EasyChair Preprint

№ 4807

Face Recognition with Particle Swarm optimization (PSO) and Support Vector Machine (SVM)

Abubakar Ibrahim Muhammad, R. P. Singh and Yusuf Ibrahim

December 25, 2020

Face Recognition with Particle Swarm Optimization (PSO) and Support Vector Machine (SVM)

Abubakar Ibrahim MUHAMMAD
Electronics & Comm. Engr. Department
Rimt University Punjab India
ibngarko@yahoo.com

R.P. SINGH
Dept of Research, Innovation & Incubation
Rimt University Punjab India
rpmgg31@gmail.com

Yusuf IBRAHIM (PhD)
Computer Engineering Department
Ahmadu Bello University, Zaria
yusuf2007ee@gmail.com

***Abstract*— In today's world, face recognition application is still a concerned research area of interest. This is because of the vital role it plays in almost every sectors like security, education, business and so on. It is used in verifying and detecting the correct face in either an image or a video stream. Almost all algorithms types deployed or used in the process of classification problem in face recognition research area, do not guarantee better or optimal performances. Sometimes, optimization becomes necessary even within the selected or desired classification method. In this research, Local Binary Pattern (LBP) was used for feature extraction while Support Vector Machines (SVM) was used as a classifier. However, it is important to find good values of the SVM hyper parameters C and γ in order to obtain good results. Particle Swarm Optimization (PSO) is proposed to optimize and choose the best SVM hyper parameters. Results obtained using the LBP-PSO-SVM pipeline gave a [C, γ], a mean-squared-error of 0.028571 and an overall best accuracy of 98.33%. The proposed framework outperformed the use of SVM alone, with less time of implementation and less complexity.**

Keywords —face recognition, particle swarm optimization, Support Vector Machine

I. INTRODUCTION

Face recognition application being the most concerned research area of interest in today's world advance technology. Because of the vital role it plays in almost every sectors like security, education, business and so on, verifying and detecting the correct object either in an image or video steam become necessarily important to solve problems. Almost all the type of Algorithms deployed or used in the process of classification problem in face recognition research area, does not give 100% accuracy. Therefore it is because of these hindrances all research areas developed in this sectors need a more attention to make a better ways for enhancing the technology of facial recognition.

The face is the primary focus of attention in our today's modern societies and social intercourse, it plays a major role in conveying message such as identity and emotion. Although the ability to infer intelligence or character from facial appearance is debatable, the human ability to recognize faces is remarkable. We can recognize thousands of faces throughout our lifetime and identify familiar faces at a glance even after years of separation. This skill is relatively robust despite significant changes in the visual stimulus caused by viewing conditions, expression, aging, and distractions, such as glasses, beards, or changes in hair style. Face recognition systems are part of facial image processing applications. The significance of these systems as a research area has increased recently. Facial recognition becomes the most concerned area of interest in the modern technology and the way to enhances security within every society having good reliability and authenticity to our usual system's activities

Facial recognition systems are usually applied and preferred by people and security personnel in metropolitan areas. These systems can be utilized for crime prevention, video surveillance, identity verification, and other similar security activities. [1]. Face recognition is one of the most relevant applications of image analysis. It's a true challenge to build an automated system which equals human ability to recognize faces. Although humans are quite good identifying known faces, we are not very skilled when we must deal with a large amount of unknown faces. The computers, with an almost limitless memory and computational speed, should overcome human's limitations.

Facial recognition been the most paramount, significances to the development of modern technology and the role it plays in the various sectors such as security, education and commercial body, there involved a lot of challenges while addressing the issues. These challenges might have different approach depending on the type of the problem. As a field of research area which great scientist or engineers deployed different methods or algorithm to solve whatever the problem might be, they also need to know a better way in order to achieve accuracy with little amount of time frame. Face

recognition involves different steps that yield a better result or good outcomes starting from face detection step, pre-processing of the image, feature extraction, classification and face recognition.

Engineering started to show interest in face recognition in the 1960's or there about. One of the first researches on this subject was Woodrow W. Bledsoe. In 1960, Bledsoe, along other researches, started Panoramic Research in Palo Alto, California. The majority of the work done by this company involved are all related contracts from the U.S. Department of Defense and various intelligence agencies [2]. In recent years, face recognition has attracted much attention and its research has rapidly expanded by not only engineers but also neuroscientists in the field of medical and clinical science, since it has many potential applications in computer vision communication, image processing, robotics, artificial intelligence, machine learning, and automatic access control system.

