Intelligent Facial Expression Recognition Using Particle Swarm Optimization Based Feature Selection

Adam Robson
School of Computer and Information Science
Northumbria University
Newcastle upon Tyne, United Kingdom
adamrobson27@gmail.com

Li Zhang
School of Computer and Information Science
Northumbria University
Newcastle upon Tyne, United Kingdom
li.zhang@northumbria.ac.uk

***Abstract*—Particle Swarm Optimization (PSO) has become a popular method of feature selection in classification problems, due to its powerful search capability and computational simplicity. Classification problems, such as facial emotion recognition, often involve data sets containing high volumes of features, not all of which are useful for classification. Redundant and irrelevant features have the potential to negatively impact the performance and accuracy of facial emotion recognition systems. The feature selection process identifies the most relevant features to achieve improved classification performance. While the use of PSO as a feature selection method in facial emotion recognition systems has seen some successes, it is still susceptible to the issue of premature convergence. This work presents seven PSO variants which mitigate against the premature convergence problem through the incorporation of three random probability distributions (Cauchy, Gaussian and Lévy). At each iteration of the proposed PSO models, probability distributions are used to increase search diversity and reduce the number of redundant features used for classification. The seven PSO variants presented in this study have demonstrated positive results when tested on real world data sets, outperforming the standard PSO model and other related work within the field.**

Keywords— Particle Swarm Optimization, classification, facial expression recognition, feature selection.

I. Introduction

Facial emotion recognition is an important field and could change the way in which we interact with the next generation of computer systems. The ability of a system to accurately classify and appropriately respond to the emotional state of the user has numerous implications on the way in which humans interact with computers. Such systems will likely process and classify emotions using a machine learning component and will involve large data sets, containing high volumes of features. For classifiers to work to the highest possible accuracy and efficiency, the features contained within the data sets must be as optimized, containing the most robust and significant characteristics required for classification.

This study presents the use of modified Particle Swarm Optimization (PSO) for feature selection in a facial emotion recognition system. PSO is a swarm intelligence optimization algorithm. Proposed in 1995 by Russel Eberhart and James Kennedy, it is inspired by the collaborative behaviour and swarming in biological populations, such as flocks of birds or schools of fish. In the PSO algorithm, the swarm consists of particles is used to explore the search space, and each particle consists of three components: a representation of a possible solution, a velocity and a representation of the closest that the current particle has come to the target criteria (referred to as a personal best). PSO contains three globally accessible variables: target criteria, a representation of the particle which has come closest to the termination criteria (referred to as the global best), and a termination value indicating when the PSO algorithm will stop if the target criteria are not met. A fitness function is used to evaluate the current position of a particle and calculate its personal best. If a particle personal best is better than the global best, both the global best and global particle solutions are updated. Particles move within the search space in search for optimal solutions, updating its position and velocity according to its own experience and that of neighbouring particles. Particles will continue to explore the search space until the maximum number of moves, or the termination criterion, is met.

PSO has established itself as a powerful and widely used technique in classification problems across a variety of domains [1-6], due to its computational simplicity and powerful search capabilities. This includes implementations in feature selection systems designed to remove redundant and irrelevant features and improve classification efficiency. While PSO has become a widely adopted optimization method for feature selection, it is still susceptible to the issue of premature convergence, as noted in Mistry et al. [1]. Premature convergence occurs when the runtime of a PSO increases and the momentum/velocity of particles reduces, decreasing the search diversity and causing a tendency to converge in a local optimum area or at a single point. To increase search diversity, this study proposes seven PSO algorithm variants which incorporate different combinations of three random probability distributions: Cauchy, Gaussian and Lévy.

II. Experiment Design

The functionality of the experimentation system can be described in two parts; Part 1: Feature Extraction, and Part 2: Feature Selection and Classification. Part 1 initially

