Efficient Breast Cancer Identification Using Sunflower Optimization-Based Feature Selection with Ensemble Machine Learning

Abstract

Breast cancer is one of the leading causes of mortality among women worldwide, and its early and accurate diagnosis is critical for effective treatment. In this study, we propose a hybrid classification framework that combines Feature Selection (FS) using the Sunflower Optimization (SFO) algorithm with multiple classification models including Support Vector Machine (SVM)**,** Random Forest (RF)**,** Naïve Bayes (NB), and Convolutional Neural Networks (CNN). The SFO algorithm, inspired by the phototropic behavior of sunflowers, is employed to identify the most informative subset of features from the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, thereby reducing dimensionality and enhancing classifier performance. Each selected feature subset is then evaluated using classical machine learning classifiers (SVM, RF, NB) and a deep learning model (CNN) to compare classification accuracy and generalization capability. The models are assessed based on performance metrics including accuracy, precision, sensitivity, specificity, and F1-score. Experimental results demonstrate that the SFO-based feature selection significantly improves the predictive accuracy across all classifiers, with CNN achieving the highest performance, reaching an accuracy exceeding 98,64%. The proposed approach proves to be an effective, robust, and interpretable solution for breast cancer classification tasks, especially in scenarios where dimensionality reduction and classification reliability are essential.

Keyword: Breast Cancer, Feature selection, Sunflower Optimization Algorithm, Classification.

1. Introduction

Breast cancer remains one of the most prevalent and life-threatening cancers affecting women globally. According to the World Health Organization (WHO), early detection and accurate diagnosis of breast cancer significantly increase the chances of successful treatment and survival [1]. In clinical practice, diagnostic tools such as mammography, ultrasound, and biopsy are widely used, In recent years, Computer-Aided Diagnosis (CAD) systems have played a vital role in assisting radiologists by automating the detection and classification of breast cancer using machine learning techniques [2]. However, CAD systems still face several limitations, including high-dimensional data, redundant or irrelevant features, increased computational cost, and the risk of overfitting. One critical component in addressing these challenges is feature selection, which aims to identify the most relevant attributes from medical datasets to improve classification accuracy and system efficiency. To tackle the feature selection problem, researchers over the globe have proposed numerous methods in the literature, emplying metaheuristic algorithms such as Grey Wolf Optimizer[3][4][5], Bat Algorithm[6], Harris Hawks Optimizer[7][8], Particle Swarm Optimization[9], Chaotic Cuckoo Search Optimization[10], Binary dragonfly optimization[11], Binary Teaching Learning-Based Optimization algorithm[12], Whale Optimization Algorithm[13].

In this study, we propose the use of the Sunflower Optimization Algorithm [14] for feature selection because it is a nature-inspired metaheuristic that effectively explores and exploits the search space to identify the most relevant features. The SFO effectively searches for optimal feature subsets by simulating sunflower movement toward sunlight, thereby enhancing the quality of input features for classification. We applied the proposed SFO-based feature selection method to the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [15], which consists of various clinical features extracted from digitized images of breast masses. Following the feature selection phase, we employed multiple classifiers, including SVM, RF, NB, and a CNN to evaluate the performance of the selected features. The classification models are evaluated using standard performance metrics such as accuracy, sensitivity (recall), specificity, precision, and F1-score to ensure a comprehensive analysis. Comparative results demonstrate that SFO-based feature selection improves classifier performance across the board, with the CNN model achieving the highest accuracy, suggesting that deep learning benefits significantly from optimized input features. This hybrid approach combining metaheuristic optimization and multi-model classification presents a robust, accurate, and interpretable framework for breast cancer detection, with potential applicability to other medical diagnosis domains involving high-dimensional data.

The paper is organized into five sections. Section 1 provides an introduction to the study. Section 2 reviews related work and discusses advanced techniques for breast cancer diagnosis using the WDBC dataset and metaheuristic algorithms. Section 3 outlines the proposed method. Section 4 focuses on experiments and their discussion. Finally, Section 5 presents the paper's conclusion.

2. Related work

In this section, a comprehensive literature review is presented focusing on feature selection (FS) techniques applied to the Wisconsin Diagnostic Breast Cancer (WDBC) dataset for classification tasks. The review emphasizes the use of various metaheuristic algorithms, which have been increasingly adopted to enhance classification performance by selecting the most relevant features. Several nature-inspired optimization methods have been explored in recent studies to reduce feature dimensionality while maintaining or improving diagnostic accuracy.

