Nigerian Research Journal of Engineering and Environmental Sciences 11(1) 2026 pp. 166-178
p ISSN: 2635-3342; e ISSN: 2635-3350

Nigerian Research Journal of Engineering and Environmental Sciences
Journal homepage: www.rjees.com

Original Research Article

Enhanced K-Nearest Neighbour Firefly Model for Breast Cancer Classification System

[1]Olusi, T.R., [2]Balogun, M.O., [3]Adepoju, T.M., [4]Sotonwa, K.A., [5]Adetunji, B.A., [5]Olabiyisi, O.S. and [3]Omidiora, O.E.

[1]Department of Computer Science, I.I.C.T. Kwara State Polytechnic, Ilorin, Nigeria.

[2]Department of Electrical and Computer Engineering, Kwara State University, P.M.B 1530, Malete, Ilorin, Kwara State, Nigeria.
[3]Department of Computer Engineering, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
[4]Department of Computer Science, Lagos State University, Lagos State, Nigeria.
[5]Department of Computer Science, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
*olusititilayo@gmail.com; monsurat.balogun@kwasu.edu.ng; tmadepoju@lautech.edu.ng; kehinde.sotonwa@lasu.ng
http://doi.org/10.5281/zenodo.21047523

ARTICLE INFORMATION

Article history:
Received 02 May 2026
Revised 29 May 2026
Accepted 06 Jun. 2026
Available online 30 Jun. 2026

Keywords:
K-Nearest Neighbour
Enhanced K-Nearest Neighbour
Breast cancer
Classification
Firefly

ABSTRACT

Most of the existing classification algorithms for breast cancer are computationally intensive, have memory limitations, and are sensitive to the local structure of data. In this research, there was a development of an enhanced K-Nearest Neighbour (KNN)-based Firefly model for the categorisation of breast cancer. Breast images were acquired from the mammogram images database of the University of Ilorin Teaching Hospital. KNN and K-Nearest Neighbour-Eagle Strategy Firefly Algorithm (KNN-ESFA), which formed the enhanced KNN, were used to classify the images. The model was implemented in the Python programming language version 6.0. Evaluation of KNN and enhanced KNN was carried out using sensitivity, specificity, accuracy, Prediction Time (PT), Memory Usage (MU), and Receiver Operating Characteristics (ROC) as performance metrics. The enhanced KNN and KNN techniques were validated by carrying out a t-test to compare the difference between the two techniques at a 5% significance level. The sensitivity, specificity, accuracy, PT, MU, and ROC for KNN gave 67.00%, 74.00%, 67.92%, 309.22s, 10KB, and 0.92, respectively, while the corresponding values for enhanced KNN were 95.00%, 98.00%, 96.50%, 5.92s, 5.31KB, and 0.98, respectively. The t-test indicated that there was a wide range of differences (p=0.000) between KNN and KNN-ESFA for PT. The developed model improved breast cancer classification compared to existing methods. Therefore, the enhanced K-Nearest Neighbour-based Firefly classification model can be used in the medical field for accurate and timely breast cancer detection and classification.

1. INTRODUCTION

In contemporary society, women face numerous health challenges, and among the most prevalent is breast cancer, the leading cause of cancer-related deaths among women (Healy et al., 2023). Breast cancer is a type of cancer that develops in the cells of the breast (Cuthrell et al., 2023). It can occur in both men and women, although it is much more common in women. Breast cancer forms when cells in the breast tissue begin to grow uncontrollably. Over time, these abnormal cells can invade nearby tissues and spread to other parts of the body (Sharma et al., 2020).

Both the diagnosis of diseases and the classification of breast cancer exemplify the effective utilization of data mining tools to address real-world challenges (Arem et al., 2018). Breast cancer initiates when cells within the breast undergo uncontrolled growth (Carter, 2017). Such cells gradually metamorphose into a tumour, which can eventually be detected and viewed using an X-ray and, at other times, the cells may be perceived as lumps (Arem & Loftfield, 2018). The cells become dangerous (cancerous) when they penetrate the tissues in the breast or spread to other parts of the body (Nolan et al., 2023).

Accurate classification not only aids in the early detection of cancer but also facilitates the selection of the most effective treatment strategies tailored to individual patient needs (Mathew et al., 2020). Mammography is a specialized medical imaging for scanning the breasts. A mammogram, which is also called a mammography examination, helps in timely detection and diagnosis of breast cancer (Aymaz, 2024; Yala et al., 2021). Mammogram image segmentation helps detect any area where breast cancer is found, leading to improved diagnosis (Abdelrahman et al., 2021). Occasionally, there may be difficulty in reading images associated with digital mammograms because there could be low contrast as well as variations in tissue types (Dratsch et al., 2023). Hence, it is imperative to have a Computer-Aided Detection (CAD) system that is capable of processing and analyzing medical images (Jochelson et al., 2021). These systems play a pivotal role in enhancing the accuracy of disease detection by significantly reducing the occurrence of errors made by radiologists (Coffey et al., 2022).

CAD systems employ sophisticated algorithms and image processing techniques to analyze medical images such as X-rays, CT scans, and MRIs. By doing so, they can highlight areas of concern or abnormalities that might otherwise be overlooked or misinterpreted by human observers (Badawy et al., 2017). One of the primary advantages of CAD is its ability to act as a second pair of eyes, effectively complementing the expertise of radiologists (Sannasi Chakravarthy and Rajaguru, 2020a). Even the most skilled radiologists can sometimes miss subtle signs of disease due to factors like fatigue, distraction, or sheer volume of cases. CAD systems, on the other hand, can tirelessly scrutinize images with consistent attention to detail, helping to catch abnormalities that might escape human notice (Sannasi Chakravarthy and Rajaguru, 2020b). By minimizing detection errors, CAD systems contribute to more accurate diagnoses and timely interventions, ultimately improving patient outcomes. Moreover, CAD can also serve as a valuable decision support tool, providing radiologists with additional information and insights to aid in their diagnostic process (Mohapatra et al., 2022).

