

Mammogram Images Classification Using Machine Learning Techniques: A Comparative Study

Waleed Mohammed*, Ashraf Mohammed*, Kamal Bashir*, Ibrahim Mohammed*, Ali Ahmed†, Mohamed Mosadag*

*College of Computer Science and Information Technology, Karary University, Omdurman 12304, Sudan

†Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
waleed.abbo11@gmail.com

Abstract

Cancer is an ordinary disease of today's world, as one in three people is diagnosed with it at some stage in their lives. Of all the diseases categorized under cancer, the best chance at early diagnosis and detection is given to breast cancer. It has been noted that early detection increases the likelihood of a successful treatment. Classification of mammographic images is critical to early diagnosis of the disease. Two classifiers were used in this study, various machine learning classifiers applied to a breast image dataset. The classifiers include: Naive Bayes (NB), Logistic Regression (LR), Multilayer Perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), AdaBoost, Decision Tree (DT), C4.5, Random Forest (RaF) and the six statistical features were extracted from the MIAS dataset. To evaluate results we use different performance metrics such as Accuracy, AUC, F-Measure, Precision, Recall. RaF has the highest AUC followed by MLP, indicating its effectiveness in classification.

Key words-Breast cancer; image processing; data mining; classification; machine learning

I. INTRODUCTION

Cancer is a malignant tumour which begins within the cells of an organ or tissue and tends to grow and divide in an uncontrolled manner. Of the known types of cancer, there are over a hundred of them[1]. Estimations have it that the global incidences of cancer will rise to 20 million by 2020[2]. Similar to most diseases, the treatment of cancer also differs with the type of cancer, its stage and location; however, early detection and preventive measures can avert 90-95% of cancer incidences. Breast cancer alone is one of the most global diseases[3]. In most cases the initial sign is an asymptomatic nodular mass in the breast. Still, pain may occur in about 10% of the patients even without such a tumor present. Like the other breast cancer diseases, the general condition of this illness could be managed significantly if diagnosed early. Thus, timely detection not only broadens the possibilities of salvation, but also increases its degree several times[4][5]. For early detection of breast cancer, they are commonly recommended and are often used together: which in connection with individual diseases are determined using personal (independent) tests of the mammary glands, clinical tests for the mammary glands (doctors emphasized), mammography[6]. Among them, X-RAY mammography is present day one of the most widely used technique of breast cancer detection and early detection[7][8]. However, this method has limitations in terms of sensitivity and specificity: mammography can fail to identify 4-34% of breast cancer while identifying about 70% of benign tumors. Secondly, as mammography involves ionizing radiation, it is not always preferred to the conventional screening. The method is also insufficient in its own right, and the calculated higher centiles are also generally wrong. Given such threats, other techniques of identifying the disease are being established most of which are less invasive and less painful. More recently, studies employing

different microwave frequencies have proffered good results in differentiating small malignant tumours from non malignant breast tissues[9][10]. Therefore, more and more investigation is directed on this non-invasive microwave frequency to sense tumor. Machine learning has revolutionized various fields by providing powerful tools to analyze large datasets, recognize patterns, and make predictions with high accuracy [11][12][13][14][15][16]. Machine learning applications in healthcare, finance, marketing, and robotics enable computer systems to execute data-based decision-making and automate tasks only possible through human work. Machine learning in healthcare allows the system to predict disease tailoring and treatment plans and enhance patient results using predictive models. The financial sector employs machine learning for fraud identification to execute trading algorithms and determine credit values. Recommendation systems operating on machine learning facilities enhance personalized content delivery on e-commerce and social media platforms[17]. Traditional machine-learning tools are valuable assets to help detect breast cancer in mammograms. The classification task uses traditional algorithms Naive Bayes (NB) together with Logistic Regression (LR) and Multilayer Perceptron (MLP) and Support Vector Machine (SVM) along with K-Nearest Neighbors (KNN) and AdaBoost followed by Decision Tree (DT), C4.5 and Random Forest (RaF). This set of traditional techniques remains widely used and effective in this domain. Mammogram image features based on texture aspects and edge patterns with shape measurements are extracted by these methods, which feed information into classification systems to determine image categories between benign and malignant outcomes. Interpretable outcomes and limited computing expenses make traditional machine-learning techniques highly valuable for clinical settings with resource constraints. Such techniques help build implementation-friendly models that can be deployed easily because large annotated datasets needed by

deep learning models might not exist[18].