In [16], an intelligent breast cancer diagnosis method was proposed based on a hybrid Information Gain directed Simulated Annealing Genetic Algorithm Wrapper (IGSAGAW) for feature selection. Features are first ranked using the Information Gain

(IG) algorithm, and the top-ranked features are selected for classification using a Cost-Sensitive Support Vector Machine (CSSVM). This approach reduces the complexity of the feature selection process, effectively extracts an optimal feature subset, and enhances classification accuracy while minimizing misclassification cost. The method was validated on the Wisconsin Original Breast Cancer (WBC) and Wisconsin Diagnostic Breast Cancer (WDBC) datasets. Experimental results show that the proposed hybrid algorithm outperforms existing methods. The ultimate goal is to support real-world clinical diagnostic systems and aid physicians in making more accurate decisions, with potential applicability to other disease diagnoses.

A novel metaheuristic optimizer was used in [17], the Chaotic Crow Search Algorithm (CCSA), designed to enhance the performance of the standard Crow Search Algorithm (CSA) in feature selection tasks. CCSA integrates ten different chaotic maps into the optimization process to improve exploration and exploitation capabilities. The algorithm is evaluated on 20 benchmark datasets for feature selection, aiming to maximize classification performance while minimizing the number of selected features. Comparative analysis demonstrates that CCSA outperforms CSA and several state-of-the-art optimization algorithms. Notably, the sine chaotic map proves to be the most effective in enhancing CSA's performance, highlighting its suitability for boosting optimization efficiency.

In [18], Authors present a novel approach called EGWO-SVM (Enhanced Grey Wolf Optimization–Support Vector Machine) for breast cancer classification. The method leverages the Grey Wolf Optimizer (GWO), known for its simplicity, fast convergence, and effective balance between exploration and exploitation, to select the most relevant features from tumor data. By integrating an enhanced version of GWO with SVM, the approach identifies an optimal subset of features to accurately distinguish between benign and malignant tumors. The model was evaluated using the WDBC dataset and compared with several state-of-the-art methods. The EGWO-SVM achieved a classification accuracy of 98.24%, outperforming all compared approaches and demonstrating its effectiveness in breast cancer detection.

In [19], an enhanced Grey Wolf Optimizer (GWO) integrated with a Two-Phase Mutation (TPM) mechanism for solving feature selection problems using a wrapper-based approach was suggested. The continuous search space is mapped to binary using a sigmoid function, aligning with the binary nature of feature selection. The first mutation phase focuses on minimizing the number of selected features while maintaining accuracy, while the second phase adds informative features to improve classification performance. To reduce computation, mutation is applied with low probability. The k-Nearest Neighbor (k-NN) classifier is used for evaluation, employing K-fold cross-validation to mitigate overfitting. The method is benchmarked against algorithms like Flower Pollination, PSO, WOA, and Bat Algorithm on 35 datasets. Results and statistical analyses confirm the superior performance and effectiveness of the proposed method.

In [20], a robust Grey Wolf Optimization-Random Forest (RGWO-RF) hybrid approach is proposed for the classification of breast cancer. The methodology involves a two-step process: feature selection and classification. A modified Grey Wolf Optimizer (GWO) is employed to identify the most significant features from the dataset. These optimal features are then used by the Random Forest (RF) classifier, chosen for its robustness and high accuracy, to perform the classification. The proposed approach is evaluated using the WDBC dataset. Experimental results demonstrate that the integration of RGWO for feature selection with the RF classifier significantly improves classification accuracy and confirms the effectiveness and robustness of the approach in detecting breast cancer.

In [21], Authors proposed an effective feature engineering approach for extracting and refining features using the WDBC Dataset, aiming to evaluate its impact on classification performance. Six popular machine learning models are compared, including logistic Regression, Random Forest, Decision Tree, K-Nearest Neighbors, Multi-Layer Perceptron (MLP), and XGBoost. Experimental results reveal that the Decision Tree classifier outperforms others when combined with the proposed feature engineering, achieving an average accuracy of 98.64%.

A hybrid feature selection method called Correlation-ModifiedGWO (CMGWO) for breast cancer classification was suggested in [22]. It combines a correlation-based filter to remove redundant features and a Modified Grey Wolf Optimizer to select the most relevant ones. The approach improves classification accuracy using selected features. Tests were conducted on the WDBC dataset with classifiers like RF, SVM, and NB. The best result, 99.12% accuracy, was achieved using Random Forest.

