EasyChair Preprint

№ 5666

Evolutionary Algorithms for Face Recognition with Mask

Ekansh Chauhan, Manpreet Sirswal, Richa Singh, Nikhil Bagla, Bhaskar Kapoor and Deepak Gupta

June 2, 2021

Evolutionary Algorithms for Face Recognition with Mask

Ekansh Chauhan[1,*], Manpreet Sirswal[1], Richa Singh[1], Nikhil Bagla[1], Bhaskar Kapoor[1], Dr. Deepak Gupta[1]

[1] Maharaja Agrasen Institute of Technology

[*] Corresponding Author. Email-id: echauhan09@gmail.com (Ekansh Chauhan)

echauhan09@gmail.com, manpreetsirswal9102@gmail.com, richa62370@gmail.com, nikhilbagla707@gmail.com, bhaskarkapoor@gmail.com, deepakgupta@mait.ac.in

A B S T R A C T

Due to the Covid-19 pandemic, wearing masks in public places has become a necessity. But it also comes with its challenges, existing face recognition systems are trained to recognize faces with all the features and therefore are failing to work efficiently due to masks. To provide a potential solution to this problem and to recognize faces with masks two evolutionary algorithms, Crow Search Algorithm (CSA) and Cuttle Fish Algorithm (CFA), are used for feature selection which select an optimal subset of features from the existing dataset with vast number of features. In the last step four machine learning classifiers (Support Vector Machine, Random Forest classifier, K-Nearest Neighbor, and Decision tree classifier) are practiced on each subset of features received by both the feature selection algorithms. Experimental results show that CSA removed most of the irrelevant features by selecting only 41% of the original featured and CFA selected 60% of the features. Highest accuracy of classification was received by CSA of 86.5% with Random Forest classifier. Therefore, it shows that CSA and CFA can be used in various other real time applications due to their reduced computational cost and high accuracy.

KEYWORDS: Covid-19, Face detection, Evolutionary algorithm, Crow-search, Cuttlefish.

INTRODUCTION

Due to Covid-19 pandemic people are wearing face masks all over the world. As proven by WHO [1] it helps in impeding the spread of virus, but however wearing masks has created some problems too. Face recognition techniques have failed terribly. It has created a problem of face identification and recognition, and hence has hindered all its applications such as face attendance, security investigation etc. Our security checks rely heavily on face recognition. Public security checks like railway stations rely heavily on traditional face recognition systems but due to face masks they are not operating effectively [2] and removing masks can increase the risk of infection as covid-19 spreads through contact. This also makes other identification systems like finger print also unsafe to use. As face recognition works

without any touch, it is the safest unlocking system. To solve the above-mentioned problems, it is necessary to improve existing Face recognition approach that relies on all facial features [3].

Therefore, in this paper an approach to recognize faces with masks is proposed. Evolutionary algorithms are used for feature selection. Evolutionary algorithms are bio-inspired algorithms and they are an active of research for solving different types of optimization problems. Evolutionary algorithms are based on natural evolution of animals. It consists of some main processes namely Initialization, evaluate fitness, selection and reproduction, etc. as shown **in Figure 1** [4].

Using the above ideas we have used two evolutionary algorithms for feature selection, Crow search algorithm (CSA) and Cuttle fish algorithm (CFA). Using a filter-based approach, these algorithms will select optimal subsets of features from a large set of dataset.

The major highlights of the paper are:

1) The prime objective of this study is to recognize faces with mask.

2) Two evolutionary algorithms are implemented, Crow search algorithm (CSA) and Cuttle fish algorithm (CFA).

3) The two evolutionary algorithms CSA and CFA are used to select relevant features from the pool of extracted image features.

4) First subset of features is created using data augmentation and Feature extraction on the image dataset.

5) Classification is performed using four different classification models. The rest of the application is as follows: section 2 presents the methodology used along with the two feature selection methods and their implementations in detail. Section 3 discusses the results. Section 4 finally concludes the paper.