Face detection is an important part of face recognition as the first step of automatic face recognition. However, face detection is not straightforward because it has lots of variations of image appearance, such as pose variation (front, non-front), occlusion, image orientation, illuminating condition and facial expression [3].

II. MOTIVATION

The activities and development of every nation within the globe need to be monitored, interactions, behaviours and transactions of individual in the societies demand total attention, because of the complex nature of our environment, a lot of improvements and progress occurs in daily. Therefore means of verifications and identification of person is necessary or even becomes compulsory.

Just within the last few decades, verification and identification happens in only two electronic forms. The first form can be carried like a magnetic card. The second form is a type of password to be memorized. These two forms are not secure because both can be given away, taken away, or lost. Moreover, many people can find means to circumvent or forge these credentials. Fortunately, a technique that makes surveillance and monitoring systems function like a pair of eyes has been established. Within computer vision, face recognition has become increasingly relevant in today's society. The recent interest in face recognition can be attributed to the increase of commercial interest and the development of feasible technologies to support the development of face recognition. Major areas of commercial interest include biometrics, law enforcement and surveillance, smart cards, and access control [4]. Now, the need to maintain global security Information is of paramount, in every organization or individual wants to improve their existing security system. Most of the people need better security system which gives complete security solution. From time to time we hear about the crimes of credit card fraud, computer break-in by hackers, or security breaches in company, in shops, in government buildings. In most of these crimes the criminals were taking advantage of that hacking the information from commercial or academic access control system. The systems do not grant access by who we are, but by what we have, such as ID cards, keys, passwords, PIN numbers. These means they are really defining us or they just want to authenticate us. It goes without Permission of owner's, duplicates, or acquires these identity means, he or she will be able to access our data or our personal property any time they want. Recently, technology became available to allow verification of true individual identity.

III. FACE RECOGNITION SYSTEM

Now as the vital fact on machine learning, all the algorithms developed by researchers are subset of machine learning. A Machine Learning (ML) system is different from that of computer program which too many scientist in other discipline found it difficult to understand

Fig. 1. Example of Computer Pragram

The Machine Learning based computer system takes both input data and the result (predicted or target output) which are feed to the learning system model, and produces the Result as program which can be used for others subsequent tacks. While that of computer program, a program is written and both the program and the input data are fed to the computer for specific task and result is obtained.

Fig. 2 Example of Machine Learning Program

IV. ORL DATABASE

The evolution trends in databases and methodologies for facial and expression recognition can be useful for assessing the next-generation topics that may have applications in security systems or personal identification systems that involve "Quantitative face" assessments. An essential part of the constant enhancements made in the field of automated facial and expression recognition has been the collection of facial databases for benchmarking purposes. Since

the 1990s there has been a drive in developing new methods for automatic face recognition as a result of the significant advances in computer and sensor technology [5]. Currently, there are several databases used for facial recognition which vary in size, pose, expressions, lighting conditions, occlusions and the number of imaged subjects The AT & T dataset formerly known as ORL Database contain images of different faces, contains a set of face images taken between April 1992 and April 1994 at the lab at AT & T Laboratory of Cambridge University.

Fig. 3. Example of ORL database

V. CLASSIFICATION

A. PARTICLE SWARM OPTIMIZATION (PSO)

Recently, a lot of research effort has been put in place towards improving human-computer interaction so that computers can have the intelligence to perceive the emotional state of the human user and react accordingly, just like a human would do. There are several applications in different technological or various sectors. Intelligent welfare robots could be developed to provide support and comfort to bed-ridden and highly disabled people who are confined to a room in their houses. This is important given the present modern life style where the population of children is declining, the middle-aged are getting busier with work schedules and where the senior citizens and the disabled are increasingly being left to fend for themselves [6].

A recent algorithm that has been found to be very efficient and effective in solving a variety of problems that involve simplifying and finding optimal solution of problem, optimizing or searching for the optimal value using the Particle Swarm Optimization (PSO) algorithm. Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Dr. Eberhart and Dr. Kennedy in 1995, inspired by social behavior of bird flocking or fish schooling [7], who's initially intended for the simulation of social behavior of birds or insect as they fly in a group searching for food. Particle Swarm Optimization either in its original form or with some modifications was soon found to be applicable in solving a variety for problems solving and better performances. Examples of its application include solving of some classical problems, electrical power systems and neural networks training and etc. It has been applied to clustering problems such as image clustering, data clustering and gene clustering.