II. THE PROPOSED METHOD FOR MAMMOGRAM IMAGES CLASSIFICATION

Phases one through six of the proposed system as shown in Fig(1) structure are image collection, image cropping, feature extraction, machine learning classifications, testing and assessment, and, in the end, the result comparison.

A. Image Collection Phase

The Mammography Image Analysis Society (MIAS), a consortium of UK research groups devoted to mammography studies, built a digital database of mammography pictures throughout the image-gathering stage. A well-known available mammogram image database The datasets used in the study are sourced from the Mammographic Image Analysis Society (MIAS), with any image having a resolution of 1024 1024 pixels. represents 322 digital mammogram images of 161 pairs at a resolution of 50 as a portable gray map style with associated truth data from the MIAS, UK [18], is selected. The MIAS data is used to feed as standard inputs to the proposed method in this work. Images are associated with annotation labels as benign or malignant from a total of 119 patients and also indicates the positions of any abnormal[19].

B. Image Cropping Phase

In this phase, the concept of a region of interest (ROI) means the small portion of an image that You are willing to filter or apply any operation.

C. Feature Extraction Phase

This step involves using six functions to extract feature values from each mammography picture after cropping the ROI using a [radius] based on the MIAS dataset from the [x] location to the [y] position. Further information about these six feature extraction-related functions can be found in the paragraphs that follow:

1) *Mean*: It is a metric that shows the average intensity of the nearby pixels in image

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

Where: μ is the mean (average) of the dataset.
N is the total number of elements in the dataset.
 x_i represents each individual element in the dataset.
The summation symbol $\sum$ denotes the sum of all the elements from $i = 1$ to N.

2) *Standard Deviation*: Standard deviation measures how spread out the numbers are from the mean.

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (2)$$

Where: σ is the standard deviation.
N is the total number of elements in the dataset.
 x_i represents each individual element in the dataset.

μ is the mean of the dataset.

The summation symbol $\sum$ represents the sum of the squared differences between each data point and the mean.

3) *kewness*: The measurement of skewness describes the absence of symmetry and its intensity in statistical data distributions and collections. Asymmetry pattern in a distribution leads to the statement that its shape is symmetric. Symmetric datasets must produce skewness values close to zero, while the normal distribution yields a skewness value that equals zero. A distribution shifts toward the left when the skewness value becomes negative and toward the right when the skewness value turns positive.

$$Skewness = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^3 \quad (3)$$

Where: Skewness measures the asymmetry of the probability distribution of a real-valued random variable.

N is the number of data points. x_i is each individual data point.

μ is the mean of the dataset.

σ is the standard deviation of the dataset.

4) *Kurtosis*: The degree of data distribution pointiness emerges from kurtosis measurement, which reveals data similarities to a normal distribution pattern. The peak of a data set with a high kurtosis value shows a clearly defined and concentrated distribution at the mean value.

$$Kurtosis = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \mu}{\sigma} \right)^4 - 3 \quad (4)$$

Where: Kurtosis measures the "tailedness" of the probability distribution of a real-valued random variable.

N is the number of data points.

x_i is each individual data point.

μ is the mean of the dataset.

σ is the standard deviation of the dataset.

5) *Contrast*: The identification of objects through their representation in pictures and displays depends on the variations in brightness and color patterns known as contrast. Visual perception depends on contrast, which explains the color and brightness differences between objects and all visible elements around them.

$$Contrast = \frac{\max(x_i) - \min(x_i)}{\max(x_i) + \min(x_i)} \quad (5)$$

Where: x_i represents the pixel intensities in the image.

$\max(x_i)$ is the maximum pixel intensity.

$\min(x_i)$ is the minimum pixel intensity.

6) *Smoothness*: Calculates the proximate intensity deviations in an area, is often used to describe the degree of variation in pixel intensities (or values in a dataset). A smooth image or dataset has less fluctuation in values, meaning the differences between neighboring pixels or data points are minimal.

$$Smoothness = \frac{1}{1 + Contrast} \quad (6)$$

Alternatively, smoothness in the context of texture features can also be defined as:

$$Smoothness = \frac{1}{\sum_{i=1}^N (x_i - \mu)^2} \quad (7)$$

Where: x_i represents each pixel or data point.
 μ is the mean of the dataset or image.
 N is the total number of data points or pixels.