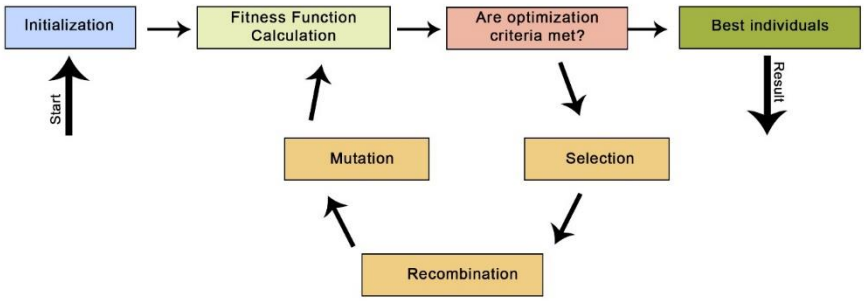

FIGURE 1. Process cycle of Genetic Algorithms

METHODOLOGY

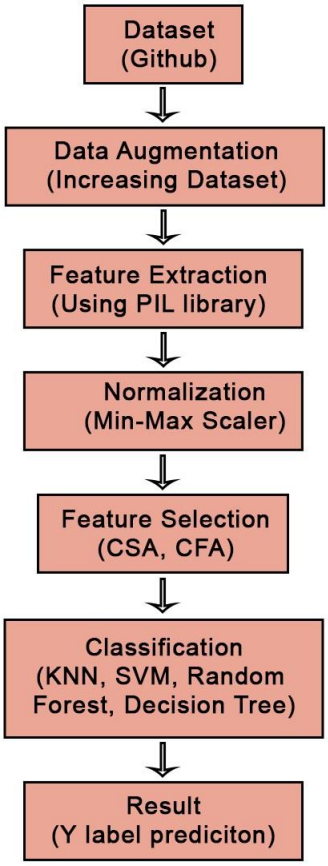

FIGURE 2. Process Flow chart

DATASET

The dataset was taken from a GitHub repository called “X-zhangyang/” [5], it consisted of 405 masked images of different people. The reason for using this source for dataset is that these sources contain images of very diverse people from different countries which is very important for a face recognition system to recognize faces from around the world. All the images from this source are available to the general public and researchers too.

DATA AUGMENTATION

To increase the size of our dataset from 405 images to 2806 images, several data augmentation techniques like flip, rotation, scale, crop and translation were used. Data augmentation makes minor alterations in our existing dataset and increases the diversity in our dataset by applying random transformations [6]. The classifier will take those images as distinct images and hence will increase the meaningful data. In our dataset, data augmentation was done using Keras pre-processing layers [7]. A sample of before and after images of data augmentation are given in Fig 3.

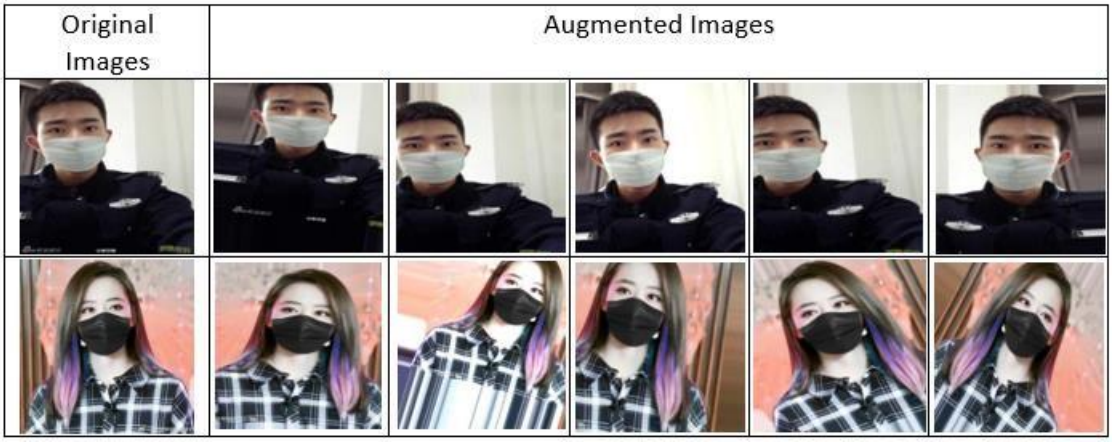

FIGURE 3. Sample images of Dataset

FEATURE EXTRACTION AND NORMALIZATION

Feature extraction is the process of converting image dataset to comma separated values (csv) file. Feature extraction derives values (features) from the existing image dataset in terms of pixels or texture. We used PIL library in python for feature extraction. During feature extraction of 2806 images, 65,536 features were derived. After feature extraction, data normalization was performed. In data normalization, the data is organized and converted to be in coherence with the values of other attributes or features, which increases the consistency and accuracy of the model. Each feature is scaled within a given range. Sklearn library was used to perform data normalization. Since MinMaxScaler was used the feature range was (0,1) i.e. the values of all features were translated between 0 and 1.