CADs offer a promising solution by leveraging vast datasets to train models capable of recognizing patterns indicative of breast cancer with remarkable accuracy. These algorithms can analyze mammographic images with speed and consistency, aiding radiologists in identifying subtle signs of malignancy that might evade human detection (Kalita et al., 2022). By augmenting radiologists' expertise, machine learning algorithms have the potential to enhance the sensitivity and specificity of mammography, leading to earlier detection of breast cancer and improved patient outcomes (Kim et al., 2020). The integration of these algorithms into clinical practice holds the promise of optimizing screening protocols, reducing unnecessary biopsies, and ultimately, saving lives (Jahwar et al., 2022).

Machine learning algorithms for mammography stand poised to revolutionize breast cancer detection, offering a powerful tool in the fight against this prevalent and life-threatening disease among which are Convolutional Neural Networks (CNNs), Support Vector Machines (SVMs), K-Nearest Neighbours (KNN), Random Forest, decision Tree and so no (Ardakani et al., 2023). KNN is one of the simplest as well as straightforward methods of data mining. KNN is often referred to as Memory-Based Classification as the training samples need to be in the memory at run-time. KNN deals with both constant and discrete traits (Khamis et al., 2017).

The output is dependent on the use to which K-NN is put, for instance, if it is employed for the purpose of regression or that of classification. When it comes to KNN categorization, class membership is often an output. The categorization of an object is based on its general acceptability by the neighboring objects, and the chosen object is considered as the commonest among its k closest neighbors (k is a positive integer, normally small) (Khamis et al., 2017). As versatile as KNN in image classification, it is still prone to problems such as Sensitive to Noise and Outliers, memory usage and computational complexity among others. The limitations were solved by using firefly algorithm to optimize the performance of KNN on mammographic images.

The Firefly Algorithm (FA) is a metaheuristic optimization algorithm inspired by the flashing behavior of fireflies (Yang, 2008). FA aims to solve optimization problems by mimicking attractive and repulsive interactions between fireflies in their search for mates. The main idea behind the Firefly Algorithm is that fireflies are attracted to each other's light, with brighter fireflies exerting a stronger attraction. At the same time, they are also repelled by other fireflies, with closer and brighter ones exerting a stronger repulsion (Emary et al., 2015). The importance of accuracy in the classification of medical data cannot be overemphasized; a hybrid approach often enhances how accurate the categorization of breast cancer will be (Kadhim et al., 2023).

In this research, KNN is combined with the Firefly algorithm to enhance the accurate classification of breast cancer. This research is motivated by the features of the Eagle Strategy Firefly Algorithm and KNN. The two algorithms were hybridized to produce an optimized Computer-Aided Detection (CAD) tool that can be used to perceive as well as categorize breast cancer into normal, benign, or malignant.

2. MATERIALS AND METHODS

2.1. Methodology Workflow

The six sequential stages to achieve the aim of this work are explained as follows.

- Obtaining the mammographic images from the University of Ilorin Teaching Hospital database.
- Pre-processing procedure/data cleaning, which was achieved by removing noise, artifacts, and suppressing pectoral muscle using the split and merge technique.
- The data reduction phase was carried out by removing highly dense mammograms. At this phase, bilateral evaluation of the pairs of breast images to figure out suspected regions was carried out by setting the pixel value (Y_S) to 0.1 as the threshold for the image, which helped in identifying the highly dense breast in line with (Adepoju *et al*. 2015).
- In the segmentation phase, the watershed algorithm was used on the enhanced image.
- Implementation of extraction was made possible with the use of Gray Level Co-occurrence Matrix (GLCM) optimized with classical firefly. This was employed to extract all possible features that helped in detecting a Region of Interest (ROI) in a mammogram.
- KNN with Eagle Strategy Firefly Algorithm was used for categorizing breast cancer as Normal, Benign, and Malignant.

The schematic representation of the Proposed System for recognition and classification of breast cancer is shown in Figure 1.

