Research on Violin Audio Feature Recognition Based on Mel-Frequency Cepstral Coefficient-Based Feature Parameter Extraction

Ming Zeng,[1] Huahong Zeng[2*]
[1]Royal Birmingham Conservatoire, Birmingham, B4 7XR, United Kingdom
[2]College of Educational Science, Quanzhou Normal University, Quanzhou 362000, China
E-mail: z_huahong@outlook.com
[*]Corresponding author

Keywords: violin, Mel-frequency cepstral coefficient, feature parameter, support vector machine, recognition rate

Received: April 1, 2024

This paper focuses on the feature recognition of violin audio. After introducing the preprocessing method, the common feature parameters, linear predictive cepstral coefficient (LPCC) and mel-frequency cepstral coefficient (MFCC), were explained. Then, MFCC + △MFCC was used as the feature parameter. The parameters of support vector machine (SVM) were optimized using the firefly algorithm (FA). The FA-SVM method was used to recognize different violin audios. It was found that the identification rate of the FA-SVM approach was above 95% for different violin notes. The recognition effect was better when using MFCC + △MFCC as the feature parameter compared with LPCC and MFCC. The FA-SVM method achieved the highest recognition rate of 97.42%. The results demonstrate the reliability of the FA-SVM method based on MFCC feature parameter extraction. This method can be applied in practical audio recognition.

Povzetek: Raziskava se osredotoča na prepoznavanje zvočnih lastnosti violine z uporabo metod ekstrakcije parametrov Mel-frekvenčnega kepstralnega koeficienta (MFCC). Optimiziran algoritem podpore vektorskih strojev (SVM) je dosegel najboljše rezultate.

1 Introduction

More and more approaches have been applied in speech recognition [1]. While computer technology has progressed, electronic music has also been developed, and computers have been applied to assist in the creation of music [2]. If different types of music can be automatically recognized by computer, intelligent creation of music can be realized more conveniently. Music, like speech, consists of sound signals, so it is possible to recognize music signals using speech recognition technology. Compared with speech, music signals have fewer acoustic differences, so it is more feasible to classify and recognize them. Currently, many methods have been applied in audio recognition [3]. Table 1 presents a summary of related works.

Table 1: A summary table of related works

	Method	Result
Noor et al. [4]	Use time-domain, frequency-domain, and joint video features to detect road event audio.	The accuracy of the method was improved by 7% compared to the method combined with temporal and spectral features alone.
Cohen-McFarlane et al. [5]	AlexNet classifier	The method performed better in classifying octaves, with an accuracy of 98%.
Singh et al. [6]	A support vector machine (SVM) optimized by the grasshopper-ride optimization algorithm	The method achieved a maximum accuracy of 0.96, a minimum false alarm rate of 0.022, and a false rejection rate of 0.0119.
Elbir et al. [7]	A deep neural network model	The performance of the proposed model was enhanced in music genre classification, music similarity, and music recommendation.

Based on Table 1, it can be observed that research on audio recognition has become relatively mature. However, there is still limited research specifically focused on music audio, with most of the attention being given to multi-instrument audio recognition. In terms of single instrument note recognition, piano receives the most attention while other instruments are less studied. Violin is an essential string instrument in both symphony orchestras and as a challenging solo instrument. Therefore, conducting research on violin audio feature recognition holds significant practical value. This paper focused on the recognition of violin audio features, used a SVM for classification based on Mel-frequency cepstral coefficient (MFCC) features, and further enhanced the classification capability of the SVM using the firefly algorithm (FA). The authors verified the reliability of the method through experiments. This work provides theoretical support for the development of music processing and composition.

2 Pre-processing and MFCC feature parameter extraction

Before the feature parameter extraction, the violin audio signal needs to be pre-processed. The sampling rate is configured at 8 kHz. The precision is set as 16 bit. The format is wav. Then, it is pre-emphasized. The transfer function used is:

$H(z) = 1 - az^{-1}$ (1)

where a is the pre-emphasis coefficient. The value of a is approximately 1, ranging from 0.9 to 1, with a typical value of 0.94; therefore, this article also set it as 0.94.