B. SUPPORT VECTORS MACHINES (SVM)

Support Vector machine is the most successfully and best classification algorithm in machine learning, it's a nice method because there is principle of derivation for and there is also an optimization package that one can used in order to get a solution and the solution has an intuitive interpretation

Linear Separable

Linear separability is an important topic in the field of artificial intelligence specifically that of machine learning area of research. When two more separable data sets are passed into a linear model against noise and most likely will not over fit, this gives an optimal accuracy and robust than that of inseparable data. Perceptron in neural networks gives such part behavior without no constrain and difficulty in their process. The figure below gives clear example of linear separable data.

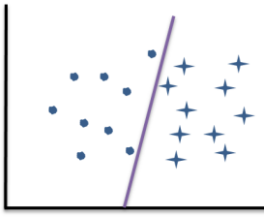


Figure 4**:** linear separable data Sample

The main ideal of SVM start from where the limitation of perceptron (linear separable algorithm) begins. Most of real scenario cases are not linearly in nature but the issue of understanding the separable data gives more room to scientist the courage to further their research in the field of nonlinear separable data.

The face recognition and authentication are the important measures in almost security sectors or environment. For the past two decades a numerous face recognition approaches were deployed and developed. The performance of these different algorithms and methods are compared based on the important parameter recognition rate; so called classification rate. If the classification rate reduces, then the misclassification rate will increase and vice versa. The facial recognition is done based on the minimum distance measure

between the testing data of the set feature vectors and training of dataset of the feature vectors. Different distance measuring techniques are available: such as Euclidean distance, cosine similarity measure, k-NN classifier and etc.

Support Vector Machine (SVM) was first coined in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92. Support vector machines (SVMs) is a supervised learning algorithm method used for different application like classification and regression etc. SVMs belong to a family of generalized linear classifiers. In another terms, Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data. Support Vector machines can be defined as systems which use hypothesis space of a linear functions in a high dimensional feature space, trained with a learning algorithm from optimization theory that implements a learning bias derived from statistical learning theory. It is also being used for many applications, such as hand writing analysis, face analysis and so forth, especially for pattern classification and regression based applications.

Margin Maximization

The issue of margin maximization is an interesting subject matter when talking about Support Vector Machine. Maximization is term which refer extending or widening a particular size or a particular space area. In relation to margin of a linear separable data. It's one of the fundament building block of the SVM process. As one can witness and from figure shown below, this gives an intuitive idea of the margin maximization scenario. There are five possible lines that separate the sample data in two different classes. Therefore the blue line which gives the highest distance between the two classes has significant effect the analysis of SVM

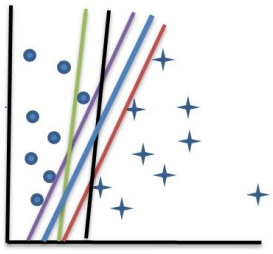


Figure 5: linear separable data with different lines of classifiers

Before dwelling into the concepts of Margin maximization, let us know why the need for the margin. Suppose we have some linear separable sample data as shown in the figure 2.2.8, and there is a need for classifying the da ıght not know the best classifier among them, whether its pu ɔr classifier, blue color classifier, green color classifier, red color or black color classifier can fit the data and give the best result for linear the classification. The issue of choosing the best classifier comes into consideration having the largest equal distances from the nearest sample data.

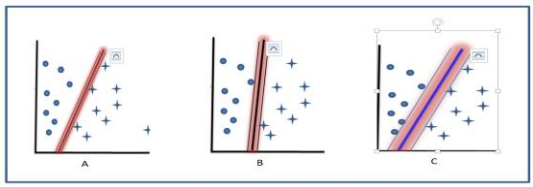


Figure 6: Margin with Different thickness

The blue color hyper plane being the fattest among the three from the above diagram gives the highest optimal and possible margin one can obtained

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. The non-probabilistic aspect is its key strength. This aspect is in contrast with probabilistic classifiers such as the Naïve Bayes, Regression and etc. SVM separates data across a decision boundary (plane) determined by only a small subset of the data (feature vectors). The data subset that supports the decision boundary are aptly called the support vectors. The remaining feature vectors of the dataset do not have any influence in determining the position of the decision boundary in the feature space. In contrast with SVMs, probabilistic classifiers develop a model that best explains the data by considering all of the data versus just a small subset. Subsequently, probabilistic classifiers likely require more computing resources

A Support Vector Machine is an algorithm that works as follows. It uses a nonlinear mapping to transform the original training data into a higher dimension. Within this new dimension, it searches for the linear optimal separating hyper plane. With an appropriate nonlinear mapping to a sufficiently high dimension, data from two classes can always be separated by hyper plane. The SVM finds this hyper plane using support vectors and margins [10].

