducing the Lagrange multipliers $\alpha_i$, we obtain the objective function:

$$L(w,b,e,\alpha) = J(w,e) - \sum_{k=1}^{N} \alpha \left\{ W^T \varphi(x_i) + b + e_i - y_i \right\} \tag{11}$$

According to the optimal solution of Karush–Kuhn–Tucker (KKT) conditions, take the partial derivatives of (12) with respect to $w$, $b$, $e$ and $\alpha$, respectively, and let them be zero, we obtain the optimal conditions as follows:

$$L(w,\ b,\ e,\ \alpha) = J(w,\ e) - \sum_{k=1}^{N} \alpha \left\{ W^T \varphi(x_i) + b + e_i - y_i \right\} \tag{12}$$

$$w = \sum_{i=1}^{l} \alpha_i \varphi(x_i), \sum_{i=1}^{l} \alpha_i = 0, \alpha_i = \gamma e_i \tag{13}$$

So, the following linear equations are obtained:

$$\begin{bmatrix} 0 & -l_{u\times l} \\ l_{l\times u} & \varphi(x_i)\varphi(x_i)^T + \dfrac{I}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{14}$$

where $l_{l\times u} = [l, l, \ldots, l] = l_{u\times l}^T$, $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_l]$, which is satisfied with Mercer's condition, the LS-SVM can be obtained as:

$$f(x) = \sum_{i=1}^{l} \alpha_i \varphi(x_i)\varphi(x_i)^T + b \tag{15}$$

Comparing the Eqs. (9) and (13), we can conclude that LS-SVM is a special case of SVM, which reorganized SVM by equation and non-negative slack variables. In addition, a least-squares loss function has been introduced into LS-SVM, which can map the original sample from a higher dimensional linearly separable space [39]. An obvious gain from LS-SVM can be concluded that the LS-SVM model can correspond to solve the linear equations rather than the QP problem.

### *2.6 Implementation*

It can be seen that the pipeline of the entire system from the flowchart of the proposed method shown in Fig. 2. The structured data of the brain data have been passed to wavelet selection and wavelet entropy, which were trained by db4 wavelet methods. The processed data will be calculated through $10 \times 6$-fold cross validation by different algorithms and classification results will be compared in the next step. The statistical results will indicate the different performances of the classification model including MLP, SVM and LS-SVM we presented.

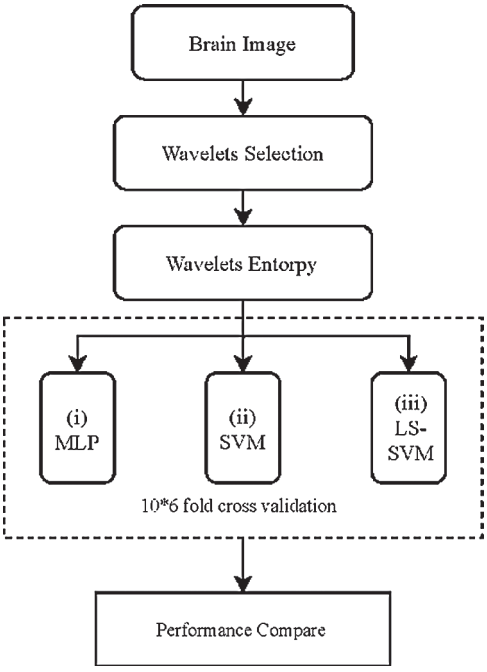


**Figure 2:** Flowchart of the proposed method

## 3 Dataset and Experiment

Our study consisted of 180 subjects: 60 healthy control (HC) subjects, 60 left-sided hearing loss (LHL) patients, and 60 right-sided hearing loss (RHL) patients. This study was approved by the Ethics Committee of local participating hospitals, and a signed informed consent form was obtained from every subject prior to entering this study.