FEATURE SELECTION

Feature selection is the process of removing noise from data, i.e. the features that contribute a negligible amount to the output or are irrelevant and redundant. Therefore, it is process of selection a subset of data from the whole dataset, which contains all the required features. This is done to reduce computational cost and increase the accuracy as having irrelevant or redundant features in the dataset can decrease the accuracy of the model. In this project Crow search algorithm [8] and Cuttle fish algorithm [9] are used for feature selection. They both are evolutionary algorithms and hence are inspired from natural evolution of animals.

Crow Search Algorithm (CSA)

"Askarzadeh proposed an algorithm called Crow Search Algorithm (CSA) (Fig 5), a metaheuristic bio-inspired optimizer inspired by the intelligence of crows. Crows are considered the most intelligent birds worldwide, they are famous for their intelligence and sharp memory. Studies have proven that crows are capable of remembering faces". They are famous for snooping on other birds and learning their food hiding places, and then stealing that food when the owner birds are not around. This is the prime idea behind CSA. Since crows are insatiable in nature, they attempt to take each other's food as well. What's more, to keep their food from being taken they utilize probabilistic approach. For an optimization problem "crows are considered as search agents, the environment is assumed to be the search space, each position of the environment is considered to be a feasible solution, fitness function defines the quality of the food source and the global solution of the optimization problem is the best food source in the environment". Dependence of the algorithm on the flight length is shown in Fig 4 [10] and all the parameters used are given in **Table 1**.

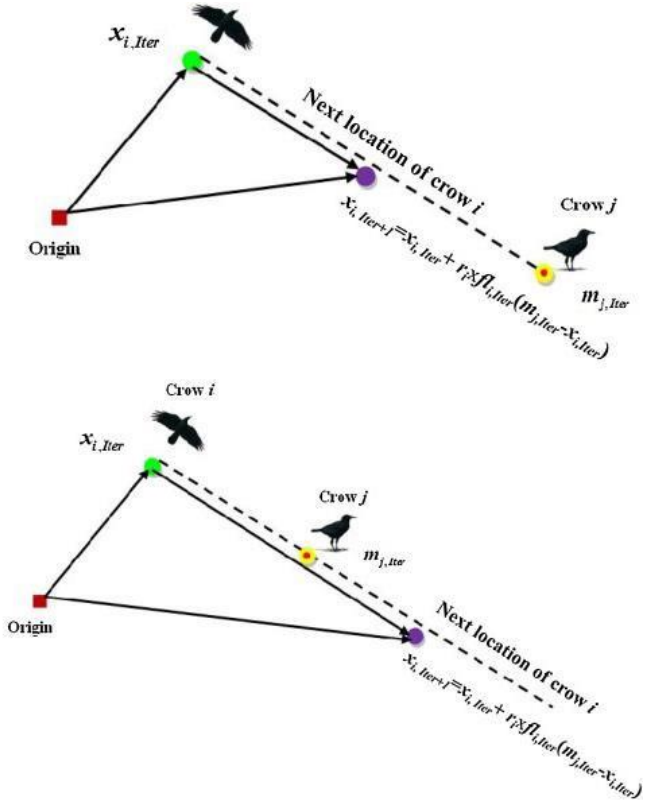

FIGURE 4. Dependence of CSA on flight length. Crow i can go to every position on the dash line

Table 1- Parameters used in Crow search algorithm

Parameters	**Value**
Total crows	20
Awareness Probability	0.1
Iterations	200
Total features	65536
Weight Factor)	1.6
Flight length	0.2

