Bees Local Phase Quantisation Feature Selection for RGB-D Facial Expression Recognition

Seyed Muhammad Hossein Mousavi

Atiye Ilanloo

Abstract Feature selection can be defined as an optimisation problem and solved by bioinspired algorithms. The Bees Algorithm (BA) returns great performance in the feature selection optimisation task. On the other hand, local phase quantisation (LPQ) is a frequency domain feature that has excellent performance on depth images. Here, after extracting LPQ features from RGB (colour) and depth images from the Iranian Kinect Face Database (IKFDB), the Bees feature selection algorithm is applied to select the desired number of features for final classification tasks. IKFDB is recorded with Kinect sensor V.2 and contains colour and depth images for facial and facial microexpression recognition purposes. Here, five facial expressions, Anger, Joy, Surprise, Disgust and Fear, are used for final validation. The proposed Bees LPQ method is compared with Particle Swarm Optimisation (PSO) LPQ, PCA LPQ, Lasso LPQ, and just LPQ features for classification tasks with Support Vector Machines (SVM), K-Nearest Neighbourhood (KNN), Shallow Neural Network and Ensemble Subspace KNN. The returned results show a promising performance of the proposed algorithm (99 % accuracy) in comparison with others.

Keywords: Bees Algorithm, Feature Selection, Local Phase Quantisation, Optimisation, Kinect, Facial Expressions, Depth Images

Seyed Muhammad Hossein Mousavi
Independent Researcher, Tehran, Iran, e-mail: mosavi.a.i.buali@gmail.com
Atiye Ilanloo

Faculty of Humanities- Psychology, Islamic Azad University-Rasht, Gilan, Iran, e-mail: elanlooatiye@gmail.com

1 Introduction

1.1 Facial Expressions Recognition

After face detection [1] and face recognition [2], analysing facial muscles' movements comes to mind. Recognising these movements and micromovements is called facial expression recognition (FER) [2] and facial microexpression recognition (FMER) [2], which have high importance in image processing and psychology. Practical applications are human-computer interaction (HCI), such as face image processing, facial video surveillance, and facial animation. Recognising facial expressions is harder than recognising faces, as they could have micromovements and differ from one person to another, which is the main problem in FER. Additionally, each expression could be misunderstood in different races, as skin wrinkles affect the recognition direction. Facial expressions are made and calculated by Facial Action Coding Systems (FACS) [3], and they are categorised into seven primary facial expressions of joy, anger, disgust, sadness, fear, surprise and neutral. For example, for Joy expression, Action Units (AU) of + and 12 are involved. FMER is FER with more precise calculations, as few changes appear on the face. Some of these expressions are illustrated in Fig. 1 alongside their corresponding action units.

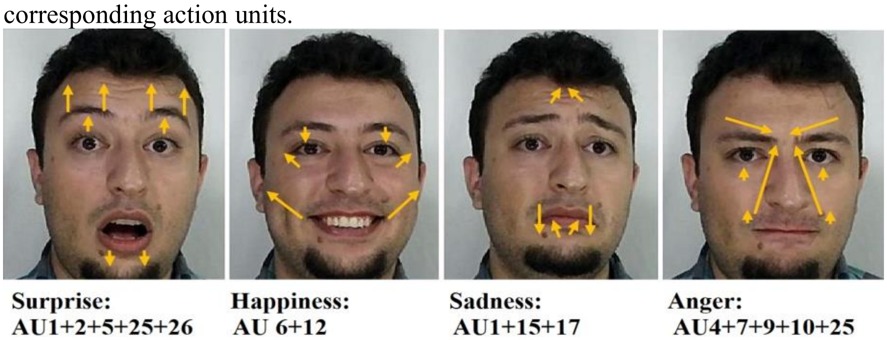

Fig 1. Four main facial expressions – These samples are taken from IKFDB [2]. The owner of the IKFDB [2] dataset is the first author of this chapter. AUs are added to the image by arrows. Also, the license can be found at: https://creativecommons.org/licenses/by-sa/4.0/

1.2 Bioinspired Algorithms and Bees Algorithm

Bioinspired or nature-inspired algorithms [4, 5] are normally mathematical models of animal and insect social behaviour in a manner that leads to problems with an optimal solution during iterations by a specific number of populations in each generation. These algorithms could be employed in multiple optimisation tasks, such as regression [6], clustering [7, 9], feature selection [8], minimum spanning tree

(MST), and hub location allocation (HLA). One of these bioinspired algorithms that has high efficiency is called the Bees Algorithm (BA) [10, 11, 18]. The Bees Algorithm has many applications in which the Bees Algorithm simply implements the social behaviour of honey bees in searching for food in flowers. Agents, scouts and Forager bees are involved in global and local searches to reach the best solution. Waggle dancing is performed by scout bees that found the best sites. Those that landed on elite sites recruit more new members. The list of best bees based on local and global information goes to the next generation and the cycle stops when termination criteria conditions are met.