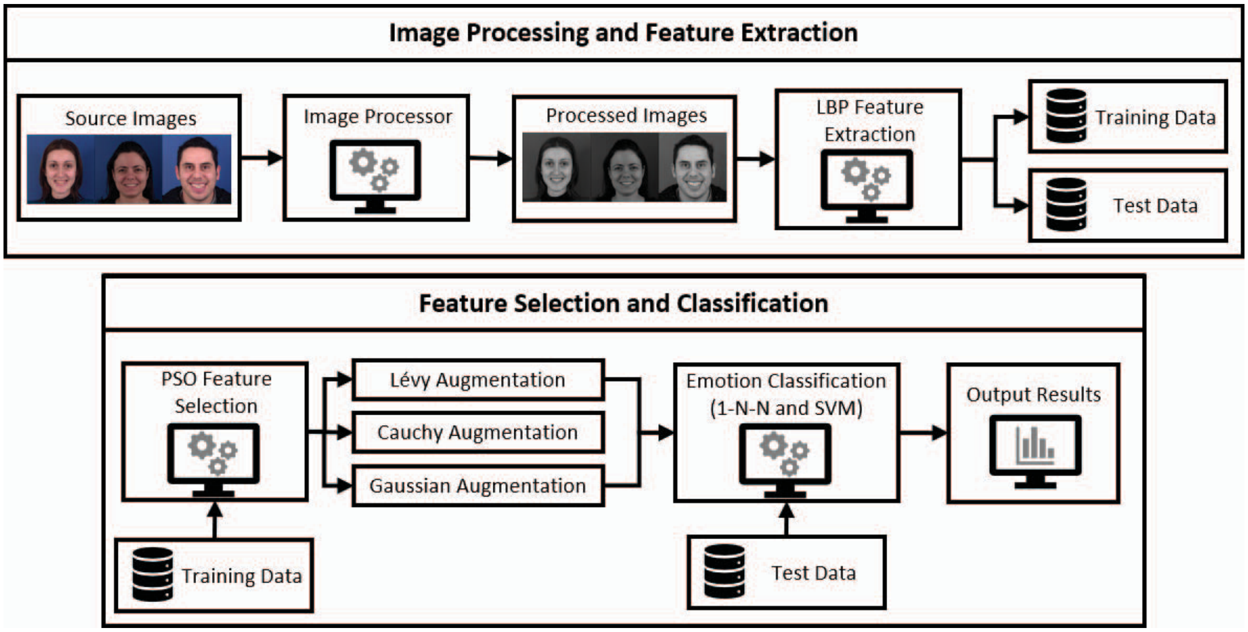

Fig. 1. System Overview: Image Processing and Feature Extraction (Part 1), and Feature Selection and Classification (Part 2)

standardises all source images by converting them to greyscale images, then a Local Binary Pattern (LBP) is used to extract features and generate the training and testing datasets. For Feature Selection, Part 2 uses the PSO implementations, outlined in section III to analyse both datasets. Features are selected based upon the scores returned by the fitness evaluation function, described in section IV. A full system overview can be seen in Figure 1.

This study focuses on the extraction of features and classification of the six basic emotions: Happiness, Sadness, Anger, Surprise, Fear and Disgust. This is the most common set of emotions classified by work in this field and therefore results generated in this study will be comparable to other research. Datasets employed for the evaluation of this research include the Multimedia Understanding Group (MUG) Facial Expression Database [7]. Images displaying the peak emotion intensity for Happiness, Sadness, Anger, Surprise, Fear and Disgust were selected to create datasets consisting of 30 images for training and 20 images for testing. Classification is achieved through the comparison of the patterns generated by the features selected in the training data, with the features selected from the test data. The system uses two classifiers to provide emotion recognition, a k-Nearest-Neighbour (k-N-N) and a Support Vector Machine (SVM), both of which are outlined in section VI.

III. PSO Variant Combinations

To maintain consistency and validity of results, the same set of experiments was conducted on each of the seven proposed PSO algorithm variants, using the same data sets. Experiments were designed with the intention of yielding a set of results comparable to several other studies within the field [8][9][10][11], all of which present unique PSO variants for feature selection in emotion classification systems, with data sets and experiment designs similar to those presented in this study. The standard PSO and each of the proposed PSO variants are described below:

1) Standard PSO: This will be used as the baseline for the other seven algorithm variants.

2) Proposed Algorithm Variant 1: Standard PSO with Lévy Distribution: A standard PSO, combined with the Lévy Distribution.

3) Proposed Algorithm Variant 2: Standard PSO with Lévy and Cauchy Distribution: This version of the proposed algorithm incorporates both Lévy and Cauchy distributtions with the standard PSO.

4) Proposed Algorithm Variant 3: Standard PSO with Lévy and Gaussian Distribution: This variant uses both Lévy and Gaussian random walks to manipulate the global best solution and update it as necessary.