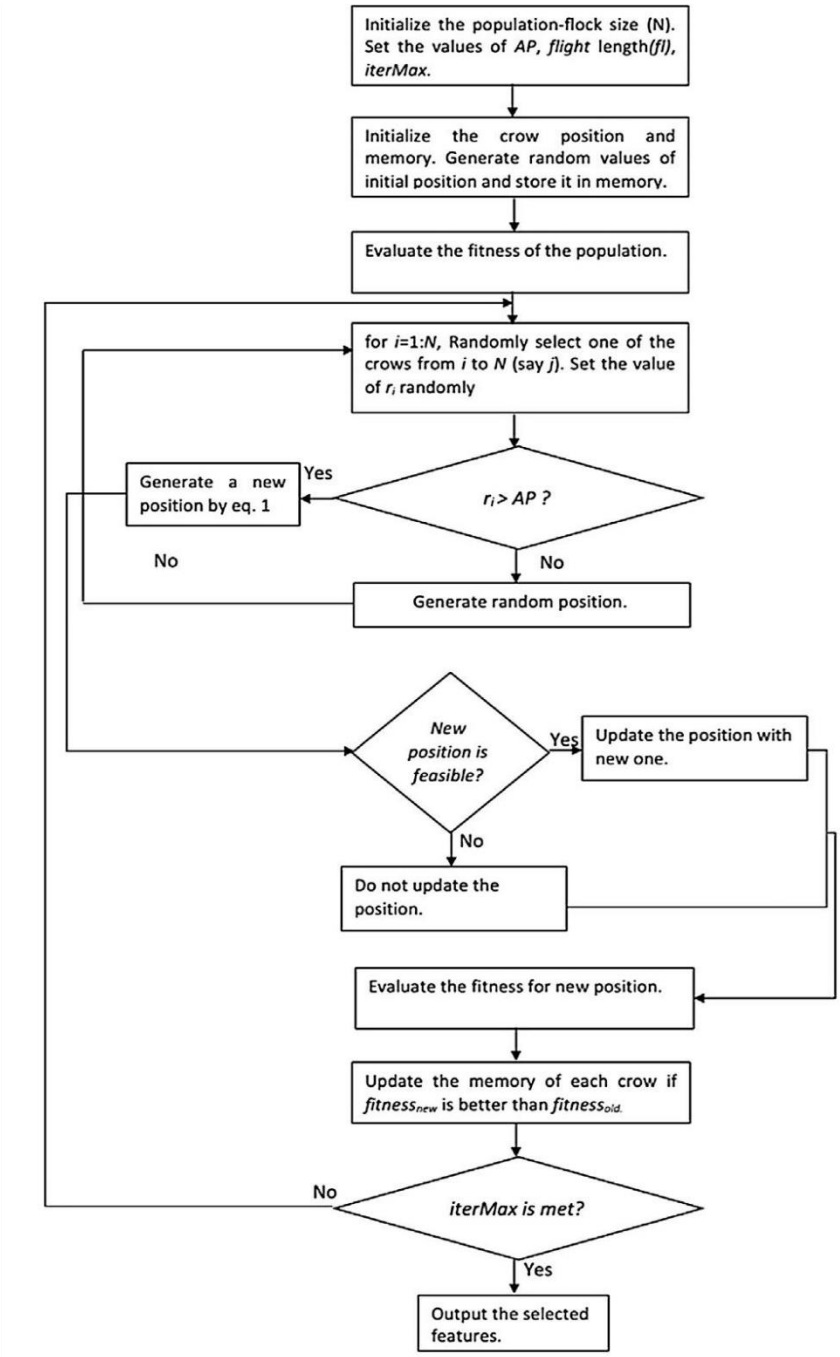

FIGURE 5. Crow search algorithm

Cuttle Fish Algorithm (CFA)

According to [9]**,** "In 2013, Eesa, Brifcani, Orman proposed a meta-heuristic bio-inspired optimization algorithm called Cuttlefish Algorithm (CFA) (Fig 7) to solve the numerical global optimization problems". CFA is based on the color changing characteristics of cuttle fish. The concept behind cuttle fish algorithm is that using reflection of light, cuttle fish creates mesmerizing patterns using chromatophores, leucophores and iridophores (Fig 6). These three are the layers of cells that a cuttle fish possesses. The amalgamation of these three layers forms the six different possibilities of reflection of light. All these 6 cases are split into four groups, called G1, G2, G3 and G4. All of these groups are independent of each other. All the parameters used are shown in **Table 2.**

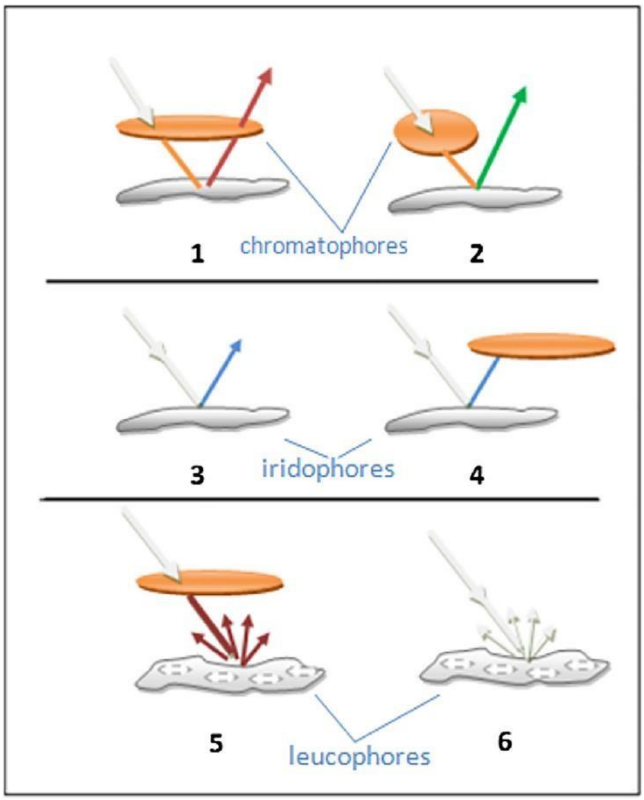

FIGURE 6. Six distinct cases of light reflection through cell layers.