The inclusion criterion was moderate-to-severe sudden sensorineural unilateral hearing loss. The exclusion criteria for all participants were known as neurological or psychiatric diseases, brain lesions such as tumors or strokes, taking psychotropic medications, and contraindications to MR imaging. This study was approved by the Ethics Committee of local participating hospitals, and a signed informed consent form was obtained from every subject prior to entering this study. Fig. 3 shows one illustration of RHL patients. This image is the result of preprocessing and transforming the data. We use the FMRIB software library to analyze the MRI brain image data of the patient. In order to ensure the scientificalness of the image processing, experienced otologists have been invited to extract pathologic features.

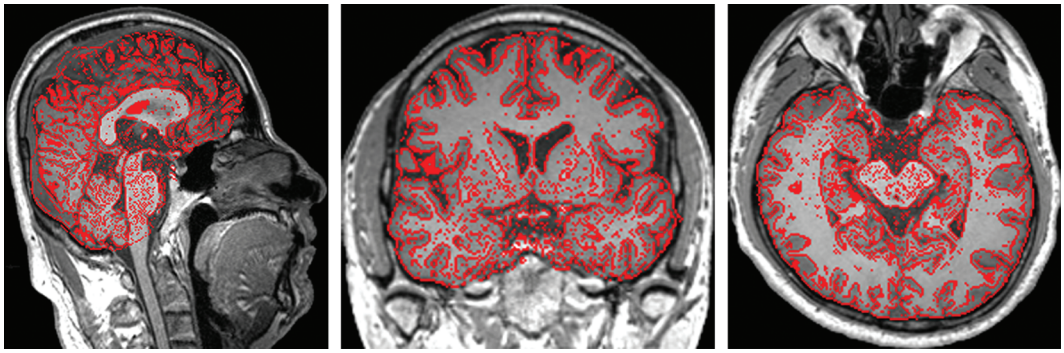

**Figure 3:** Illustration of one RHL patient

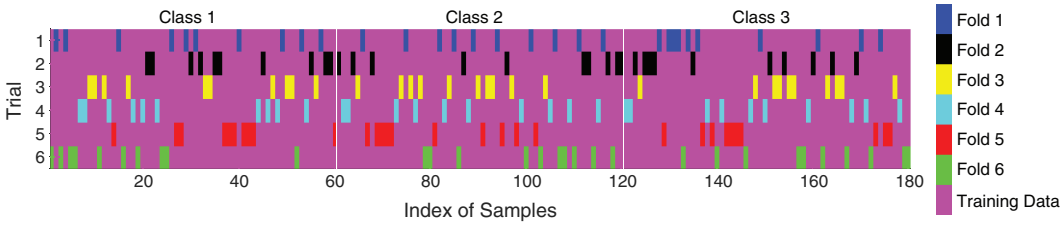


**Figure 4:** Index of stratified cross validation

The experiments were performed over a 6-fold cross validation approach. Each fold contains 10 HCs, 10 LHLs, and 10 RHLs. We ran the 6-fold cross validation 10 times, and the results were analyzed in terms of individual sensitivity and overall accuracy. Fig. 4 gives the illustration of indexes of one 6-fold stratified cross validation. During the experiments, the samples have been divided into 6 groups, and 5 groups samples have been chosen as training set, the rest samples as test set. The process has been performed 10 times, and finally the best optimal results of the model can be calculated. It can avoid some local minima and the fitting phenomenon.

## 4 Results and Discussions

Our experiment is based on Windows 10 and Matlab 2018a. The CPU is Intel@Core$^{TM}$ i7, and the GPU is Geforce 750M whose cuda version is 9.1.