3. Proposed method

For identifying optimal feature subsets that enhance classification performance while reducing dimensionality, building upon this body of work, we propose a novel feature selection approach based on the SFO algorithm. The objective is to effectively identify the most informative features from the WDBC dataset. Following the FS process, multiple machine learning classifiers are employed to evaluate the diagnostic performance of the selected feature subsets, demonstrating the robustness and generalizability of the proposed SFO-based method.

In the proposed work, the Wisconsin Diagnostic Breast Cancer (WDBC) dataset in CSV format is first uploaded and imported into the system for analysis. The dataset then undergoes a thorough preprocessing stage, which includes handling missing or inconsistent values, normalizing feature scales, and encoding categorical labels into numerical values to ensure compatibility with machine learning models. After preprocessing, feature selection is performed using the SFO algorithm, a nature-inspired metaheuristic technique that efficiently identifies the most relevant and informative features. By selecting an optimal subset of features, SFO enhances classification performance and reduces computational overhead. The selected features are then used for classification through an ensemble of machine learning models, including SVM, CNN, NB, and RF. This ensemble approach leverages the individual strengths of each classifier to achieve more accurate and stable predictions. Finally, the proposed methodology is evaluated using various performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, ensuring a comprehensive assessment of the classification effectiveness. Fig.1 represent a flowchart of the proposed work.

3.1 Preprocessing

Before applying the SFO for feature selection, the WDBC dataset underwent essential preprocessing steps to ensure data quality and improve model performance. The WDBC dataset consists of 569 instances, each with 30 numerical features extracted from digitized images of fine needle aspirate (FNA) of breast masses, along with a diagnosis label indicating whether the tumor is benign or malignant. First, we addressed missing values and outliers, although the dataset is relatively clean and well-structured. Next, we performed label encoding, converting the categorical diagnosis labels into binary numerical values suitable for machine learning algorithms. To ensure that all features contributed equally to the classification task and to avoid dominance by features with larger scales, we applied feature normalization using Min-Max scaling, which transforms the feature values to a uniform range between 0 and 1. This step was particularly important for improving the convergence of gradient-based classifiers and optimizing the performance of distance-based models such as SVM. Finally, the dataset was split into training and testing subsets, preserving class distribution to ensure balanced evaluation. These preprocessing steps laid a solid foundation for effective feature selection and accurate classification in the subsequent stages of our CAD system.