- The equation of point $i(\boldsymbol{X_n})$ which gives the distance from the point to plane is

$$\boldsymbol{W^T X_n + b = 0}$$

$$\boldsymbol{|W^T X_n + b| = 0}$$

- The weight vector $\boldsymbol{W}$ in the above equation is ⊥ to the plane in the $\boldsymbol{X}$ space
- Let consider the two points on the plane with $\boldsymbol{X^I}$and
- Having the equation $\boldsymbol{W^T X^I + b = 0}$ and $\boldsymbol{W^T X^{II} + b = 0}$
- Combining the two equation $\boldsymbol{W^T X^{II} = 0}$ and $\boldsymbol{W^T X^I = 0}$

$$\boldsymbol{W^T (X^{II} - X^I) = 0}$$

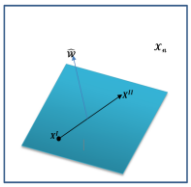


Figure 7: Support Vector

- The distance between and the plane $\boldsymbol{X_n}$
- Take any point $\boldsymbol{X}$ on the plane
- Projection of $\boldsymbol{X_n - X}$ on $\boldsymbol{W}$

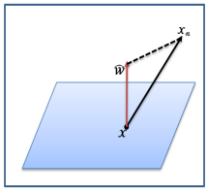


Figure 8: Unit Vector Type equation here.

$$\widehat{\boldsymbol{W}} = \frac{\boldsymbol{w}}{||\boldsymbol{w}||} \text{ = Distance = } \boldsymbol{|W^T (X_n - X)|}$$

$$\textbf{Distance} = \frac{1}{||\boldsymbol{w}||}\{|\boldsymbol{W^T X_n - W^T X}|\}$$

$$\frac{1}{||\boldsymbol{w}||}\{|\boldsymbol{W^T X_n + b - W^T X - b}|\} = \frac{1}{||\boldsymbol{w}||}$$

$$\text{Maximize } \frac{1}{||\boldsymbol{W}||} \textit{ Min. } \boldsymbol{|W^T X_n + b| = 1}$$

$$\text{Minimize } \frac{1}{2}\boldsymbol{WW^T} \quad \text{Subject to } \boldsymbol{n = 1,2, \dots, N}$$

But Notice $\boldsymbol{|W^T X_n + b|} = \boldsymbol{y_n (W^T X_n + b)}$

$$\text{Subject to } \boldsymbol{y_n (W^T X_n + b) \geq 1} \quad \boldsymbol{for \ \ n = 1, 2, \dots, N}$$

$$\boldsymbol{W \in \mathbb{R}^d, b \in \mathbb{R}}$$

Now because of the constrained optimization, it's not helpful and also there exist of an inequality constrained. Therefore to obtain an unconstrained optimization problem, we make use of *Lagrange multiplier* for this conversion (from constrained to unconstrained optimization). This will in turn gives the best separating plane having the best possible margin.

2.2.4.3 Lagrange Formulation

For the regularization of the in sample error term,

Min. $\mathbf{E}_{\text{in}}(\boldsymbol{W}) = \frac{1}{N}(\mathbf{Zw - y})^{\mathbf{T}}(\mathbf{Zw - y})$ *Subject to*; $\boldsymbol{W^T W \leq C}$

$\nabla \boldsymbol{E}_{\text{in}}$ Is normal to the constrain

$\boldsymbol{X^{II}}$

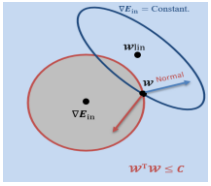


Figure 9: Quadratic Programing

	Optimization	Constrain
Regularization	$\boldsymbol{E}_{in}$	$\boldsymbol{W^T W}$
SVM	$\boldsymbol{W^T W}$	$\boldsymbol{E}_{in}$

For the Lagrange formulation problem, it converts the constrained optimization problem to unconstrained one by introducing a *Lagrange multiplier* $\boldsymbol{\alpha}$