### *4.1 MLP vs. SVM*

The results of using MLP and SVM are listed in Tabs. 1 and 2, where C1, C2, C3 represent LHL, RHL and HL. On the whole, the experiment results through $10 \times 6$-fold stratified cross validation has shown that the overall accuracy of these two algorithms is higher than 80%. It can be concluded that SVM obtains better results than MLP, because the overall accuracy of SVM is $81.22 \pm 2.03$, which is better than that of MLP. The prime reason is that, SVM could be seen as a single hidden layer neural network, it can allow kernel function to convert the datasets with nonlinear problem into a linearly separable datasets in kernel space. In addition, the classification accuracy of MLP highly dependent on various parameters, such as the learning rate, the structure of hidden layer and the number of nodes. SVM, by contrast, can obtain exact solutions by adjusting a few parameters. In our experiment, all the parameters were chosen via the trial-and-error method.

**Table 1:** Results of $10 \times 6$-fold stratified cross validation using MLP

| Run | C1 | C2 | C3 | Overall |
|---|---|---|---|---|
| 1 | 81.67 | 78.33 | 85.00 | 81.67 |
| 2 | 80.00 | 81.67 | 86.67 | 82.78 |
| 3 | 86.67 | 73.33 | 81.67 | 80.56 |
| 4 | 81.67 | 80.00 | 83.33 | 81.67 |
| 5 | 81.67 | 76.67 | 78.33 | 78.89 |
| 6 | 76.67 | 88.33 | 70.00 | 78.33 |
| 7 | 81.67 | 85.00 | 76.67 | 81.11 |
| 8 | 76.67 | 81.67 | 80.00 | 79.44 |
| 9 | 78.33 | 80.00 | 80.00 | 79.44 |
| 10 | 80.00 | 78.33 | 86.67 | 81.67 |
| Total | $80.50 \pm 2.94$ | $80.33 \pm 4.22$ | $80.83 \pm 5.11$ | $80.56 \pm 1.46$ |

**Table 2:** Results of $10 \times 6$-fold stratified cross validation using SVM

| Run | C1 | C2 | C3 | Overall |
|---|---|---|---|---|
| 1 | 86.67 | 83.33 | 80.00 | 83.33 |
| 2 | 83.33 | 78.33 | 80.00 | 80.56 |
| 3 | 83.33 | 78.33 | 80.00 | 80.56 |
| 4 | 80.00 | 83.33 | 80.00 | 81.11 |
| 5 | 76.67 | 85.00 | 75.00 | 78.89 |
| 6 | 78.33 | 85.00 | 83.33 | 82.22 |
| 7 | 80.00 | 80.00 | 83.33 | 81.11 |
| 8 | 78.33 | 81.67 | 78.33 | 79.44 |
| 9 | 81.67 | 76.67 | 80.00 | 79.44 |
| 10 | 86.67 | 85.00 | 85.00 | 85.56 |
| Total | $81.50 \pm 3.46$ | $81.67 \pm 3.14$ | $80.50 \pm 2.84$ | $81.22 \pm 2.03$ |

### *4.2 LS-SVM*

Finally, we applied our proposed LS-SVM to this task, and the statistical results were shown in Tab. 3. Our LS-SVM method achieved an overall accuracy of three classes as $84.89 \pm 1.77$, higher than MLP and SVM approaches. Comparing with SVM, LS-SVM starts from the loss function of machine learning, and optimizes the objective function by L2-norm. More importantly, LS-SVM replaces the inequality constraint of the SVM algorithm with an equality constraint, which can obtain the solution by Karush–Kuhn–Tucker (KKT) conditions. Fig. 5 shows the comparison between MLP, SVM, and LS-SVM approaches.

In this study, we proved LS-SVM may perform the best for identifying the hearing loss disease. In the future, we shall try to test LS-SVM on a larger HL dataset, meanwhile on other diseases, such as lung/kidney lesions, Alzheimer's disease, multiple sclerosis, alcoholism, tumor identification, etc.