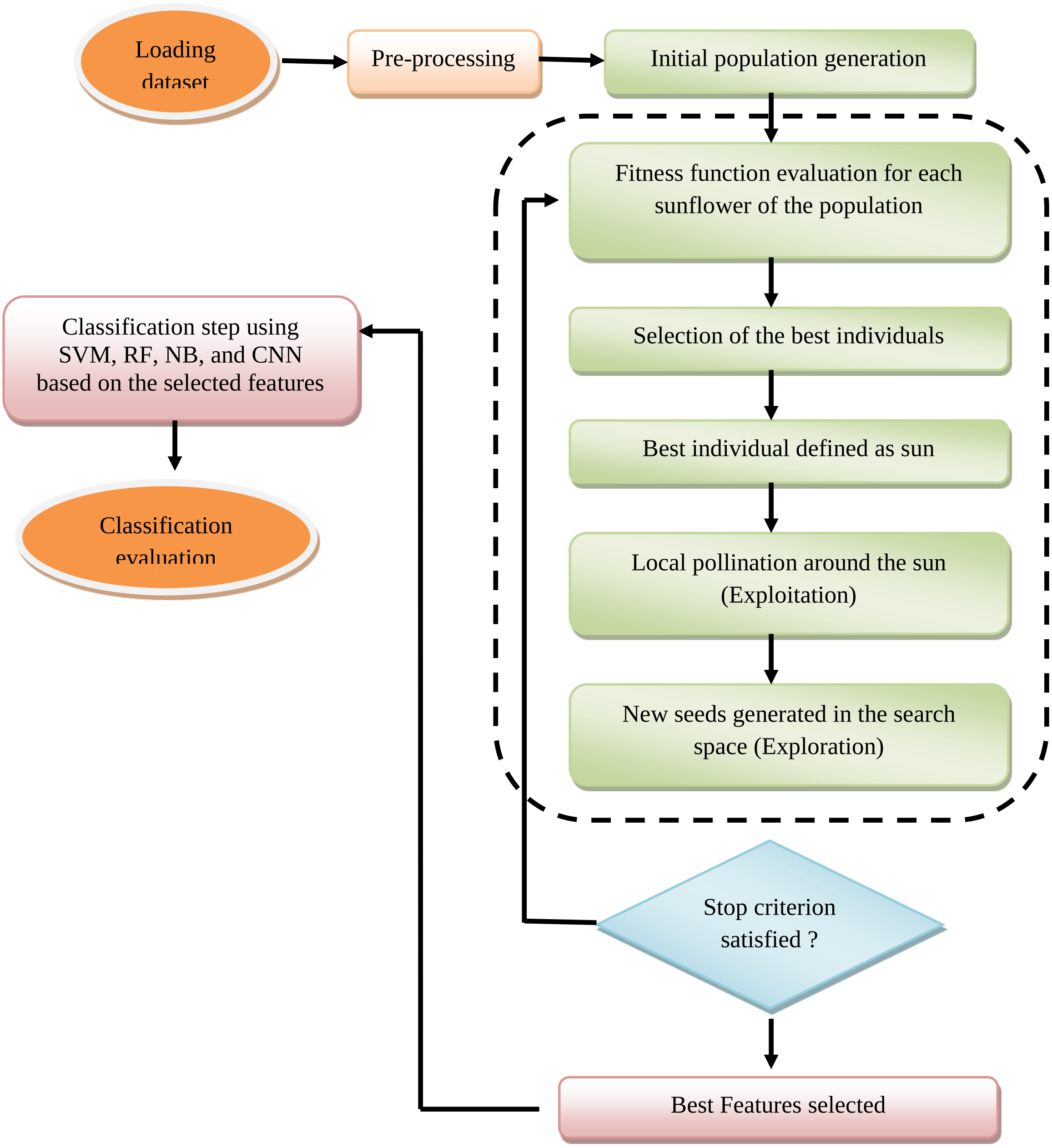

Fig.1. Flowchart of the proposed work.

3.2 Feature selection methodology using SFO

In this subsection we focus on two essential points, the first, an overview about the nature inspired of SFO was explained, then a description about how SFO select the best features.

3.2.1 Sunflower Optimization Algorithm

The Sunflower Optimization algorithm is a nature-inspired metaheuristic introduced by G. F. Gomes et al. in 2019 [14]. This algorithm models the behavior of sunflowers as they orient themselves toward the sun. Each morning, sunflowers instinctively adjust their positions to maximize exposure to sunlight, seeking the optimal direction for absorbing solar radiation. Throughout the day, they track the sun's movement, and at sunset, they reset their orientation to prepare for the next day. In the context of optimization, sunflowers that are already well-aligned with the sun (near-optimal solutions) make smaller movements and remain relatively stable. Conversely, sunflowers positioned far from the ideal direction (poorer solutions) take larger steps to move closer to the optimal orientation. This dynamic mimic the balance between exploration and exploitation, driving the population toward the global optimum. Fig.2 represent an initial population of flowers and identification of the sun. Then, all the sun flowers will be oriented toward the sun as we see in Fig.3. Fig.4 illustrate how best flower pollinate around the sun.

The steps of the SFO algorithm are:

1. Generate the population X_i^t randomly, i = 1, ..., n.
2. The fitness function f (X_i^t) of sunflowers is evaluated.
3. Retain the best solutions in the sunflower population X*.
4. Modify all sunflowers headed for the best one (called sun) as Equation (1).

$$s_i = \frac{X^* - X_i}{\|X^* - X_i\|} \quad , \qquad i=1,2\ldots,n \tag{1}$$

5. Determine the direction for each sunflower by Equation (2).

$$d_i = \lambda * \ P_i(\|X_i + X_{i-1}\|) * \ \|X_i + X_{i-1}\|, \tag{2}$$

In which,
λ: Inertial displacement of the sunflower plants.
P_i: Pollination probability.
$X_i,\ X_{i-1}$: Current position and nearest neighbor position

6. Examine the highest step of individual as Equation (3).

$$d_{max} = \frac{\|X_{max} - X_{min}\|}{2 * N_{pop}} \tag{3}$$

where,
X_{max}, X_{min}: The lower and upper limits.
N_{pop}: the number of populations.
The position of new generated individual (sunflower) is updated using the Equation (4).

$$\vec{X}_{i+1} = X_i + d_i * \ \vec{s}_i \tag{4}$$

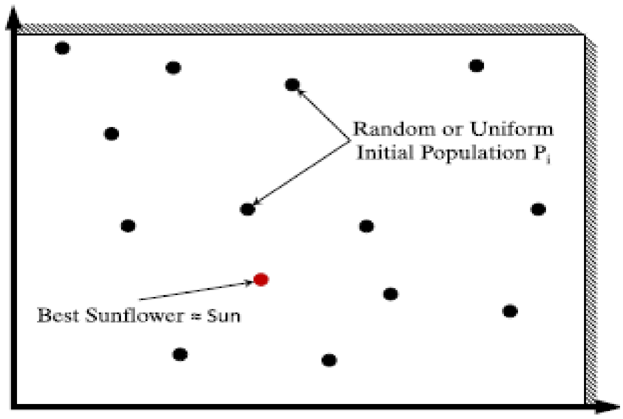

Fig.2. Initial population of flowers and identification of the sun(s).