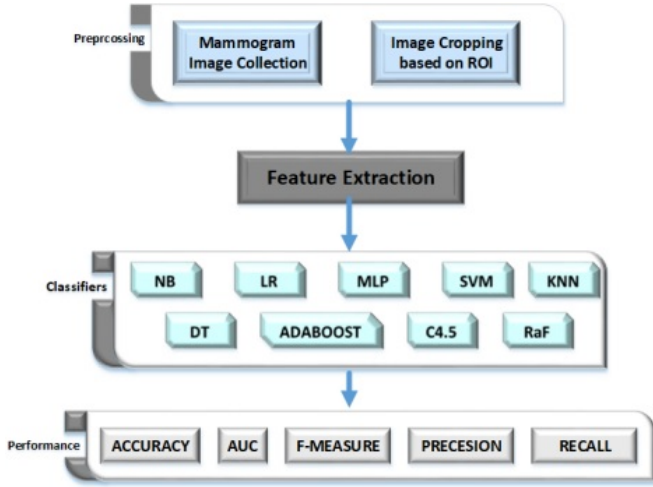


Fig. 1: The proposed method for Mammogram Images Classification

D. Classification Phase

The study uses eight commonly implemented supervised learning techniques. Techniques in this study. These include NB, RaF, KNN, MLP, SVM, C4.5, ADABOOST, DT and LR. The WEKA Weka serves as the platform for developing all defined classifiers. The model uses optimal performance parameters for the operation What is specific to certain learning algorithms is their own set of parameters. Thus, In this examination, the following configurations were applied to the learning models. Learners based on preliminary investigation. The NB learner is required to use the useKernelEstimator set to true for implementing kernel estimation on numerical values. The learner uses kernel distribution estimation instead of standard normal estimation to handle numeric attributes. The RaF learning algorithm requires modifications to its values. The research uses 200 as the assigned numTrees value instead of the default 100. The change leads to an expanded number of trees resulting from the system. The number of trees needs to be increased to determine more accurate out-of-bag predictions. The author followed earlier research findings, which warned about extended computation times when elevating numTree values. The selection of this value considered the concerns mentioned in past research. In the case a configuration of the KNN learning algorithm uses Weight by 1/distance as the distance weighting parameter while adjusting the kNN parameter to 5 and enabling cross-validate. The Parameter is tuned to true. Considering the MLP, we set the hiddenLayers parameter to 3. This means that One hidden layer contains three nodes when we define this network. The validationSetSize Parameter is also adjusted to 10. Applying this approach, the MLP classifier devotes 10% of training data for validation. The iterative training procedure needs a validation set to establish termination points. Two

configurations are made. The SVM receives two configuration settings: the first configures the c complexity constant value at 5.0, and the second activates the use Laplace Parameter. The method for creating Logistic Models has been enabled to be true in the current configuration. In this work, the Laplace parameter for C4.5 is activated (set to true) to enhance the probability estimates. The final parameter for LR activation required setting the errorOnProbabilities value to True while setting the useCrossValidation value to False. The useCrossValidation parameter is set to false alongside valid for the errorOnProbabilities parameter to minimize root mean squared error. The root mean squared error.

E. Performance Evaluation

TABLE I: CONFUSION MATRIX

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Many approaches are used in assessing performance. These include Accuracy, Sensitivity and Specificity. These measures are calculated from the confusion matrix. A confusion matrix is a two-way table used to evaluate a classification model where actual test values are already known. The structure of the confusion matrix is illustrated in Table(I): Where True Positive (TP): The number of positive samples correctly classified as positive, False Negative (FN): The number of positive samples incorrectly classified as negative, False Positive (FP): The number of negative samples incorrectly classified as positive, True Negative (TN): The number of negative samples correctly classified as negative.