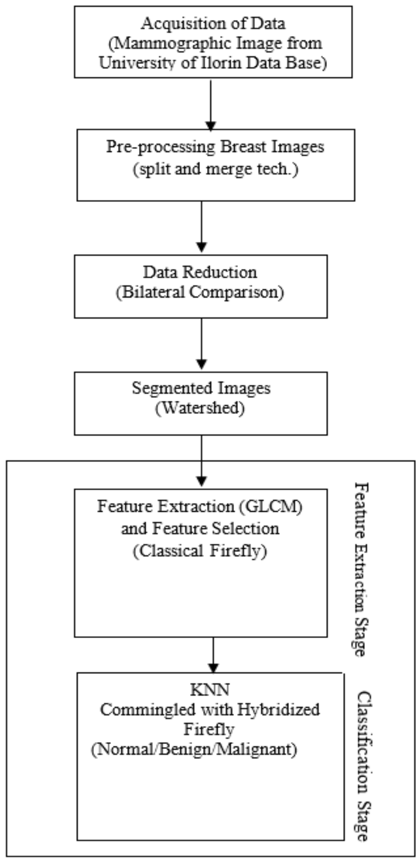

Figure 1: Flow diagram of the developed breast cancer classification system

2.2. Breast Cancer Classification

The input to the model is a data set from a Mammographic Image gotten from University of Ilorin Teaching Hospital Oke-Oyi in Kwara State, Nigeria. The database which was preprocessed contains One Thousand Five Hundred and Eighty-Five (1585) breast images. Some samples of the mammographic images are presented in Figure 2. K-fold cross-validation was used for the implementation of this system. The dataset was categorized into testing and training sets. Categorization is a data mining activity which allocates a specific entity to one out of many pre-determined categories based on the traits of the entity, using k-fold for this work. Based on Yadav et al. (2016), K-fold cross-validation is a statistical technique for evaluation as well as relating learning algorithms through separation of data into two sections. For this work, the first set is employed to train the required model; the second is employed as a means of validating the model to obtain an accurate result. Yadav et al. (2016) specified that in characteristic cross-validation, both training and validation sets must cross over in consecutive circles so that the possibility of every data set being validated against is equal. Shown in Figure 3 is the flowchart for KNN Eagle Strategy Firefly. The images were divided into five (5) folds with 317 images in each of the folds. In carrying out the test, a fold that contains 317 images was used for testing, while the remaining four (4) folds 1268 were used for training. These were repeated until all five folds were used for testing and training. The algorithm and flowchart of the KNN Eagle Strategy Firefly are presented in Algorithm 1 and Figure 3.

T.R. Olusi et al. / Nigerian Research Journal of Engineering and Environmental Sciences
11(1) 2026 pp. 166-178

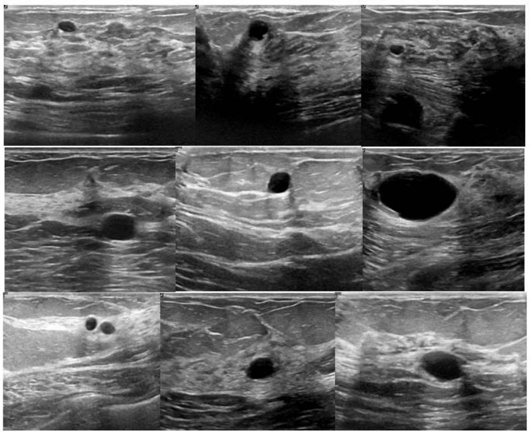

Figure 2: Samples of acquired mammographic images

Algorithm 1: KNN eagle strategy firefly algorithm

1. /* A is a 1-D array of 1585 elements*/
2. Input and store data in Array A
3. Apply split and merge technique

I=0;
M=317 ; /*(1585/5)
Do while i < 5
Test on [$A_{(i*m+1)}$ to $A_{(i+1)*m}$]
If i>0
Train on [$A_{(i+1)*m+1}$ *to* $A_{5*m)}$] + [A_1 to $A_{(i*m)}$]
Else
Train on [$A_{(i+1)*m+1}$ to $A_{5*m)}$]
Endif
f (x)= $A(i)$
Initialise $X^{n=0}$
while ($\|X^{n+1} - X^n\|$ > tolerance)
Random search by performing Levy walk
Evaluate the $f(x)$
Intensive local search with a hypersphere
via the Firefly Algorithm
if (a better solution is found)
Update the current best
end if
Update $n = n + 1$
Calculate means and standard deviations
Calculate distance
Classify image
end while
end do

4. Output result
5. Stop

T.R. Olusi et al. / Nigerian Research Journal of Engineering and Environmental Sciences
11(1) 2026 pp. 166-178

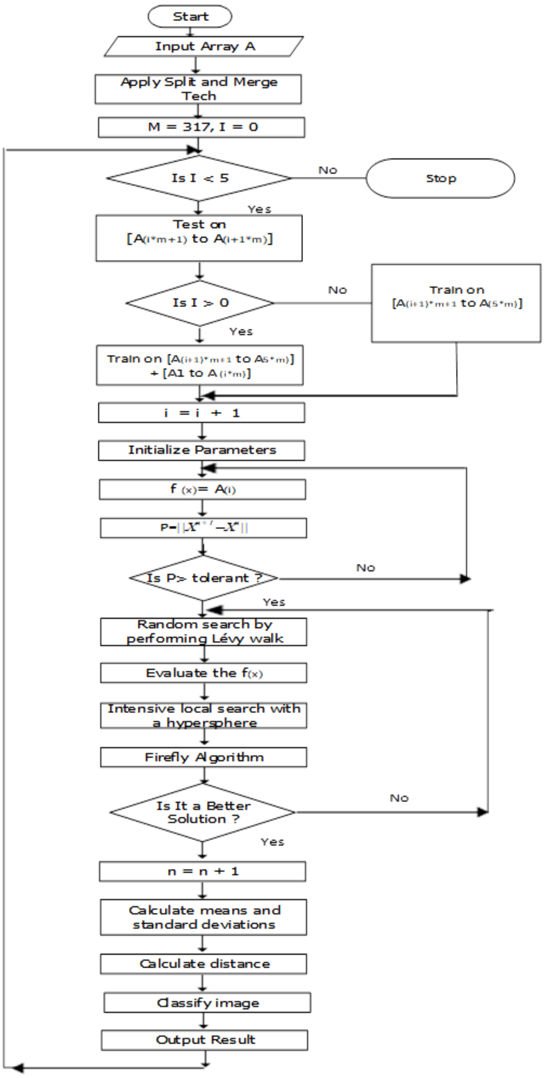

Figure 3: Flowchart for KNN eagle strategy firefly

2.3. Interactive Graphic User Interface for the Database

A collaborative Graphic User Interface (GUI) was developed for a mammogram-image database using the Python programming language, aiming to enhance user-friendliness. The GUI facilitated easier navigation and interaction with the system. The implementation involved utilizing both KNN and KNN with the Eagle Strategy Firefly algorithm. These systems were integrated into the desktop environment for execution. Upon launching the main file with a .py extension, users were prompted with a login interface. Upon successful authentication, users could browse and select new data for classification. The selected data was then inputted into the respective KNN models. Figures 4 to 7 showcase samples of the developed collaborative GUI.

