Received: October 12, 2022. Revised: November 14, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Feature Selection on Gammatone Cepstral Coefficients for Bird Voice Classification using Particle Swarm Optimization

Pulung Nurtantio Andono[1]* Guruh Fajar Shidik[1] Dwi Puji Prabowo[2]
Dzuha Hening Yanuarsari[2] Yuslena Sari[3] Ricardus Anggi Pramunendar[1]

[1]Department of Informatics Engineering, Faculty of Computer Science
Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[2]Department of Visual Communication Design, Faculty of Computer Science
Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[3]Department of Information Technology, Faculty of Engineering
Universitas Lambung Mangkurat, Banjarmasin 70123, Indonesia
**Corresponding author: pulung@dsn.dinus.ac.id*

Abstract: Indonesia is home to several endangered species. This paper suggests a way for automatically categorizing bird sound patterns. This experiment used publicly available bird data to obtain bird sound patterns. However, bird sounds also contain noise, and the features are inappropriate, so data processing is required to reduce noise and select valuable features. YAMNet sound classification network data processing and feature selection incorporating Particle Swarm Optimization (PSO) may be used to decrease noise in bird voice data and choose appropriate features. All steps are completed before and after the gammatone frequency cepstral coefficient (GFCC) approach is used as feature extraction. This is ended by the KNN classification method to get the classification performance results. The experimental findings demonstrate that our proposal offers a performance for bird identification recognition of 78.33%, which is 1.27% better than our prior study. This finding outperforms previous research that solely used the dimensionality reduction-based classification algorithm, without picking the most important features.

Keywords: Bird recognition, KNN, GFCC, PSO.

1. Introduction

Birds significantly affect human lives for stress recovery and as an attention restoration tool, also significantly as a warning signal for changes in hazardous weather and other situations [1]. However, birds are becoming increasingly vulnerable to extinction because of Illegal hunting and habitat change [2, 3]. According to the World Bank, 160 of the world's 4,584 bird species are endangered. According to the red list of the international union for conservation of nature (IUCN), 9% of Indonesia's 1,771 indigenous bird species are threatened by extinction. Thus, it could impact the long-term viability of ecotourism and birdwatching.

The government has implemented conservation measures in several sites to reduce the number of endangered bird species and preserve bird population habitats. The community and researchers must work together to learn about the species, classification, morphological traits, habitat, distribution, a threat to animal status, and how to conserve it [4].

Recognizing all types of birds is difficult due to a lack of knowledge, especially when done manually in the open and depending on the limited abilities of the human sense of hearing. However, it can be challenging to identify birds visually, particularly in a heavily forested jungle habitat. Therefore, identifying bird species based on sound may be preferable [1]. Limitations induce differences in grouping performance [5], and the researcher's task is to help to learn about recognizing all types of birds in their environment [6]. Several researchers have tried to do this. They are trying to identify some endemics in their area because the threat of extinction is not limited to one country but can impact other countries [7]. The results are less impressive when the same methods are used to classify various bird species in the open.

As a result, this study presents a new workflow of distinguishing diverse bird species in multiple nations utilizing sound pattern data to fill in the gaps left by prior studies. The sound pattern is created by integrating numerous parts extracted from the original sound using cutting-edge algorithms [8–10]. The k-nearest neighbor (KNN) technique is used to classify the sounds in this study because it is rarely used in bird sound categorization. This classification technique is consistent with the classification method we used in previous studies [6]. In contrast, the feature selection method chosen for this study is a particle swarm optimization method that selects the best features based on how the particle collection achieves the goal. This strategy is deemed suitable because it aims to achieve a higher level of precision.

This study's contribution is validated by comparing the results of bird sound classification in order to: (1) demonstrate the effect of a feature selection method on classification performance; (2) investigate the relationship between a feature selection method and bird signal classification performance; and (3) identify the parameters necessary to achieve bird signal classification performance.

This document is structured as follows: In section 2, previous researches related to the current study is mentioned. Section 3 outlines our suggested model and discusses our experimental strategy. Section 4 describes the experiment's findings and detailed evaluation. Finally, section 5 concludes.

2. State-of-the-art

A group of ornithologists undertook a study to categorize birds in the wild. Automatic bird sound species classification has been carried out in ornithology and conservation monitoring. They want to increase the classification and categorization accuracy of bird sounds.

Briggs et al. [10] developed the closest neighbour classification (KNN) technique, which uses the Kullback Leibler Divergence and Hellinger matrices to classify birds based on their aural patterns. The KNN algorithm extracts spectral density, and Mel-frequency cepstral coefficient (MFCC) features based on Hellinger Matric. The accuracy rating is 92.10%. Raghuram et al. [11] have developed a novel framework for the identification of the voices of 35 species of birds, which consists of 27 characteristics, including four-pitch features, four energy characteristics, four duration characteristics, thirteen MFCC characteristics, and two tempo characteristics. This research used the classification methods naive bayes (NB), neural network (NN), random forest (RF), and support vector machines (SVM) (SVM). The RF technique has an accuracy of up to 83.33 %, compared to 69.90 %, 78.60 %, and 70.75 %for the NB, NN, and SVM methods, respectively.

Kahl et al. [12] used audios of 100 chosen species in their convolutional neural network (CNN) categorization method. To produce features, CNN turns audio recordings into visual representations. The mean average precision (mAP) performance in this investigation was 0.605. Supriya et al. [13] used the SVM with GMM classification based on MFCC feature to recognize birds. The results reveal that the GMM classification algorithm outperforms the SVM, with GMM scoring 95% and SVM scoring 86%. While Jancovic and Kokuer [14] used 48 species, decomposed using the sinusoid, and represented by frequency and magnitude values. The Markov model is combined with a hybrid deep learning algorithm to develop the classification strategy. The performance received a score of 98.7%. They were also done by Ramashini et al. [15], that employed the nearest centroid (NC) classification strategy to classify five local bird species and compared it to SVM and KNN. This study also uses a reduction strategy based on linear discriminant analysis to reach the best performance results (LDA).