1.3 Kinect Sensor and Depth Data

There are many infrared sensors, of which the Kinect sensor [12] is one of the most efficient and cheapest for educational purposes. Here, data recorded with the Kinect Version 2 sensor are used in the validation section. Kinect V.2 is capable of recording colour data with 1920*1080 resolutions and depth data with 512*424 resolutions with 30 frames per second (FPS). Colour data are based on red, green and blue (RGB) channels, and depth data are grey-like images, which are sometimes called 2.5-dimensional (2.5-D) images. Each pixel of the depth image in the Kinect sensor represents the distance between the sensor and the subject in millimetres. The range is 0.8 meters to 5 meters, and for instance, a value of 2000 pixels means that the object is 2 meters from the sensor. Depth images could be transformed into 3-D images and aid colour images to increase recognition accuracy. As depth images are generated from infrared particles, the Kinect sensor could operate in absolute darkness.

1.4 Feature Selection

In dealing with big data [13] or a massive number of samples (especially image samples), and after the feature extraction [14] step, selecting the most impactful features out of data is essential. Feature selection [15] or dimensionality reduction helps to decrease computational time and volume plus eliminating outliers in the dataset. Outliers lead the classification [16] task into misclassification, and removing them is vital. In this paper, the Bees Algorithm is employed as a feature selection tool. Here, the Local Phase Quantisation (LPQ) [17] feature is extracted, which provides 256 features and Bees feature selection and selects fewer than 128 features without a classification accuracy drop. The paper consists of five main sections: introduction, prior related research, proposed method, validation and results and conclusion.

2 Prior Related Research

As this research is about feature selection, famous and benchmark feature selection or dimensionality reduction techniques belonging to traditional and bioinspired categories will be explained. These algorithms could be applied directly to extracted images or signal features or could be applied to unfolded versions of images, signals or any type of numerical matrix to eliminate outliers and reduce computational time.

Principal component analysis (PCA) [19] generates new matrixes, named principal components (PAs). Each PA is a linear mixture of the original matrixes. All the PAs are orthogonal to one another, so there is no extra information. By selecting the best principles components, the best features will be chosen.

On the other hand, Lasso [20] is a regularisation [20] method for estimating generalised linear models. Lasso includes a red line that limits the size of the approximated coefficients. Therefore, it is similar to ridge regression [21]. Lasso is a shrinkage approximator: it makes coefficient approximations that are biased to be tiny. By shrinking features, the best features remain.

The chi-square test [22] is a statistical test employed to contrast observed data and expected data. The objective is to specify if a change between observed data and expected data is because of chance or because of a relationship between the variables that are being studied. Having two variables, it is possible to obtain the observed count and expected count. Chi-square measures how the desired count and observed count deviate from each other. If the target variable is independent of the feature variable, it is possible to ignore that specific feature. If they are dependent, the feature is very significant.

Selecting the best features based on the highest Data Envelopment Analysis (DEA) [23] for Data Management Units (DMUs) [23] is a management-related feature selection technique. In that case, those features or DMUs with the best efficiency will be selected, and the rest will be considered leftovers or outliers [24].

Among bioinspired feature selection techniques, few have proper performance, including PSO feature selection [25], firefly feature selection [26] and bee feature selection [27].