5) Proposed Algorithm Variant 4: Standard PSO with Lévy, Cauchy and Gaussian Distribution: This variant of the proposed algorithm implements all three distributions to the standard PSO algorithm.

6) Proposed Algorithm Variant 5: Standard PSO with Cauchy and Gaussian Distribution: This variant of the proposed algorithm incorporates both Cauchy and Gaussian distributions into the PSO algorithm.

7) Proposed Algorithm Variant 6: Standard PSO with Cauchy Distribution: This version of the proposed algorithm includes the Cauchy function.

8) Proposed Algorithm Variant 7: Standard PSO with Gaussian Distribution: The final variant is a standard PSO combined with the Gaussian distribution.

IV. Fitness Function

The fitness evaluation function for each particle, p, is shown in Equation (1), where f denotes fitness, a indicating accuracy for each expression and t representing the number of selection features. w_a and w_b denote the weights for the classification performance and the number of selected features, respectively. The fitness function consists of two criteria: the number of selected features and the classification performance. The standard PSO and the proposed models are applied to each of the six emotions separately, to allow for the identification of features for each distinct expression.

$$f(p) = w_a \times a_p + w_b \times t_p \tag{1}$$

To avoid any potential bias toward specific emotion categories during optimization, the accuracy shown in equation (1) indicates the accuracy of each separate emotion expression, as opposed to a combined accuracy across all expression categories. Predefined weights for classification accuracy (w_a) and the number of selected features (w_b), are used with $w_a = 1 - w_b$. Additionally, parameters w_a and w_b indicate the relative importance of classification performance and the number of features selected. In this study, classification performance is considered to be of more important than the number of selected features, and therefore w_a (0.9) is assigned with a higher value than that of w_b (0.1). A detailed analysis, review and comparison of the results of the selected studies and this research study can be found in section VIII.

V. Random Probability Distributions

A. Gaussian Probability Distribution

In probability theory, Gaussian distribution gives a representation of data that clusters around a mean and the graph of the correlating probability density peaks at the mean value. Gaussian was chosen as one of the random walks, which shows different characteristics to those of other random probability distributions used within this research. The Gaussian distribution has also been utilized in similar research studies, for example work conducted by Lee and Lee in 2013 [12] and Mishra et al. in 2017 [13]. Both studies implement Gaussian distribution as a mutation to alleviate premature convergence issues, with [12] using Gaussian to manipulate the global best solution and [13] manipulating individual particle positions. These studies generated interesting results through their implementation of Gaussian mutations, demonstrating both increased search diversity and improved performance.

B. Lévy Distribution

The Lévy distribution is a non-Gaussian random distribution algorithm, introduced by French mathematician Paul Pierre Lévy and it is a statistical description of motion. A Lévy distribution is composed of a cluster of both long and short steps. The ability of Lévy is to provide varying jump distances of either long jumps, further away from the mean value, or short jumps, closer to the mean value, could potentially increase the search variance, as the other random probability distributions used, Gaussian and Cauchy, do not provide the same versatility on their own. The Lévy distribution has been applied to PSO before in 2014 by Hakli and Uguz [14], and in 2017 by Barisal et al. [15]. Both studies applied Lévy in an attempt to avoid the premature convergence and improve the overall global search capability of the standard PSO. Promising performance increases were noted in both works.

C. Cauchy Distribution

Cauchy distribution is a continuous probability function similar to Gaussian, except it has an undefined mean and an infinite variance, and therefore does not have finite moments of order. Cauchy was selected for experimentation as it is heavier tailed than Gaussian, which means that it has a tendency to produce values that fall farthest from its mean, which makes it more likely to escape premature convergence and increase search diversity. A system proposed by Wu and Law [16] used both Cauchy and Gaussian distributions when the PSO was deciding the next move of the particle. Q. They compared the Cauchy Gaussian PSO with a Standard PSO and a Gaussian PSO and demonstrated that a Cauchy and Gaussian PSO generated a better set of results than a Standard PSO, or a Gaussian PSO.

