Hyperparameter Tuning of the Shunt-murmur Discrimination Algorithm Using Bayesian Optimization

Fumiya Noda* and Keisuke Nishijima* and Ken'ichi Furuya*
* Oita University, Oita, Japan
E-mail: v19e3015@oita-u.ac.jp

***Abstract*—Patients undergoing hemodialysis generally have shunts implanted in their bodies; a number of other problems, such as vascular stenosis, can be encountered. Patients undergoing hemodialysis can inspect the effective functioning of their shunts by listening to the shunt murmur. However, this manual inspection is difficult and requires experience. In this paper, we propose a method of exploring the hyperparameters of the shunt-murmur discrimination algorithm using Bayesian optimization. The resistance index(RI) obtained from the ultrasound system is used as a class label. The normalized cross-correlation coefficients, Mel frequency cepstrum coefficients, and frequency power percentage were the features to be trained by a random forest (RF). Bayesian optimization was used to explore the hyperparameters of the RF, achieving a significant accuracy improvement.**

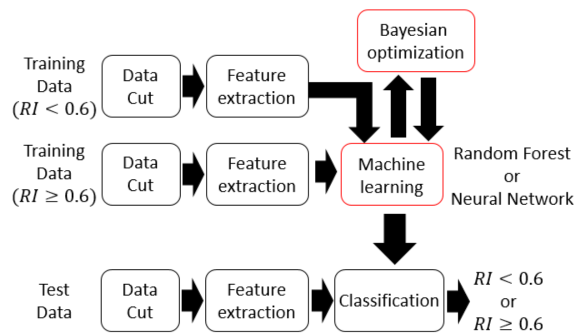

Fig. 1. Flow of processing

I. Introduction

Patients with kidney diseases such as kidney failure undergo hemodialysis to eliminate waste products and excess fluid from their blood. In this process, an arteriovenous fistula (AVF), called a shunt, is created together with an anastomosis between an artery and a vein. However, problems such as stenosis and blockage can occur owing to factors such as aging of patients, and prolonged use of a shunt. If the shunt's function is interfered with, patients cannot undergo hemodialysis unless they undergo reoperation. Therefore patients must routinely inspect the shunt for effective functioning. Although listening to a shunt murmur can be used to investigate a functioning shunt, relevant knowledge and experience are required to make appropriate conclusions. Therefore, it is desirable to provide a system that automatically determine a functioning shunt. From Murakami [3], the number of patients with shunt dysfunction increases if the resistance index (RI) value exceeds 0.6. Murakami further described the increasing trend observed.

In [1][2], discrimination experiments were conducted using SVM and random forest (RF) based on features and RI values from frequency analysis of shunt murmurs. An RI value is used as a class label, particularly, classes with less than, and greater than or equal to 0.6 RI values. RI is an indicator of poor blood flow to the periphery and is measured using an ultrasound device. However, owing to low discrimination accuracy, it cannot be used in a real environment. An additional problem is that, the hyperparameter search of the identification algorithm is time-consuming.

In this study, we use Bayesian optimization to perform hyperparameter search of shunt murmur's discrimination algorithm. Furthermore, we propose a method for identifying shunt states. RF and neural network (NN) are used as shunt murmur's identification algorithms. Because Bayesian optimization uses a surrogate model to perform the search efficiently. We assumed that this can reduce the search time compared to the grid search used in conventional methods. To examine its effectiveness, we performed a cross-test and evaluated it in terms of discrimination accuracy and search time. The data and features used in the experiments are the same as those used in existing studies[1].

II. Proposed Methods

In this study, we use frequency analysis of shunt murmurs and RI values to automatically identify shunt stenosis using RF and NN. We propose a method to search for RF and NN hyperparameters using Bayesian optimization. The processes involved is shown in Fig.1. RI is a measure of the difficulty of blood flow to the periphery and is expressed by (1).

$$RI = \frac{PSV - EDV}{PSV} \tag{1}$$

where PSV indicates the maximum systolic blood flow and EDV indicates end-diastolic blood flow.