Sukri et al. [16] developed a new pre-processing pipeline to segment bird sounds by measuring power per unit of frequency to get specific characteristics. The NN approach is used for the classification of bird noises. The accuracy results suggest that this approach is capable of classifying bird noises. Chandu et al. [17] used the CNN method based on the Alexnet architecture to classify 400 bird recording data samples with four different bird species. In this work, feature extraction was performed utilizing spectrogram creation, with early processing conducted before the extraction procedure to improve the sound.

Ramashini et al. presented a flow classification of bird sounds using the linear discriminant analysis (LDA) approach as the best feature selection method and the closest centroid (NC) method as a classification method in their study [1]. The findings reveal that combining LDA and NC approaches produce virtually identical results when applied to LDA and NN. When LDA is used to choose the best features, these accuracies improve by 3.4%. Pahuja dan Kumar [18] used a different method to classify

Received: October 12, 2022. Revised: November 14, 2022.

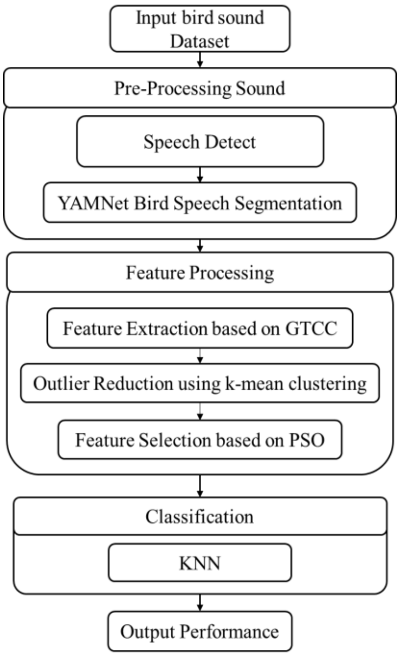

Figure 1. Research proposed

bird sounds. The stages used are pre-processing, sound segmentation, sound balancing, and proper cleaning. The clean data were extracted using a short-term fourier transform (STFT) spectogram extraction method and classified using MLP-ANN. The workflow provides an accuracy of 81.4%.

From these several studies, many researchers focus on the same research to achieve better accuracy performance. However, most researchers use privately selected datasets with various publicly available datasets. The same methods are used with different workflows for different datasets. Therefore, this study proposes a new workflow to recognize multiple types of birds based on the sound patterns they produce.

3. Research proposed

According to prior research, data, features, and classification algorithms affect the effectiveness of classification systems. The findings of the classification strategy demonstrate that important features do not always indicate poor precision. The suggested may utilize these attributes to display the appropriate voice quality based on the context. However, these parameters are affected by the data source. This study presents a unique identification methodology for bird calls that incorporates the original vocal features (see Fig. 1).

3.1 Data acquisition and pre-processing

In this research, it is believed that the voice of birds has different characteristics, but the environment that impacts it has the necessary characteristics. In order to identify the optimal model, this research depends on datasets from public sources. The original dataset of 21,375 bird sounds from 264 bird species (CCB) was donated by the Cornell Lab of Ornithology's center for conservation bioacoustics. Two steps of data processing are the preparation of data and classification of bird calls based on the characteristics of bird calls. In the first step, data collection is constructed based on the characteristics of files of identical types. The experiment extracted audio utilizing time-coded annotations from the newly provided set of information for correct space validation. The files in the training set are subsequently processed to produce new data sets with distinct information. The speech boundary in the audio stream is identified, and the voice is grouped or segmented, with the sound of birds being taken into consideration and other voices being removed. During this phase, the YAMNet neural network AudioSet ontology is used to extract the portion of the audio stream that is considered to be a bird voice [6, 19, 20].

3.2 Gammatone frequency cepstral coefficients feature characteristic

Feature extraction is used to get bird speech features. Specific characteristics are employed using Gammatone frequency cepstral coefficients (GFCC) [6]. GFCC is based on a collection of Gammatone filter banks. The Gammatone filter bank output is converted into a cochleagram, a time-frequency signal representation. The GFCC features are calculated using a cochleagram. The stages of the GFCC are shown as follows:

3.2.1. Gammatone filter

Gammatone filters imitate the human hearing system's mechanics. Eq. (1) may be used to specify the center frequency (f_c) of a Gammatone filter. The gain value is controlled by the variable a, the filter order is defined by the value of n set to a number less than four, and the value of b is determined by (2). For the Gammatone filter bank to provide a representation comparable to the FFT-based spectrogram, a collection of Gammatone filters with changing center frequencies are regarded as channels with varying center frequencies.

$$g(t) = at^{n-1}e^{-2\pi bt}cos(2\pi f_c + \varphi) \quad (1)$$

$$b = 25.17\left(\frac{4.37f_c}{1000} + 1\right) \quad (2)$$

3.2.2. Windowing

The GFCC requires a window to cover K points in each frame and shift each L point. $x(t; f_c(m))$ defines each frame, with the frequency center (f_c) in the m filter. Each frame's Cochleagram representation is calculated on average across the t window (3). Where one indicates the magnitude of the complex number and the other represents the dependent factor in frequency. For 16 kHz signals producing 100 frames per second, M is the number of filter bank channels with K values of 400, L of 160, and M of 32.

$$\bar{x}(n; m) = \frac{1}{K}\sum_{0}^{K-1} \gamma|x(nL + i; f_c(m))| \quad (3)$$