VI. Classifiers

A. k-N-N Classifier

The k-Nearest-Neighbour classifier (k-N-N), is a simple algorithm for predicting the classification of a test case based on the values of training data provided. The k-N-N classifier has non-parametric statistics, which means that the data passed to the classifier does not need to conform to a normal distribution. It also assumes that the data provided is non-characteristic in its structure and has no parameters. Classification is achieved through levels of similarity between the test data provided and the training data provided in the feature space. For example, the process for 1-N-N to make a classification is as follows: in order to classify x from the test data, it will find its closest neighbour within the training data, labelling it $x\hat{}i$ and then assign x the value of $x\hat{}i$ [17]. Works such as Guru et al. [17] and Tran et al. [18] show us that the k-N-N classifier is a useful function for benchmarking the performance of feature extraction and recognition systems. The k-N-N classifier was also selected because it is a simple to configure, parameter free classifier, that can efficiently test the quality of features.

B. Support Vector Machine

A Support Vector Machine (SVM), introduced in 1992 by Boser et al. [19], is a supervised learning algorithm that analyses data for classification and regression analysis. SVMs take two sets of structured data, with every data element labelled to identify the record as being in one of two categories. Training data is used by the SVM to construct a model which identifies records from the Testing data to one of the two categories. The SVM model has been widely used for diverse classification problems [20]. For instance, Abdulrahman and Eleyan [21] presented a facial expression recognition system which implemented a SVM for emotion classification. They implemented two feature extraction methods, Principal Component Analysis (PCA) and Local Binary Pattern (LBP), with Local Binary Pattern being used with different parameters (LBP1 = 16 regions and LBP2 = 64 regions). SVM was selected as a classifier for this study as it is powerful and robust enough to deal with errors and still present useful data. SVM has also been used widely within the field of facial emotion and expression recognition, which allows for an appropriate and clear comparison to the results of this study with related research.

VII. RESULTS

A. Overall Emotion Classification Results

After conducting the experiments outlined in the previous section, the resulting data is presented here. Throughout the experiments conducted, the performance of the standard PSO is used as the baseline for which to compare with those of other methods. Table I shows the overall average performance of each proposed algorithm. Tables II – VII show the results for each of the six emotions used for recognition. Table VIII shows the results of this study compared to similar work within the field. The 1-N-N classifier returned a consistent average across each variant algorithm, and the classifier's highest performer was the PSO combined with Gaussian distribution (7), with the PSO integrated with Lévy and Gaussian distributions (8) showing the second highest level of accuracy and the other variants very close behind. Overall, the highest performer shown with the SVM classifier was the PSO incorporated with Lévy, Cauchy and Gaussian distribution (4), with all other proposed algorithm variants returning a slightly lower level of accuracy. On evaluation of the overall average performances, the proposed algorithm variants have outperformed the standard PSO implementation. Differences in performance were evident within the individual emotion datasets, on average all proposed algorithm variants yielded similar levels of accuracy, precision and recall.

This text can be removed.

TABLE I. AVERAGE PERFORMANCE

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	43	57	90.48	91.91	86.42	72.22	95.84	78.05
PL	40	56	92.75	94.18	89.42	74.56	98.1	80.66
PLC	40	57	92.78	94.39	90.13	75.72	98.12	81.6
PLCG	38	56	92.87	94.44	89.32	76.89	97.94	82.01
PC	41	59	92.51	93.97	88.17	75.06	97.76	80.34
PCG	40	58	92.71	94.24	89.76	74.94	98.1	81.01
PG	38	57	93.04	94.14	89.73	74.11	98.14	80.52
PLG	43	57	90.48	91.91	86.42	72.22	95.84	78.05

[a] Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), [b] Minimum Number of Selected Features, [c] Maximum Number of Selected Features, [d] 1-N-N Accuracy, [e] SVM Accuracy, [f] Average Precision, [g] Average Recall, [h] Average Specificity, [i] F-Measure

B. Individual Emotion Classification Results

To generate results comparable with other work within the field, this study conducted classification experiments with all proposed algorithm variants, on six basic emotions commonly used for testing; Happiness, Sadness, Anger, Surprise, Fear and Disgust. The results of the conducted classification experiments are presented in this section. Table II shows the overall results generated for Emotion One: Happiness. PSO combined with Lévy and Gaussian (8) shows the best results in all areas apart from Average Precision and Average Specificity, where PSO integrated with Lévy and Cauchy (3) returned better results. The highest scores for the Average Precision are obtained by (8) and (3).