**Table 3:** Results of $10 \times 6$-fold stratified cross validation using our method of LS-SVM

| Run | C1 | C2 | C3 | Overall |
|---|---|---|---|---|
| 1 | 88.33 | 80.00 | 90.00 | 86.11 |
| 2 | 81.67 | 88.33 | 83.33 | 84.44 |
| 3 | 81.67 | 88.33 | 81.67 | 83.89 |
| 4 | 85.00 | 88.33 | 85.00 | 86.11 |
| 5 | 85.00 | 88.33 | 88.33 | 87.22 |
| 6 | 85.00 | 88.33 | 85.00 | 86.11 |
| 7 | 88.33 | 71.67 | 88.33 | 82.78 |
| 8 | 81.67 | 83.33 | 81.67 | 82.22 |
| 9 | 86.67 | 88.33 | 85.00 | 86.67 |
| 10 | 85.00 | 86.67 | 78.33 | 83.33 |
| Total | $84.83 \pm 2.54$ | $85.17 \pm 5.52$ | $84.67 \pm 3.58$ | $84.89 \pm 1.77$ |

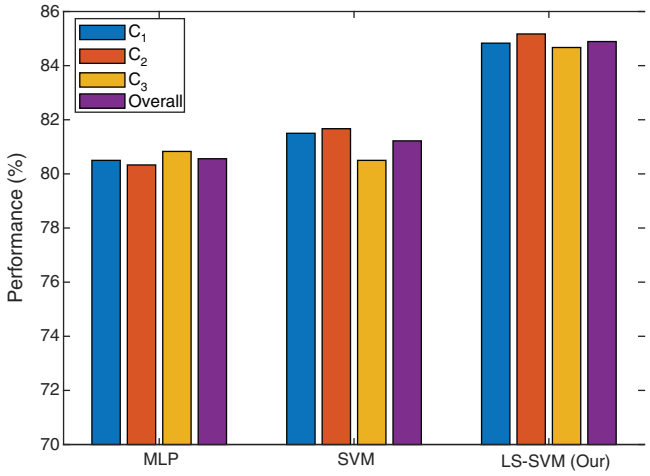


**Figure 5:** Algorithm comparison

**Table 4:** Wavelet selection among dbN families

| Wavelet | C1 | C2 | C3 | Overall |
|---|---|---|---|---|
| db2 | $79.50 \pm 4.78$ | $80.17 \pm 4.68$ | $79.67 \pm 4.57$ | $79.78 \pm 0.88$ |
| db3 | $82.50 \pm 2.75$ | $81.33 \pm 2.70$ | $82.33 \pm 2.63$ | $82.06 \pm 1.05$ |
| db4 (Ours) | $\mathbf{84.83 \pm 2.54}$ | $\mathbf{85.17 \pm 5.52}$ | $\mathbf{84.67 \pm 3.58}$ | $\mathbf{84.89 \pm 1.77}$ |
| db5 | $83.17 \pm 2.28$ | $82.83 \pm 2.49$ | $83.67 \pm 2.92$ | $83.22 \pm 1.01$ |

### *4.3 Wavelet Selection*

In this experiment, we compared the performance of db2, db3, db4, and db5. The results are shown in Tab. 4 and Fig. 6. Wavelet transform is a great tool for image processing [40,41]. It can be observed that the db4 wavelet we proposed has the best effect among the four methods. The accuracy of the db4 is $84.83 \pm 2.54$, $85.17 \pm 5.52$, $84.67 \pm 3.58$, corresponding to C1, C2

and C3, and the Overall accuracy is $84.89 \pm 1.77$. The experiment results demonstrate that db4 has the best effect both in three orientations and Overall accuracy. It can be observed that the best performance can be represented by the vertical axis. The error bar of wavelet selection clearly shows that the db4 wavelet stabilized around 85% which has the best performance in all orientations. The main reason may be that, db4 method is more conducive to extract medical image information, which is beneficial to categorize patients.