A. Bayesian Optimization

Bayesian optimization is a method to create a proxy model for predicting the objective function using Gaussian process and Tree-structured Parzen Estimators (TPE). It selects the candidate search point with the highest expectation as the next search point using the acquisition function. In this study, we used TPE[4] as a proxy model and expected improvement (EI) as an acquisition function. TPE models $p(x|y)$ and is composed of two distributions divided by a threshold y^*.

$$p(x|y) = \begin{cases} l(x) & (y < y^*) \\ g(x) & (y \geqq y^*) \end{cases} \tag{2}$$

y^* is given prior by the quantile γ.

$$p(y < y^*) = \gamma$$

The $l(x)$ and $g(x)$ are distributions of y being large and small, respectively. These probability density functions are estimated by kernel density estimation from sampled points. If minimizing the evaluated value, $l(x)$ and $g(x)$ are distributions of parameters x that are less and greater than some loss, respectively. EI is defined as

$$EI_{y^*}(x) = (\gamma + \frac{g(x)}{l(x)}(1-\gamma))^{-1} \tag{3}$$

To minimize the evaluated value, the acquisition function selects a parameter x that has a higher probability of reducing the loss than the threshold y^*. This method of searching for the next parameter x to calculate the loss is used in TPE. For search efficiency, this method enables finding optimal solutions with few search than grid search.

B. Random Forest

An RF [5] is a machine learning algorithm that combines multiple decision trees to estimate a class. It is considered as a type of ensemble method because it combines multiple decision trees to build a significantly powerful model. A major problem with decision trees is that they overfit the training data. In RF, the degree of overfitting can be reduced by creating multiple decision trees that overfit in different directions and averaging the results. The name random forest originates from the introduction of random numbers in the process of building the decision tree to distinguish decision trees. In creating each decision tree, bootstrap sampling is performed. This will enable duplicates from the total training data and randomly extract data to develop a training data set. This results in developing decision trees of random forests for different datasets. In addition, because features are randomly extracted to create a decision tree, it is feasible to investigate their importance.

C. Neural Network

A neural network[6] is a mathematical model of the neurons and their relationships in human brains. In neural networks, the elementary units that model neurons in human brains are called artificial neurons. Neural networks are also called multilayer perceptrons. An example is shown in Fig.2. The leftmost, rightmost and middle columns are called the input, output, and middle or hidden layers, respectively. From the input layer to the output layer, Layer 0, Layer 1, and Layer 2 in order, so Fig.2 is called a three-layer network. A perceptron that receives two input signals x_1, x_2 and outputs y can be expressed as (4) and (5). Where b is the bias and w is the weight.

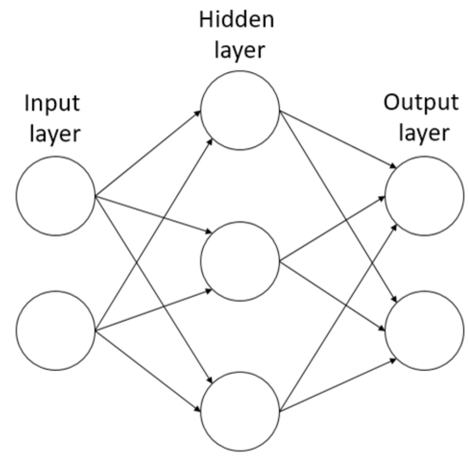

Fig. 2. Example of neural network

$$y = h(b + w_1x_1 + w_2x_2) \tag{4}$$

$$h(x) = \begin{cases} 0 (x \leq 0) \\ 1 (x > 0) \end{cases} \tag{5}$$

The $h(x)$ in (5) is called the activation function. Decomposing (4), we obtain (6) and (7).

$$a = b + w_1x_1 + w_2x_2 \tag{6}$$

$$y = h(a) \tag{7}$$

In the perceptron, (6) is the activation function. The NN learns by updating (optimizing) the weights (w) reduce the value of the loss function.