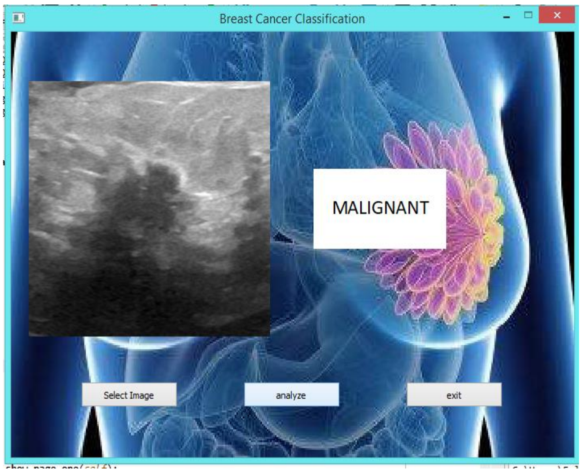

Figure 4: Classified result for malignant type for KNN eagle strategy firefly

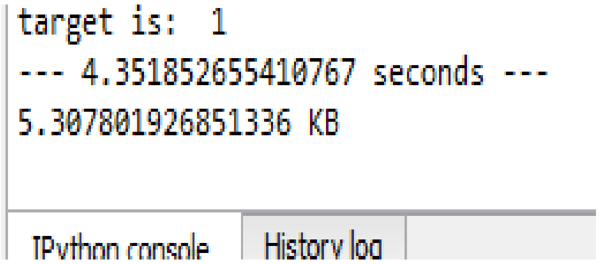

Figure 5: Processing time and memory usage for malignant classification using KNN eagle strategy firefly model

Figure 6: Classified result for benign type for KNN eagle strategy firefly

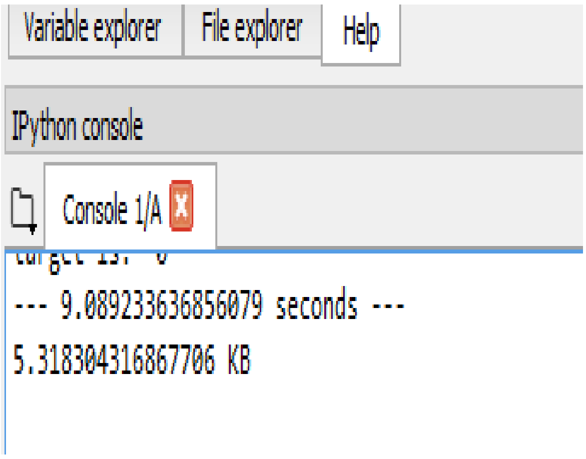

Figure 7: Processing time and memory usage for benign classification using KNN eagle strategy firefly model

3. RESULTS AND DISCUSSION

3.1. System Evaluation

The KNN model was executed for the three classes. For benign classification, the processing time with KNN was approximately 313.80 seconds. The same process was repeated for malignant and normal classes, resulting in processing times of 307.50 seconds and 306.37 seconds, respectively. The average results indicate that the processing time with KNN was relatively high. This slow processing can be attributed to the large size of the data (1.6 Megabytes), which was stored in the working storage area during execution. In contrast, when using the KNN Eagle Strategy Firefly classifier, the processing times were notably lower. Specifically, benign classification took 9.10 seconds, malignant classification took 4.32 seconds, and normal classification took 4.30 seconds, as shown in Figure 8.

The average results presented in Figure 8 demonstrate that KNN-Firefly effectively addressed the limitations of traditional KNN. This improvement is attributed to Firefly's utilization of the Eagle Strategy technique, which combines levy flight search with the Firefly algorithm. The inspiration behind the Eagle Strategy stems from the foraging behaviour of eagles, wherein they swiftly employ efficient tactics upon spotting prey (Fister et al., 2013). Specifically, when cancer is detected in the breast, the heightened light intensity of the affected area attracts the attention of KNN Eagle Strategy, enabling rapid cancer detection. This contrasts with the traditional KNN algorithm, which lacks this capability, thus leading to reduced classification processing time. Consequently, this enhancement mitigates the computational costs associated with KNN, as depicted in Figure 8, which clearly illustrates the disparities in processing time between KNN and the enhanced KNN variant.