Table 2- Parameters used in Cuttle Fish algorithm

Parameters	**Value**
Cell Population	20
No. of Dimensions	0.1
Group Size	10
Iterations	65
Visibility degree used in case 1 & 2	1
Reflection degree used in case 3 & 4	1
Reflection degree used in case 5	1
Lower bound of initial weights	-1
Upper bound of initial weights	1

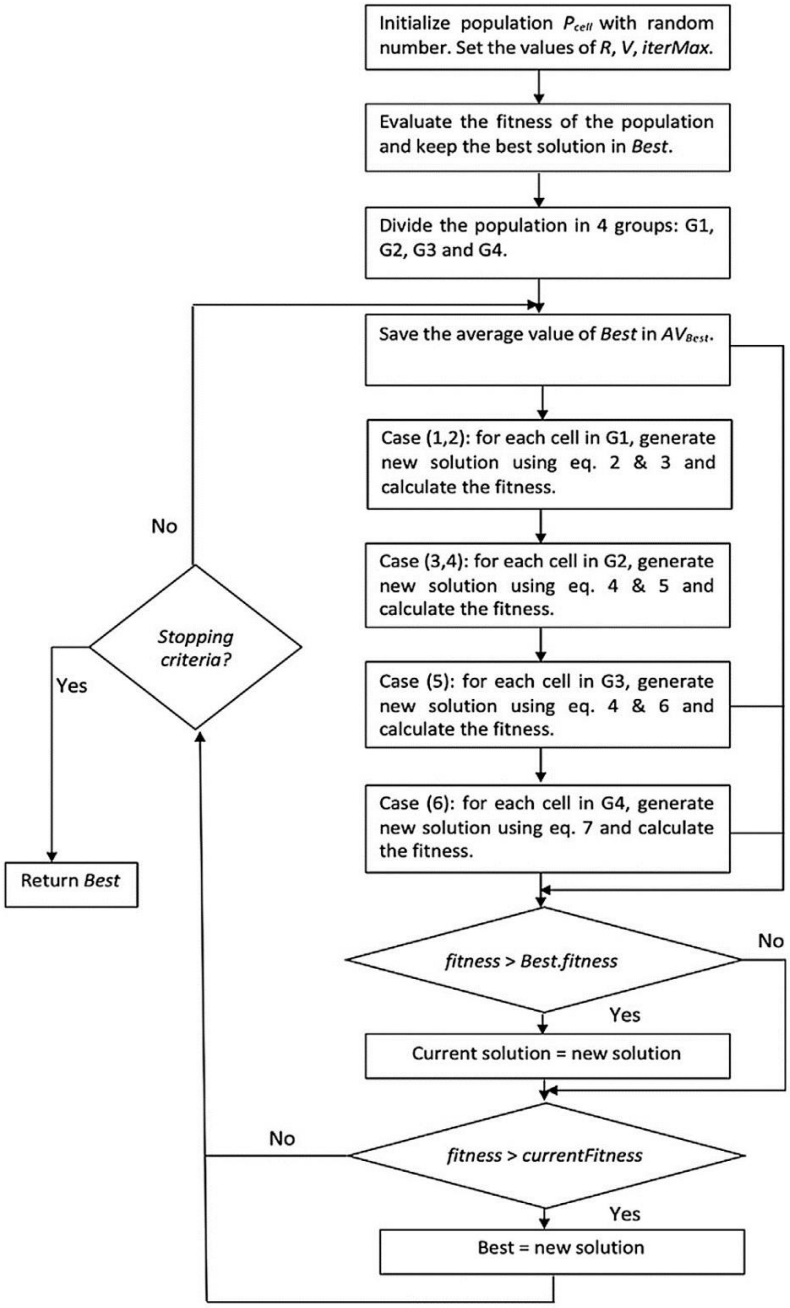

FIGURE 7. Crow search algorithm

2.5 CLASSIFICATION:

For the classification of images, four different classifiers are used namely, Random Forest Classifier, Support Vector Machine (SVM), K-Nearest Neighbours (k-NN) and Decision Tree
Classifier.