3 Proposed Method

The proposed method consists of multiple parts: data acquisition, preprocessing, LPQ feature extraction, Bee feature selection, labelling, classification and comparison. These operations are illustrated in Fig. 2. For validation, the Iranian Kinect Face Data Base (IKFDB) [2] is employed, which consists of 40 subjects. This dataset contains facial images for facial recognition, age estimation facial expressions and facial micro-expression recognition tasks. Additionally, data are recorded with Kinect sensor V.2 in colour and depth formats. All seven main facial expressions along the side with pitch, roll and yaw as video frames are available in this database. Here, five expressions of joy, anger, fear, disgust and surprise in colour and depth formats are used for the main experiment. After reading the data images from the input, the preprocessing stage, including grey level conversion, intensity adjustment, histogram equalisation and image resizing, takes place. The next step is extracting the LPQ feature from the frequency domain from all colour and depth samples. For each image sample, 256 LPQ features will be extracted. The fourth step consists of feature selection by the Bees Algorithm in the desired number of features. Here, the range of 2 to 255 features could be selected. As the classification task is a supervised learning method, labelling is needed for the fifth step. Classification algorithms are Support Vector Machines (SVM) [5], K-Nearest Neighborhood (KNN) [28], Shallow Neural Network [2] and Ensemble Subspace KNN [29]. Next, comparison by PCA, Lasso, PSO feature selection and LPQ using only the mentioned classification algorithms takes place. Finally, confusion matrixes and receiver operating characteristic (ROC) [30] curve plots were generated.

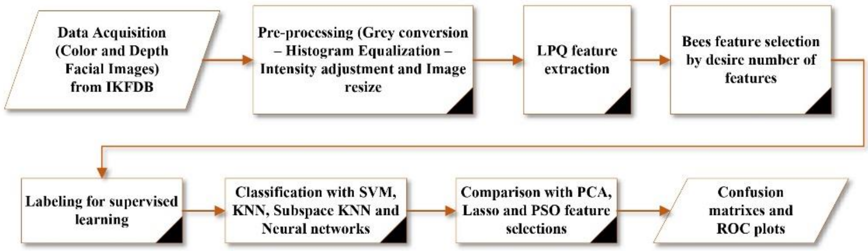

Fig 2. Proposed method flowchart

3.1 Local Phase Quantisation

LPQ [17] is a frequency neighbourhood-based feature based on Fourier transform [31]. It manipulated the blurring effect in magnitude and phase channels. The phase channel is capable of deactivating low-pass filters that exist in some images. LPQ features are perfect for use on depth data in the frequency domain. Figure 3 illustrates the LPQ algorithm process. Furthermore, this feature has perfect performance in dealing with low-pass Gaussian filters.

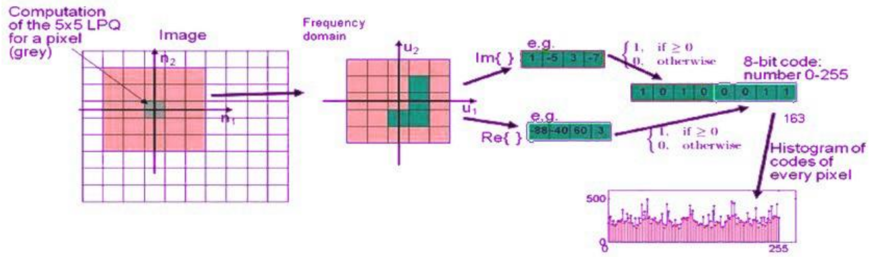

Fig 3. LPQ algorithm workflow

3.2 Bee Feature Selection

In feature selection, we address the number of features (NF), the weight of features (w) and the mean square error (MSE) [32], which should be minimised to select the feature. Additionally, if x_i are values of NF, then $\hat{x_i}$ would be selected features out of NF. Therefore, considering the number of features entering the system, "y' would be the output and "t" would be the target. To calculate the final error (objective function), e_i must be calculated, which is $t_i - y_i$. Thus, the final error is:

$$\text{Objective Function} = \text{Min}\, MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2 + w * NF \quad (1)$$

This goes for all features, and finally, those features with the lowest MSE will be selected. In the combination of Bees and feature selection, each feature vector is considered a bee with a different fitness function. Those bees that could fit into the final iteration would be selected alongside their related features with lower error, as mentioned. The pseudocode of Bees feature selection is presented below.

Start
Load LPQ features
Generating the initial population (features of F)
Define FN (number of features) and w (wight of features)
Evaluating the population based on the fitness function
Sorting
***While** max iteration is not satisfied*
Select elite patches and non-elite best patches for local search
Recruit forager bees for elite patches and non-elite best patches
Evaluate the fitness value of each patch
Sorting (select NF of F)
Allocate the rest of the bees for global search
Evaluate the fitness value of non-best patches
Sorting (Select NF of F)
End While
Select best first NF's

End

4 Validation and Results

As mentioned earlier for validation, IKFDB [2] data are used, which consist of colour and depth frames of seven main facial expressions. Here, 1000 colour and depth samples of five expressions (each 200 samples) are used in this experiment.