After pre-emphasis, the signal is processed frame by frame by windowing to obtain a short-term stationary signal. The frame length is designated as N, the frame signal is set as $x(n)$, and the window function is $w(n)$. The signal $y(n)$ after windowing is written as:

$y(n) = x(n)w(n)$. (2)

Commonly used window functions [8] include:

① rectangular window: $w(n) = \begin{cases} 1, 0 \le n \le N-1 \\ 0, else \end{cases}$,

② Hanming window: $w(n) = \begin{cases} 0.5 - [1 - \cos(2\pi n/(N-1))], 0 \le n \le N-1 \\ 0, else \end{cases}$.

The rectangular window has good spectral smoothing performance; however, it tends to lose waveform details, while the Hamming window does not have this problem, so this paper uses the Hamming window to process the violin audio signal. The frame length is set as 32 ms, and the frame shift was set as 16 ms.

The most basic parameters of music signals include amplitude and frequency [9], but they perform poorly in signal identification. Currently, the commonly used feature parameters include MFCC, linear predictive coefficient (LPC), linear predictive cepstral coefficient (LPCC), etc. [10]. This paper mainly studied LPCC and MFCC.

LPCC is mainly used to represent the response of the sound track [11]. LPC can be acquired by minimizing the mean square error between the sampled value of the actual audio and the sampled value derived from linear prediction. For $s(n)$, its predicted value is written as:

$\bar{s}(n) = \sum_{i=1}^{t} a_i s(n-i)$, (3)

where a_i is the t-th order LPC. The forecasting error is:

$e(n) = s(n) - \bar{s}(n)$. (4)

The quadratic sum of all errors is written as: $E = \sum_{n=t}^{N-1} [s(n) - \sum_{i=1}^{t} a_i s_{n-i}]^2$. The LPC feature is obtained when the value of E is minimum. LPCC is obtained after cepstrum operation.

MFCC has extensive applications in speech recognition [12], and Figure 1 shows its extraction process. The signal after framing is processed by fast discrete Fourier transform [13] and then passes through a triangular filter to form a Mel frequency spectrum. After logarithmic energy processing and then discrete cosine transform (DCT) [14] to obtain a set of 13-dimensional Mel cepstrum coefficients. Then, the first-order differential cepstrum coefficient △MFCC is calculated based on MFCC. They are combined to obtain MFCC+△MFCC as the final characteristic parameter.

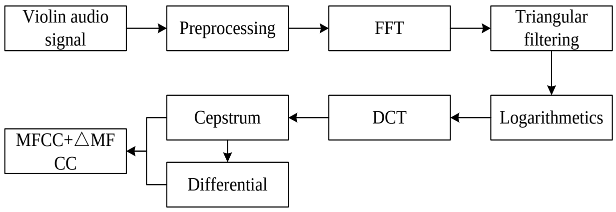

Figure 1: MFCC feature parameter extraction process.

The equation of FFT is: $X(k) = \sum_{n=0}^{N-1} x(n) e^{-\frac{2j\pi}{N}}$, where $k = 0,1,\cdots,N-1$, $x(n)$ refers to the preprocessed signal, and N is the window width.

The equation of Mel filter is: $Mel(f) = 2595lg\left(1 + \frac{f}{700}\right)$.

The equation of DCT is: $C_{mfcc}(i) = \sqrt{\frac{2}{N}} \sum_{l=1}^{L} \log m(l) \, cos\left\{\left(l - \frac{1}{2}\right)\frac{i\pi}{L}\right\}$,

where $m(l)$ is the energy of the l-th subband, N is the quantity of filters, and $C_{mfcc}(i)$ is the i-th MFCC. The first 12 MFCCs are selected.

As MFCC can reflect the static feature of signals, to obtain the dynamic feature of signals, the first-order difference parameter △MFCC is computed on the basis of $C_{mfcc}(i)$. The equation for the first-order differential cepstrum coefficient is:

$$\Delta C_{mfcc}(i) = \frac{1}{\sqrt{\sum_{l=-k}^{k} l^2}} \sum_{l=-k}^{k} l \, C_{mfcc}(i+n),$$

where k is a constant, 2 usually.