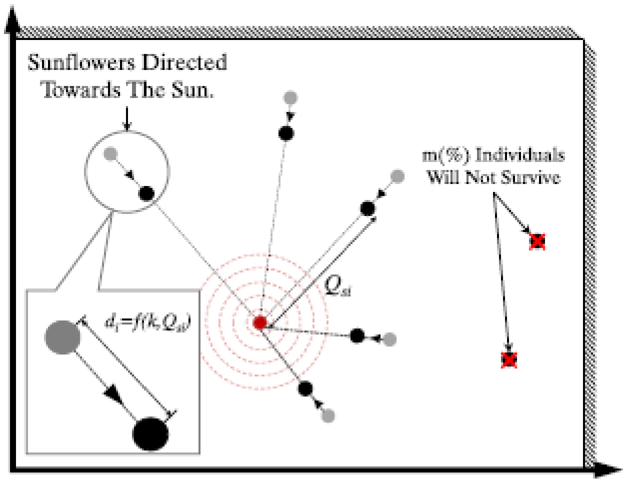

Fig.3.All the sun flowers will be oriented toward the sun.

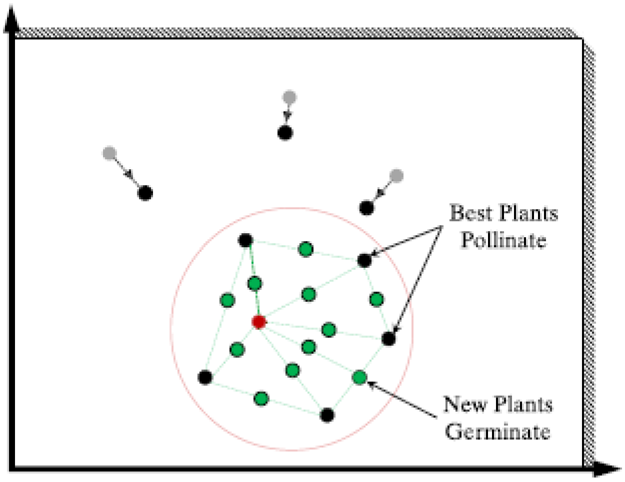

Fig.4. Best flowers pollinate around the sun.

3.2.2 Feature selection using SFO

In the context of feature selection, each sunflower in the population represents a potential subset of features selected from the dataset. The quality or fitness of each sunflower (feature subset) is evaluated using a classification-based objective function, typically measuring classification accuracy on a validation set. During each iteration, sunflowers adjust their positions in the feature space by simulating sunlight attraction, pollination, and reproduction. The most promising individuals those leading to higher classification performance exert stronger attraction on neighboring sunflowers, guiding the search toward more optimal feature subsets. Pollination allows for random exploration to prevent premature convergence and maintain diversity in the population. Through this balance of exploitation (intensification) and exploration (diversification), SFO efficiently navigates the high-dimensional feature space and converges toward a minimal subset of the most discriminative and relevant features. This process reduces data redundancy and improves classifier performance by eliminating irrelevant or noisy features. The original 30 features in the WDBC dataset were reduced to the 20 most informative features, selected based on their contribution to classification accuracy. This reduction eliminated redundant and irrelevant information that could negatively impact model performance. Ultimately enhancing the effectiveness and robustness of the breast cancer CAD system.

3.3 Classification step

After selecting the most relevant features using the SFO, the reduced feature set was passed to the classification phase to evaluate the effectiveness of the selected features in distinguishing between benign and malignant breast tumors. We employed four different classifiers to ensure robustness and comparative performance analysis, SVM, RF, NB, and a CNN. SVM is known for its effectiveness in handling high-dimensional spaces and its ability to construct optimal hyperplanes for binary classification tasks. RF, an ensemble-based method, provides high accuracy and resistance to overfitting by combining multiple decision trees. NB, a probabilistic classifier, offers simplicity and fast computation, especially suitable for medical datasets with conditional independence assumptions. Lastly, CNN, although typically used for image data, was adapted to process the feature vectors through a 1D convolutional architecture, enabling it to learn deep abstract representations from the selected features. The classification results demonstrated that the use of SFO significantly improved the predictive accuracy across all classifiers, with CNN and SVM achieving the highest performance. This confirms that the proposed SFO-based feature selection not only reduces dimensionality but also enhances the overall robustness, generalization ability, and diagnostic reliability of the breast cancer classification system.

