

# FOHC: Firefly Optimizer Enabled Hybrid approach for Cancer Classification

Abhilash Pati<sup>1,\*</sup>, Amrutanshu Panigrahi<sup>2,\*</sup>, Bibhuprasad Sahu<sup>3</sup>, Ghanashyam Sahoo<sup>4</sup>, Manoranjan Dash<sup>5</sup>, Manoranjan Parhi<sup>6</sup>, and Binod Kumar Pattanayak<sup>7</sup>

<sup>1,2,4,5,6,7</sup> Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India

<sup>3</sup>Vardhaman College of Engineering (Autonomous), Hyderabad, Telangana, India

<sup>4</sup>GITA Autonomous College, Bhubaneswar, Odisha, India

<sup>1,\*</sup>er.abhilash.pati@gmail.com, <sup>2,\*</sup>amrutansup89@gmail.com, <sup>3</sup>prasadnikhil176@gmail.com, <sup>4</sup>ghanarvind@gmail.com,

<sup>5</sup>manoranjandash@soa.ac.in, <sup>6</sup>manoranjanparhi@soa.ac.in, and <sup>7</sup>binodpattanayak@soa.ac.in

**Abstract**—Early detection and prediction of cancer, a group of chronic diseases responsible for a large number of deaths each year and a serious public health hazard, can lead to more effective treatment at an earlier stage in the disease's progression. In the current era, machine learning (ML) has widely been used to develop predictive models for incurable diseases such as cancer, heart disease, and diabetes, among others, taking into account both existing datasets and personally collected datasets, more research is still being conducted in this area. Using recursive feature elimination (RFE), principal component analysis (PCA), the Firefly Algorithm (FA), and a support vector machine (SVM) classifier, this study proposed a Firefly Optimizer-enabled Hybrid approach for Cancer classification (FOHC). This study considers feature selection and dimensionality reduction techniques RFE and PCA, and FA is used as the optimization algorithm. In the last stage, the SVM is applied to the pre-processed dataset as the classifier. To evaluate the proposed model, empirical analysis has been carried out on three different kinds of cancer disease datasets including Brain, Breast, and Lung cancer obtained from the UCI-ML warehouse. Based on the various performance parameters like accuracy, error rate, precision, recall, f-measure, etc., some experiments are carried out on the Jupyter platform using Python codes. This proposed model, FOHC, surpasses previous methods and other considered state-of-the-art works, with 98.94% accuracy for Breast cancer, 95.58% accuracy for Lung cancer, and 96.34% accuracy for Brain cancer. The outcomes of these experiments represent the effectiveness of the proposed work.

**Keywords**- Cancer Classification; Support vector machine (SVM); Firefly Algorithm (FA); Principal component analysis (PCA); Recursive feature elimination (RFE).

## I. INTRODUCTION

Machine learning (ML) has been widely used in medical research in the detection of cancerous cells. The development of ML algorithms, which is an extension of artificial intelligence (AI), is of great benefit to society in terms of contributing to the well-being of human beings [1]. The use of various ML algorithms simplified the detection process, allowing for a higher success rate in the treatment process and an increase in the patient's survival rate. When ML is used, the system incorporates the ability to learn based on the dataset that was used in the training process. There are a variety of methods available for creating systems that can learn. Supervised, Unsupervised, and reinforcement learning are the three broad categories of ML. The learning process for classification techniques is supervised, whereas the learning process for regression techniques is unsupervised [2, 3].

The biopsy and microarray datasets are two examples of accessible datasets for cancer diagnosis, whereas the biopsy dataset is the most popular choice. The genetic information of

the individuals whose lab test results are included in the biopsy dataset is redacted [4-6]. However, the researcher will not be able to use the biopsy dataset to diagnose the cancer disease accurately because genetic information is crucial in accurate cancer diagnosis. Since there are blanks in the biopsy data, the researcher looks to the microarray dataset for answers. Massive amounts of data on gene expression are generated in only one experiment utilizing microarray data [7 – 9]. It offers a tremendous opportunity for determining the genetic basis of disease associations.

Nevertheless, high dimensionalities are observed in the gene expression data, which are unimportant in the diagnosis of diseases [10], [11]. The high-dimension microarray dataset contains redundant and noisy information which can play a vital role in decreasing the classification accuracy [12], [13], [19]. In addition to the above-said issue, the small sample size is another problem that arises when dealing with the microarray dataset. This is due to the presence of fewer sample details as compared to several features.

#### A. *Motivation and Objective*

Cancer patients have a far better prognosis and survival rate if their disease is detected and treated promptly. The more precisely malignancies are classified, the fewer individuals will be subjected to unneeded procedures. As a result, there is a lot of effort put into determining the best ways to diagnose cancer and categorize people as either normal or abnormal. ML is widely accepted as the approach of choice in cancer classification and forecast modeling due to its unique advantages in essential feature discovery from complicated cancer datasets.