3.2.3. Discrete continue transform (DCT)

Uncorrelated cepstral coefficients were obtained using DCT. The range u starts from 0 to 31 and is similar to the MFCC operation (4). The GFCC technique generates 39 features, each consisting of 13 GFCC values and 26 GFCC deltas.

$$g(n; m) = \left(\frac{2}{M}\right)^{0.5} \sum_{i=0}^{M-1} \left\{\frac{1}{3} log(\bar{x}) \, cos\left[\frac{\pi u}{2M(2i-1)}\right]\right\} \quad (4)$$

3.3 K-means clustering outlier reduction

Occasionally, duplicate data is produced, but if the feature sets are thought to be connected, the feature vector may be reduced without compromising a great deal of information. K-means may minimize the quantity of duplicated data for a label while retaining a substantial amount of information [6]. By combining the produced data and picking the data with the most influential members, K-means may accomplish outlier reduction. The data is created each time the speech signal is subjected to feature extraction, so it does not subtract from the data and, since it is gathered in the same group, yields data of the best quality.

The K-means method categorizes the data into numerous groups. The data in one group share certain characteristics with the data in other groups, as well as possessing unique characteristics. They decreased the objective function by limiting variance inside a cluster and boosting variation across clusters. Eq. (5) is utilized as the goal function. The frequency of the signal used is f, and the centroid is ce, which is determined randomly.

$$dist = \sqrt{(f - ce)^2} \quad (5)$$

3.4 Feature selection based-on particle swarm optimization

The stages of the PSO algorithm are depicted in the following steps [21, 22].

a) Initialize the population (swarm) t and velocity v of the particle p_t so that when t is 0 then the position $xpos_t$ of each particle $xpos_t \in p_t$ in a wide space is a random value.
b) The fitness values of all particles are evaluated using the fitness function based on the optimization problem's objective. A fitness function is a calculation that determines the suitability or objective calculation of a solution. The classification method is used in this study to evaluate fitness performance or f.
c) Taking the fitness f value of each particle and comparing it to the pbest fitness value. If the current value is greater than pbest, the current fitness value is set to pbest. Otherwise, pbest remains unchanged.
d) Determine the best fitness value in a population from all particle fitness values. Then, compare that value to gbest and set it to gbest if it is better, otherwise leave it alone.
e) Update the velocity and position of each particle is define by (6) and (7). Each particle t, has a vector $xpos_t$ $(xpos_1, xpos_2, \dots, xpos_d)$ is used to represent the position and a vector v_t $(v_1, v_2, \dots, v_d)$ is used to represent velocity. Swarm t in dimension d as large as the feature size. w is an inertial weight between 0 and 1. c_1 and c_2 as two acceleration weights for particles and swarms that have values ranging from 0 to 2. r_1 and r_2 are two random numbers that are evenly distributed between 0 to 1. The position of a particle is limited in the range $xpos_{min}$ to $xpos_{max}$, the velocity of a particle is also limited in the range v_{min} to v_{max}.

$$v_{t+1} = w \times v_t + c_1 \times r_1 \times (p_{best} - xpos_t) + c_2 \times r_2 \times (g_{best} - xpos_t) \quad (6)$$

$$xpos_{t+1} = xpos_t \times v_{t+1} \quad (7)$$

f) If the best solution that matches the specified minimum error is found or the maximum number of iterations is reached, stop; otherwise, repeat steps (b) to (e).

3.5 K-nearest neighbors classification (KNN)

KNN is a simple algorithm that maintains all existing examples and classifies new cases according to similarity measurements (5). There are tagged training data sets available. When introducing fresh data without labels, compare features to the training set in order to identify the k-most comparable feature. KNN offers excellent accuracy, is insensitive to outliers, and requires no data input assumptions. KNN, on the other hand, are computationally and spatially complicated.

3.6 Performance evaluation

Calculating the accuracy is used to evaluate the methods of classification performance. The proper classification of all data obtained is characterized as accuracy. Eq. (8) is used to get the accuracy value, with *t* and *n* being the number of correctly classified sample data and n denoting the total sample data.

$$accuracy = t/n \times 100 \quad (7)$$

3.7 Research design

Standard learning methods like NB, SVM, and DT are used to classify data and give an accuracy score compared to the KNN method. The parameter default settings restrict the number of parameters used in this investigation. The GFCC is a feature extraction approach that employs a 0.03-times-the-frequency-value hamming function with regular repeats of windows in actual vectors. Size An integer provides the overlap length between adjacent windows with a value equal to 0.02 times the frequency value. A value of 0.02 is used to compute the window limit for voice detection. There are 39 combinations of characteristics in the GFCC feature. 625,381 data were extracted from 21,375 original bird voice data using GFCC and k-means clustering concurrently, yielding a total of 625,381 data. The default parameter settings for each learning technique are likewise applied. This study used the PSO