4. Experimental results

In this experimental study, the WDBC dataset was employed to evaluate the performance of the proposed approach. The Sunflower Optimization algorithm was executed with an initial population of 20 candidate solutions and iterated for 50 generations to ensure robust convergence. To validate the model's effectiveness, the dataset was divided into two subsets: 80% of the data was allocated for training the classifier, while the remaining 20% was reserved for testing its generalization capability.

4.1. Experimental results with different ML classifiers

In order to prove the effectiveness of the proposed method, we calculate the accuracy of different ML algorithm. As we see in Table 1, the classification results reveal the impact of applying the SFO for FS on the WDBC dataset. Notably, the SFO-SVM model improved from 96.3% to 98.25%, showing that the selected features enhanced the SVM's ability to distinguish between classes. The SFO-CNN model also benefited significantly, increasing from 97.37% to 98.64%, indicating that deep learning models can effectively leverage the refined feature set to boost performance. SFO-NB experienced a modest improvement from 94.4% to 94.73%, suggesting that even a simple probabilistic model can benefit slightly from optimal feature selection. Overall, the results confirm that the Sunflower algorithm enhances classification performance, particularly for models like SVM and CNN, by eliminating irrelevant or redundant features.

Table 1. Classification accuracy results with feature selection and without feature selection.

Without feature selection		**Feature Selection using Sunflower algorithm**	
Classifer	**Accuracy (%)**	**Classifer**	**Accuracy (%)**
SVM	96,3	SFO-SVM	98,25
RF	97,8	SFO-RF	96,49
NB	94,4	SFO-NB	94,73
CNN	**97,37**	**SFO-CNN**	**98,64**

4.2. Comparing between different ML classifiers using several performance metrics

The detailed performance metrics provide deeper insights into the effectiveness of the classifiers when combined with Sunflower-based feature selection. From the Table 2, SVM achieved the highest sensitivity (100%), indicating it correctly identified all malignant cases without false negatives, a critical factor in medical diagnosis. However, its specificity (93.7%) suggests some benign cases were misclassified. CNN offered the best precision (98.59%), meaning it made very few false positive predictions, and also had the highest F1-score (97.90%), reflecting a balanced performance between precision and recall. RF delivered high sensitivity (99.1%) but had slightly lower specificity (92.1%) and precision (95.5%), indicating a tendency toward over-predicting the positive class. NB maintained a solid balance, with high sensitivity (99.1%) and specificity (98.9%), though its precision (93.9%) and F1-score (96.4%) were slightly lower compared to other models. Overall, CNN emerged as the top performers by achieving an accuracy equal to (98,64%) while the results underscore the effectiveness of the Sunflower Optimization algorithm in enhancing diagnostic performance through relevant feature selection.

Table 2. Comparing between different classifiers using several metrics of evaluation.

	SFO-SVM	**SFO-RF**	**SFO-NB**	**SFO-CNN**
Sensitivity	100	99,1	99,1	97,22
Specificity	93,7	92,1	98,9	97,62
Precision	96,4	95,5	93,9	**98,59**
F1-score	98,2	97,3	96,4	**97,90**

4.3. Comparing with existing methods

The comparison of classification accuracy across different feature selection methods highlights the effectiveness of the proposed approach. Our method, combining Sunflower Optimization for feature selection with a CNN classifier, achieved the highest accuracy of 98.64%, outperforming several existing techniques. The FS-Enhanced GWO by Kumar [18], which used SVM, achieved a close 98.24%, while FS-RGWO [20] with Random Forest also performed competitively at 98.6%. In contrast, the Crow Search Algorithm (CSA) [17] paired with KNN yielded a significantly lower accuracy of 90.28%, indicating a less effective feature selection or model compatibility. These results emphasize the strength of the proposed method in selecting optimal features that enhance deep learning-based classification performance, surpassing other well-known metaheuristic-based approaches.

Table 3. Comparing accuracy between the proposed approach and existing work.