In this paper, a Firefly Optimizer-enabled Hybrid approach for Cancer classification (FOHC), has been introduced for the prediction of breast cancer based on principal component analysis (PCA), recursive feature elimination (RFE), firefly algorithm (FA), and support vector machine (SVM) classifier for dimensionality reductions, feature selection, optimization, and classifications respectively. The proposed model is evaluated against various UCI-ML sourced datasets on Breast Cancer, Lung Cancer, and Brain Cancer.

#### B. *Key Contributions*

This paper's main contributions are:

- The proposed model is developed using classification approaches, feature selection, dimensionality reduction, and optimization techniques.
- Introduced a model for predicting various forms of cancers.
- Described the proposed model with a comparison to other approaches based on different evaluative measures.

#### C. *Paper Structure*

The remainder of the paper is structured as follows: Previous studies on the topic are discussed in Section 2, while the methodology and materials employed in this study are presented in Section 3. Section 4 presents the results of the work's empirical research, while Section 5 offers a summary and an outlook on what lies ahead.

## II. EXISTING WORKS

Zheng et al. [4] employed a hybrid strategy including K-means and SVM techniques on the WBCD dataset, verifying their results with K-fold cross-validation, and claimed to have attained an accuracy of 97.38 percent.

Shah and Jivani [5] used the Weka tool to apply the classification techniques DT, NB, and KNN to the WDBC dataset. They found that NB performed significantly better than

DT and KNN, achieving 95.99 percent accuracy using the NB classification technique.

Rana et al [7] have applied the classification techniques SVM, KNN, NB, and LR on the WBCD dataset using the MATLAB tool, and the results show that KNN implemented with Euclidean distance achieved the highest accuracy of 95.68 percent, which outperforms other implemented classification algorithms.

Talukdar and Kalita [15] conducted a study in which they used two classification techniques, namely, J48 DT and Zero, to predict breast cancer and discovered that J48 DT is more accurate than ZeroR in predicting breast cancer.

Asri et al [16] used four classifiers on the WBCD dataset, namely, SVM, C4.5 DT, NB, and KNN, and claimed to have achieved the highest accuracy of 97.13 percent using the SVM classifier. The authors used the Weka tool to conduct their research.

Keles [17] found that the maximum accuracy for the classifiers he tested on the Antenna dataset was 92.2% using the Weka tool and the RF, IBK (Instance-Based Learning), RF, Bagging, and Random Committee approach.

Israni [18] used PCA and SVM classification in the study of breast cancer. When compared to the other models studied, he found that the proposed hybrid model achieved the highest accuracy (92.78 percent).

## III. PROPOSED WORK: FOHC

This section presents the various materials and methods which are used to develop the proposed FOHC model. The microarray dataset of various cancer diseases is being considered as a starting point for this investigation. Various cancer datasets sourced from the University of California; Irvine (UCI-ML) are being considered for the current research work FOHC. To select the reduced features, the techniques such as PCA and RFE are used, and to obtain the best features from them, FA is used. Finally, the SVM is employed in the classification of cancers. In this work, we used Jupyter, the Python environment, to conduct experiments using Python codes.

#### A. *Dataset Description*

Biopsy and microarray datasets may help diagnose cancer. The biopsy dataset comprises lab test results from a group of patients but no genetic information. The biopsy dataset lacks genetic information for cancer diagnosis. The researcher uses microarray data to fill up biopsy data gaps. For the current research work, three different kinds of datasets are considered, namely Breast Cancer, Lung Cancer, and Brain Cancer datasets [20]. The dimensions of these datasets are listed in Table 1.

TABLE I. DATASET DESCRIPTION

Dataset	Size
Breast Cancer	31 x 569
Lung Cancer	62 x 2178
Brain Cancer	29 x 1008

B. Methodology

The PCA and the RFE are all considered in the current research work, FOHC, for dimensionality reduction and feature selection, respectively, while the FA is used for optimization. The SVM is then used to classify the cancer data at the end of the process. The Jupyter Notebook platform was used in the work experiment. After the data has been thoroughly pre-processed, PCA, a dimensionality reduction strategy, is used to reduce the dimensions and select the useful features, and then RFE, a feature selection approach, is used to further reduce the dimensions and select the useful features after PCA. The FA is used to optimize and select the best features, and SVM is used to categorize the outcomes and make predictions based on that categorization.

- **Principal Component Analysis (PCA):** PCA is a dimensionality reduction technique that relies on the linear dimensionality reduction methodology. It converts a collection of variables that are correlated ( $m$ ) into a lesser  $n$  ( $m \ll n$ ) number of variables, that are uncorrelated, which are called principal components while preserving as much variance as feasible. The correlation between variables is taken into account using PCA. PCA tries to integrate highly correlated variables and locate the directions of greatest variance in higher-dimensional data if the correlation is very high. PCA is an unsupervised ML approach that finds relevant variables that may be used for subsequent regression, grouping, and classification tasks in the context of ML [21].

- **Recursive Feature Elimination (RFE):** Feature selection (FS) tackles over-fitting concerns and improves model performance by deleting features from the data that are not important or redundant. RFE is a well-liked FS method because it is effective at determining which features (columns) in a training dataset are more significant in predicting the target variable. Although the number of valid features is usually unknown in advance, a particular number of features must be maintained. There are two key configuration variables when using RFE: the number of features available and the process for choosing features. Although it is possible to study both of these hyper-parameters, the effectiveness of the approach is not significantly affected by their right setting [8], [22].