2.5.1 Random Forest Classifier:

According to [11]**, "**Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. Random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting". The major parameters used for random forest are n_estimators=50, sample_split=2, bootstrap=2. All the other parameters are shown in **table 3.**

2.5.2 Support Vector Machines (SVM):

It is a supervised Machine Learning model which is used for classification problems. It works by creating a hyperplane or a set of hyperplanes in an infinite dimensional space [12]. The parameters used for support vector machine are: Kernel= linear, gamma = 1, C = 1.0, epsilon =1. They are also shown in **table 3**.

Table 3- Parameters used in Classifiers

Model	**Tuning parameters**
KNN	n_neighbors=6, metric='minkowski'
SVM	kernel = linear, gamma = 1, C = 1.0, epsilon =1, shrinking =true
Random Forest	n_estimators=500
Decision Tree	criterion='gini', splitter='best', max_depth=10

2.5.3 K-Nearest Neighbors (KNN):

According to [13] "*k*-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until function evaluation. Since this algorithm relies on distance for classification, normalizing the training data can improve its accuracy dramatically". The

parameters used for K-NN are n_neighbors=6, weights= “uniform”, metric= “minkowski”.

2.5.4 Decision tree Classifier:

Decision tree is a classification technique that uses a very straightforward idea. It organizes a series of test questions and conditions in a tree structure [14]. The parameters used for Decision Tree Classifiers are max_depth=10, min_samples_split=20 and all other default parameters are shown in Table 3.

RESULTS

For Feature selection, crow search algorithm (CSA) gave the best results, it selected only 26869 features from the set of 65536 features, i.e. 41% (Fig 8) and Cuttlefish Algorithm (CFA) selected 39321 features from 65536 features i.e. 60% (Fig 9) and comparison bar plot has been generated to clearly observe the difference in the number of features selected by both the algorithms (Fig 10). Based on that, clearly CSA removed more irrelevant features, as shown **in Table 4.**

When the data extracted from CSA was put into different classifiers the accuracy was as follows KNN=84.6%, Random Forest =86.5%, SVM= 83.2% and Decision tree= 81.8%.

While the data extracted from CFA was put into different classifiers the accuracy was as follows KNN=81.4%, Random Forest =82.4%, SVM= 78.2% and Decision tree= 76.1% as shown in **Table 5 and Fig 11.**

So, out of CFA and CSA, CSA performed more efficiently. And out of the four classifiers used Random Forest gave the best accuracy.

Table 4- Features selected by evolutionary algorithms

Feature Selection Method	No. of features selected	Total Number of features
Crow Search Algorithm	26869 (approx. 41%)	65536
Cuttle Fish Algorithm	39321 (approx. 60%)	65536

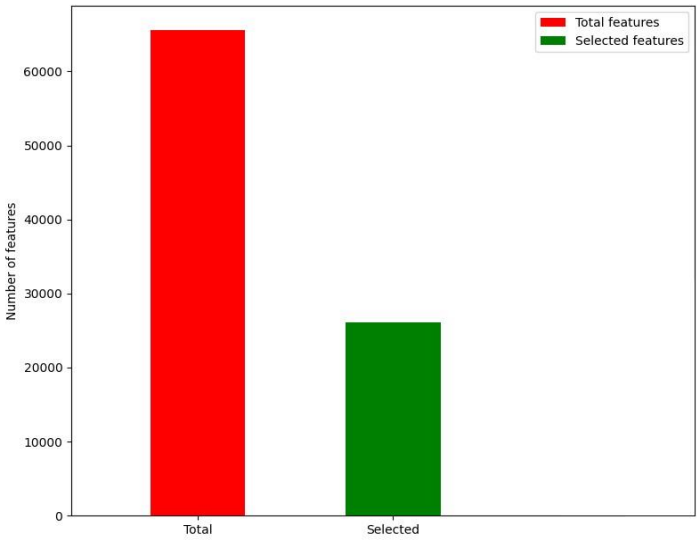

FIGURE 8- Bar graph of number of features selected by CSA.