Methods	Classifier	Accuracy (%)
Proposed	**CNN**	**98,64**
FS- Crow search algorithm (CSA) [17]	**KNN**	**90,28**
FS-Enhanced GWO [18]	**SVM**	**98,24**
FS-RGWO [20]	**RF**	**98,60**

4.4. Comparing the classification results using ROC-AUC between different Machine Learning classifiers

The ROC curve can help to understand more clearly the power of a machine learning system. Fig.5 demonstrates that CNN is the optimal classifier. The Area Under the Curve (AUC) measures a classifier's ability to discriminate between classes and summarizes the ROC curve. Higher AUC indicates greater classifier performance. Fig.5 demonstrates that CNN outperforms SVM, RF, and NB classifiers in terms of ROC-AUC metric, with an AUC criterion of 100%.

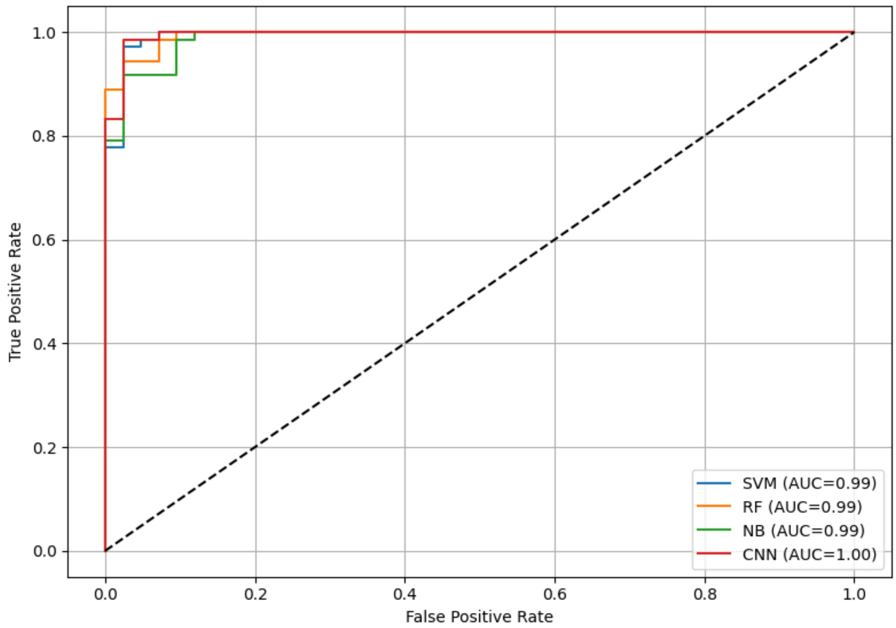

Fig.5. ROC-AUC metric for SVM, RF, NB, and CNN.

5. Conclusion

In this study, we presented an efficient breast cancer identification system that combines Sunflower Optimization-Based Feature Selection with Ensemble Machine Learning models. The proposed approach successfully reduces data dimensionality by selecting the most informative features from the WDBC dataset, resulting in enhanced classification performance. The experimental evaluation demonstrates that the selected features significantly improve the accuracy and robustness of various classifiers, particularly CNN and SVM, which achieved outstanding performance. By integrating a metaheuristic feature selection strategy with both deep and traditional machine learning models, our method provides a reliable, interpretable, and computationally efficient framework for medical diagnosis. These promising results highlight the potential of the proposed system as a valuable tool for early and accurate breast cancer detection. As future work, we plan to validate the system on larger and more diverse datasets, and explore its application in other critical domains such as multi-class tumor classification and real-time clinical decision support systems.

6. References

[1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36

Cancers in 185 Countries. CA Cancer J Clin. 2021 May ;71(3):209-249. 2021. doi: 10.3322/caac.21660. Epub 2021 Feb 4. PMID: 33538338.

[2] Kosmia Loizidou ∗ , Rafaella Elia, Costas Pitris, « Computer-aided breast cancer detection and classification in mammography: A comprehensive review", Computers in Biology and Medicine 153 (2023) 106554, 2023. https://doi.org/10.1016/j.compbiomed.2023.106554

[3] Saabia, A.B.R.; AbdEl-Hafeez, T.; Zaki, A.M. Face recognition based on GreyWolf Optimization for feature selection. In International Conference on Advanced Intelligent Systems and Informatics; Springer: Berlin/Heidelberg, Germany; pp. 273–283, 2018.

[4] Sreedharan, N.P.N.; Ganesan, B.; Raveendran, R.; Sarala, P.; Dennis, B. Grey Wolf optimisation-based feature selection and classification for facial emotion recognition. IET Biom, 7, 490–499, 2018.