TABLE II. EMOTION ONE: HAPPINESS

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	46	57	93.5	95.83	87.37	88	97.4	87.54
PL	43	57	93.56	96.56	89.74	90	97.87	89.7
PLC	44	58	93.89	96.94	92.48	89.33	98.47	90.72
PLCG	38	58	93.33	96.17	88.46	89	97.6	88.5
PC	43	61	92.72	96.39	88.18	90.67	97.53	89.29
PCG	42	58	93.89	96.33	88.54	90	97.6	89.11
PG	38	57	93.06	96.28	89.42	88.67	97.8	88.88
PLG	46	65	94.56	97.11	91.52	91.33	98.27	91.32

[a] Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), [b] Minimum Number of Selected Features, [c] Maximum Number of Selected Features, [d] 1-N-N Accuracy, [e] SVM Accuracy, [f] Average Precision, [g] Average Recall, [h] Average Specificity, [i] F-Measure

Table III shows the results for Emotion Two: Sadness. The proposed model (4), i.e. PSO combined with Lévy, Cauchy and Gaussian, achieves the best performance. Results for Emotion Three: Anger are shown in Table IV.

TABLE III. EMOTION TWO: SADNESS

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	44	56	93.33	94	85.47	78	97.2	81.34
PL	41	55	93.56	94.39	89.26	76	98.07	81.94
PLC	43	57	93.5	94.28	88.84	77	97.73	82.11
PLCG	43	56	93.22	94.67	89	78.33	97.93	83.04
PC	36	57	92.33	93.44	86.03	73.33	97.47	78.99
PCG	40	66	92.89	94.22	87.84	76.67	97.73	81.67
PG	38	52	93.39	93.28	84.51	74	97.13	78.71
PLG	39	53	92.56	93.67	86.44	74.33	97.53	79.64

[a] Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), [b] Minimum Number of Selected Features, [c] Maximum Number of Selected Features, [d] 1-N-N Accuracy, [e] SVM Accuracy, [f] Average Precision, [g] Average Recall, [h] Average Specificity, [i] F-Measure

TABLE IV. EMOTION THREE: ANGER

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	44	57	92.11	93.67	95.24	65.67	99.27	77.43
PL	40	55	93.33	93.61	93.76	66.33	99.07	77.57
PLC	36	56	93.11	93.78	95.55	66	99.33	77.91
PLCG	38	52	93.44	94.22	94.47	69.67	99.13	80.04
PC	40	59	92.44	93.56	94.22	66	99.07	77.18
PCG	36	57	92.72	93.33	94.52	64.33	99.13	76.26
PG	35	54	93.94	94.33	97.61	67.67	99.67	79.83
PLG	39	53	92.56	94.11	97.19	66.67	99.6	78.9

a. Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), b. Minimum Number of Selected Features, c. Maximum Number of Selected Features, d. 1-N-N Accuracy, e. SVM Accuracy, f. Average Precision, g. Average Recall, h. Average Specificity, i. F-Measure

PSO integrated with the Gaussian distribution (7) is the highest performer for Average 1-N-N Accuracy, Average SVM Accuracy, Average Precision and Average Specificity. Table V shows the results for Emotion Four: Surprise, PSO incorporated with Cauchy and Gaussian distributions (6) showing the best results for Average SVM Accuracy (97.39%), Average Precision (98.53%), Average Specificity (99.73%) and F-Measure (91.52%), although differences between other high performers are marginal in most cases.

TABLE V. EMOTION FOUR: SURPRISE

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	42	58	93.67	96.56	95.69	83	99.27	88.8
PL	39	60	93.44	97.11	95.34	87	99.13	90.87
PLC	39	59	93.22	97.22	95.8	87.33	99.2	91.3
PLCG	36	56	94	96.78	94.26	87.33	98.67	90.37
PC	42	61	94.78	96.67	94.22	86.33	98.73	89.79
PCG	41	58	94.11	97.39	98.53	85.67	99.73	91.52
PG	38	58	94.61	97.22	96.79	86.33	99.4	91.09
PLG	37	57	94.33	96.11	93.35	83	98.73	87.73

a. Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), b. Minimum Number of Selected Features, c. Maximum Number of Selected Features, d. 1-N-N Accuracy, e. SVM Accuracy, f. Average Precision, g. Average Recall, h. Average Specificity, i. F-Measure

Results for Emotion Five, Fear, are shown in Table VI, with PSO integrated with the Gaussian function (7) performing nominally better than other PSO variants.