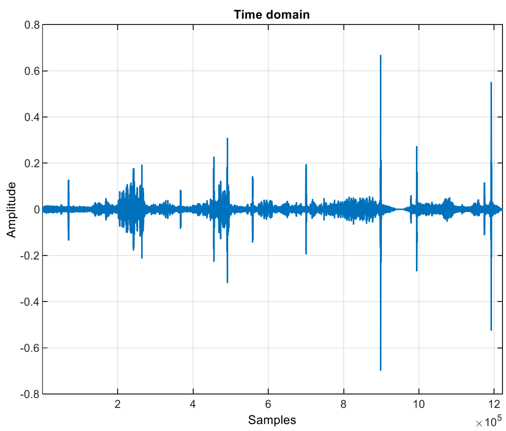

Figure. 2 Original bird audio

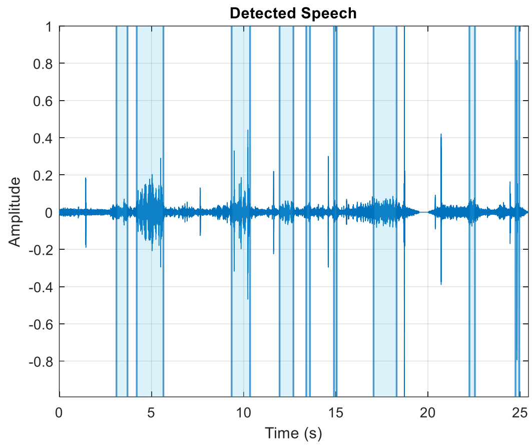

Figure. 3 Bird voice and other

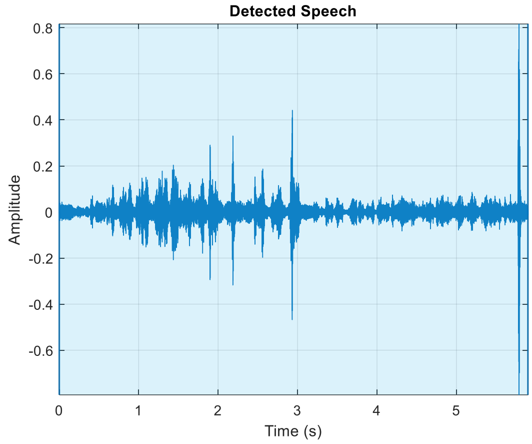

Figure. 4 Clear bird voice

optimization method to choose these characteristics, using parameter swarms of 10 to 100 particles evaluated 5 to 30 times. The KNN algorithm decides which method achieves the maximum level of accuracy. This research used the student edition of MATLAB R2021b to enhance the signal, feature extraction, and classification.

4. Result and discussion

In this section, we present basic processing step, feature extraction stage, and classification stage are all covered.

4.1 Data acquisition and pre-processing result

There are 21,375 bird sounds from 264 species in this dataset of native bird noises. Each sound has a particular wavelength, resulting in a large amount of data being separated into multiple windows to distort the frame and the signal. Windowing is used to avoid gaps and is accomplished by replicating each sample frame from its start point to its endpoint. It increases the sound signal's consistency at the frame's start and end points.

Received: October 12, 2022. Revised: November 14, 2022.

Table 1. Feature extraction applied using data reduction

No	gtcc1	gtcc2	gtcc3	gtcc4	.	gtcc39
1	3,25	-6,00	-6,00	-6,00		-2,96
2	1,63	0,72	0,26	-0,08		-2,59
3	1,62	0,95	0,65	0,02		-2,71
4	1,56	0,81	0,49	-0,59		-2,60
5	1,66	0,95	0,27	0,09		-1,81
6	1,63	0,67	0,42	-0,16		-2,16
7	1,77	0,77	0,32	-0,22		-2,96
.	.	.				.
625.381	1,33	0,04	0,21	0,09		0,23

The data clearly shows that not all data is derived from bird sounds. As a result, a search for the bounds of reasonableness is carried out to collect only features in the form of bird noises (Fig. 2). According to the findings, as many as a few data segments do not contain bird sounds (Fig. 3). Non-bird sound signals are removed. Bird sound signals are combined to provide a complete signal that includes bird noises (Fig. 4).

4.2 Feature processing

Signal quality is enhanced by extracting characteristics from signal data and turning them into numbers that can be recognized by machine learning algorithms. The GFCC technique recovers more than 21,375 features of bird sounds. There is data reduction since each signal feature extraction result yields real data, which totals 16,924,040 data extracted with 39 data features. The K-means clustering approach is used to minimize data by determining the most prominent total cluster member for each clustered data set, yielding 625,831 records with 39 characteristics (Table 1).

4.3 Particle swarm optimization based on feature selection

The original features were chosen using the PSO optimization approach to obtain the best features for this investigation. According to the experimental findings, a swarm of 80 swarms had the highest performance outcomes. It is indicated that the initial evaluation produced the most extraordinary performance, 71.64%.

4.3.1. Swarm parameter on PSO

The swarm number parameter significantly impacts the evaluation's results. A swarm of 80 particles gave the most significant findings (71.64%) in Fig. 5. The result offers a value with a minimum range of 63.888% and a maximum range of 71.643%

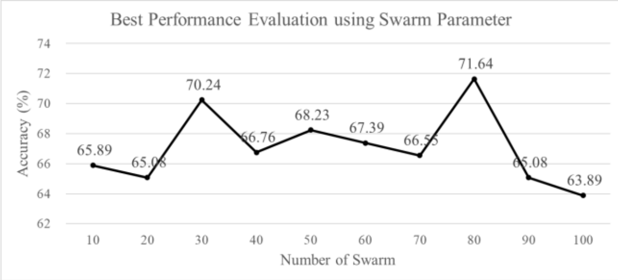

Figure. 5 Best accuracy using swarm parameter

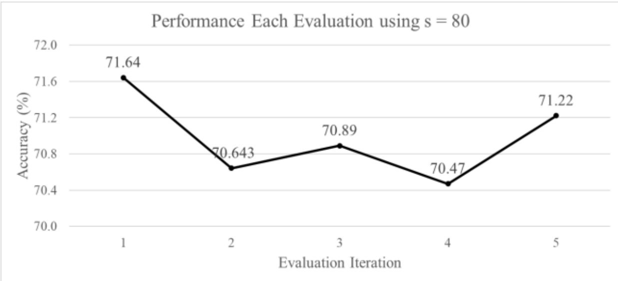

Figure. 6 Performance each evaluation using s 80

based on the number of swarms of 100 and 80. The performance data from each swarm demonstrates that having many swarms does not allow for improved performance. However, many swarms will multiply the opportunities attained, allowing them to achieve their objectives swiftly, which impacts the time spent.