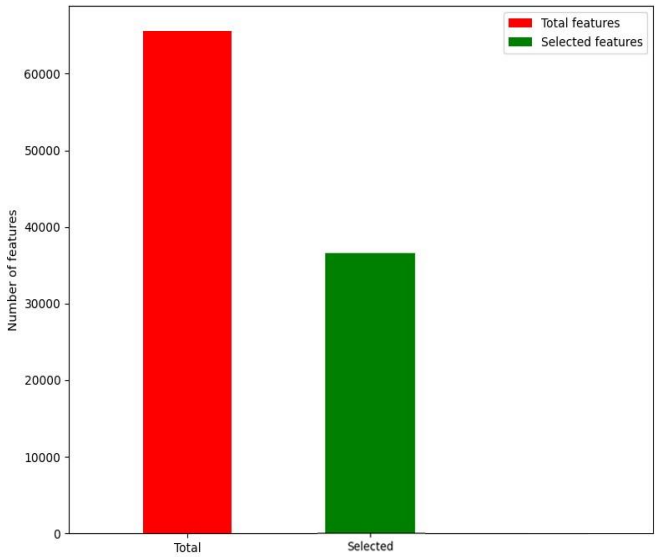

FIGURE 9- Bar graph of number of features selected by CFA.

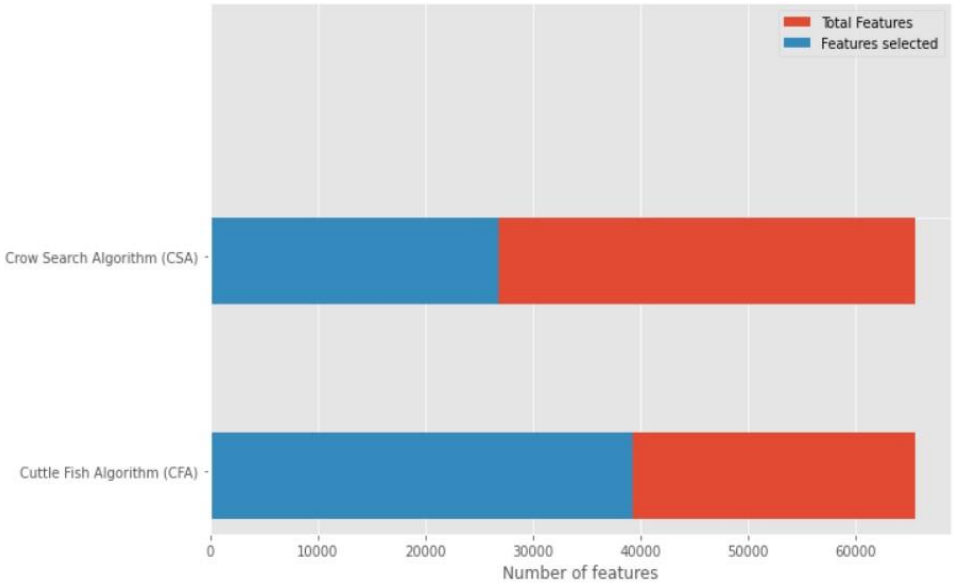

FIGURE 10- Comparison between number of features selected by CSA and CFA.

Table 5- Accuracy of model for each classifier

Method	**Classifier**	**Accuracy (%)**
Crow Search Algorithm (CSA)	k-NN	84.6
	Random Forest	86.5
	SVM (Linear)	83.2
	Decision Tree	81.8
Cuttle Fish Algorithm (CFA)	k-NN	81.4
	Random Forest	82.4
	SVM (Linear)	78.2
	Decision Tree	76.1