TABLE VI. EMOTION FIVE: FEAR

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	38	58	91.83	93.44	94.67	64.67	99.2	76.65
PL	37	55	92.06	93.22	95.87	62	99.47	75.11
PLC	38	52	91.44	92.94	94.15	61.67	99.2	74.4
PLCG	36	58	92.39	93.78	95.44	66	99.33	77.91
PC	40	61	91.83	93.28	93.89	64	99.13	75.97
PCG	40	54	92.33	93.56	94.2	66.33	99	77.34
PG	39	62	92.5	93.39	96.21	63	99.47	75.96
PLG	38	56	92.22	93.5	95.7	64	99.4	76.62

a. Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), b. Minimum Number of Selected Features, c. Maximum Number of Selected Features, d. 1-N-N Accuracy, e. SVM Accuracy, f. Average Precision, g. Average Recall, h. Average Specificity, i. F-Measure

Table VII shows the results generated for Emotion Six: Disgust. PSO combined withLévy and Cauchy (3) performed better than all other algorithm variants overall, with the other six proposed variants close behind and relatively low variances present throughout performances.

TABLE VII. EMOTION SIX: DISGUST

Exp[a]	Min FS (%)[b]	Max FS (%)[c]	1NN (%)[d]	SVM (%)[e]	AP (%)[f]	AR (%)[g]	AS (%)[h]	FM (%)[i]
PSO	43	58	90.44	89.94	72.05	66	94.73	68.54
PL	41	53	90.56	90.17	72.55	66	95	68.79
PLC	41	59	91.5	91.17	73.96	73	94.8	73.18
PLCG	38	58	90.83	91	74.31	71	95	72.18
PC	42	55	90.94	90.5	72.48	70	94.6	70.83
PCG	41	57	90.33	90.61	74.94	66.67	95.4	70.14
PG	41	57	90.72	90.33	73.86	65	95.4	68.65
PLG	37	58	91.28	90.61	73.62	68.67	95	70.71

a. Experiments (PSO: Standard PSO, PL: PSO combined with Lévy, PLC: PSO combined with Lévy and Cauchy, PLCG: PSO combined with Lévy, Cauchy and Gaussian, PC: PSO combined with Cauchy, PCG: PSO combined with Cauchy and Gaussian, PG: PSO combined with Gaussian, PLG: PSO combined with Lévy and Gaussian), b. Minimum Number of Selected Features, c. Maximum Number of Selected Features, d. 1-N-N Accuracy, e. SVM Accuracy, f. Average Precision, g. Average Recall, h. Average Specificity, i. F-Measure

A discussion of results generated from this research, along with a comparison of results with other work within the field is provided in section VIII.

VIII. DISCUSSION

The PSO algorithm variants proposed in this study have yielded a higher level of performance than the standard PSO implementation they were benchmarked against. This section compares the results of this study to several related studies within the field [8][9][10][11], with a comparison shown in Table VIII.

TABLE VIII. COMPARISON WITH RELATED STUDIES

Emotion	Zhang et al.	Wu et al.	Jain et al.	Xia et al.	PL[a]	PLC[b]	PLCG[c]	PC[d]	PCG[e]	PG[f]	PLG[g]
1.) Happiness	80	87.7	98.55	72	96.56	96.94	96.17	96.39	96.33	96.28	97.11
2.) Sadness	60	78.4	77.22	73.5	94.39	94.28	94.67	93.44	94.22	93.28	93.67
3.) Anger	90	82.9	76.71	71.5	93.61	93.78	94.22	93.56	93.33	94.33	94.11
4.) Surprise	77	87.9	99.06	66	97.11	97.22	96.78	96.67	97.39	97.22	96.11
5.) Fear	65	66.7	94.37	59.5	93.22	92.94	93.78	93.28	93.56	93.39	93.5
6.) Disgust	83	67.7	81.51	62	90.17	91.17	91	90.5	90.61	90.33	90.61
7.) Average	75.83	78.55	85.84	67.42	94.18	94.39	94.44	93.97	94.24	94.14	94.19

[a.] PSO combined with Lévy, [b.] PSO combined with Lévy and Cauchy, [c.] PSO combined with Lévy, Cauchy and Gaussian, [d.] PSO combined with Cauchy, [e.] PSO combined with Cauchy and Gaussian, [f.] PSO combined with Gaussian, [g.] PSO combined with Lévy and Gaussian