Some samples of IKFDB in different expressions and in colour-depth form are presented in Fig. 4. Additionally, the proposed Bees LPQ, PSO LPQ [25], PCA LPQ [19], Lasso LPQ [20] and solo LPQ using SVM [5], KNN [28], Shallow Neural Network [2] and Ensembles Subspace KNN [29] classification algorithms are compared. Table 1 presents the PSO and Bees parameters that are used in the experiment. Parameters are determined based on the size of the dataset, the number of samples, and many experiments. Increasing the number of iterations and population was not necessary, as it increased the runtime and did not improve the final accuracy. Table 2 presents the classification comparison results for different feature selection algorithms with half of the features (128), a quarter of the features (64) and an octant of features (32). Obviously, all features (256) are considered for solo LPQ in this table.

Fig 5 represents confusion matrixes and ROC curves for solo LPQ with all 256 features, the Lasso algorithm with 64 features and the Bees Algorithm with 64 features for the SVM classifier. Fig 6 illustrates the same but for the KNN classifier. Fig. 7 presents the performance of the Bees Algorithm for the feature selection task over 100 iterations.

Fig 4. Samples of the IKFDB [2] – The owner of the IKFDB [2] dataset is the first author of this chapter. The only change is to crop the eyes of the participants. Also, the license can be found at: https://creativecommons.org/licenses/by-sa/4.0/

Table 1. Bees and PSO Parameters

Algorithm	Bees	PSO
Iteration	100	100
Population	10 Bees	20 Particles
Decision Variable	20	20
Decision Variable Size	[1, 20]	[1, 20]
Lower Bound (LB)	-10	-10
Upper Bound (UB)	10	10
Mutation Rate	0.2	0.2
Inertia Weight	-	1
Inertia Weight Damping Ratio	-	0.99
Personal Learning Coefficient	-	1.5
Global Learning Coefficient	-	2
Selected Sites (SS)	Bees * 0.5	-
Select Elite Sites	SS * 0.4	-
Recruited Bees for Selected Sites	Bees * 0.5	-
Recruited Bees for Elite Sites	SS * 2	-
Neighbourhood Radius	0.1 * (UB – LB)	-
Neighbourhood Radius Damp Rate	0.2 0.96	-

Table 2. Comparison Classification Results with Different Numbers of Features (256, 128, 64 and 32)

	LPQ	PCA	Lasso	PSO	Bees
SVM	256 : 98.8 %	128 : 98.4 %	128 : 90.6 %	128 : 99.1 %	128 : 99.6 %
		64 : 98.0 %	64 : 89.3 %	64 : 98.6 %	64 : 98.9 %
		32 : 90.8 %	32 : 84.5 %	32 : 97.4 %	32 : 97.3 %
KNN	256 : 98.0 %	128 : 98.1 %	128 : 93.0 %	128 : 98.8 %	128 : 99.2 %
		64 : 97.9 %	64 : 93.9 %	64 : 98.2 %	64 : 98.1 %
		32 : 97.5 %	32 : 92.8 %	32 : 97.6 %	32 : 97.7 %
Shallow	256 : 97.6 %	128 : 97.3 %	128 : 96.7 %	128 : 99.1 %	128 : 99.4 %
NN		64 : 96.6 %	64 : 90.7 %	64 : 97.8 %	64 : 98.7 %
		32 : 95.8 %	32 : 91.2 %	32 : 96.9 %	32 : 97.9 %
Ensemble	256 : 98.3 %	128 : 97.9 %	128 : 95.6 %	128 : 99.5 %	128 : 99.8 %
Subspace		64 : 97.7 %	64 : 93.1 %	64 : 98.3 %	64 : 98.9 %
KNN		32 : 98.1 %	32 : 89.1 %	32 : 98.0 %	32 : 98.5 %