- **Firefly Algorithm (FA):** The firefly algorithm is a novel bioinspired computational technique for optimization in

which the search process is modeled after firefly social behavior and luminescent communication. The fluctuation of light intensity and the formulation of attraction are two critical challenges with the firefly algorithm. The brightness function  $B$  identifies every firefly  $f_i$  from the swarm [14]. There are 3 different rules present for the FA such as:

- Any firefly  $f_i$  from the swarm will be responsible for attracting the others.
- The attractiveness depends upon the brightness, and any 2 flashing  $f_i$  will fly to the brighter one.
- $B(f_i)$  can be defined by using its fitness function  $F(f_i)$ .

The attractiveness between two fireflies  $f_i$  and  $f_j$  can be defined based on distance  $d_{ij}$  (Equations (1), (2) and (3)):

$$d_{i,j} = |f_i - f_j| \quad (1)$$

$$= \sqrt{\sum (f_i - f_j)^2} \quad (2)$$

$$= \alpha_0 e^{-\delta d_{i,j}} \quad (3)$$

Where  $k$  is the dimension index of the firefly,  $\alpha$  is the sum of brightness for one firefly concerning the other,  $\alpha_0$  is the initial brightness, and  $\delta$  is the light density coefficient. The movement of one firefly  $f_i$  towards  $f_j$  can be defined by the following equations (4) and (5) with  $Rand$  as the random number from the range (0,1):

$$f_{i,j} = (1 - \alpha) f_{i,k} + \alpha f_{j,k} + \delta_{i,k} \quad (4)$$

$$\delta_{i,k} = (Rand - 0.5) \quad (5)$$

- **Support Vector Machine (SVM):** SVM is a supervised ML technique [9], [13] which can be used for both classification and regression. It simply separates the dataset linearly into two different categories. The data set will include  $m$  number of attributes along with  $n$  number of sample details. Out of  $m$  attributes, few attributes are considered outliers which must be removed to enhance the classification and prediction result performance. Let  $D$  be a data set with the attributes  $\{D_1, D_2, D_3, \dots, D_m\}$  and the class label  $C_i$  belonging to  $[1, -1]$  or  $[1, 0]$ . The decision boundary must then be accurately determined while taking into account hyperparameters such as cost and gamma. Reducing the gamma parameter will raise the accuracy %, but lowering the cost value will result in less accuracy.

The proposed work FOHC is the combination of the techniques PCA, RFE, FA, and SVM in a sequential order to achieve better-classified outcomes. The workflow and algorithm of the proposed work are represented using Figure 1 and Algorithm 1 respectively.

**Algorithm 1.** Pseudocode for the proposed work

**Input:** Dataset D with n attributes

**Output:** List of attributes as per the relevance factor to be fit into SVM and binary classification

**Proc PCA ()**

- Calculate the dot product of D and DT
- Perform the Eigen analysis
- Calculate Eigenvector values.
- For (i=1 to n)
- Choose PC<sub>i</sub>
- endfor

- Choose a Specific number of PCs as per Eigenvalue.
- Return D\* = {PC1, PC2, PC3,.....}

**Proc RFE(D\*)**

- Fit the model using SVM
- For (i= 1 to k)
  - Calculate Feature\_Importance<sub>i</sub>
  - F={i from 1 to k| Feature\_Importance<sub>i</sub> }
  - F<sub>min</sub>= Minimum (F)
  - Eliminate F<sub>min</sub> from F

- End for
- D\*\* □ Delete feature having F<sub>min</sub>
- Proc RFE ()

**Proc FA (D\*\*)**

- Define maximum iteration M.
- Determine the initial population
- Define the objective function
- Define the attractiveness (D) of every firefly f<sub>i</sub>
- While (i<M) do
  - For (j=1 to n)
  - For (k=1 to n)
  - If (D<sub>j</sub> < D<sub>k</sub>)
  - Move f<sub>k</sub> towards f<sub>j</sub>
  - endif
  - Update the attractiveness and position
  - End for k
  - End for j

- Rank the fireflies as per D and calculate the best position
- Calculate the global best
- End while
- Return updated D\*\*
- Apply SVM for binary classification and performance measures

**IV. EMPIRICAL ANALYSIS**

The basic objective of any performance evaluation including a prediction or classification process is to find the confusion matrix and determine any model's correctness and accuracy [23

– 25]. T1 stands for True Positive, T2 for True Negative, F1 for False Positive, and F2 for False Negative; these letters represent the results of the confusion matrix. Accuracy (ACC), error rate (ER), precision (PRE), recall (REC) or sensitivity, specificity (TNR), f-measure (F-M), Mathew's Correlation Coefficient (MCC), true positive rate (TPR), false positive rate (FPR), and false negative rate (FNR) are just some of the evaluation metrics on which predictions in this work can be agreed upon. The definitions of these parameters are summarized in Table 2 [26, 27]. After the data has been thoroughly pre-processed, PCA and RFE are used to further reduce the data's dimensionality and select the most relevant features for further analysis. The characteristics are then optimized using FA, and SVM is utilized for classification and prediction.