According to the results of $C_{mfcc}(i)$ and $\Delta C_{mfcc}(i)$, the finally extracted features are 12-dimensional MFCC and 12-dimensional △MFCC, totaling 24 dimensions.

3 SVM-based recognition method

An SVM can handle small-sample, nonlinear problems [15] and have good applications in text classification and speech recognition [16]. Assume there exists a set of samples that can be divided in a linear manner (x_i, y_i), $i =$

$1,2,\cdots,n$, $y \in \{+1,-1\}$. The classification hyperplane is expressed as: $wx+b=0$, where w and b are the weight and bias. To make the classification interval $\frac{2}{\|w\|}$ maximum, it needs to satisfy: $y_i(wx_i+b)-1 \geq 0$. The Lagrange function for this problem is expressed as:

$$L(w,b,a)=\frac{1}{2}(w\cdot w)-\sum_{i=1}^{n} a_i\{y_i(wx_i+b)-1\}, \quad (5)$$

where a_i is the Lagrange coefficient. The above equation is solved by making it zero, then $w-\sum_{i=1}^{l} a_i\, y_i x_i=0$, $\sum_{i=1}^{l} a_i\, y_i=0$. Thus, the original problem changes to a dual problem. The problem is solved as follows.

$$\max_{a} \sum_{i=1}^{n} a_i-\frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j\left(x_i\cdot x_j\right) \quad (6)$$

$$s.t. \sum_{i=1}^{l} a_i y_i=0, a_i \geq 0, i=1,2,\cdots,n \quad (7)$$

If the optimal solution is a_i^*, then $w^*=\sum_{i=1}^{n} a_i^* y_i x_i$. The final optimal classification function is:

$$f(x)=sgn\{\sum_{i=1}^{n} a_i^* y_i (x_i\cdot x)+b^*\}. \quad (8)$$

In the linearly inseparable case, slack variable ϑ and penalty factor C are added. Then, solving the optimal hyperplane problem is written as:

$$min\left(\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\vartheta_i\right), \quad (9)$$

$$s.t. y_i(wx_i+b) \geq 1-\vartheta_i, \vartheta_i \geq 0, i=1,2,\cdots,n. \quad (10)$$

The rest of the solution process is the same as the linear SVM. In an SVM, a kernel function is usually added to improve its performance. In this paper, a Gaussian kernel function is added: $K(x_i,x_j)=exp\left(-\frac{\|x_i-x_j\|^2}{\sigma^2}\right)$, where σ is the kernel function parameter. During the calculation, parameters C and σ of the SVM are usually determined based on empirical values, which significantly influences the recognition effect of the SVM; therefore, this paper selects the FA [17] to optimize parameters.

Suppose that in the D-dimensional space, the position of firefly i is: $x_i=(x_{i1},x_{i2},\cdots,x_{iD})$. The relative brightness of the firefly is defined as: $L=L_0\times e^{-\gamma\times r_{ij}^2}$, where L_0 is the original brightness of the firefly, γ is the light absorption coefficient, and r_{ij} is the fireflies' distance: $r_{ij}=\sum_{d=1}^{D}\left(x_{id}-x_{jd}\right)^2$. The degree of attraction between the fireflies is: $\beta_{ij}=\beta_0\times e^{-\gamma\times r_{ij}^2}$, where β_0 is the initial attraction degree of the fireflies. In the FA, the fireflies move toward the brighter individual to find the optimal solution. The position update formula is written as: $x_i^{t+1}=x_i^t+\beta\times\left|x_i-x_j\right|+\alpha\times\left(rand-\frac{1}{2}\right)$, where t is the number of iterations, α is the step length factor, and $rand$ is the random number in [0,1].

Finally, the flow of the FA-SVM method based on MFCC feature parameter extraction is shown in Figure 2. The extracted MFCC+ΔMFCC is used as the input of the SVM. Optimal parameters C and σ are obtained using the FA, Finally, the FA-SVM model is established to realize the recognition of violin audio.