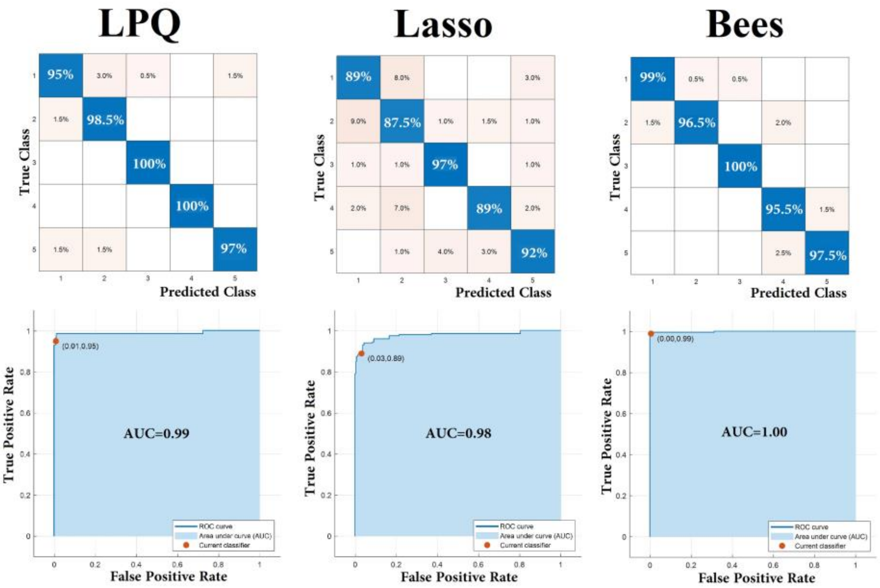

Fig 5. Confusion matrixes and ROC curves for LPQ, Lasso and Bees Algorithms with 256, 64 and 64 features with SVM classifier

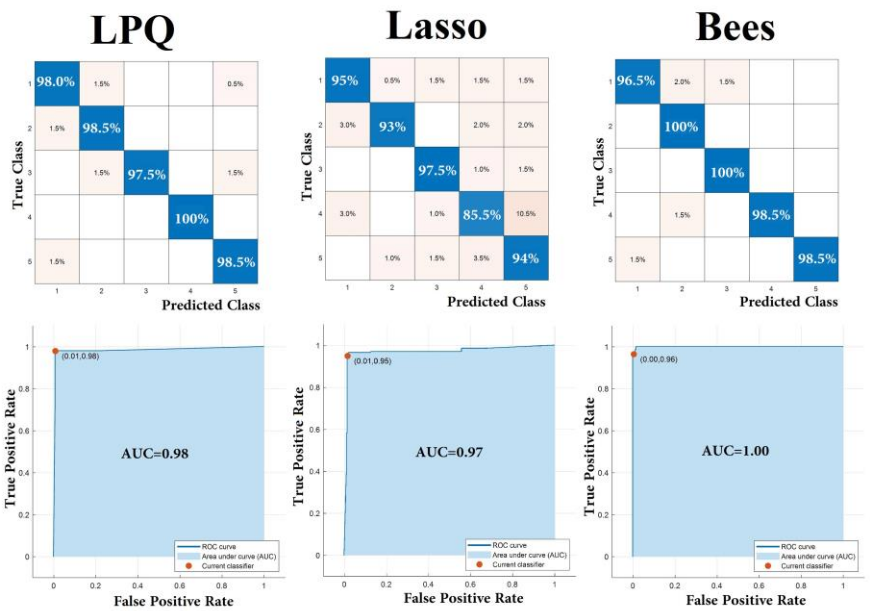

Fig 6. Confusion matrixes and ROC curves for the LPQ, Lasso and Bees Algorithms with 256, 64 and 64 features with the KNN classifier

Looking at Table 2 (results), the best results belong to Bees feature selection, PSO feature selection and PCA feature selection for different numbers of features. Additionally, Lasso feature selections are placed at the end of the ranking according to its weak performance. Additionally, solo LPQ with all 256 feature places is in the middle of the ranking table.

PSO and Bees compete with each other with different classifiers, and the best result ever belongs to the Bees feature selection algorithm with 128 features and by the subspace KNN classifier (99.8 % accuracy). However, the weakest performance goes to Lasso with 32 features and with the SVM classifier (84.5 % accuracy). Clearly, accuracy is the number of correctly classified samples divided by the number of all samples multiplied by 100. Here, facial expressions are labelled numerically as 1: Anger, 2: Joy, 3: Fear, 4: Disgust and 5: Surprise. Additionally, the area under the ROC curve for Bees features and for all classifiers returned higher values (closer to 1) by having higher true positive rates and lower false positive rates. Additionally, Fig. 7 depicts fast convergence for BA feature selection, which occurred in the initial iterations (iterations 20 to 25). This means that BA feature selection could obtain proper features very quickly. However, more iterations mean having the best features possible.