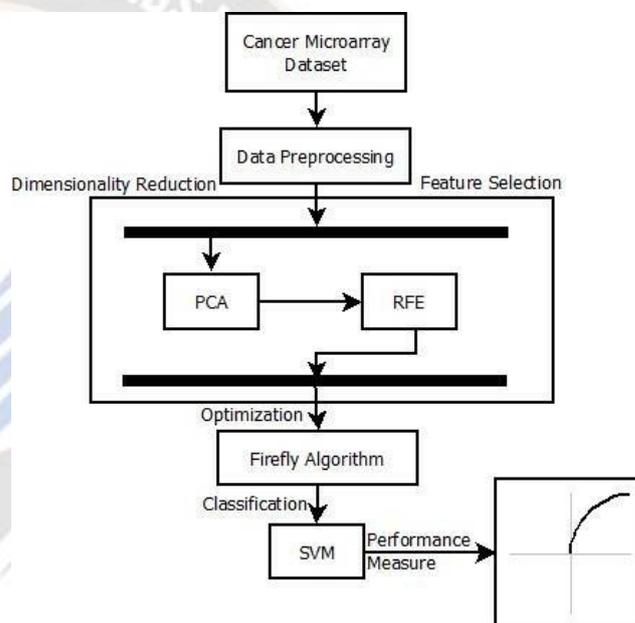


Figure 1. Workflow of the proposed work.

The datasets were separated into two sets in this study, training data, and testing data with a distribution ratio of 0.2. After experimenting with various classifiers like logistic regression, naive Bayes, k-nearest neighbors, decision trees, and the random forest along with SVM, it was discovered that SVM is the best of them all for this proposed work. Following that, the FOHC model is implemented, and several evaluation measures are computed.

We have carried out 3 different approaches in this research SVM with RFE only, SVM with PCA only, and SVM with RFE, PCA, and FA, which is then termed FOHC, the proposed work. Table 3 shows the empirical analysis of the proposed method in contrast to other state-of-the-art ML algorithms. Figures 2-7 show the performance measurement analysis for the different datasets concerning various models and in comparison with the proposed FOHC.

TABLE II. PERFORMANCE PARAMETER DEFINITION

Measures	Calculation Formula
ACC	$(T_1+T_2)/(T_1+T_2+F_1+F_2)$
ER	$(F_1+F_2)/(T_1+T_2+F_1+F_2)$
PRE	$T_1/(T_1+F_1)$
REC	$T_1/(T_1+F_2)$
F-M	$(2 *PRE*REC)/(PRE+REC)$
TPR	$(T_1*100)/(T_1+F_2)$
TNR	$(T_2*100)/(T_2+F_1)$
FPR	$(F_1*100)/(T_2+F_1)$
FNR	$(F_2*100)/(T_1+F_2)$
MCC	$\{(T_1*T_2)-(F_1*F_2)\} / \sqrt{\{(T_1+F_1)(T_1+F_2)(T_2+F_1)(T_2+F_2)\}}$

TABLE III. PERFORMANCE MEASURE COMPARISONS IN CONTRAST TO A DIFFERENT MODEL

Method	Breast Cancer			Lung Cancer			Brain Cancer		
	SVM-RFE	SVM-PCA	FOHC	SVM-RFE	SVM-PCA	FOHC	SVM-RFE	SVM-PCA	FOHC
ACC(%)	91.49	92.55	98.94	92.02	93.09	95.58	94.64	95.12	96.34
ER (%)	8.51	7.45	1.06	7.98	6.91	4.42	5.36	4.88	4.66
PRE	0.9	0.92	0.99	0.92	0.93	0.96	0.94	0.95	0.97
REC	0.88	0.89	0.98	0.88	0.89	0.94	0.95	0.94	0.96
F-M	0.89	0.9	0.99	0.9	0.91	0.93	0.94	0.93	0.96
TPR (%)	87.67	89.04	98.59	87.84	89.04	92.65	98.59	89.04	90.28
TNR (%)	93.91	94.78	99.15	94.74	95.68	96.56	99.15	95.68	97.39
FPR (%)	6.09	5.22	0.85	5.26	4.35	2.63	4.26	4.79	3.45
FNR (%)	12.33	10.96	1.41	12.16	10.96	12.33	9.89	11.13	9.72
MCC	0.82	0.84	0.98	0.83	0.85	0.87	0.96	0.86	0.89

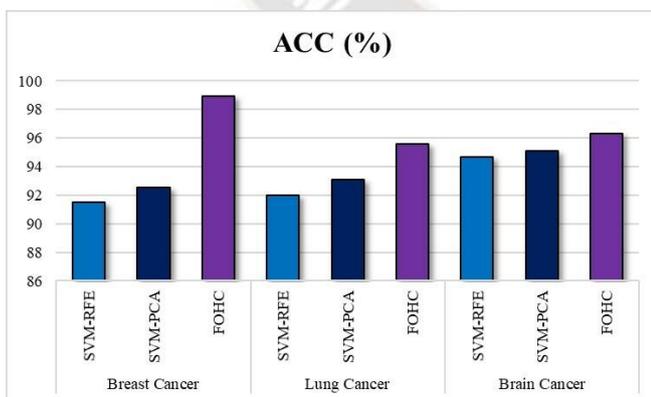


Figure 2. Accuracy-based comparison with FOHC.