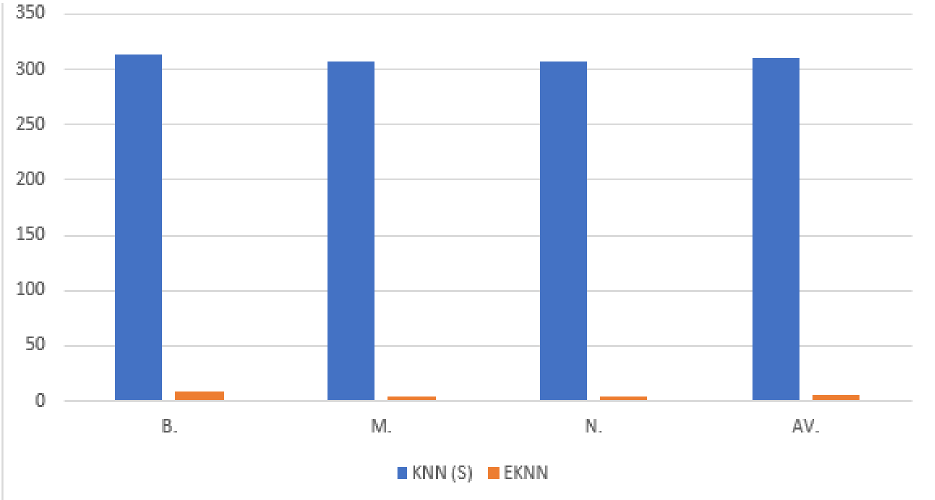

Figure 8: Classification efficiency of KNN and KNN-Firefly

3.2. Confusion Matrix for KNN

The outcomes of employing the KNN classifier on datasets categorized as normal, benign, and malignant are presented in Figure 9. In the normal class category, 57 out of 106 normal data images were correctly classified as normal, resulting in a true-positive value of 57, false positives of 8, true negatives of 25, and false negatives of 16. The sensitivity for this class stands at 67%, with a specificity of 74% and an F1-score of 0.70. For the benign class, there were 71 true positives, 12 false positives, 20 true negatives, and 3 false negatives. The sensitivity for this class is 61%, with a specificity of 76% and an F1-score of 0.68. In the malignant class, there were 80 true positives, 5 false positives, 14 true negatives, and 7 false negatives, resulting in a sensitivity of 87%, a specificity of 62%, and an F1-score of 0.63, with an accuracy of 67.92%, as depicted in Figure 9. These metrics were notably low due to KNN's characteristic of storing all training data, making it sensitive to irrelevant features, which contributed to the drawbacks of the KNN algorithm this is in line with (Demidova, 2021).

3.3. Confusion matrix for KNN with Eagle Strategy Firefly

The dataset comprises 318 images, evenly distributed across three classes: "normal," "malignant," and "benign." Figure 10 illustrates various metrics obtained from the evaluation of the classification stage. In the normal class category, 69 out of 106 normal data were correctly classified as normal, yielding a true-positive value of 69, with only 1 false positive, 32 true negatives, and 4 false negatives. This results in an impressive sensitivity of 95% and a specificity of 98%, accompanied by an F1-score of 0.96. These high values indicate the robustness of the classifier in accurately identifying normal cases. For the benign class, 73 instances were correctly classified as benign, with 3 false positives, 29 true negatives, and 1 false negative. This yields a sensitivity of 98%, a specificity of 91%, and an F1-score of 0.94. The classifier demonstrates excellent performance in identifying benign cases, with high sensitivity and specificity values contributing to its overall accuracy.

T.R. Olusi et al. / Nigerian Research Journal of Engineering and Environmental Sciences
11(1) 2026 pp. 166-178

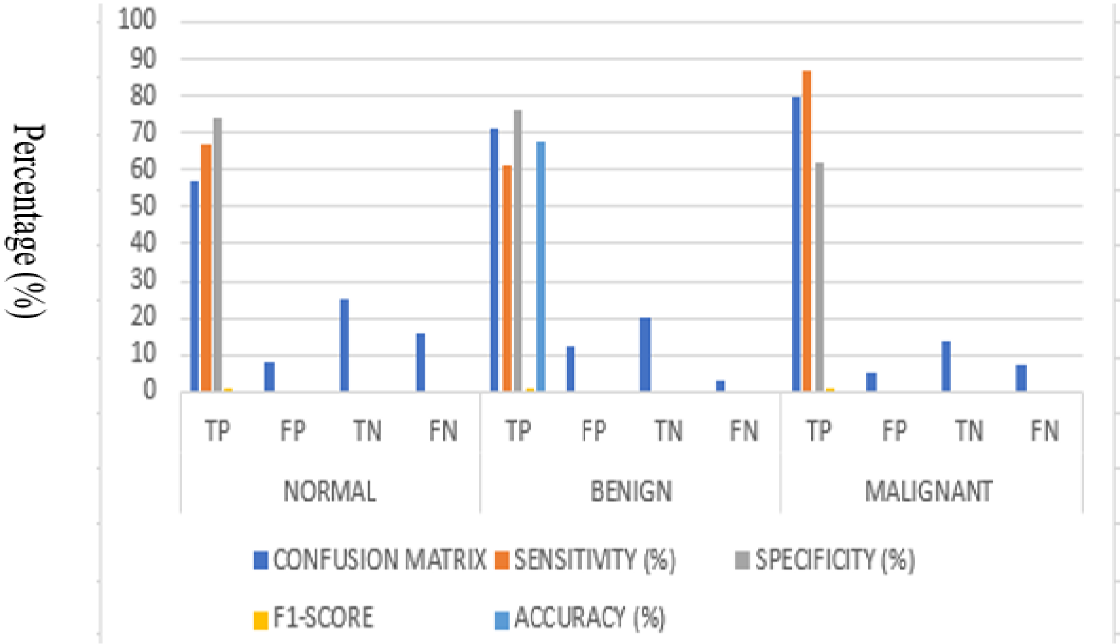

Figure 9: Confusion matrix for KNN

In the malignant class, 86 cases were classified correctly as malignant, with only 1 false positive, 18 true negatives, and 1 false negative. This results in a sensitivity of 98%, a specificity of 88%, and an F1-score of 0.92. The classifier exhibits remarkable accuracy in detecting malignant cases, with high sensitivity indicating its ability to capture true positives effectively. The overall accuracy of the classifier stands at an impressive 96.5%, as depicted in Figure 10. This indicates the overall effectiveness of the KNN Eagle Strategy Firefly classifier in accurately categorizing images into their respective classes. The improvement seen in the KNN Firefly classifier compared to traditional KNN can be attributed to the utilization of the Eagle Strategy Firefly. This strategy incorporates an automatic subdivision of the population into subcategories with specified typical distances. This innovative approach enhances the performance of the KNN classifier by optimizing its ability to discern patterns and relationships within the data, leading to improved classification accuracy and efficiency.