[5] Gupta N, Gupta D, Khanna A, Reboucas F et al (2019) Evolutionary algorithms for automatic lung disease detection. Meas J Int Meas Confed 140:590–608. https:// doi. org/ 10. 1016/j. measu rement. 2019. 02. 042, 2019.

[6] A.M Taha, A. Mustapha and S.D. Chen, "Naïve Bayes-Guided bat algorithm for feature selection", The Scientific World Journal, vol 2013, http://dx.doi.org/10.1155/2013/325973 ,2013.

[7] Ibrahim, H.T.; Mazher,W.J.; Jassim, E.M. Feature Selection: Binary Harris Hawk Optimizer Based Biomedical Datasets. Intel. Artif. 2022, 25, 33–49, 2022.

[8] Long, W.; Jiao, J.; Xu, M.; Tang, M.; Wu, T.; Cai, S. Lens-imaging learning Harris hawks optimizer for global optimization and its application to feature selection. Expert Syst. Appl. 2022, 202, 117255, 2022.

[9] Chohra, A.; Shirani, P.; Karbab, E.B.; Debbabi, M. Chameleon: Optimized feature selection using particle swarm optimization and ensemble methods for network anomaly detection. Comput. Secur, 117, 102684, 2022.

[10] Wang, L.; Gao, Y.; Li, J.; Wang, X. A feature selection method by using chaotic cuckoo search optimization algorithm with elitist preservation and uniform mutation for data classification. Discret. Dyn. Nat. Soc, 2021, 1–19, 2021.

[11] Mafarja, M.; Aljarah, I.; Heidari, A.A.; Faris, H.; Fournier-Viger, P.; Li, X.; Mirjalili, S. Binary dragonfly optimization for feature selection using time-varying transfer functions. Knowl. Based Syst, 161, 185–204, 2018.

[12] Allam, M.; Nandhini, M. Optimal feature selection using binary teaching learning-based optimization algorithm. J. King Saud Univ. Comput. Inf. Sci. 34, 329–341, 2022.

[13] Zamani H, Shahraki M. Feature selection based on whale optimization algorithm for diseases diagnosis. Int J Comput Sci Inform Secur. https:// doi. org/ 10. 13140/ RG.2. 2. 29065. 88161, 2016.

[14] Guilherme Ferreira Gomes, Sebastiao Simões da Cunha Jr, Antonio Carlos Ancelotti Jr, A sunflower optimization (SFO) algorithm applied to damage identification on laminated

composite plates, Engineering with Computers ,2018. https://doi.org/10.1007/s00366-018-0620-8

[15] D. Dua and C. Gra®, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml] (University of California, School of Information and Computer Science, Irvine, CA, 2019.

[16] N. Liu, E. Qi, M. Xu and G. Liu, A novel intelligent classification model for breast cancer diagnosis, Inform. Process. Manage. (56) 609–623 , 2019.

[17] G. L. Sayed, A. E. Hassanien and A. T. Azar, Feature selection via a novel chaotic crow search algorithm, Neural Comput. Appl. (31) 171–188, 2019.

[18] S. Kumar and M. Singh, "Breast Cancer Detection Based on Feature Selection Using Enhanced Grey Wolf Optimizer and Support Vector Machine Algorithms," Vietnam Journal of Computer Science , Vol. 8, No. 2, 177–197, 2021. Doi: 10.1142/S219688882150007X.

[19] M. Abdel-Basset, D. El-Shahat, I. El-henawy and S. Mirjalili, A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection, Exp. Syst. Appl. (139) 112824 , 2020.

[20] Ali Mezaghrani, Mohamed Debakla, and Khalifa Djemal « Robust Method for Breast cancer classification Based on Feature selection using RGWO Algorithm » First International conference, on Aftificial lntelligence: Theories and Applications ICAITA 2022.Mascara, Algeria, November 7-8' 7022;

[21] Strelcenia, E.; Prakoonwit, S. Effective Feature Engineering and Classification of Breast Cancer Diagnosis: A Comparative Study. BioMedInformatics, 3, 616–631, 2023. https://doi.org/10.3390/ biomedinformatics3030042.

[22] Ali Mezaghrani, Mohammed Debakla, Khalifa Djemal. Nove1 feature selection method for accurate breast cancer classification using Correlation coefficient and Modified GWO Algorithm, Computer science journal of moldova, v o1.32, no .2(9 5), 2024,