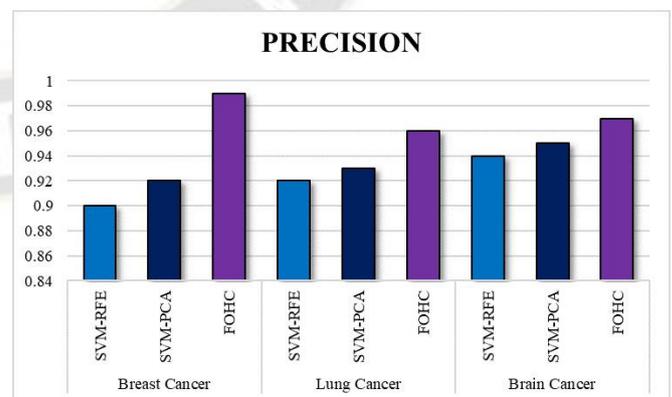


Figure 3. Precision-based comparison with FOHC.

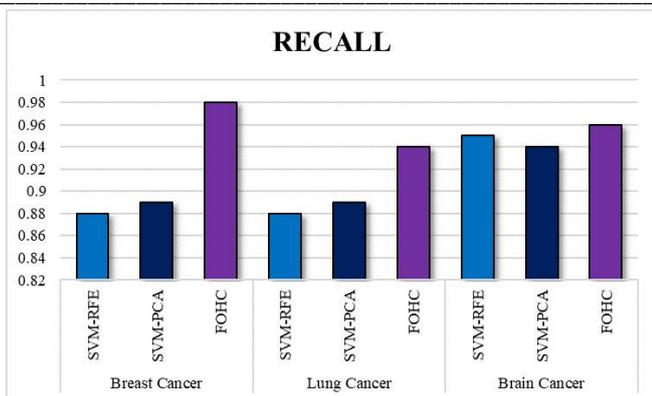


Figure 4. Recall-based comparison with FOHC.

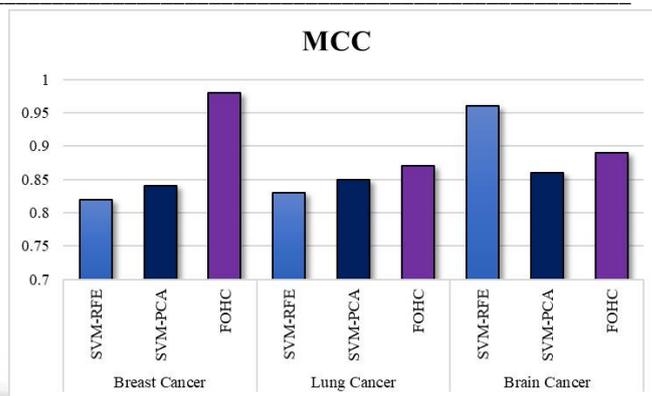


Figure 6. MCC-based comparison with FOHC.

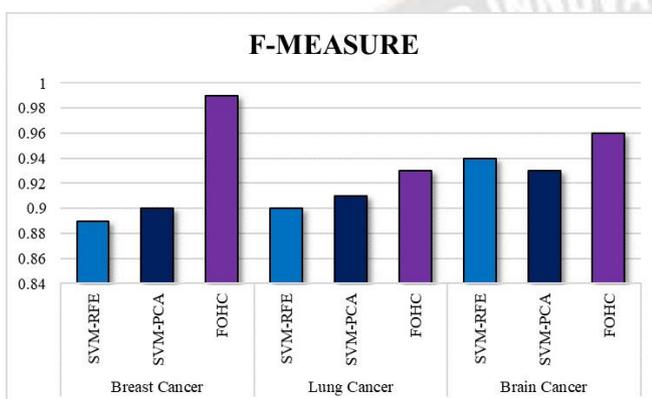


Figure 5. F-Measure-based comparison with FOHC.

It can be observed that the proposed work, FOHC, achieves 98.94%, 95.58%, and 96.34% of accuracies in Breast, Lung, and Brain cancer datasets respectively, which is comparatively higher than the other two approaches. Besides, this proposed work also outperforms other approaches in terms of evaluative parameters like precision, recall, f-measure, TPR, TNR, FPR, FNR, ER, MCC, etc. A comparison of this proposed work with some existing works considered in this paper on breast cancer classification specifically is shown in Table 4 and Figure 8.