III. FEATURE EXTRACTION

A. MFCC

Mel frequency cepstrum coefficients (MFCC) is a low-dimensional spectral information defined in the Keflency region. To determine MFCC, a filter called Mel-filter bank is defined and multiplied by the spectra to obtain a low-dimensional overview of the spectrum. The Mel filter bank is a filter that is fine in the low frequency range and coarse in the high frequency range on the Mel scale. The Mell scale[7] used to obtain the Mel-filter bank is given by

$$Mel(f) = 2595 \log 10(1 + \frac{f}{700}) \tag{8}$$

The rough form of the spectrum obtained by multiplying the Mel-filter bank with the spectra is transformed to the Keflency

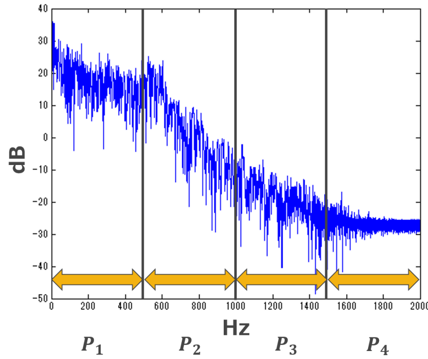

Fig. 3. Ratio of frequency power

region by a discrete cosine transform, and the resulting low-dimensional feature is called MFCC. In this study, up to 16 MFCC dimensions were used as features.

B. Ratio of Frequency Power

A stenosis shunt murmurs is characterized by its frequency. Therefore, the Fourier transform, that is a frequency analysis method was used to extract the features. Initially, Fourier transform is applied to the shunt murmur signal to obtain the frequency spectrum. Thereafter, we partition the frequency power for each band from 1-2,000 Hz into four parts at 500 Hz each.

$$p_1 = \sum_{f=1}^{500} \{20 \log 10(\mathrm{abs}(X(f)))\} \tag{9}$$

$$p_2 = \sum_{f=500}^{1000} \{20 \log 10(\mathrm{abs}(X(f)))\} \tag{10}$$

$$p_3 = \sum_{f=1000}^{1500} \{20 \log 10(\mathrm{abs}(X(f)))\} \tag{11}$$

$$p_4 = \sum_{f=1500}^{2000} \{20 \log 10(\mathrm{abs}(X(f)))\} \tag{12}$$

The sum of the values obtained from (9) to (12) is used to calculate the sum of the frequency power from 1-2,000 Hz.

$$P_{total} = p_1 + p_2 + p_3 + p_4 \tag{13}$$

Finally, we calculate the ratio for each band against the total sum calculated in (13).

$$P_1 = \frac{p_1}{P_{total}}, P_2 = \frac{p_2}{P_{total}}, P_3 = \frac{p_3}{P_{total}}, P_4 = \frac{p_4}{P_{total}} \tag{14}$$

This process is shown in Fig.3.

C. Normalized Cross-Correlation Coefficient

Normalized cross-correlation coefficients[8] can be obtained by the normalized cross-correlation analysis method. It shows the strength of cross-correlation between two images and is an effective method to investigate the independence and similarity between systems. Herein, the normalized cross-correlation coefficients are obtained by treating the time-frequency results obtained from the wavelet transform as images. The wavelet transform[9] is a technique for time-frequency analysis by shifting and scaling small waves called wavelets.
The function $\psi(t)$, whose mean is 0 and is localized around the origin $t = 0$, is called a wavelet. $\psi(t)$ is shifted or scaled on the t-axis to generate the basis $\psi_{a,b}(t)$.