From iteration 1 through iteration 5, the outcomes of each evaluation are displayed in Fig. 6. It is evident that iteration 1 had the greatest influence. The minimum and maximum ranges of the number shown in Fig. 6 are 70.471% and 71.643%, respectively. The PSO iteration was increased up to five times so that the herd could only move to the target spot in five phases. This demonstrates that each step requires a large position change to obtain high precision, implying that accuracy may be achieved with few positional adjustments.

The most precise evaluation of 1 comes from the experiment shown in Fig. 6, and Fig. 7 shows the dispersion of particle positions. The distribution generates a number with a minimum range of 0.01 and a maximum range of 0.98. The average score for the top position is 0.596, with a standard deviation of 0.307 points. The outcomes display each particle's location inside the swarm, and the sheer quantity of these particles enables a thorough search. This affects how quickly or slowly people get to their destination, with 59.6% of people being placed nearby.

Each particle's velocity is shown in Fig. 8. The results show a minimum velocity range of -0.738 and a maximum velocity range of 0.917. The ideal average velocity result is 0.114, with a standard deviation of 0.414. It is also demonstrated that the speed of each particle varies from one another to the

Received: October 12, 2022. Revised: November 14, 2022.

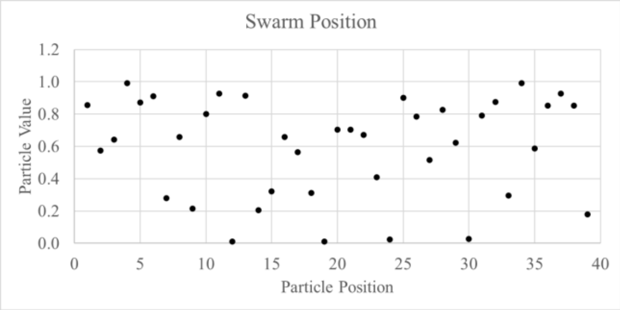

Figure. 7 1st swarm position

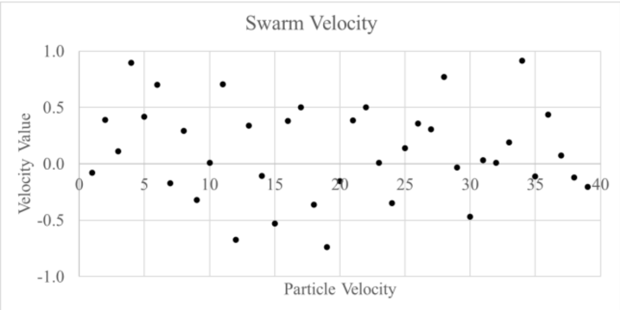

Figure. 8 1st swarm velocity

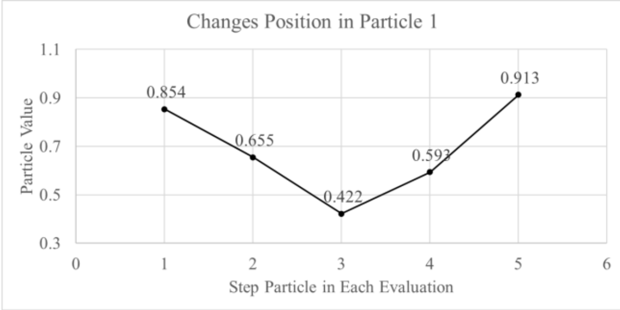

Figure. 9 Movement step on the 1st particle

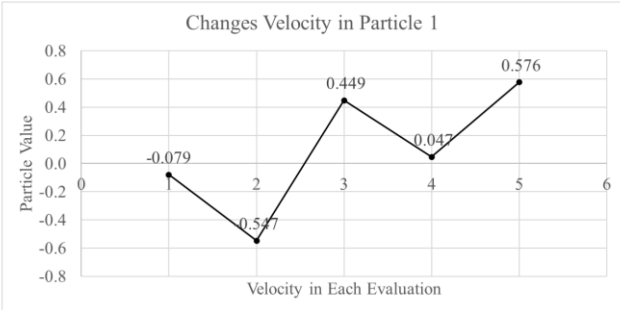

Figure. 10 Velocity changes at 1st particle

number of particles that spread out over one another (shown in Fig. 8). Even though it only utilizes an average speed of 0.114, this enables the ability to search for performance objectives to be carried out carefully and randomly in order to acquire the proper goals.

The repetitions of each evaluation (Fig. 9) for a maximum of 5 evaluations show that the status of each swarm has altered for the evaluation stage. These results show that the minimum and highest ranges are, respectively, 0.422 and 0.913. An average outcome of 0.687 with a standard deviation of 0.199 characterizes the best position transfer. These findings demonstrate that, on average, the optimal location is at 0.687, which enables the particles to be near to one another.

Additionally, the varying ideal swarm velocity for each evaluation is shown (Fig. 10). These results indicate that the minimum and highest velocity ranges are -0.547 and 0.576, respectively. The most favorable outcome for the average velocity change is 0.089, with a standard deviation of 0.448. The fact that the average is just 0.089 suggests that the herd is making an effort to anticipate achieving the expected result.

It was acquired in this experiment by choosing some features, and the chosen features were 27 features produced by GTCC. The 7th, 9th, and 12th GTCC features, the 14th, 15th, 18th, 19th, and 24th GTCCDelta features, and the 30th, 33rd, and 39th GTCCDeltaDelta features are all unused features. The best results were found for features 1, 2, 3, 4, 5, 6, 8, 10, 11, 13, 16, 17, 20, 21, 22, 25, 26, 27, 28, 29, 31, 32, 34, 35, 36, 37, and 38 when PSO used random to choose each swarm. Based on the classification procedure, the classification performance from these swarms was achieved. The Friedman test was used to do a statistical analysis of this experiment. According to the findings ($Q = 9.79$, $p = 0.3677$), using the number of swarm parameters does not significantly improve accuracy performance.