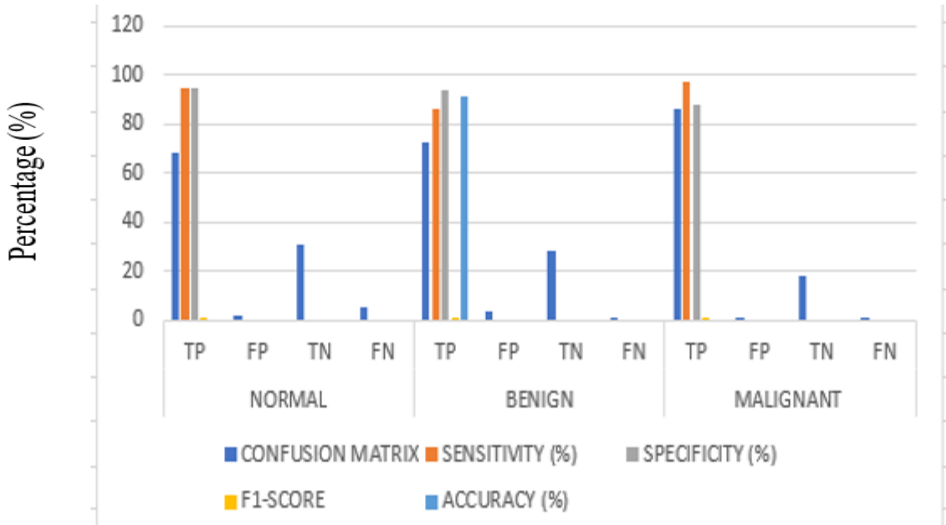

Figure 10: Comparison of KNN with eagle strategy firefly using the confusion matrix

3.4. Comparison of Receiver Operating Characteristic (ROC) KNN with KNN Eagle Strategy Firefly

The Receiver Operating Characteristic (ROC) curve is adopted for evaluating the output quality of classifiers. It is a plot of true positive rate against the false positive rate. A higher AUC value indicates a better classifier (Prosperi et al., 2023). For the KNN Eagle Strategy Firefly classifier, the AUC (Area under the ROC Curve) is reported to be 0.98, as depicted in Figure 11. Comparatively, the AUC for traditional KNN is noted to be 0.92. This significant difference in AUC values suggests that the KNN Eagle Strategy Firefly classifier outperforms traditional KNN in terms of discriminative ability. Furthermore, the accuracy of the classifiers can be inferred from their respective AUC values (Lavazza et al., 2023). Since the AUC of the KNN Eagle Strategy Firefly is 0.98, which is notably higher than the AUC of traditional KNN (0.92), it follows that the accuracy of the KNN Eagle Strategy Firefly classifier (96.5%) is also higher than that of traditional KNN (67.92%). The superiority of the KNN Eagle Strategy Firefly classifier over traditional KNN is thus established.

This is supported not only by the higher AUC value but also by the corresponding higher accuracy. The ROC analysis provides a robust quantitative measure to compare the performance of different classifiers, and in this case, it unequivocally indicates the superiority of the KNN Eagle Strategy Firefly classifier for breast cancer classification. According to Li et al., 2023 a larger AUC signifies a better-performing model. With an AUC of 0.98, the KNN Eagle Strategy Firefly classifier surpasses traditional KNN (AUC of 0.92), further confirming its superiority in breast cancer classification.

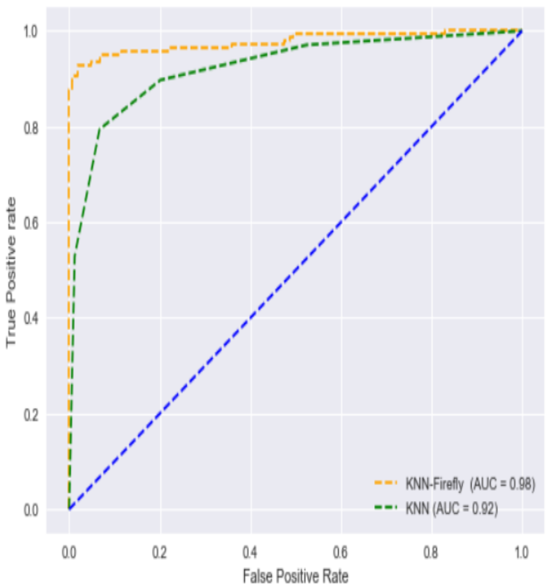

Figure 11: ROC curve for comparing the performance of KNN with KNN Eagle strategy firefly