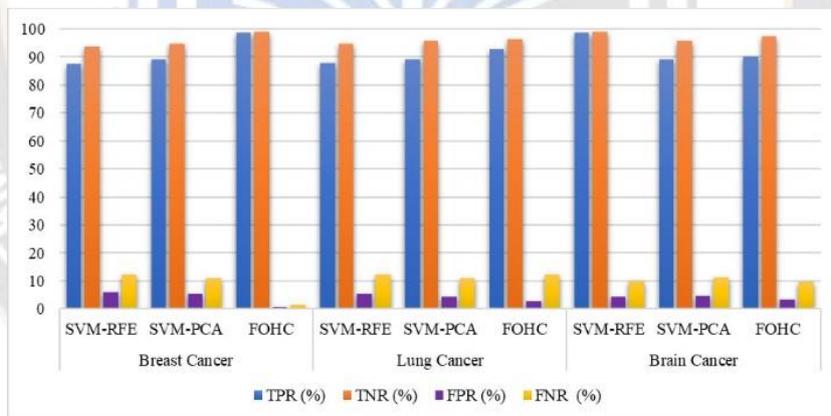


Figure 7. TPR, TNR, FPR, and FNR-based comparison with FOHC.

TABLE IV. PERFORMANCE MEASURE COMPARISONS IN CONTRAST TO A DIFFERENT MODEL

Work	Evaluative Measures (in %)			
	ACC	PRE	REC	F-M
K-SVM [4]	97.38	X	X	X
BC – NB [5]	95.99	X	X	X
KNN – Euclidean [7]	95.68	X	X	X
J48 DT [15]	95.38	97.3	95.4	X
SVM [16]	97.13	98.0	97.0	97.5
RFA [17]	92.2	X	X	X
SVM – PCA [18]	92.78	X	X	X
<b>Proposed FOHC Model</b>	<b>98.94</b>	<b>99.0</b>	<b>98.0</b>	<b>98.5</b>

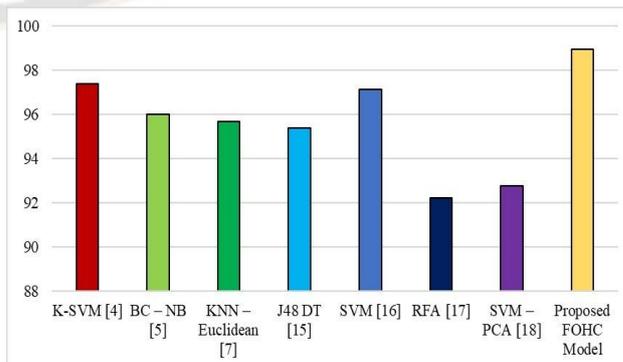


Figure 8. Comparison of FOHC with the various existing model in contrast to Accuracy.

The effectiveness of a classification model is shown using a receiver operating characteristic (ROC) curve, which takes into account all possible levels of categorization. The link between the TPR and the FPR is shown by this curve. The degree of categorical ambiguity is shown using a ROC curve. When the criteria for categorization are lowered, there is an increase in the number of both false positives and true positives. A region that is under the ROC curve (AUC i.e. Area Under the Curve). The AUC is a measurement that evaluates performance across all levels of categorization. The AUC measures how likely it is that a model would give a positive example a higher score than a negative example. The area under the ROC curve is measured in two dimensions and ranges from (0,0) to (1,1). The ROC of the suggested FOHC model can be shown in Figure 9, which also displays the AUC value of 0.916.

and state-of-the-art works, the suggested model, FOHC, achieves higher accuracy rates (98.94% for Breast cancer, 95.58% for Lung cancer, and 96.34% for Brain cancer).

The suggested approach will help doctors make more accurate predictions and diagnoses for their breast cancer patients, and it may be useful for future research in prediction that makes use of other datasets. The results may be better if the number of dimensions was reduced before feature selection, and using deep learning techniques could lead to more precise predictions.