4.3.2. Evaluation parameter on PSO

A 25-iteration evaluation yields the optimum result, according to experiments employing parameters ranging from 5 to 30 iterations. When using the evaluation parameter maximum of 30, the findings show continual growth. The performance results of 25 iterations of each swarm produced 78.331% more after an increase of 6.687%. (Shown in Fig. 11). With a standard deviation of 2.448%, a minimum accuracy of 71.643%, a maximum accuracy of 78.331%, and an average performance of 75.964%.

These findings indicate that Fig. 12 depicts the displacement of the particle location based on the most significant performance value or a maximum of 25 repetitions. It was demonstrated in a maximum of 25 evaluations that the swarm made movements that varied in the distance; occasionally, the swarm only moved nearby. This enables the swarm to adjust to one another while searching for the ideal posture to take when traveling to the intended location. Statistically, Fig. 12 demonstrates the displacement

Received: October 12, 2022. Revised: November 14, 2022.

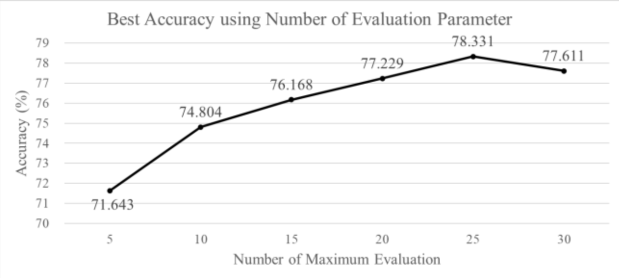

Figure. 11 Best accuracy using evaluation parameter

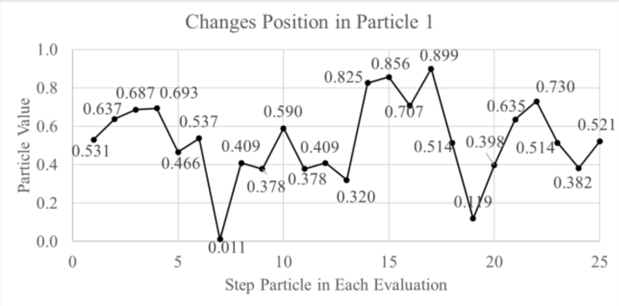

Figure. 12 Movement step on the best particle

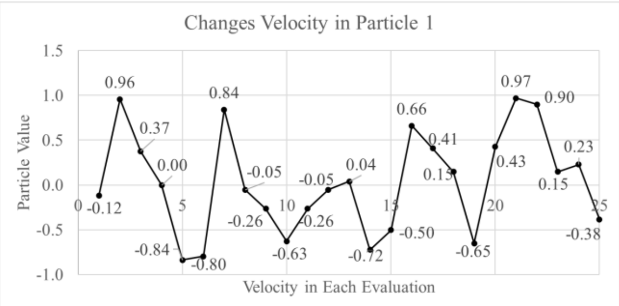

Figure. 13 Velocity changes on the best particles

position values ranging from 0.011 for the lowest position to 0.899 for the highest rank, with 0.526 for the average and 0.212 for the standard deviation. This shows the swarm changing locations to reach its destination.

Along with location changes, the PSO parameters also experience variations in speed. After 25 evaluations of the velocity variation, it was determined that the swarm occasionally increased and occasionally decreased its velocity. This enables the particles to modify their speed in response to the circumstances the swarm is experiencing. According to statistics, the velocity change is 0.033 on average, 0.970 at its highest point, -0.838 at its lowest point, and 0.563 at its standard deviation (shown in Fig. 13). This demonstrates that each particle in a swarm impacts the average speed needed to reach the target because the average velocity displayed is just 4% of the required velocity.

Fig. 14 shows the impact of changes in location and speed as well as the length of time required to conduct this experiment in terms of assessment. The

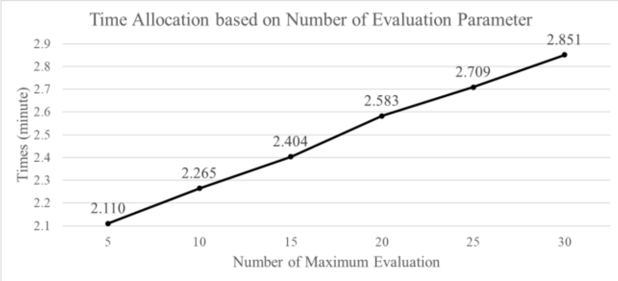

Figure. 14 Time allocation for classification based on number of evaluation parameter

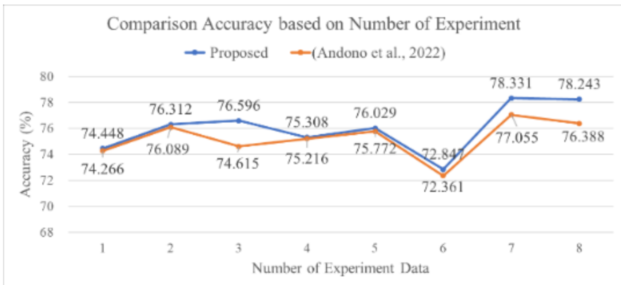

Figure. 15 Accuracy comparison between proposed and previous research [6] (in %)

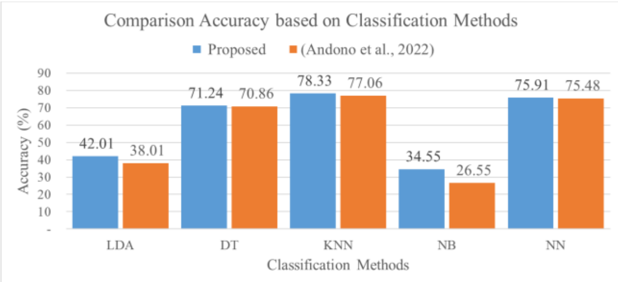

Figure. 16 Accuracy comparison based on classification methods (in %)

graph illustrates how an increase of evaluations results in an increase in time required. a highly likely scenario given that the large number of evaluations drove the herd to go farther and took a lot of time.