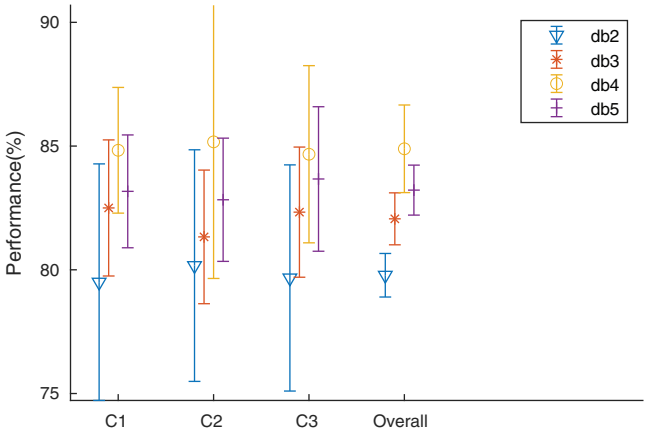


**Figure 6:** Error bar of wavelet selection

**Table 5:** Comparison of SOTA approaches

| Method | C1 | C2 | C3 | Overall |
|---|---|---|---|---|
| WE + GA [6] | $81.25 \pm 4.91$ | $80.42 \pm 5.57$ | $81.67 \pm 6.86$ | $81.11 \pm 1.34$ |
| CSO [7] | $\mathbf{85.50 \pm 6.85}$ | $84.50 \pm 4.97$ | $83.50 \pm 5.80$ | $84.50 \pm 0.81$ |
| LS-SVM (Ours) | $84.83 \pm 2.54$ | $\mathbf{85.17 \pm 5.52}$ | $\mathbf{84.67 \pm 3.58}$ | $\mathbf{84.89 \pm 1.77}$ |

Bold means the best.

### *4.4 Comparison of State-of-the-Art Approaches*

Finally, we compared our LS-SVM algorithm with three state-of-the-art approaches. The results are detailed in Tab. 5 and Fig. 7. Genetic Algorithm (GA) is an efficient parallel algorithm, which is well known as the strong robustness and broad applicability. Cat Swarm Optimization (CSO) is a novel optimization method based on the hunting strategy of cats, which is composed of seeking mode and tracing mode. Comparing with WE + GA and CSO methods proposed by Nayeem [6] and Gao [7] respectively, our LS-SVM algorithm performance was the best in C2, C3, and overall accuracy. This is mainly because LS-SVM has excellent nonlinear mapping ability at high-dimensional space transformation. It can be seen that CSO performed better in the first column of Tab. 5, one reason may be that the CSO is effective for global search by a few parameters. To a certain extent, it can overcome the genetic algorithm premature convergence problem.

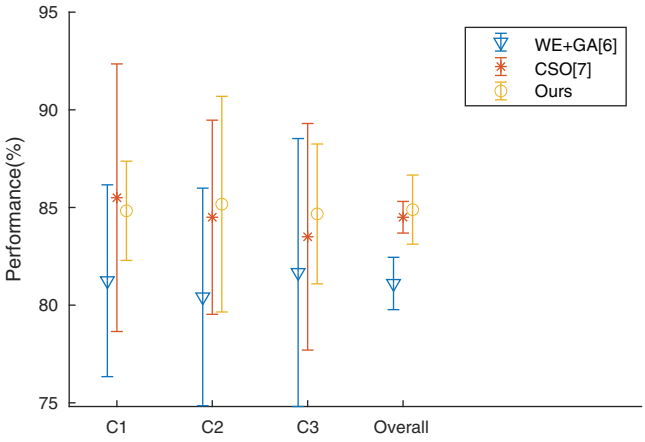


**Figure 7:** Error bar of algorithm comparison

## 5 Conclusions

This study investigated two classical machine learning methods: MLP and LS-SVM approach, to the task of hearing loss identification via wavelet entropy. Our results show that the least-square support vector machine performs better than plain support vector machine and multilayer perceptron approaches for our task.