1) *Accuracy*: Accuracy is the proportion of correctly classified instances out of all instances. It is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

2) *AUC (Area Under the Curve)*: AUC refers to the area under the Receiver Operating Characteristic (ROC) curve. It is a performance metric that provides an aggregate measure of a classifiers ability to distinguish between classes. The range of AUC is from 0 to 1:

3) *F-Measure (F1-Score)*: The F1-Score is the harmonic mean of Precision and Recall, which balances the trade-off between them. It is defined as:

$$F1 - Score = 2 \frac{Precision \cdot Recall}{Precision + Recall} \quad (9)$$

Where: Precision = $TP / (TP + FP)$ and Recall = $TP / (TP + FN)$

4) *Precision*: Precision is the proportion of positive predictions that are actually correct. It is defined as:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

5) *Recall*: Recall is the proportion of actual positives that are correctly identified. It is defined as:

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

TABLE II: THE CLASSIFICATION PERFORMANCE FOR ALL CLASSIFIERS USING DIFFRENT PERFORMANCE EVALUATION METRICS

	NB	LR	MLP	SVM	KNN	ADABOOST	DT	C4.5	RaF
ACCURACY	0.581	0.596	0.641	0.618	0.488	0.596	0.526	0.611	0.595
AUC	0.59	0.527	0.697	0.591	0.465	0.41	0.492	0.5	0.723
F-MEASURE	0.443	0.294	0.492	0.448	0.338	0	0.263	0	0.162
PRECESION	0.419	0.307	0.542	0.452	0.344	0	0.295	0	0.227
RECALL	0.493	0.32 0.49 0.47 0.353 0 0.25 0 0.133	0.49	0.47	0.353	0	0.25	0	0.133

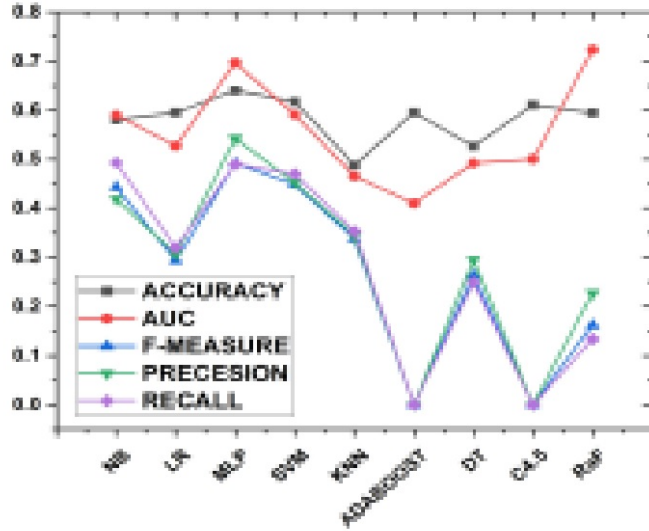


Fig. 2: The classification performance for all classifiers using different performance evaluation metrics

III. RESULTS AND DISCUSSIONS

The experiment in this study was conducted steps as follows:

- 1) Collecting mammogram images.
- 2) Cropping each image based on (X, Y) coordinates and radius (R) to detect the region of interest.
- 3) Extracting six statistical features from the images.
- 4) Applying the nine classifiers using the extracted statistical features (the experiment was performed three times) and calculating the classification results.

The data partitions were configured through 10-fold cross-validation. The ten subsets in each dataset remain distinct (the folds) and are about the same size. Ten times, Nine folds among the available resources get chosen to perform model training before the remaining fold conducts testing. The tenth portion serves for evaluation purposes, while the nine sections work to develop models. All learners receive subject assessments based on the software defect datasets. The empirical results analysis is conducted through MATLAB version 3.6.13 and WEKA version 3.6.13. The breast image dataset evaluation using Table(II) and Fig(2) reveals significant findings about the different machine learning classifiers' performance. The experimental assessment involved NB, LR, MLP, SVM, KNN, DT, C4.5, and RaF as the assessed classifiers. The performance assessment of the models included Accuracy, AUC (Area Under the Curve), F-Measure,