4.4 Comparison with previous research [6]

The outcomes explain that the proposed strategy consistently delivers the optimum performance. Based on Fig. 15, we reject the null hypothesis since the p-value is 0.025 less than 0.05 (or F is 8.130 is better than 5.591 for Fcrit). At a 95% level of confidence, we conclude that there is a significant difference in the yields for the optimization using the feature selection approach.

Based on Fig. 16, it is possible to conclude that the suggested technique enhances the overall classification method. The naive bayes (NB) technique yields the most significant improvement at 8%, followed by the linear discriminant analysis (LDA) method at 4%. While other methods, including decision tree (DT), neural network (NN), and K-nearest neighbours (KNN) based on [6], yield just a modest improvement (0.38 %, 0.43 %, and 1.27 %, respectively). This technique modifies the fundamental PSO method with the classification-based fitness function.

4.5 Discussion

The result indicated that the employment of the feature selection approach could result in improved performance. Based on the findings in Fig. 16, performance see that the suggested approach improves the overall classification method. This is consistent with the results shown in Figs. 5 to 16. These results demonstrated that the proposed PSO allows the selection of the most appropriate features. It provides a variety of choices to examine, in keeping with the idea that the swarm's movement is not constant. The swarm, however, can accomplish its objective effectively because it shares a common purpose. Due to this idea, the PSO optimization technique has a significant influence on offering the optimal features, affecting both accuracy performance and time use (according to Figs. 11 to 14).

The performance of the classification approach is improved when PSO is used for feature selection. With a classification performance of 77.055% provided by the features produced by the GTCC feature extraction method used by Andono et al. [6], some of the 39 GTCC features may have flaws or are less relevant for each dataset, allowing the feature selection method to improve recognition performance. When the naive bayes approach is used in Fig. 16, it may result in a considerable improvement due to the significant number of irrelevant features' poor recognition performance. The KNN approach may be used to enhance the effectiveness of other classification techniques, though.

The use of selection features based on the PSO optimization method appears to improve several other classification methods to be applied to bird voice recognition, with the most significant improvement being 8%. This is also increasing by 1.27% using the KNN classification method. If parameter settings are used more effectively, they will likely perform better than this suggestion. It should be mentioned, however, that the best feature selection technique, which uses the classification approach to obtain the fitness value in PSO, has proven successful and merits being applied in the actual application of bird signal identification.

5. Conclusions

One of the most challenging issues confronting many nations is protecting endangered bird species, which is why we are doing this research. This objective may be accomplished by implementing an automatic bird speech recognition system, especially when used outdoors. Our suggested technique may decrease the number of extracted records and data features from GTCC, and the testing results offer a wide range of characteristics. Our solution improves the performance of previous classification techniques when applying feature selection, achieving a recognition performance result of 78.33%. KNN as a classification technique led to a rise of 1.27%, while naive bayes as a classification method led to a very high increase of 8%. Other classification methods also saw improvements in accuracy performance.

Compared to the prior feature technique, the suggested combination produces superior outcomes. These results indicate that, even with our restricted resources, our proposed approach can enhance the precision of bird voice recognition. Future research is anticipated to identify and assess the performance-improving potential of other optimization techniques.

Conflicts of interest

In accordance with the International Journal of Intelligent Engineering and Systems's policy and my ethical duty as a researcher, I certify that this manuscript has not been previously published, copied, or submitted to another journal. I have reported them in full to the International Journal of Intelligent Engineering and Systems and obtained the consent of all authors to handle any possible conflicts of interest arising from this research.

Author contributions

The authors consist of Pulung Nurtantio Andono (PNA), Guruh Fajar Shidik (GFS), Dwi Puji Prabowo (DPP), Dzuha Hening Yanuarsari (DHY), Yuslena Sari (YS), and Ricardus Anggi Pramunendar (RAP) with the following contributions: Conceptualisation, PNA and RAP; methodology, PNA and RAP; software, RAP; investigation, DPP, DHY and YS; formal analysis, PNA, RAP, GFS, and YS; resources, RAP and DPP; data curation, DPP and DHY; writing—original draft preparation, PNA and RAP; writing—review and editing, GFS and YS; visualisation, RAP; supervision, GFS; project administration, DPP and DHY; funding acquisition, RAP.

Acknowledgments

This study is funded by the Indonesian Ministry of Research and Higher Education (DPRM-DIKTI) and the Faculty of Computer Science at Universitas Dian Nuswantoro in Semarang as well as Universitas Lambung Mangkurat in Banjarmasin. The authors are also appreciative for the reviewers' insightful remarks and recommendations, which helped enhance the presentation.

References

[1] M. Ramashini, P. E. Abas, and L. C. de Silva, "A novel approach of audio-based feature optimisation for bird classification", *Pertanika J Sci Technol*, Vol. 29, No. 4, pp. 2383–2407, 2021, doi: 10.47836/PJST.29.4.08.

[2] V. Nijman, S. L. Sari, P. Siriwat, M. Sigaud, and K. A. I. Nekaris, "Records of four Critically Endangered songbirds in the markets of Java suggest domestic trade is a major impediment to their conservation", *BirdingASIA*, Vol. 27, No. August, pp. 20–25, 2017.