In the future, we shall try to collect more data of hearing loss subjects, and tested the performance of using deep learning methods, such as transfer learning via classification deep neural network, e.g., AlexNet or DenseNet, both of which can be used as the based model for transfer learning algorithm to identify hearing loss. We can also employ augmentation techniques to increases the amount of training data to improve classification accuracy in test data.

**Data Availability:** The datasets used in this study can be obtained by contacting the first author.

**Funding Statement:** This research was supported by grants from the Ph.D. Programs Foundation of Henan Polytechnic University (B2016-38).

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

1. Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E., Klein, R. et al. (1998). Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The epidemiology of hearing loss study. *American Journal of Epidemiology, 148(9),* 879–886.
2. Jamesdaniel, S., Rosati, R., Westrick, J., Ruden, D. M. (2018). Chronic lead exposure induces cochlear oxidative stress and potentiates noise-induced hearing loss. *Toxicology Letters, 292,* 175–180. DOI 10.1016/j.toxlet.2018.05.004.
3. Hubbard, H. I., Mamo, S. K., Hopper, T. (2018). Dementia and hearing loss: interrelationships and treatment considerations. *Seminars in Speech and Language, 39(3),* 197–210. DOI 10.1055/s-0038-1660779.
4. Muniak, M. A., Ayeni, F. E., Ryugo, D. K. (2018). Hidden hearing loss and endbulbs of held: evidence for central pathology before detection of ABR threshold increases. *Hearing Research, 364,* 104–117. DOI 10.1016/j.heares.2018.03.021.