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \tag{15}$$

where, a is the parameter of scaling, called scale, and b is called shift. The inner product of $\psi_{a,b}$ and the signal $f(t)$ is the wavelet transform.
In this study, the normalized cross-correlation coefficients were obtained by comparing them with the shunt sounds of two patients with low RI values. The normalized cross-correlation coefficients are calculated by (16).

$$R = \frac{f(t,\omega) - \overline{f(t,\omega)} \times (g(t,\omega) - \overline{g(t,\omega)})}{\sqrt{(f(t,\omega) - \overline{f(t,\omega)})^2 \times (g(t,\omega) - \overline{g(t,\omega)})^2}} \tag{16}$$

where t denotes time, ω denotes frequency, and $\overline{f(t,\omega)}$ and $\overline{g(t,\omega)}$ denote the average luminance. Furthermore, $f(t,\omega)$ and $g(t,\omega)$ are given by (17) and (18).

$$f(t,\omega) = \frac{1}{\sqrt{a}} \int f_1(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \tag{17}$$

$$g(t,\omega) = \frac{1}{\sqrt{a}} \int f_2(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \tag{18}$$

where $\overline{\psi(-)}$ is the complex conjugate of $\psi(-)$.

IV. EXPERIMENT

Bayesian optimization is used to search for RF and NN hyperparameters. We aim to examine the effectiveness of the proposed method by comparing the execution time and the percentage of correct answers.

A. Experimental conditions and methods

Shunt murmurs used for identification is the shunt murmurs presented in [1]. The shunt murmurs at the anastomosis of 60 patients with AVF was used for identification (RI value less than 0.6 in 30 patients and 30 with a RI of 0.6 or higher). Five 0.8-second data were extracted from each person from the recorded shunt murmurs. A total of 300 data was used. A fifth-order cross-test was conducted using 80% of the data as the training data and 20% as the test data, and the percentage of correct answers was used as the evaluation index. The experimental conditions are shown in TableI. The percentage of correct answers is determined using (19).

$$Accuracy = \frac{The\ number\ of\ correct\ answer}{The\ total\ number\ of\ data} \tag{19}$$

Normalized cross-correlation coefficients, percentage of frequency power, and MFCC were used as training features. The training and test data were obtained from a different individual.

TABLE I
EXPERIMENTAL CONDITION

Sampling frequency	48 kHz (microphone)
Data length	0.8 sec
Number of test data	60
Number of training data	240
Feature	MFCC Ratio of frequency power Normalized cross correlation coefficient
Classification algorithm	Random Forest Neural Network

TABLE II
RF PARAMETERS

Parameters	Range
Number of decision trees	1–22
Maximum depth of the decision trees	1–22
random state	1–999
Number of searches	1000

To compare the execution time, the hyperparameters were searched by grid search and Bayesian optimization under the conditions in TableI and TableII. Subsequently, the range of hyperparameters was extended and discrimination experiments were performed for RF and NN, respectively. The extended hyperparameters are shown in TableIII and TableIV.

B. Experimental Results

To compare the execution time, the hyperparameter search is performed using grid search and Bayesian optimization under the conditions of TableI and TableII, and the results are shown in TableV. Although from TableV the hyperparameter search with grid search was prolonged, the optimal solution can be determined. In the hyperparameter search with Bayesian optimization, although the search time was significantly reduced, it did not determine the optimal solution. However, because we obtained a high percentage of correct answers in less time, it is considered as an effective method.

Thereafter, we extended the range of hyperparameters to discriminate between RF and NN, and the results are shown in TableVI. We used Bayesian optimization to search for RF hyperparameters and obtained a high percentage of correct answers in a small search time. From the NN, additional hyperparameter search is necessary for further studies. The results of the correctness of the search are shown in Fig.4.

V. DISCUSSION

From the results in TableV, although Bayesian optimization reduced the search time, it did not determine the optimal solution. We can be attributed to wide range of random numbers. Therefore we conducted the experiment with a narrow range of random numbers as shown in TableIII. Extending the range of RF hyperparameters by using Bayesian optimization can improve the discrimination accuracy with reduced search time. However, the result of NN hyperparameter search using Bayesian optimization is worse than RF discrimination accuracy. This can be owing to insufficient range of hyperparameters to be searched. Further investigation of the discrimination algorithm is required to improve its accuracy.