Precision and Recall evaluation metrics. MLP produces the optimal results when evaluating all available classifiers. The model delivers the maximum value of 0.697 AUC, establishing its impressive capabilities to detect faulty and functioning instances. The model exhibits a superior F-Measure scoring of 0.492 and precision, which reached a value of 0.542 because it strikes an excellent compromise between precision and recall. MLP shows its ability to recognize intricate data relationships in this evaluation because of its deep learning foundation. The AUC value of RaF reaches 0.723, surpassing all other classifiers, indicating its highest discrimination capability. The strong discriminatory power of RaF demonstrates that it serves well as a prediction tool for breast images. The metrics of F-Measure at 0.162 and Precision at 0.227 reveal RaF has good class separation capabilities but demonstrates limitations in Precision and recall balance control. The results indicate that Adaboost provides the worst performance in all tested metrics. This classifier provides absolutely no meaningful predictions since its AUC measures 0.41 while its F-Measure reaches 0.00. The inferior metrics from Adaboost point toward its badly suited nature for handling this particular dataset since imbalanced data and a limited number of meaningful weak learners might be to blame. LR and NB show moderate performance. LR's AUC score is 0.527, whereas NB's reaches 0.590 but falls behind MLP and RaF in AUC performance. For simple baseline models, the Accuracy level between 58-59% indicates that these models would not be the most suitable for defect prediction tasks. DT and C4.5 show near-random performance, with AUC values of 0.492 and 0.500, respectively. Analysis shows that the models struggle with generalization ability because they apply training patterns exclusively to their prediction decisions. KNN demonstrates a moderate to weak performance in the dataset through its AUC value of 0.465.

IV. CONCLUSION

Breast cancer is one of the diseases that top the chart that strikes the life of women, so it is essential to have an early detection and diagnosis of the disease. In this paper, we conduct a comprehensive comparative study on mammogram image classification using nine different classifiers. Based on the metrics' strong and consistent performance, MLP appears to be the best classifier for breast image prediction. RaF also does quite well in terms of AUC. Adaboost, DT, and KNN have relatively lower predictive performances in the dataset and thus might not be suitable. The results show that advanced machine learning techniques, MLP or RaF, are better than methods like NB, LR, or DT. In future work, we will use more dataset deep learning methods.

REFERENCES

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 68, no. 6, pp. 394–424, 2018.
- [2] R. Kumar, A. Sharma, and R. K. Tiwari, "Application of microarray in breast cancer: An overview," *Journal of Pharmacy and Bioallied Sciences*, vol. 4, no. 1, pp. 21–26, 2012.
- [3] S. Hartmann, T. Reimer, and B. Gerber, "Management of early invasive breast cancer in very young women (< 35 years)," *Clinical breast cancer*, vol. 11, no. 4, pp. 196–203, 2011.
- [4] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019," *CA: a cancer journal for clinicians*, vol. 69, no. 1, pp. 7–34, 2019.
- [5] C. Brekelmans, C. Seynaeve, M. Menke-Pluymers, H. Brüggewirth, M. Tilanus-Linthorst, C. Bartels, M. Kriege, A. van Geel, C. Crepin, J. Blom, et al., "Survival and prognostic factors in brca1-associated breast cancer," *Annals of oncology*, vol. 17, no. 3, pp. 391–400, 2006.
- [6] A. N. Hisham and C.-H. Yip, "Overview of breast cancer in malaysian women: a problem with late diagnosis," *Asian Journal of Surgery*, vol. 27, no. 2, pp. 130–133, 2004.
- [7] B. Stewart and C. P. Wild, "International agency for research on cancer," *World cancer report*, vol. 2014, 2014.
- [8] L. L. Lipscombe, P. J. Goodwin, B. Zinman, J. R. McLaughlin, and J. E. Hux, "The impact of diabetes on survival following breast cancer," *Breast cancer research and treatment*, vol. 109, pp. 389–395, 2008.
- [9] L. Yang, D. M. Parkin, J. Ferlay, L. Li, and Y. Chen, "Estimates of cancer incidence in china for 2000 and projections for 2005," *Cancer Epidemiology Biomarkers & Prevention*, vol. 14, no. 1, pp. 243–250, 2005.
- [10] M. Hassan, A. Malik, D. Fofi, B. Karasfi, and F. Meriaudeau, "Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study," *Measurement*, vol. 149, p. 106804, 2020.
- [11] K. Bashir, T. Li, C. W. Yohannese, and Y. Mahama, "Enhancing software defect prediction using supervised-learning based framework," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 1–6, IEEE, 2017.
- [12] K. Bashir, T. Li, and C. W. Yohannese, "An empirical study for enhanced software defect prediction using a learning-based framework," *International Journal of Computational Intelligence Systems*, vol. 12, no. 1, pp. 282–298, 2018.
- [13] K. Bashir, T. Li, C. W. Yohannese, and M. Yahaya, "Smotefris-inffc: Handling the challenge of borderline and noisy examples in imbalanced learning for software defect prediction," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 1, pp. 917–933, 2020.
- [14] K. Bashir, T. Ali, M. Yahaya, and A. S. Hussein, "A hybrid data pre-processing technique based on maximum likelihood logistic regression with filtering for enhancing software defect prediction," in *2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 921–927, IEEE, 2019.
- [15] K. Bashir, T. Li, C. W. Yohannese, M. Yahaya, and T. Ali, "A novel preprocessing approach for imbalanced learning in software defect prediction," in *Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018)*, pp. 500–508, World Scientific, 2018.
- [16] K. Bashir, T. Li, and M. Yahaya, "A novel feature selection method based on maximum likelihood logistic regression for imbalanced learning in software defect prediction," *Int. Arab J. Inf. Technol.*, vol. 17, no. 5, pp. 721–730, 2020.
- [17] K. Bashir, S. Pirasteh, H. Abdelrhman, M. Mosadag, and A. Mohammed, "An enhanced feature selection approach for breast cancer prediction using a hybrid framework," *Journal of Karary University for Engineering and Science*, 2024.
- [18] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, et al., "Mammographic image analysis society (mias) database v1. 21," 2015.
- [19] A. Kharrat, K. Gasmi, M. B. Messaoud, N. Benamrane, and M. Abid, "A hybrid approach for automatic classification of brain mri using genetic algorithm and support vector machine," *Leonardo journal of sciences*, vol. 17, no. 1, pp. 71–82, 2010.