## REFERENCES

- [1] S. K. Chaturvedi, F. J. Strohschein, G. Saraf, and C. G. Loisselle, "Communication in cancer care: Psycho-social, interactional, and cultural issues. A general overview and the example of India," *Frontiers in Psychology*, vol. 5, no. NOV. Frontiers Research Foundation, 2014. DOI: 10.3389/fpsyg.2014.01332.
- [2] P. R. Deshpande, M. Sheriff, A. Nazir, S. Bommareddy, A. Tumkur, and A. N. Naik, "Patient-reported quality of life outcomes in Indian breast cancer patients: Importance, review of the researches, determinants and future directions," *Journal of Cancer Research and Therapeutics*, vol. 9, no. 1. pp. 11–16, Jan. 2013. DOI: 10.4103/0973-1482.110341.
- [3] Prof. Nikhil Surkar. (2015). Design and Analysis of Optimized Fin-FETs. *International Journal of New Practices in Management and Engineering*, 4(04), 01 - 06. Retrieved from <http://ijnpme.org/index.php/IJNPME/article/view/39>
- [4] N. Fatima, L. Liu, S. Hong, and H. Ahmed, "Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis," *IEEE Access*, vol. 8. Institute of Electrical and Electronics Engineers Inc., pp. 150360–150376, 2020. DOI: 10.1109/ACCESS.2020.3016715.
- [5] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," *Expert Systems with Applications*, vol. 41, no. 4 PART 1, pp. 1476–1482, 2014, DOI: 10.1016/j.eswa.2013.08.044.
- [6] C. Shah and A. G. Jivani, "Comparison of data mining classification algorithms for breast cancer prediction," 2013. DOI: 10.1109/ICCCNT.2013.6726477.
- [7] A. Pati, M. Parhi, B. K. Pattanayak, B. Sahu, and S. Khasim, "CanDiag: Fog Empowered Transfer Deep Learning Based Approach for Cancer Diagnosis," *Designs*, vol. 7, no. 3, p. 57, Apr. 2023, doi: 10.3390/designs7030057.
- [8] M. Rana, P. Chandorkar, A. Dsouza, and N. Kazi, "BREAST CANCER DIAGNOSIS AND RECURRENCE PREDICTION USING MACHINE LEARNING TECHNIQUES." [Online]. Available: <http://www.ijret.org>
- [9] H. Lei and V. Govindaraju, "Speeding Up Multi-class SVM Evaluation by PCA and Feature Selection."
- [10] A. Pati, M. Parhi, B. K. Pattanayak, D. Singh, D. Samanta, A. Banerjee, S. Biring, and G. K. Dalapati, "Diagnose Diabetic Mellitus Illness Based on IoT Smart Architecture," *Wireless Communications and Mobile Computing*, vol. 2022, p. e7268571, Aug. 2022, doi: <https://doi.org/10.1155/2022/7268571>

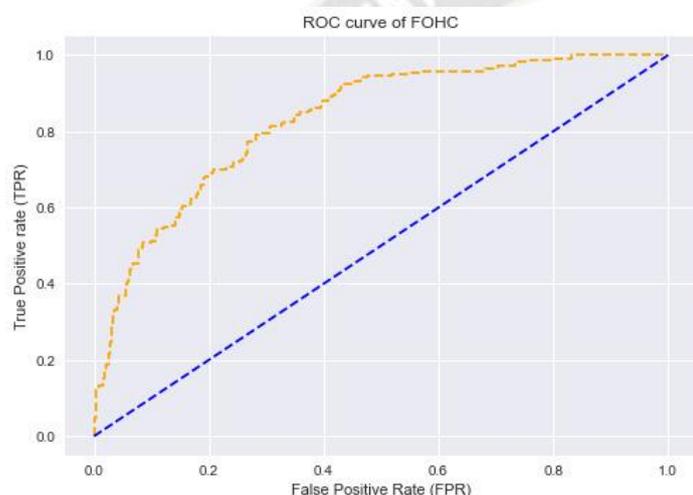


Figure 9. ROC of the proposed FOHC.

## V. CONCLUSION AND FUTURE SCOPE

This study describes the FOHC, or Firefly Optimizer-enabled Hybrid approach for Cancer classification, which is a multidisciplinary approach. When using classifiers to improve the performance and accuracy of various cancer disease classifications, the goal is to remove useless and unsuitable features from the dataset and only select those features that are most beneficial from the classification point of view. PCA, RFE, FA, and SVM are all used in the proposed FOHC model. Uncorrelated variables are reduced through the use of the PCA; redundant features are selected through the use of the RFE; FA is used for optimization, and pre-processed data are classified through the use of the SVM. After that, the results are compared to those obtained using PCA-SVM and RFE-SVM. Some experiments are carried out using the Breast Cancer, Lung Cancer, and Brain Cancer datasets obtained from the UCI-ML repository and the Python programming language, with the results indicating that the proposed FOHC technique outperforms the other techniques as well as the works considered during the literature survey. When compared to existing methods