TABLE III
RF PARAMETERS(RANGE EXPANSION)

Parameters	Range
Number of decision trees	1–100
Maximum depth of the decision trees	1–10
random state	580, 916
Split criterion	gini, entropy
Minimum number of samples required to split an internal node	2–9
Minimum number of samples required to be at a leaf node	1–9
Number of features to consider when looking for the best split	$sqrt$, $log2$
Number of searches	1000

TABLE IV
NN PARAMETERS

Parameters	Range
Batch size	6
Epoch number	10
Optimization algorithm	Adam
Number of hidden layers	1–3
Number of units	4–128
Activation function	relu
Dropout rate	0.2–0.5
Number of searches	1000

VI. CONCLUSIONS

In this study, to automatically identifying shunt stenosis, we proposed a method to examine shunts using Bayesian optimization by performing RF and NN hyperparameter searches. Although expanding the range of the RF hyperparameters using Bayesian optimization improves the discrimination accuracy, the search time is reduced. However, the result of NN hyperparameter search with Bayesian optimization is worse than the discrimination accuracy of RF. This indicates that shunt stenosis can be identified in a reduced search time if the range of hyperparameters is set appropriately. Because the current discrimination accuracy is difficult to use clinically, it is necessary to examine the discrimination algorithm and features to improve the accuracy in the further studies.

ACKNOWLEDGMENT

This research and development work was supported by JSPS KAKENHI grant numbers 19K12044, 18K11377 and the MIC/SCOPE #192110002.

TABLE V
RF RESULT

	RF(Grid search)	RF(Bayesian Optimization)
Time	About 3 days	About 24 min
Accuracy	71.3%	66.7%

TABLE VI
BAYESIAN OPTIMIZATION RESULT

	RF	NN
Time	About21min	About 60 min
Accuracy	73.0%	67.0%

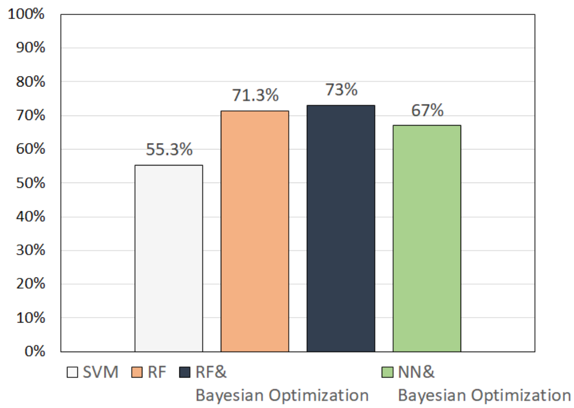

Fig. 4. Result of calculating accuracy

REFERENCES

[1] D.Higashi *et al.*, "Classification of Arteriovenous Fistula Stenosis Using Shunt Murmurs Analysis and Support Vector Machine", CISIS2018, pp.884-892, 2018.
[2] F.Noda *et al.*, "Classification of Arteriovenous Fistula Stenosis Using Shunt Murmur Analysis and Random Forest ", CISIS2019, pp.723-732, 2019.
[3] K.Murakami *et al.*, "Effectiveness of ultrasonic pulse Doppler method in shunt Management",Kidney and Dialysis, pp.39-43, 2003.
[4] J.Bergstra *et al.*, "Algorithms for Hyper-Parameter Optimization", NIPS'11:Proceedings of the 24th International Conference on Neural Information Processing Systems, pp2546-2554, 2011.
[5] Andreas C. Muller,Sarah Guido *et al.*, "Introduction to Machine Learning with Python: A Guide for Data Scientists",O'Reilly Media, pp.82-97, 2016.
[6] K.Saito, "Deep Learning from scratch-Python theory of deep learning and its implementation", O'Reilly Japan, pp.39-122, 2016.
[7] S.Shikano *et al.*, "Speech recognition system", Ohmsha, Ltd., 2001.
[8] K.Sasaki *et al.*, "Functional assessment of vascular access based on time-frequency analysis of shunt murmurs", IEICE, IEICE Technical Report vol.114, no.54, pp.25-30, 2014.
[9] H.Nakano *et al.*, "Signal and Image processing by the wavelet", Kyouritsu publisher, 1999.