Waleed Mohammed is a Lecturer in College of Computer Science and Information Technology, Karary University, got his B.S. degree in Computer Science from Karary University, Sudan, in 2011, MSc. degree in Information Technology from Elnileen University, Sudan, in 2019. His area of research interests Includes Data Mining, Machine Learning, Data Base



Ashraf Mohammed is a Lecturer in College of Computer Science and Information Technology, Karary University, got his B.S. degree in Computer Science from Karary University, Sudan, in 2010, MSc. degree in Information Technology from Elnileen University, Sudan, in 2019. His area of research interests Includes Data Mining, Machine Learning, Data Base.



Kamal Bashir is an Assistant Professor in College of Information Science and Technology, Karary University, got his B.S. degree in Computer Science from Karary University, Sudan, in 2009, MSc. degree in Computer Science from Khartoum University, Sudan, in 2013, PhD. degree in Computer Science and Technology, China in 2020. His area of research interests Includes Data Mining, Machine Learning, Software Quality Assessment.



Ibrahim Mohamed is an Associate professor in College of Information Science and Technology, Karary University, got his B.S. degree in Computer Science from Sudan University of Science and Technology, Sudan, in 1998, MSc. degree in Computer Science from Al Gezira University, Sudan, in 2004, PhD. degree in Computer Science and Technology, from Ain Shams University, Egypt, in 2015. His area of research interests Includes Ai expert systems in medical domain, data mining and biometrics.



Ali Ahmed received the B.Sc. degree in computer engineering from Karary University, Omdurman, Sudan, in 2001, the M.Sc. degree in computer science from Khartoum University, Khartoum, Sudan, in 2003, and the Ph.D. degree in computer science from UTM University, Malaysia, in 2013. He did his Postdoctoral study at UTM University, in 2014. Since 2014, he has been working as an Assistant Professor in Sudan (Karary University). He is currently a Professor of computer science with King Abdulaziz University, Rabigh, Saudi Arabia. He is also a Professor with FCITR, King Abdulaziz University. His research interests include machine learning, information retrieval, and medical image retrieval.



Mohamed Mosadag is an Assistant Professor in College of Information Science and Technology, Karary University, got his B.S. degree in Computer Science from Karary University, Sudan, in 2010, MSc. degree in Computer Science from Karary University, Sudan, in 2016, PhD. degree in Computer Science and Technology, from Dongla University, Sudan, in 2024. His area of research interests Includes Machine Learning, Data Base.