The first study used to provide a comparison of results is a system proposed by Zhang et al. [8], which concentrates on intelligent neural network based facial emotion recognition for a humanoid robot. Zhang et al. implemented a system which utilized the standard vision APIs of the NAO humanoid robot platform and a neural network based recognizer. Zhang et al. created an extensive training dataset for the proposed system, along with an equally large testing dataset. Wu et al. [9] is the second study used for comparison, implementing spatiotemporal Gabor filters for automatic facial emotion recognition in their research study and using a SVM for classification. The third study used for comparison is Facial Expression Recognition with Temporal Modelling of Shapes, presented by Jain et al. [10]. In their study, Jain et al. created a facial emotion recognition system which used video sequences and latent-dynamic conditional random fields. Their proposed system performed temporal modelling of shapes and evaluated these by using image sequences and a SVM classifier. The final study used for comparison, Facial Expression Recognition based on SVM [11], in which Xia presents an in-depth study of a multi-classification SVM. Xia [11] presents a system which uses Gabor vectors for facial feature extraction, and the SVM to achieve classification. From the collated results presented in Table VIII, it can be seen that the overall average performance of each proposed algorithm variant outperforms the average performance of other research in the field. It should also be noted that Jain et al. out performs all proposed algorithm variants for Happiness (1), Surprise (4) and Fear (5). The results presented by Jain et al. demonstrate a relatively high variance (9.3%) and a lack of consistent performance across the all emotions. There are large differences between the best performing emotion, Surprise (99.06%), and the worst performing emotion, Anger (76.71%). The levels of accuracy in emotion recognition presented in this research, are high in comparison with those of similar research within the field. The level of consistency provided by the proposed algorithm variants in terms of accuracy, precision and recall is extremely positive and demonstrates that the system is robust enough to handle different types of emotions.

While all proposed algorithm variants performed with high levels of accuracy, recall and precision when compared to the standard PSO implementation, the Lévy distribution is present in the majority of the highest performers. It is believed that this is due to the increased diversity and global exploration of the search space that the Lévy distribution offers. The combination of several random walk strategies is shown to increase the search diversity of the standard PSO, and also increase both local and global exploration. Overall, the results produced by this research study suggests that the premature convergence problem of the standard PSO can be avoided through manipulation of the global best, resulting in a better overall level of performance, Further testing is required more diverse datasets, to fully benchmark the performance of the proposed PSO algorithm variants in different areas.

IX. CONCLUSION

Overall, the proposed PSO algorithm variants presented in this study provide a higher level of accuracy than the standard PSO implementation. The proposed algorithms also outperformed other work in the field in most cases. The summary of the research contributions of this study is as follows:

1) All of the proposed algorithms outperformed the standard PSO implementation, via the evaluation of several datasets to benchmark the performance of the algorithms.

2) The Lévy distribution is present in the majority of the highest performing algorithms, due to the increased search diversity and global exploration of the search space offered by this particular random walk strategy.

3) Combinations of several random walk strategies have also demonstrated an increase in the search diversity and an increase in the local and global exploration of the original PSO implementation. Therefore, the proposed PSO variants outperform the original PSO consistently.

4) The proposed PSO variants also outperform most other related facial expression research reported in the literature. These proposed algorithms can also be applied to other datasets for feature optimization and dimensionality reduction.

While further testing is still required, the results generated by the experiments presented in this study can be viewed in a positive light and all of the proposed PSO algorithm outperformed the standard PSO model. This research also shows great potential when compared to the results of other current research within the field and demonstrated a high level of consistency in terms of performance. As the proposed algorithms variants show improved levels of accuracy, precision and recall, they could be applied to a variety of datasets to improve optimization and reduce dimensionality in machine learning or decision support systems such as object recognition, disease diagnosis (e.g. skin cancer detection, breast cancer detection, retinal disease detection etc.), stock market prediction, housing price or sale prediction and hyper-parameter tuning of diverse classification and deep learning models.

REFERENCES

[1] Mistry, K., Zhang, L., Neoh, S.C., Lim, C.P. and Fielding, B. (2017). A micro-GA Embedded PSO Feature Selection Approach to Intelligent Facial Emotion Recognition. IEEE Transactions on Cybernetics. 47 (6) 1496–1509. (IF: 7.384, Journal Ranking 4%).