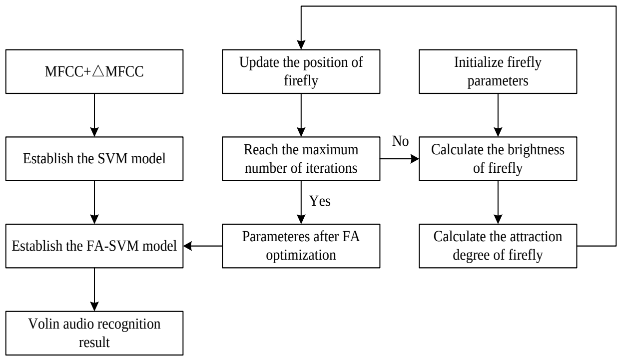

Figure 2: The FA-SVM method for violin audio feature recognition.

4 Results and analysis

A violin audio was recorded in a quiet laboratory room. The same performer recorded the audio at the same time every day using the same equipment to maintain sample stability. The recording lasted for 30 days, and 30 sets of samples were obtained. They were saved in wav format. MFCC feature parameter extraction was performed after pre-processing; then, they were recognized using the FA-SVM method. The C and σ intervals of the SVM were [0.01,100] and [0.01,100], respectively. The firefly population sizes were 10, 20, and 30. The count of iterations was 10, 20, and 30. The performance of the FA-SVM approach was assessed using five-fold cross-validation (Figure 3), and Table 2 displays the outcomes.

Round 1	Train	Train	Train	Train	Test
Round 2	Train	Train	Train	Test	Train
Round 3	Train	Train	Test	Train	Train
Round 4	Train	Test	Train	Train	Train
Round 5	Test	Train	Train	Train	Train

Figure 3: The flow of the five-fold cross-validation.

Table 1: Effect of population size and number of iterations on recognition rate.

[Population size, number of iterations]	Recognition rate/%
[10, 10]	97.36±2.12
[10, 20]	95.44±1.98
[10, 30]	93.27±2.01
[20, 10]	97.11±1.11
[20, 20]	**98.64**±2.31
[20, 30]	96.33±1.06

[30, 10]	96.77±2.25
[30, 20]	95.46±1.36
[30, 30]	94.36±1.45

According to Table 2, it was determined that when the firefly population size was 20 and the count of iterations was 20, the recognition rate for violin audio was the highest, 98.64%, and the optimal parameters obtained were $C = 86.77$, $\sigma = 91.26$.

Some notes were recognized by the FA-SVM method based on the MFCC feature parameter (MFCC + △MFCC), and the outcomes are displayed in Table 3.

Table 3: Recognition results of the FA-SVM method.

Name of note	Recognition rate/%
C	95.64±1.27
D	96.77±1.36
E	97.36±1.45
F	96.78±1.26
G	98.12±1.41
A	97.66±1.33
B	95.36±1.51
#C	96.77±1.07
#F	98.22±1.37
#G	97.33±1.28

It was observed from Table 3 that the identification rate of the FA-SVM approach for different notes was above 95%. The highest recognition rate was 98.22%, the lowest was 95.36%, and the average was 97%, indicating that the FA-SVM method based on the MFCC feature parameter could accurately recognize violin audio and distinguish different violin notes. To further examine the performance of the approach, LPCC, MFCC, and MFCC+△MFCC were compared, and the corresponding recognition rates are displayed in Figure 4.

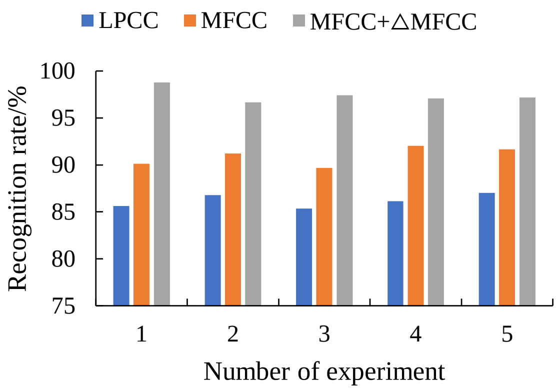

Figure 4: Recognition effect of the approach based on different feature parameters.