5. Jakubíková, J., Kabátová, Z., Pavlovcinová, G., Profant, M. (2009). Newborn hearing screening and strategy for early detection of hearing loss in infants. *International Journal of Pediatric Otorhinolaryngology, 73(4),* 607–612. DOI 10.1016/j.ijporl.2008.12.006.
6. Profant, O., Škoch, A., Balogová, Z., Tintěra, J., Hlinka, J. et al. (2014). Diffusion tensor imaging and MR morphometry of the central auditory pathway and auditory cortex in aging. *Neuroscience, 260,* 87–97. DOI 10.1016/j.neuroscience.2013.12.010.
7. Liu, F., Nayeem, A., Pereira, A. (2017). Hearing loss detection based on wavelet entropy and genetic algorithm. *Advances in Intelligent Systems Research, 153,* 49–53.
8. Gao, R., Liu, J. (2019). Hearing loss identification by wavelet entropy and cat swarm optimization. *AIP Conference Proceedings,* 2073.
9. Lipschitz, N., Kohlberg, G. D., Scott, M., Greinwald, J. H. (2020). Imaging findings in pediatric single-sided deafness and asymmetric hearing loss. *Laryngoscope, 130(4),* 1007–1010. DOI 10.1002/lary.28095.
10. Zhang, Y., Wu, L. (2008). Improved image filter based on SPCNN. *Science in China Series F: Information Sciences, 51(12),* 2115–2125. DOI 10.1007/s11432-008-0124-z.
11. Zhang, Y., Wu, L. (2009). Segment-based coding of color images. *Science in China Series F: Information Sciences, 52(6),* 914–925. DOI 10.1007/s11432-009-0019-7.
12. Zhang, Y., Wu, L., Wang, S., Wei, G. (2010). Color image enhancement based on HVS and PCNN. *Science China Information Sciences, 53(10),* 1963–1976. DOI 10.1007/s11432-010-4075-9.
13. Herzig, D., Nakas, C. T., Stalder, J., Kosinski, C., Laesser, C. et al. (2020). Volumetric food quantification using computer vision on a depth-sensing smartphone: preclinical study. *JMIR mHealth and uHealth, 8(3),* 11.
14. Mao, X., Ding, L., Zhang, Y., Zhan, R., Li, S. (2019). Knowledge-aided 2-D autofocus for spotlight SAR filtered backprojection imagery. *IEEE Transactions on Geoscience and Remote Sensing, 57(11),* 9041–9058. DOI 10.1109/TGRS.2019.2924221.
15. Zhang, Y. D., Jiang, Y., Zhu, W., Lu, S., Zhao, G. (2018). Exploring a smart pathological brain detection method on pseudo Zernike moment. *Multimedia Tools and Applications, 77(17),* 22589–22604. DOI 10.1007/s11042-017-4703-0.
16. Kong, F., Govindaraj, V. V., Zhang, Y. D. (2018). Ridge-based curvilinear structure detection for identifying road in remote sensing image and backbone in neuron dendrite image. *Multimedia Tools and Applications, 77(17),* 22857–22873. DOI 10.1007/s11042-018-5976-7.
17. Lotfy, H. M., Monir, H. H., Erk, N., Rostom, Y. (2020). Novel feature extraction approach for achieving potential spectral resolution: green analytical application on zofenopril calcium and hydrochlorothiazide in their spectrally overlapping binary mixture. *Spectrochimica Acta. Part A, Molecular and Biomolecular Spectroscopy, 230,* 13. DOI 10.1016/j.saa.2019.117998.
18. Zhang, Y. D., Pan, C., Chen, X., Wang, F. (2018). Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling. *Journal of Computational Science, 27,* 57–68. DOI 10.1016/j.jocs.2018.05.005.
19. Zhang, Y. D., Muhammad, K., Tang, C. (2018). Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform. *Multimedia Tools and Applications, 77(17),* 22821–22839. DOI 10.1007/s11042-018-5765-3.
20. Jiang, X., Chang, L., Zhang, Y. D. (2020). Classification of Alzheimer's disease via eight-layer convolutional neural network with batch normalization and dropout techniques. *Journal of Medical Imaging and Health Informatics, 10(5),* 1040–1048. DOI 10.1166/jmihi.2020.3001.
21. Jiang, X., Zhang, Y. D. (2019). Chinese sign language fingerspelling via six-layer convolutional neural network with leaky rectified linear units for therapy and rehabilitation. *Journal of Medical Imaging and Health Informatics, 9(9),* 2031–2090. DOI 10.1166/jmihi.2019.2804.
22. Lu, S., Lu, Z., Zhang, Y. D. (2019). Pathological brain detection based on AlexNet and transfer learning. *Journal of Computational Science, 30,* 41–47. DOI 10.1016/j.jocs.2018.11.008.
23. Hong, J., Cheng, H., Zhang, Y. D., Liu, J. (2019). Detecting cerebral microbleeds with transfer learning. *Machine Vision and Applications, 30(7),* 1123–1133. DOI 10.1007/s00138-019-01029-5.