4. CONCLUSION

This study highlights the substantial performance enhancement achieved by integrating the Eagle Strategy Firefly algorithm with the K-Nearest Neighbors (KNN) classifier for breast cancer categorization. The results demonstrate a significant boost in accuracy, with the KNN Eagle Strategy Firefly model achieving an impressive 96.5% accuracy, compared to the conventional KNN accuracy of 67.92%. Moreover, the Receiver Operating Characteristic (ROC) analysis reveals a notable improvement in performance, with the KNN Eagle Strategy Firefly model achieving an ROC score of 0.98, in contrast to the KNN's score of 0.92. The substantial disparity in accuracy and ROC scores underscores the superior classification capabilities of the KNN Eagle Strategy Firefly model over traditional KNN. These findings not only hold statistical significance but also indicate the enhanced discriminatory power and robustness of the proposed model. Beyond theoretical implications, these results suggest tangible benefits for the medical field. By providing a more accurate and reliable method for breast cancer classification, the enhanced KNN Eagle Strategy Firefly model offers a promising approach for improving both diagnosis and treatment outcomes. Its superior performance highlights its potential as a valuable tool for clinicians and researchers, facilitating improved recognition and categorization of breast cancer cases. In conclusion, the development of the enhanced KNN Firefly model signifies a significant advancement in breast cancer classification. The demonstrated enhancements in accuracy and performance position this model as a viable option for adoption in medical settings, where precise classification is crucial for effective decision-making and patient care. This research not only contributes to the progression of machine learning techniques but also holds promise for enhancing the accuracy and efficacy of breast cancer diagnosis and management practices.

5. CONFLICT OF INTEREST

There is no conflict of interest associated with this work.

REFERENCES

Abdelrahman, L., Al Ghamdi, M., Collado-Mesa, F. and Abdel-Mottaleb, M., (2021). Convolutional neural networks for breast cancer detection in mammography: A survey. *Computers in biology and medicine*, *131*, p.104248.

Adepoju, T.M., Ojo, J.A., Omidiora, E.O., Olabiyisi, S.O. and Bello, T.O., (2015). Detection of tumour based on breast tissue categorization. *British Journal of Applied Science Technology*, *11*(5), pp.1-12.

Ahmed, N., Islam, M.Z. and Farjana, S., (2012). Pattern of skin diseases: Experience from a rural community of Bangladesh. *Bangladesh Medical Journal*, *41*(1), pp.50-52.

Ardakani, A.A., Mohammadi, A., Mirza-Aghazadeh-Attari, M. and Acharya, U.R., (2023). An open-access breast lesion ultrasound image database: Applicable in artificial intelligence studies. *Computers in Biology and Medicine*, *152*, p.106438.

Arem, H. and Loftfield, E., (2018). Cancer epidemiology: a survey of modifiable risk factors for prevention and survivorship. *American journal of lifestyle medicine*, *12*(3), pp.200-210.

Aymaz, S., (2024). A new framework for early diagnosis of breast cancer using mammography images. *Neural Computing and Applications*, *36*(4), pp.1665-1680.

Badawy, S. M., Hefnawy, A. A., Zidan, H. E. and GadAllah, M. T. (2017). Breast cancer detection with mammogram segmentation: a qualitative study, *International Journal of Advanced Computer Science and Applications,* 8(10), pp 2382-2393.

Carter, V.L., (2017). *Breast cancer risk factors in a sexual minority population: an examination of the 2014 and 2015 Behavioral Risk Factor Surveillance System* (Doctoral dissertation, University of Alabama Libraries).

Coffey, K. and Jochelson, M.S., (2022). Contrast-enhanced mammography in breast cancer screening. *European journal of radiology*, *156*, p.110513.

Cuthrell, K.M. and Tzenios, N., (2023). Breast cancer: updated and deep insights. *International Research Journal of Oncology*, *6*(1), pp.104-118.

Demidova, L.A., (2021). Two-stage hybrid data classifiers based on SVM and KNN algorithms. *Symmetry*, *13*(4), p.615.

Dratsch, T., Chen, X., Rezazade Mehrizi, M., Kloeckner, R., Mähringer-Kunz, A., Püsken, M., Baeßler, B., Sauer, S., Maintz, D. and Pinto dos Santos, D., (2023). Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*, *307*(4), p.e222176.

Duffy, S.W., Tabár, L., Yen, A.M.F., Dean, P.B., Smith, R.A., Jonsson, H., Törnberg, S., Chen, S.L.S., Chiu, S.Y.H., Fann, J.C.Y. and Ku, M.M.S., (2020). Mammography screening reduces rates of advanced and fatal breast cancers: results in 549,091 women. *Cancer*, *126*(13), pp.2971-2979.

Emary, E., Zawbaa, H.M., Ghany, K.K.A., Hassanien, A.E. and Parv, B., (2015), September. Firefly optimization algorithm for feature selection. In *Proceedings of the 7th balkan conference on informatics conference* (pp. 1-7).

Fister, I., Fister Jr, I., Yang, X.S. and Brest, J., (2013). A comprehensive review of firefly algorithms. *Swarm and evolutionary computation*, *13*, pp.34-46.

Higa, A., (2018). Diagnosis of breast cancer using decision tree and artificial neural network algorithms. *cell*, *1*(7), pp.23-27.

Islam, M.Z., Ahmed, M.S., Ahmed, N., Farjana, S. and Mazumder, S.K., (2014). Tetanus toxoid vaccination coverage among women of reproductive age: Experience from a rural community. *Bangladesh Medical Journal*, *41*(1), pp.37-41.

Jahwar, A.F. and Abdulazeez, A.M., (2022), May. Segmentation and classification for breast cancer ultrasound images using deep learning techniques: a review. In *2022 IEEE 18th international colloquium on signal processing and applications (CSPA)* (pp. 225-230). IEEE.

Jochelson, M.S. and Lobbes, M.B., (2021). Contrast-enhanced mammography: state of the art. *Radiology*, *299*(1), pp.36-48.

Kadhim, R.R. and Kamil, M.Y., (2023). Comparison of machine learning models for breast cancer diagnosis. *IAES International Journal of Artificial Intelligence*, *12*(1), p.415.