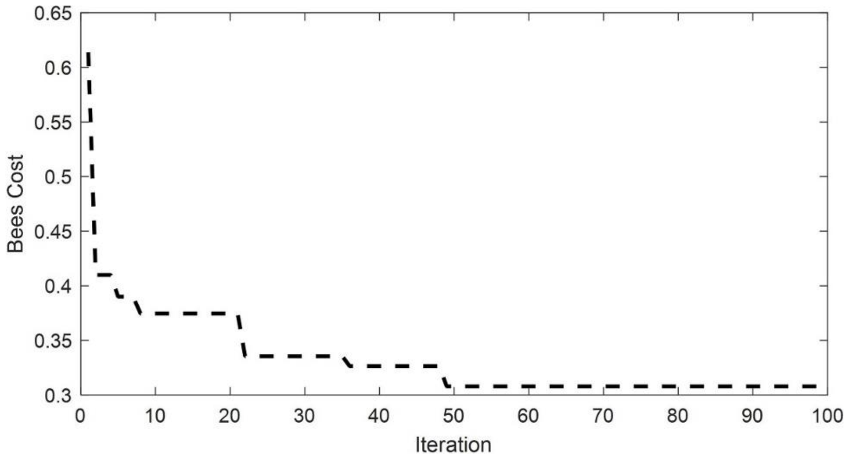

Fig 7. Bees Algorithm performance over 100 iterations for the feature selection task

4 Conclusion and suggestions

Employing bioinspired or nature-inspired algorithms for feature selection tasks could outperform traditional feature selection algorithms in face analysis and especially facial expression recognition tasks. The capability of extracting facial impactful features is a very difficult task. BA feature selection was able to perform this task very well, which was significant. Defining the feature selection cost function for the Bees Algorithm returned successful results for all four classification algorithms. Therefore, it can be concluded that the Bees Algorithm has a great impact on feature selection optimisation after feature extraction by LPQ features. However, the run time increases, but the Bees Algorithm using just a quarter of the features could reach a point that traditional algorithms are not capable of. Additionally, the proposed feature selection has fast convergence; in the case of having fewer categories with distinguished features, the number of iterations could be very small, which saves considerable time. Furthermore, using the proposed method on more facial expressions (7 to 15) and in other face analysis tasks, such as face recognition, age estimation, gender detection, and face recognition under makeup, are future works. Additionally, it is suggested to use the proposed method with more classification algorithms, such as tree and naive Bayes classifiers. In addition, comparing the proposed method with more traditional and bioinspired feature selection techniques, such as the chi-square test and Genetic Algorithm (GA), is another suggestion.

References

1. Mousavi, Seyed Muhammad Hossein. "A new way to age estimation for RGB-D images, based on a new face detection and extraction method for depth images." International Journal of Image, Graphics and Signal Processing 10 (2018): 10.
2. Mousavi, Seyed Muhammad Hossein, and S. Younes Mirinezhad. "Iranian kinect face database (IKFDB): a color-depth based face database collected by kinect v. 2 sensor." SN Applied Sciences 3.1 (2021): 1-17.
3. Ekman, Paul, and Wallace V. Friesen. "Facial action coding system." Environmental Psychology & Nonverbal Behavior (1978).
4. Singh, Abhilash, Sandeep Sharma, and Jitendra Singh. "Nature-inspired algorithms for wireless sensor networks: A comprehensive survey." Computer Science Review 39 (2021): 100342.
5. Mousavi, Seyed Muhammad Hossein, Vincent Charles, and Tatiana Gherman. "An evolutionary Pentagon Support Vector finder method." Expert Systems with Applications 150 (2020): 113284.