[2] Zhang, L., Mistry, K., Neoh, S.C. and Lim, C.P. (2016). Intelligent facial emotion recognition using moth-firefly optimization. Knowledge-Based Systems. Volume 111, Nov. 2016, 248–267. (IF: 4.529, Journal Ranking 12%).

[3] Neoh, S.C., Srisukkham, W., Zhang, L, Todryk, S., Greystoke, B., Lim, C.P., Hossain, A. and Aslam, N. (2015) An Intelligent Decision Support System for Leukaemia Diagnosis using Microscopic Blood Images. Scientific Reports, 5 (14938). Nature Publishing Group, ISSN 2045-2322.

[4] Zhang, Yang, Zhang, Li, Neoh, Siew Chin, Mistry, Kamlesh and Hossain, Alamgir (2015) Intelligent affect regression for bodily expressions using hybrid particle swarm optimization and adaptive ensembles. Expert Systems with Applications, 42 (22). pp. 8678-8697.

[5] S. Vijayarani and S. Priyatharsini, (2016). Particle Swarm Optimization Algorithm for Facial Image Expression Classification. International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 9, No. 9, pp 11-24.

[6] Y. Zhang, D. Gong, Y. Hu, W. Zhang, Feature selection algorithm based on bare bones particle swarm optimization, Neurocomputing 148 (2015) 150–157.

[7] N. Aifanti et al., (2010). The MUG Facial Expression Database, in Proc. 11th Int. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Desenzano, Italy, April 12-14 2010

[8] L. Zhang et al., (2013). Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. Expert Systems, 5160–5168.

[9] T. Wu et al., (2010). Facial expression recognition using Gabor motion energy filters. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), IEEE, pp. 42-47.

[10] S.Jain et al., (2011). Facial expression recognition using Gabor motion energy filters. IEEE International Conference on Computer Vision Workshops (ICCVWorkshops), Barcelona, pp. 1642–1649.

[11] L. Xia, (2014). Facial Recognition Based on SVM. 7th International Conference on Intelligent Computation Technology and Automation, Changsha, pp. 256-259.

[12] J.W. Lee and J.J. Lee, (2013). Gaussian-Distributed Particle Swarm Optimization: A Novel Gaussian Particle Swarm Optimization. IEEE International Conference on Industrial Technology (ICIT).

[13] Mishra et al., (2018). Direction Aware Particle Swarm Optimization with Sensitive Swarm Leader. Big Data Research. 2214-5796.

[14] H. Hakli and H. Uguz, (2014). A novel particle swarm optimization algorithm with Lévy flight. Applied Soft Computing archive. Volume 23, Pages 333-345. Elsevier Science Publishers B. V. Amsterdam, The Netherlands.

[15] A.K. Barisal et al., (2017). "A Hybrid PSO-LEVY Flight Algorithm Based Fuzzy PID Controller for Automatic Generation Control of Multi Area Power Systems: Fuzzy Based Hybrid PSO for Automatic Generation Control." IJEOE 6.2 (2017): 42-63.

[16] Q. Wu and R. Law, (2011). Cauchy mutation based on objective variable of Gaussian particle swarm optimization for parameters selection of SVM. Expert Systems with Applications 38 (2011) 6405–6411.

[17] D. S. Guru et al., (2010). Texture Features and KNN in Classification of Flower Images. IJCA Special Issue on Recent Trends in Image Processing and Pattern Recognition.

[18] Chi-Kien Tran, Chin-Dar Tseng, Pei-Ju Chao, Hui-Min Ting, Liyun Chang, Yu-Jie Huang, Tsair-Fwu Lee, (2017). "Local intensity area descriptor for facial recognition in ideal and noise conditions," Journal of Electronic Imaging 26(2), 023011.

[19] Boser et al., (1992). A training algorithm for optimal margin classifiers. Proceedings of the fifth Annual Workshop on Computational Learning Theory, Pittsburgh, USA, 144–152.

[20] P.C. Vasanth and K.R. Nataraj, (2015). Facial Expression Recognition Using SVM Classifier. Indonesian Journal of Electrical Engineering and Informatics (IJEEI). Vol. 3, No. 1, pp. 16-20. ISSN: 2089-3272.

[21] M. Abdulrahman and A. Eleyan, (2015). Facial Expression Recognition Using Support Vector Machines. Signal Processing and Communications Applications Conference (SIU).