[3] E. Buechley and C. H. Sekercioglu, "Endangered species", *Grzimek's Animal Life Encyclopedia*, No. August 2013, pp. 159–176, 2018.

[4] N. J. Bennett, R. Roth, S. C. Klain, K. Chan, P. Christie, D. A. Clark, G. Cullman, D. Curran, T. J. Durbin, G. Epstein, A. Greenberg, M. P. Nelson, J. Sandlos, R. Stedman, T. L. Teel, R. Thomas, D. Veríssimo, and C. Wyborn, "Conservation social science: Understanding and integrating human dimensions to improve conservation", *Biol Conserv*, Vol. 205, pp. 93–108, Jan. 2017, doi: 10.1016/j.biocon.2016.10.006.

[5] R. A. Pramunendar, S. Wibirama, P. I. Santosa, P. N. Andono, and M. A. Soeleman, "A robust image enhancement techniques for underwater fish classification in marine environment", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 5, pp. 116–129, Oct. 2019, doi: 10.22266/ijies2019.1031.12.

[6] P. N. Andono, G. F. Shidik, D. P. Prabowo, D. Pergiwati, and R. A. Pramunendar, "Bird Voice Classification Based on Combination Feature Extraction and Reduction Dimension with the K-Nearest Neighbor", *International Journal of Intelligent Engineering and Systems*, Vol. 15, No. 1, pp. 262–272, 2022, doi: 10.22266/IJIES2022.0228.24.

[7] J. Atanbori, W. Duan, J. Murray, K. Appiah, and P. Dickinson, "Automatic classification of flying bird species using computer vision techniques", *Pattern Recognit Lett*, Vol. 81, pp. 53–62, 2016, doi: 10.1016/j.patrec.2015.08.015.

[8] D. Stowell and M. D. Plumbley, "Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning", *PeerJ*, Vol. 2, No. 1, p. e488, 2014, doi: 10.7717/peerj.488.

[9] Supriya, R. Prakrithi, S. Bhat, Shivani, and S. Santhosh, "Classification of birds based on their sound patterns using GMM and SVM classifiers", *International Research Journal of Engineering and Technology*, Vol. 05, No. 2004, pp. 4708–4711, 2018.

[10] F. Briggs, R. Raich, and X. Z. Fern, "Audio classification of bird species: A statistical manifold approach", In: *Proc. of IEEE International Conference on Data Mining, ICDM*, pp. 51–60, 2009, doi: 10.1109/ICDM.2009.65.

[11] M. A. Raghuram, N. R. Chavan, R. Belur, and S. G. Koolagudi, "Bird classification based on their sound patterns", *Int. J. Speech Technol.*, Vol. 19, No. 4, pp. 791–804, 2016, doi: 10.1007/s10772-016-9372-2.

[12] S. Kahl, T. W. Stein, H. Hussein, H. Klinck, D. Kowerko, M. Ritter, and M. Eibl, "Large-scale bird sound classification using convolutional neural networks", *CEUR Workshop Proc*, Vol. 1866, 2017.

[13] P. R. Supriya, S. Bhat, Shivani, and S. Santhosh, "Classification of birds based on their sound patterns using GMM and SVM classifiers", *International Research Journal of Engineering and Technology*, Vol. 5, No. 4, pp. 4708–4711, 2018, [Online]. Available: www.irjet.net.

[14] P. Jancovic and M. Kokuer, "Bird species recognition using unsupervised modeling of individual vocalization elements", *IEEE/ACM Trans Audio Speech Lang Process*, Vol. 27, No. 5, pp. 932–947, 2019, doi: 10.1109/TASLP.2019.2904790.

[15] M. Ramashini, P. E. Abas, U. Grafe, and L. C. D. Silva, "Bird Sounds Classification Using Linear Discriminant Analysis", In: *Proc. of ICRAIE 2019 - 4th International Conference and Workshops on Recent Advances and Innovations in Engineering: Thriving Technologies*, Vol. 2019, No. November, pp. 27–29, 2019, doi: 10.1109/ICRAIE47735.2019.9037645.

[16] M. M. M. Sukri, U. Fadlilah, S. Saon, A. K. Mahamad, M. M. Som, and A. Sidek, "Bird Sound Identification based on Artificial Neural Network", In: *Proc. of 2020 IEEE Student Conference on Research and Development,*

SCOReD 2020, No. September, pp. 342–345, 2020, doi: 10.1109/SCOReD50371.2020.9250746.

[17] B. Chandu, A. Munikoti, K. S. Murthy, G. Murthy, and C. Nagaraj, "Automated Bird Species Identification using Audio Signal Processing and Neural Networks", In: *Proc. of 2020 International Conference on Artificial Intelligence and Signal Processing, AISP 2020*, 2020, doi: 10.1109/AISP48273.2020.9073584.

[18] R. Pahuja and A. Kumar, "Sound-spectrogram based automatic bird species recognition using MLP classifier", *Applied Acoustics*, Vol. 180, p. 108077, 2021, doi: 10.1016/j.apacoust.2021.108077.

[19] J. F. Gemmeke, D. P. W. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio Set: An ontology and human-labeled dataset for audio events", *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 776–780, 2017, doi: 10.1109/ICASSP.2017.7952261.

[20] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, "CNN architectures for large-scale audio classification", In: *Proc. of 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 131–135, 2017. doi: 10.1109/ICASSP.2017.7952132.

[21] M. Clerc and J. Kennedy, "The Particle Swarm - Explosion, Stability, and Convergence in a Multidimensional Complex Space", *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 1, pp. 58–73, 2002, doi: 10.1109/4235.985692.

[22] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms", *Applied Soft Computing Journal*, Vol. 18, pp. 261-276, 2014.