6. Singh, Abhilash, et al. "A Gaussian process regression approach to predict the k-barrier coverage probability for intrusion detection in wireless sensor networks." Expert Systems with Applications 172 (2021): 114603.
7. Mousavi, Seyed Muhammad Hossain. "A New Clustering Method Using Evolutionary Algorithms for Determining Initial States, and Diverse Pairwise Distances for Clustering." International Journal of Mechatronics, Electrical and Computer Technology (IJMEC) 9.31 (2019): 4098-4110.
8. Ji, Bai, et al. "Bioinspired feature selection: An improved binary particle swarm optimization approach." IEEE Access 8 (2020): 85989-86002.
9. Pham, D. T., et al. "Data clustering using the Bees Algorithm." (2007).
10. Pham, D. T., et al. "The Bees Algorithm—a novel tool for complex optimization problems." Intelligent production machines and systems. Elsevier Science Ltd, 2006. 454-459.
11. Pham, D. T., et al. "The Bees Algorithm." Technical Note, Manufacturing Engineering Centre, Cardiff University, UK (2005): 44-48.
12. Gonzalez-Jorge, H., et al. "Metrological comparison between Kinect I and Kinect II sensors." Measurement 70 (2015): 21-26.
13. Charles, Vincent, and Tatiana Gherman. "Achieving competitive advantage through big data. Strategic implications." Middle-East Journal of Scientific Research 16.8 (2013): 1069-1074.
14. Mousavi, Seyed Muhammad Hossein, VB Surya Prasath, and Seyed Muhammad Hassan Mousavi. "Persian classical music instrument recognition (PCMIR) using a novel Persian music database." 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE). IEEE, 2019.
15. Mousavi, Seyed Muhammad Hossein, S. Younes MiriNezhad, and Atiye Mirmoini. "A new support vector finder method, based on triangular calculations and K-means clustering." 2017 9th International Conference on Information and Knowledge Technology (IKT). IEEE, 2017.
16. Dezfoulian, Mir Hossein, et al. "Optimization of the Ho-Kashyap classification algorithm using appropriate learning samples." 2016 Eighth International Conference on Information and Knowledge Technology (IKT). IEEE, 2016.
17. Yuan, Baohua, Honggen Cao, and Jiuliang Chu. "Combining local binary pattern and local phase quantization for face recognition." 2012 International Symposium on Biometrics and Security Technologies. IEEE, 2012.
18. Ismail, Asrul Harun, et al. "Combinatorial Bees Algorithm for Vehicle Routing Problem." Macromolecular Symposia. Vol. 396. No. 1. 2021.
19. Abdi, Hervé, and Lynne J. Williams. "Principal component analysis." Wiley interdisciplinary reviews: computational statistics 2.4 (2010): 433-459.
20. Tibshirani, Robert. "Regression shrinkage and selection via the lasso." Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.
21. Kibria, B. M., and Shipra Banik. "Some ridge regression estimators and their performances." (2020).
22. Jin, Xin, et al. "Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles." International workshop on data mining for biomedical applications. Springer, Berlin, Heidelberg, 2006.

23. Shiraz, Rashed Khanjani, Vincent Charles, and Leila Jalalzadeh. "Fuzzy rough DEA model: A possibility and expected value approaches." Expert Systems with Applications 41.2 (2014): 434-444.
24. Zhang, Yishi, et al. "Feature selection using data envelopment analysis." Knowledge-Based Systems 64 (2014): 70-80.
25. Mistry, Kamlesh, et al. "A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition." IEEE transactions on cybernetics 47.6 (2016): 1496-1509.
26. Emary, Eid, et al. "Firefly optimization algorithm for feature selection." Proceedings of the 7th balkan conference on informatics conference. 2015.
27. Alomari, Osama, and Zulaiha Ali Othman. "Bees algorithm for feature selection in network anomaly detection." Journal of applied sciences research 8.3 (2012): 1748-1756.
28. Ding, Hongwei, et al. "Imbalanced data classification: A KNN and generative adversarial networks-based hybrid approach for intrusion detection." Future Generation Computer Systems 131 (2022): 240-254.
29. Ramasamy, Karthikeyan, Kiruthika Balakrishnan, and Durgadevi Velusamy. "Detection of cardiac arrhythmias from ECG signals using FBSE and Jaya optimized ensemble random subspace K-nearest neighbor algorithm." Biomedical Signal Processing and Control 76 (2022): 103654.
30. Mahdizadeh, M., and Ehsan Zamanzade. "On estimating the area under the ROC curve in ranked set sampling." Statistical Methods in Medical Research (2022): 09622802221097211.
31. Gupta, Varun, et al. "Detection of R-peaks using fractional Fourier transform and principal component analysis." Journal of Ambient Intelligence and Humanized Computing 13.2 (2022): 961-972.
32. Fraś, Mieszko, Marcin Witkowski, and Konrad Kowalczyk. "Convolutional Weighted Minimum Mean Square Error Filter for Joint Source Separation and Dereverberation." ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022.