It was seen from Figure 4 that when LPCC was used as the feature parameter, the recognition rate of the algorithm for violin audio was the lowest, about 85%; the recognition rate of the algorithm was around 90% when MFCC was used as the feature parameter, which was about 5% higher, indicating that the MFCC feature parameter of audio was better than LPCC. The FA-SVM method achieved the highest recognition rate when utilizing MFCC + △MFCC as its input. The average value of the five experiments was calculated, and the results were 86.18%, 90.94% and 97.42%, respectively. The recognition rate of MFCC+△MFCC was 11.24% higher than LPCC and 6.48% higher than MFCC. This verified the reliability of MFCC+△MFCC.

Then, the performance of the hidden Markov model (HMM) [18], Gaussian mixture model (GMM) [19], back-propagation neural network (BPNN) model [20], SVM model, genetic algorithm (GA)-optimized SVM model [21], particle swarm optimization (PSO)-based SVM model [22], and FA-SVM model were compared on the basis of the MFCC + △MFCC feature, and the results were averaged, as displayed in Figure 5.

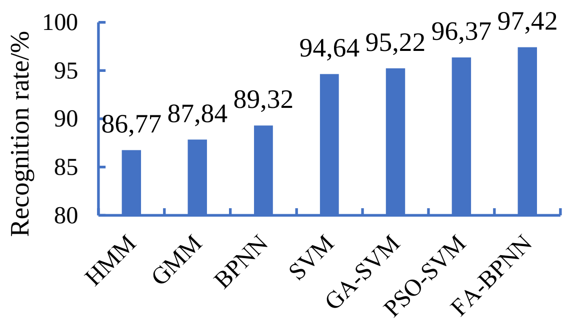

Figure 5: Comparison of recognition effects.

It was observed in Figure 5 that under the same feature parameter, the recognition rate of the HMM was the lowest, 86.77%, and the recognition rates of the GMM and BPNN were 87.84% and 89.32%, respectively, below 90%; the recognition rate of the SVM model was 94.64%, which was significantly higher than the previous methods, demonstrating the superiority of the SVM model in recognizing violin audio. Then, the FA-SVM model achieved a recognition rate of 97.42%, surpassing the HMM by 10.65% and outperforming the SVM model by 2.78%. Moreover, it was 2.2% higher than the GA-SVM model and 1.05% higher than the PSO-SVM model. These results verify that the FA-SVM model is reliable and accurate for recognizing violin audio notes.

Finally, the performance of different parts on violin note recognition were analyzed through ablation experiments, as shown in Table 4.

Table 4: The results of ablation experiments.

	Recognition rate/%
Baseline (SVM+MFCC)	88.36
Baseline+△MFCC	94.64(+6.31%)

Baseline+△MFCC+FA	97.42 (+2.78%)

From Table 4, it can be observed that firstly, in terms of feature extraction, the addition of △MFCC led to a 6.37% improvement in violin note recognition. The result demonstrated the impact of the selection of violin audio features on recognition performance and confirmed that the chosen △MFCC feature contained crucial information related to note recognition. Furthermore, after incorporating FA, there was an additional 2.78% improvement in recognition performance, highlighting the importance of parameter optimization by FA for achieving better results.

5 Discussion

With the increasingly widespread use of music in daily life, research on music is also deepening. Music retrieval, genre classification, and instrument recognition are all current focuses of research. However, in instrument recognition, there is currently more research on multi-instrument recognition, which involves classifying audio from multiple instruments. There is less research on single instrument recognition, with piano being the main focus and very little involvement in in-depth studies of violin audio. Therefore, this paper conducted a study specifically on violin audio recognition.

In order to address the shortcomings of MFCC in feature description, this paper combined MFCC and △MFCC as inputs for the recognition model. Then, in terms of selecting the recognition method, the SVM method with good performance in recognition classification was adopted, and its parameter selection was optimized by the FA algorithm. Experiments and analysis were conducted using violin audio collected in practical environments. It can be observed that the inclusion of △MFCC significantly improved the recognition rate of violin audio, showing high performance in recognizing different notes. Furthermore, compared to other existing methods, the designed FA-SVM method also demonstrated notable advantages. Both GA and PSO are parameter optimization methods, but in terms of SVM parameter optimization, FA has better performance with fast convergence and high accuracy. Therefore, it achieved good results in SVM parameter optimization as well. The ablation experiment results also confirmed the advantages of MFCC +△MFCC + FA as it achieved optimal recognition rates, suggesting the practicality of this method.