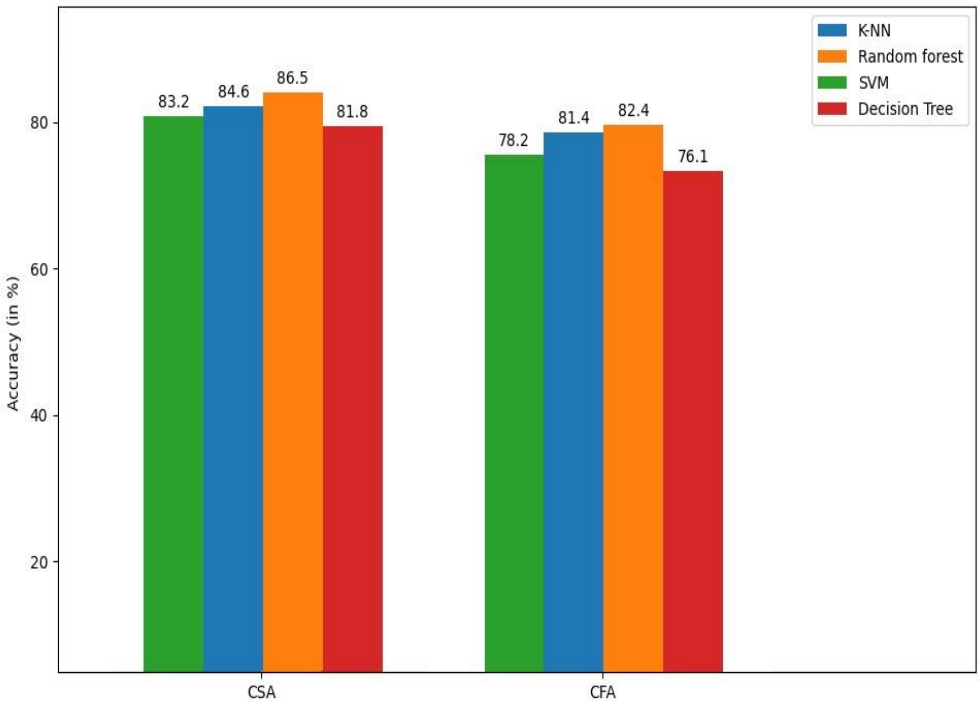

FIGURE 11- Comparison between accuracy obtained from different classifiers for two evolutionary algorithms.

CONCLUSION AND FUTURE SCOPE:

In the presented work, two evolutionary algorithms were used on the masked face dataset, for feature selection, to improve the accuracy of classification and to reduce the computational cost. After that four machine learning classifiers were applied to each subset of features obtained using the feature selection algorithms. It was observed that Crow Search algorithm gave the best accuracy with Random Forest classifier followed by KNN. Subsequently high level of accuracy was received by both the feature selection algorithms. The two evolutionary algorithms can be used in various other fields. For example, it can be implemented to identify and classify diseases in medical science such as chest CT scan images.

Several other bio-inspired algorithms can be explored for feature selection as a future work. The proposed algorithms can be applied in a wide range of research areas having global optimization problems. This work may be extended by considering more evolutionary algorithms. They can be used as feature selection

methods by combining with other classification models and deep learning techniques for obtaining more accurate results with lesser computational times.

REFERENCES:

[1] World Health Organization (WHO), "Novel Coronavirus – China," *World Health Organization, disease uotbreak news*, 2020. .

[2] B. Rai, A. Shukla, and L. K. Dwivedi, "COVID-19 in India: Predictions, Reproduction Number and Public Health Preparedness," *medRxiv*, 2020, doi: 10.1101/2020.04.09.20059261.

[3] D. Comaniciu, "Artificial Intelligence for Healthcare," 2020, doi: 10.1145/3394486.3409551.

[4] T. Bartz-Beielstein, J. Branke, J. Mehnen, and O. Mersmann, "Evolutionary Algorithms," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2014, doi: 10.1002/widm.1124.

[5] Z. Wang *et al.*, "Masked face recognition dataset and application," *arXiv*. 2020.

[6] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, 2019, doi: 10.1186/s40537-019-0197-0.

[7] Keras, "Image Preprocessing - Keras Documentation," *Keras Documentation*, 2019. .

[8] A. Askarzadeh, "A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm," *Comput. Struct.*, 2016, doi: 10.1016/j.compstruc.2016.03.001.

[9] A. Sabry Eesa, A. Mohsin, A. Brifcani, and Z. Orman, "Cuttlefish Algorithm – A Novel Bio-Inspired Optimization Algorithm," *Int. J. Sci. Eng. Res.*, 2013.

[10] N. Gupta, D. Gupta, A. Khanna, P. P. Rebouças Filho, and V. H. C. de Albuquerque, "Evolutionary algorithms for automatic lung disease detection," *Meas. J. Int. Meas. Confed.*, 2019, doi: 10.1016/j.measurement.2019.02.042.

[11] Y. Qi, "Random forest for bioinformatics," in *Ensemble Machine Learning: Methods and ApplicatiOns*, 2012.

[12] R. G. Brereton and G. R. Lloyd, “Support Vector Machines for classification and regression,” *Analyst*. 2010, doi: 10.1039/b918972f.

[13] Z. Zhang, “Introduction to machine learning: K-nearest neighbors,” *Ann. Transl. Med.*, 2016, doi: 10.21037/atm.2016.03.37.

[14] P. E. Utgoff, N. C. Berkman, and J. A. Clouse, “Decision Tree Induction Based on Efficient Tree Restructuring,” *Mach. Learn.*, 1997, doi: 10.1023/A:1007413323501.