Minimizing the objective function $\frac{1}{2}\boldsymbol{WW^T}$under the constrained/Subject to $\boldsymbol{y_n (W^T X_n + b) \geq 1}$

$$\text{minimize } \boldsymbol{\mathcal{L}(\omega, b, \alpha)} = \frac{1}{2}\boldsymbol{WW^T} - \sum_{n=1}^{N} \boldsymbol{\alpha_n (y_n (W^T X_n + b) - 1)}$$

W.r.t $\boldsymbol{\omega}$ and $\boldsymbol{b}$. & maximize W.r.t each $\boldsymbol{\alpha_n \geq 0}$

$$\boldsymbol{\nabla_\omega \mathcal{L} = W - \sum_{n=1}^{N} \alpha_n (y_n (X_n)) = 0}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \boldsymbol{\sum_{n=1}^{N} \alpha_n y_n = 0} \ \boldsymbol{W = \sum_{n=1}^{N} \alpha_n y_n X_n} \textbf{ And } \boldsymbol{b = \sum_{n=1}^{N} \alpha_n y_n}$$

But showing from the Lagrange formula, we can substitute for

$$\boldsymbol{\mathcal{L}(\omega, b, \alpha)} = \frac{1}{2}\boldsymbol{WW^T} - \sum_{n=1}^{N} \boldsymbol{\alpha_n (y_n (W^T X_n + b) - 1)}$$

Then there we obtained the expression/formula for $\boldsymbol{\alpha_n}$

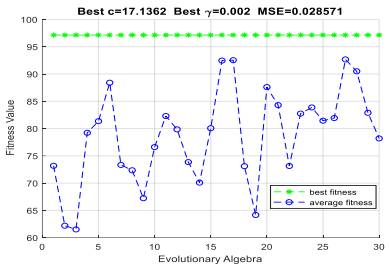
$$\boldsymbol{\mathcal{L}(\alpha)} = \sum_{n=1}^{N} \boldsymbol{\alpha_n} - \frac{1}{2}\sum_{n=1}^{N}\sum_{n=1}^{N} \boldsymbol{\alpha_n \alpha_m y_n y_m X_n X_m}$$

VI. RESULTS AND DISCUSSION

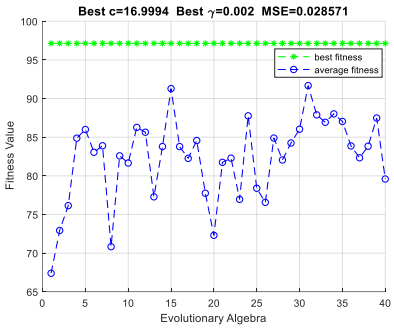
In this section, the results obtained after implemeting each item of the methodology is presented and discussed.

4.1 Results for Finding the Values of the Hyperparameters

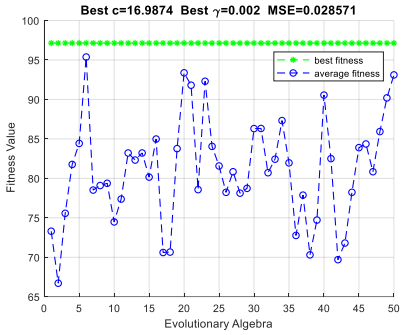
The PSO fitness curve for searching the best SVM hyperparameters (C and γ) are as shown in Figures 4.1-4.3 for different number of generation.



Figures 4.1: PSO fitness curve for 30 geneartion number



Figures 4.2: PSO fitness curve for 40 geneartion number



Figures 4.3: PSO fitness curve for 50 geneartion number

As can be seen from these results, only the value of hyperparameter C tends to change slightly as the number of generations are increased while the hyperparameter γ and the mean squared error remains unchanged. We however observed that the classification accuracy tends to be improved as the generation number is raised. The summary of the result obtained is presented in Table 4

Table 4.1: Summary of the results obtained

Approach	Maximum Generation	Optimized C	Optimized γ	MSE	Classification Accuracy
LBP-PSO-RBF-SVM	30	17.1362	0.002	0.028571	0.9583
LBP-PSO-RBF-SVM	40	16.19994	0.002	0.028571	0.9750
LBP-PSO-RBF-SVM	50	16. 9874	0.002	0.028571	0.9833

The details of how the model classifies each of the forty (40) facial images is presented in the form of confusion matrices of figures 4.1-4.3. The figures give a summary of the counts of the correct and incorrect predictions broken down by each class.