The research on violin audio feature recognition in this paper has yielded some results, providing an effective method for identifying different violin notes. This paper presents a new approach for instrument audio recognition and even speech recognition, while also providing theoretical support for optimizing SVM. The experimental results verify that the designed violin feature recognition method can be applied in practice, providing a more intelligent approach for teaching and learning the violin. It helps learners receive feedback faster and better understand the correctness of note performance. However, there are also limitations in this study, such as limited experimental samples and insufficient consideration of comprehensive features' impact on recognition results. In future work, the researchers will further expand the experimental samples to verify method performance on more comprehensive samples and consider additional features to discover more effective ones.

6 Conclusion

This paper studied the feature recognition of violin audio. MFCC and △MFCC of the audio were extracted as feature parameters. An SVM-based recognition method, namely the FA-SVM method, was designed. Experiments were conducted on the collected violin audio. It was found that the recognition rate of the algorithm was above 80% for various notes. The comparison of different feature parameters demonstrated that the recognition performance of MFCC+△MFCC significantly outperformed LPCC and MFCC. Among the different recognition methods, the FA-SVM method achieved the highest recognition rate, reaching 97.42%. The effectiveness of the MFCC+△MFCC feature-based FA-SVM approach was confirmed by the experimental results. The approach has the potential to be widely disseminated and implemented in real-world scenarios.

References

[1] Qazi KA, Nawaz T, Mehmood Z, Rashid M, Habib HA (2018). A hybrid technique for speech segregation and classification using a sophisticated deep neural network. *Plos One*, 13(3), pp. 1-15. https://doi.org/10.1371/journal.pone.0194151

[2] Liang L, Liu J (2021). An exploration of the application of computer music production software in music composition. *2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, IEEE, Dalian, China, pp. 794-796. https://doi.org/10.1109/IPEC51340.2021.9421093

[3] Scarpiniti M, Colasante F, Di Tanna S, Ciancia M, Lee YC, Uncini A (2021). Deep Belief Network based audio classification for construction sites monitoring. *Expert Systems with Applications*, 177(Sep.), pp. 114839.1-114839.14. https://doi.org/10.1016/j.eswa.2021.114839

[4] Almaadeed N, Asim M, Al-Maadeed S, Bouridane A, Beghdadi A (2018). Automatic Detection and Classification of Audio Events for Road Surveillance Applications. *Sensors*, 18(6), pp. 1-19. https://doi.org/10.3390/s18061858

[5] Cohen-McFarlane M, Goubran R, Wallace B (2020). Challenges with Audio Classification using Image Based Approaches for Health Measurement Applications. *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, IEEE, Bari, Italy, pp. 1-5. https://doi.org/10.1109/MeMeA49120.2020.9137254

[6] Singh S P, Jaiswal U C (2021). Audio classification using grasshopper-ride optimization algorithm-

based support vector machine. *IET Circuits Devices & Systems*, 15(5), pp. 434-447. https://doi.org/10.1049/cds2.12039

[7] Elbir A, Aydin N (2020). Music Genre Classification and Music Recommendation by Using Deep Learning. *Electronics Letters*, 56(12), pp. 627-629. https://doi.org/10.1049/el.2019.4202

[8] Cömert Z, Boopathi AM, Velappan S, Yang Z, Kocamaz AF (2018). The influences of different window functions and lengths on image-based time-frequency features of fetal heart rate signals. *2018 26th Signal Processing and Communications Applications Conference (SIU)*, IEEE, Izmir, Turkey, pp. 1-4. https://doi.org/10.1109/siu.2018.8404247

[9] Dimitrova-Grekow T, Klis A, Igras-Cybulska M (2019). Speech Emotion Recognition Based on Voice Fundamental Frequency. *Archives of Acoustics*, 44(2), pp. 277-286. https://doi.org/10.24425/aoa.2019.128491

[10] Uddin MA, Hossain MS, Pathan RK, Biswas M (2020). Gender Recognition from Human Voice using Multi-Layer Architecture. *2020 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, IEEE, Novi Sad, Serbia, pp. 1-7. https://doi.org/10.1109/INISTA49547.2020.9194654