- [11] B. Sahu, C. J. Badajena, A. Panigrahi, C. Rout, and S. Sethi, "An Intelligence-Based Health Biomarker Identification System Using Microarray Analysis," in CRC Press, 2020, pp. 137–161.
- [12] B. Sahu, A. Panigrahi, and S. K. Rout, "DCNN-SVM: A NEW APPROACH FOR LUNG CANCER DETECTION," CRC PRESS, 2020.
- [13] B. Sahu, A. Panigrahi, S. K. Rout and A. Pati, "Hybrid Multiple Filter Embedded Political Optimizer for Feature Selection," 2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCCSP), Hyderabad, India, 2022, pp. 1-6, doi: 10.1109/ICICCCSP53532.2022.9862419.
- [14] A. Pati, M. Parhi, M. Alnabhan, B. K. Pattanayak, A. K. Habboush, and M. K. Al Nawayseh, "An IoT-Fog-Cloud Integrated Framework for Real-Time Remote Cardiovascular Disease Diagnosis," *Informatics*, vol. 10, no. 1, p. 21, Feb. 2023, doi: 10.3390/informatics10010021
- [15] B. Sahu, A. Panigrahi, S. Mohanty, and S. S. Panigrahi, "A hybrid Cancer Classification Based on SVM Optimized by PSO and Reverse Firefly Algorithm," *International Journal of Control and Automation*, vol. 13, no. 4, pp. 506–517, 2020
- [16] Kanna, D. ., & Muda, I. . (2021). Hybrid Stacked LSTM Based Classification in Prediction of Weather Forecasting Using Deep Learning. *Research Journal of Computer Systems and Engineering*, 2(1), 46:51. Retrieved from <https://technicaljournals.org/RJCSE/index.php/journal/article/view/22>
- [17] J. Talukdar and S. Kr Kalita, "Detection of Breast Cancer using Data Mining Tool (WEKA)," *International Journal of Scientific & Engineering Research*, vol. 6, no. 11, 2015, [Online]. Available: <http://www.ijser.org>.
- [18] H. Asri, H. Mousannif, H. al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," in *Procedia Computer Science*, 2016, vol. 83, pp. 1064–1069. DOI: 10.1016/j.procs.2016.04.224.
- [19] Muñoz, S., Hernandez, M., González, M., Thomas, P., & Anderson, C. Enhancing Engineering Education with Intelligent Tutoring Systems using Machine Learning. *Kuwait Journal of Machine Learning*, 1(2). Retrieved from <http://kuwaitjournals.com/index.php/kjml/article/view/116>
- [20] M. Kaya Keleş, "Breast cancer prediction and detection using data mining classification algorithms: A comparative study," *Tehnicki Vjesnik*, vol. 26, no. 1, pp. 149–155, Feb. 2019, DOI: 10.17559/TV-20180417102943.
- [21] P. Israni, "Breast cancer diagnosis (BCD) model using machine learning," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 10, pp. 4456–4463, Aug. 2019, DOI: 10.35940/ijitee.I9973.0881019.
- [22] A. Pati, M. Parhi, & B. K. Pattanayak, (2022). HeartFog: Fog Computing Enabled Ensemble Deep Learning Framework for Automatic Heart Disease Diagnosis. In *Intelligent and Cloud Computing* (pp. 39-53). Springer, Singapore.
- [23] D. W. Aha, "Heart Disease dataset <https://archive.ics.uci.edu/ml/datasets/heart+disease>," UCI-ML repository, Jul. 22, 1988.
- [24] M. N. Taib., Universiti Teknologi MARA. Faculty of Electrical Engineering., and IEEE Control Systems Society. Chapter Malaysia., CSPA 2010: 2010 6th International Colloquium on Signal Processing & Its Applications: 21-23 May 2010, Mahkota Hotel, Melaka, Malaysia. IEEE, 2010.
- [25] Abhale, A. B. ., & Reddy A, J. . (2023). Deep Learning Perspectives to Detecting Intrusions in Wireless Sensor Networks. *International Journal of Intelligent Systems and Applications in Engineering*, 11(2s), 18–26. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/2504>
- [26] A. Adorada, R. Permatasari, P. W. Wirawan, A. Wibowo, and A. Sujiwo, "Support Vector Machine-Recursive Feature Elimination (SVM-RFE) for Selection of MicroRNA Expression Features of Breast Cancer," in 2018 2nd International Conference on Informatics and Computational Sciences (ICICoS), 2018, pp. 1–4.
- [27] Alexei Ivanov, *Machine Learning for Traffic Prediction and Optimization in Smart Cities*, Machine Learning Applications Conference Proceedings, Vol 3 2023.
- [28] A. Pati, M. Parhi, and B. K. Pattanayak, "A review on prediction of diabetes using machine learning and data mining classification techniques," *International Journal of Biomedical Engineering and Technology*, vol. 41, no. 1, p. 83, 2023, doi: <https://doi.org/10.1504/ijbet.2023.128514>
- [29] A. Pati, M. Parhi, and B. K. Pattanayak, "IHDPM: an integrated heart disease prediction model for heart disease prediction," *International Journal of Medical Engineering and Informatics*, vol. 14, no. 6, p. 564, 2022, doi: <https://doi.org/10.1504/ijmei.2022.126526>.
- [30] A. Pati, A. Panigrahi, D. S. K. Nayak, G. Sahoo, and D. Singh, "Predicting Pediatric Appendicitis using Ensemble Learning Techniques," *Procedia Computer Science*, vol. 218, pp. 1166–1175, 2023, doi: <https://doi.org/10.1016/j.procs.2023.01.095>.
- [31] S. S. Tripathy, N. Tripathy, M. Rath, A. Pati, A. Panigrahi and M. Dash, "Smart Farming based on Deep Learning Approaches," 2022 2nd Odisha International Conference on Electrical Power Engineering, Communication and Computing Technology (ODICON), Bhubaneswar, India, 2022, pp. 1-5, doi: 10.1109/ODICON54453.2022.10010165.
- [32] S. K. Rout, B. Sahu, A. Panigrahi, B. Nayak, and A. Pati, "Early Detection of Sepsis Using LSTM Neural Network with Electronic Health Record," *Smart Innovation, Systems and Technologies*, pp. 201–207, Nov. 2022, doi: [https://doi.org/10.1007/978-981-19-6068-0\\_19](https://doi.org/10.1007/978-981-19-6068-0_19).