Kalita, D.J., Singh, V.P. and Kumar, V., (2022). Detection of breast cancer through mammogram using wavelet-based LBP features and IWD feature selection technique. *SN Computer Science*, *3*(2), p.175.

Khamis, H.S., Cheruiyot, K.W. and Kimani, S., (2017). Application of k-nearest neighbour classification in medical data mining in the context of Kenya.

Kim, J.Y., Kim, J.J., Hwangbo, L., Suh, H.B., Kim, S., Choo, K.S., Nam, K.J. and Kang, T., (2020). Kinetic heterogeneity of breast cancer determined using computer-aided diagnosis of preoperative MRI scans: relationship to distant metastasis-free survival. *Radiology*, *295*(3), pp.517-526.

Kumar, M.S., Harini, M., Kalaiyarasi, M. and Rajaguru, H., (2023), December. EEG Signals: Performance Analysis of the Firefly Optimization Technique for Classifying Epilepsy Risk Levels. In *2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR)* (Vol. 1, pp. 1-6). IEEE.

Healy, L.M., (2023). Certain Breast Cancer Patients Can Safely Forgo Radiotherapy After Surgery. *Cancer Therapy Advisor*, pp.NA-NA.

Lavazza, L., Morasca, S. and Rotoloni, G., (2023), June. On the reliability of the area under the ROC curve in empirical software engineering. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering* (pp. 93-100).

Li, B., Gatsonis, C., Dahabreh, I.J. and Steingrimsson, J.A., (2023). Estimating the area under the ROC curve when transporting a prediction model to a target population. *Biometrics*, *79*(3), pp.2382-2393.

Lisany, N.F., Jamal, M.A.H.M., Chung, H.J., Hong, S.T. and Rahman, M.S., (2020). Prognostic significance of the Cdk5 gene in breast cancer: an in silico study. *Network Modeling Analysis in Health Informatics and Bioinformatics*, *9*(1), p.19.

Mathew, T.E. and Kumar, K.A., (2020). A logistic regression based hybrid model for breast cancer classification. *Indian Journal Computer of Science and Engineering*, *11*(6), pp.899-906.

Mazen, F., AbulSeoud, R.A. and Gody, A.M., (2016). Genetic algorithm and firefly algorithm in a hybrid approach for breast cancer diagnosis. *International Journal of Computer Trends and Technology*, *32*(2), pp.62-68.

Mohapatra, S., Muduly, S., Mohanty, S., Ravindra, J.V.R. and Mohanty, S.N., (2022). Evaluation of deep learning models for detecting breast cancer using histopathological mammograms Images. *Sustainable Operations and Computers*, *3*, pp.296-302.

Naranje, S.M., Erali, R.A., Warner Jr, W.C., Sawyer, J.R. and Kelly, D.M., (2016). Epidemiology of pediatric fractures presenting to emergency departments in the United States. *Journal of Pediatric Orthopaedics*, *36*(4), pp.e45-e48.

Nolan, E., Lindeman, G.J. and Visvader, J.E., (2023). Deciphering breast cancer: from biology to the clinic. *Cell*, *186*(8), pp.1708-1728.

Prosperi, A., Korswagen, P.A., Korff, M., Schipper, R. and Rots, J.G., (2023). Empirical fragility and ROC curves for masonry buildings subjected to settlements. *Journal of Building Engineering*, *68*, p.106094.

Sannasi Chakravarthy, S.R. and Rajaguru, H., (2020). Detection and classification of microcalcification from digital mammograms with firefly algorithm, extreme learning machine and non-linear regression models: A comparison. *International Journal of Imaging Systems and Technology*, *30*(1), pp.126-146.

Sharma, G.N., Dave, R., Sanadya, J., Sharma, P. and Sharma, K., (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology and research*, *1*(2), pp.109-126.

Swain, S.M., Shastry, M. and Hamilton, E., (2023). Targeting HER2-positive breast cancer: advances and future directions. *Nature reviews Drug discovery*, *22*(2), pp.101-126.

Whelan, T.J., Smith, S., Parpia, S., Fyles, A.W., Bane, A., Liu, F.F., Rakovitch, E., Chang, L., Stevens, C., Bowen, J. and Provencher, S., (2023). Omitting radiotherapy after breast-conserving surgery in luminal A breast cancer. *New England Journal of Medicine*, *389*(7), pp.612-619.

Xiahou, X. and Harada, Y., (2022). B2C E-commerce customer churn prediction based on K-means and SVM. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(2), pp.458-475.

Yadav, S. and Shukla, S., (2016), February. Analysis of k-fold cross-validation over hold-out validation on colossal datasets for quality classification. In *2016 IEEE 6th International conference on advanced computing (IACC)* (pp. 78-83). IEEE.

Yala, A., Mikhael, P.G., Strand, F., Lin, G., Smith, K., Wan, Y.L., Lamb, L., Hughes, K., Lehman, C. and Barzilay, R., (2021). Toward robust mammography-based models for breast cancer risk. *Science translational medicine*, *13*(578), p.eaba4373.

Yang, X.S., (2008). *Introduction to mathematical optimization: from linear programming to metaheuristics*. Cambridge international science publishing.

Yazdani, A., Safaei, A.A., Safdari, R. and Zahmatkeshan, M., (2019). Diagnosis of breast cancer using decision tree, artificial neural network and naive bayes to provide a native model for fars province. *Payavard Salamat*, *13*(3), pp.241-250.

Yu, X., Zhou, Q., Wang, S. and Zhang, Y.D., (2022). A systematic survey of deep learning in breast cancer. *International Journal of Intelligent Systems*, *37*(1), pp.152-216.