[11] Wang Y B, Chang D G, Qin S R, Fan YH, Mu HB, Zhang GJ (2019). Separating Multi-Source Partial Discharge Signals Using Linear Prediction Analysis and Isolation Forest Algorithm. *IEEE Transactions on Instrumentation and Measurement*, 69(6 Pt.1), pp. 2734-2742. https://doi.org/10.1109/TIM.2019.2926688

[12] Wibawa I, Darmawan I (2021). Implementation of audio recognition using mel frequency cepstrum coefficient and dynamic time warping in wirama praharsini. *Journal of Physics: Conference Series*, 1722, pp. 1-8. https://doi.org/10.1088/1742-6596/1722/1/012014

[13] Delina M, Nurhusni P A (2021). Feature extraction of noise signal in motorcycle by Fast Fourier Transform. *Journal of Physics: Conference Series*, 1869(1), pp. 1-4. https://doi.org/10.1088/1742-6596/1869/1/012198

[14] Ratnadewi, Hangkawidjaja A D, Prijono A, Suherman J (2021). Discrete cosine transform and multi class support vector machines for classification cardiac atrial arrhythmia and cardiac normal. *Journal of Physics: Conference Series*, 1858(1), pp. 1-13. https://doi.org/10.1088/1742-6596/1858/1/012093

[15] Wang B, Shahzad M, Zhu X, Rehman KU, Uddin S (2020). A Non-linear Model Predictive Control Based on Grey-Wolf Optimization Using Least-Square Support Vector Machine for Product Concentration Control in L-Lysine Fermentation. *Sensors*, 20(11), pp. 1-17. https://doi.org/10.3390/s20113335

[16] Fitri E, Ardiansyah, Junaidi A, Prabowo R, Lumbanraja F R (2021). Abstract Classification Using Support Vector Machine Algorithm (Case Study: Abstract in a Computer Science Journal). *Journal of Physics: Conference Series*, 1751, pp. 1-12. https://doi.org/10.1088/1742-6596/1751/1/012042

[17] Melin P, Sánchez D, Monica J C, Castillo O (2021). Optimization using the firefly algorithm of ensemble neural networks with type-2 fuzzy integration for COVID-19 time series prediction. *Soft Computing*, (10), pp. 1-38. https://doi.org/10.1007/s00500-020-05549-5

[18] Tamaoka T, Midoh Y, Yamamoto K, Aritomi S, Tanigaki T, Nakamura M, Nakamae K, Kawasaki M, Murakami Y (2021). Denoising electron holograms using the wavelet hidden Markov model for phase retrieval—Applications to the phase-shifting method. *AIP Advances*, 11(2), pp. 025135. https://doi.org/10.1063/5.0027599

[19] Zhang H, Jin X, Wu QMJ, Wang Y, He Z, Yang Y (2018). Automatic Visual Detection System of Railway Surface Defects With Curvature Filter and Improved Gaussian Mixture Model. *IEEE Transactions on Instrumentation and Measurement*, 67(7), pp. 1593-1608. https://doi.org/10.1109/TIM.2018.2803830

[20] Zhou B, Yan D, Xiong L, Li Y, Zhang Y (2021). Distributed Photovoltaic Short-Term Power Prediction Based On Genetic Algorithm Optimized BP Neural Network. *IOP Conference Series: Earth and Environmental Science*, 692(2), pp. 1-7. https://doi.org/10.1088/1755-1315/692/2/022104

[21] Wang Y, Huo H (2019). Speech Recognition Based on Genetic Algorithm Optimized Support Vector Machine. *2019 6th International Conference on Systems and Informatics (ICSAI)*, Shanghai, China, pp. 439-444. https://doi.org/10.1109/ICSAI48974.2019.9010502.

[22] Kaur R, Sharma RK, Kumar P (2020). Speaker Classification with Support Vector Machine and Crossover-Based Particle Swarm Optimization. *International Journal of Pattern Recognition and Artificial Intelligence*, 34(10), pp. 2051010.1-2051010.17. https://doi.org/10.1142/S0218001420510106.