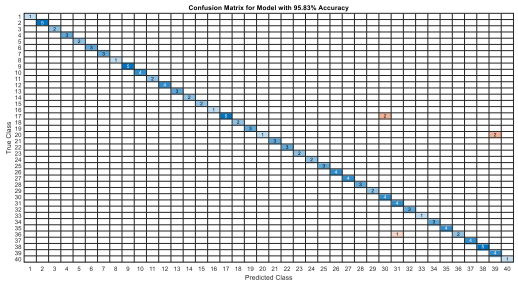


Figure 4.1: Confusion Matrix for Model with 95.8% Accuracy

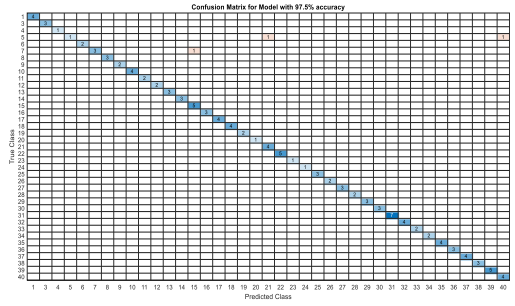


Figure 4.2: Confusion Matrix for Model with 97.5% Accuracy

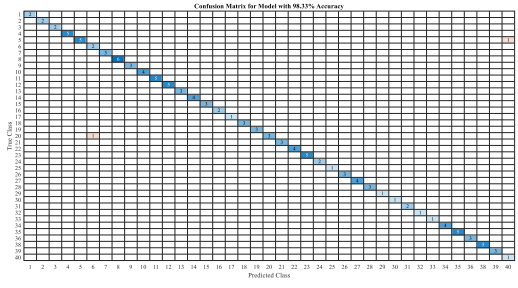


Figure 4.3: Confusion Matrix for Model with 98.33% Accuracy

From the figures 4.1-4.3, it can be seen that there are 5, 3 and 2 misclassified classes respectively.

VII. CONCLUSION

Face recognition application is still a concerned research area of interest since it plays a vital role in a number of applications. It is used in verifying and detecting the correct face in either an image or a video stream. In order to obtain good recognition models, optimization of some hyperparameters becomes necessary even within the selected or desired classification method. In this research, Local Binary Pattern (LBP) was used for feature extraction while Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel was used as a classifier. Particle Swarm Optimization (PSO) was used to optimize and choose the best RBF-SVM hyper parameters [C, γ] and the results obtained were competitive.

The development of a face recognition algorithm where features were extracted from ORL images using Local Binary Patterns and classified using kernel based support vector machine has been developed. The SVM hyperparameters were optimized using Particle Swarm Optimization. Results obtained yielded promising results in terms of Mean Squared Error and average classification accuracy.

REFERENCES

1. Z. A. Fadahl, T. A. Al-Saadi, M. K. Al-Hasnawi and N. J. Al-Khafaji(2014), ' Facial recognition Based on Back Propagation Technique', International Journal of advances in computer Network and its Security(IJCNS), Vol 4, PP (11).
2. M. Ballantyne, R. S. Boyer, and L. H. Woody bledsoe (1996), 'His life and legacy'. AI Magazine, Vol. 17(1) PP (7–20).
3. Siddharth Bhavsar, Niraj Achari, and Bhavin Pandya, (2013), 'Face Detection based on Video' International Journal of Advances in Computer Science and its Applications (IJCSIA) Vol. 3, PP (86-88).
4. M. A. Kashem, M. N. Akhter, S. Ahmed, and M. M. Alam,(2011), 'Face Recognition System Based on Principal Component Analysis (PCA) with Back Propagation Neural Networks (BPNN),' International Journal of Scientific & Engineering Research, vol. 2.
5. Sohini Roychowdhury and Michelle Emmons, 'A Survey Of The Trends In Facial And Expression Recognition Databases And Methods', 1Department of Electrical Engineering, University of Washington, Bothell, (https://arxiv.org/ftp/arxiv/papers/1511/1511.02407.pdf).
6. Bashir Mohammed Ghandi, R. Nagarajan and Hazry Desa,(2009) 'Particle Swarm Optimization Algorithm for Facial Emotion Detection', IEEE Symposium on Industrial Electronics and Applications (ISIEA 2009), PP.(595-599).
7. http://www.swarmintelligence.org/tutorials.php#:~:text=Particle%20swarm%20optimization%20%28PSO%29%20is%20a%20population%20based,evolutionary%20computation%20techniques%20such%20as%20Genetic%20Algorithms%20%28GA%29.