24. Yu, X., Zeng, N., Liu, S., Zhang, Y. D. (2019). Utilization of DenseNet201 for diagnosis of breast abnormality. *Machine Vision and Applications, 30(7),* 1135–1144. DOI 10.1007/s00138-019-01042-8.
25. Zhang, Y. D., Govindaraj, V. V., Tang, C., Zhu, W., Sun, J. (2019). High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. *Journal of Medical Imaging and Health Informatics, 9(9),* 2012–2021. DOI 10.1166/jmihi.2019.2692.
26. Medina-Daza, R. J., Parra, N. E. V., Upegui, E. (2017). Wavelet Daubechies (db4) transform assessment for WorldView-2 images fusion. *Journal of Computers, 12,* 301–308. DOI 10.17706/jcp.12.4.301-308.
27. Tabaraki, R., Nateghi, A. (2018). Removal of methylene blue, malachite green and rhodamine b in a ternary system by pistachio hull; application of wavelet neural network modeling and Doehlert design. *Analytical and Bioanalytical Chemistry Research, 5(1),* 143–157.
28. Khishe, M., Mosavi, M. R., Moridi, A. (2018). Chaotic fractal walk trainer for sonar data set classification using multi-layer perceptron neural network and its hardware implementation. *Applied Acoustics, 137,* 121–139. DOI 10.1016/j.apacoust.2018.03.012.
29. Zhang, Y. D., Wu, L. (2008). Weights optimization of neural network via improved BCO approach. *Progress in Electromagnetics Research, 83,* 185–198. DOI 10.2528/PIER08051403.
30. Zhang, Y., Dong, Z., Wu, L., Wang, S. (2011). A hybrid method for MRI brain image classification. *Expert Systems with Applications, 38(8),* 10049–10053. DOI 10.1016/j.eswa.2011.02.012.
31. Lu, Z., Lu, S., Liu, G., Zhang, Y., Yang, J. et al. (2016). A pathological brain detection system based on radial basis function neural network. *Journal of Medical Imaging and Health Informatics, 6(5),* 1218–1222. DOI 10.1166/jmihi.2016.1901.
32. Nandakumar, H., Mallick, S. P., Srivastava, S. (2020). Sensing high frequency sub-nanometer vibrations using optical coherence tomography with real-time profilometry of multiple inner layers. *Optics and Lasers in Engineering, 127,* 105992. DOI 10.1016/j.optlaseng.2019.105992.
33. Zhang, Y. D., Sui, Y., Sun, J., Zhao, G., Qian, P. (2018). Cat swarm optimization applied to alcohol use disorder identification. *Multimedia Tools and Applications, 77(17),* 22875–22896. DOI 10.1007/s11042-018-6003-8.
34. Zhang, Y. D., Zhao, G., Sun, J., Wu, X., Wang, Z. H. et al. (2018). Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm. *Multimedia Tools and Applications, 77(17),* 22629–22648. DOI 10.1007/s11042-017-5023-0.
35. Yang, J., Jiang, Q., Wang, L., Liu, S., Zhang, Y. D. et al. (2019). An adaptive encoding learning for artificial bee colony algorithms. *Journal of Computational Science, 30,* 11–27. DOI 10.1016/j.jocs.2018.11.001.
36. Lu, S., Qiu, X., Shi, J., Li, N., Lu, Z. H. et al. (2017). A pathological brain detection system based on extreme learning machine optimized by bat algorithm. *CNS & Neurological Disorders Drug Targets, 16(1),* 23–29. DOI 10.2174/1871527315666161019153259.
37. Zhang, Y. D., Sun, J. (2018). Preliminary study on angiosperm genus classification by weight decay and combination of most abundant color index with fractional Fourier entropy. *Multimedia Tools and Applications, 77(17),* 22671–22688. DOI 10.1007/s11042-017-5146-3.
38. Zavala-Ortiz, D. A., Ebel, B., Li, M. Y., Barradas-Dermitz, D. M., Hayward-Jones, P. M. et al. (2020). Support vector and locally weighted regressions to monitor monoclonal antibody glycosylation during CHO cell culture processes, an enhanced alternative to partial least squares regression. *Biochemical Engineering Journal, 154,* 107457. DOI 10.1016/j.bej.2019.107457.
39. Asadi, A., Alarifi, I. M., Nguyen, H. M., Moayedi, H. (2020). Feasibility of least-square support vector machine in predicting the effects of shear rate on the rheological properties and pumping power of MWCNT-MgO/oil hybrid nanofluid based on experimental data. *Journal of Thermal Analysis and Calorimetry, 16,* 279. DOI 10.1007/s10973-020-09279-6.

40. Sun, J., Wu, X. (2016). Infrared target recognition based on improved joint local ternary pattern. *Optical Engineering, 55(5),* 053101. DOI 10.1117/1.OE.55.5.053101.
41. Karaca, Y., Aslan, Z., Cattani, C., Galletta, D., Zhang, Y. (2016). Rank determination of mental functions by 1D wavelets and partial correlation. *Journal of Medical Systems, 41(1),* 2. DOI 10.1007/s10916-016-0606-2.