e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

Machine Learning Models and Interactive Dashboards in Breast Cancer Detection

Dr. Emmanuel Udoh[1], Dr. Tachi Udoh[2], Liz Udoh[3]

Subhani Department of Computer Information Systems, Texas A&M University Central, Killeen, USA[1]

Johns Hopkins Osler Internal Medicine, Johns Hopkins University, Baltimore, USA[2]

Neuroscience/Psychology Department, University of Kentucky, Lexington, USA[3]

Abstract: Breast cancer is one of the leading types of cancer in the world, affecting many people every year. Early diagnosis with high accuracy is vital for treatment and patient care. In this project, publicly available cancer data sets from the University of California, Irvine Repository (kaggle.com) were analysed. Data analysis reveals that cancer morphological features, such as radius, perimeter, and area, exhibit a very high correlation coefficient in the detection process. The decision tree model revealed that the concave point is a highly relevant predictor, with a threshold of 0.048 to distinguish between malignant and benign tumors. The logistic regression model achieved an accuracy of 80.95% and an F1 score of 0.75, indicating good overall classification performance; however, a precision score of 0.60 suggests a moderate capability to minimize false predictions. By leveraging machine learning models and interactive dashboards (utilizing advanced data analytics and visualization), the work supports healthcare professionals in making more informed decisions regarding tumor classification and patient care.

Keywords: Breast Cancer, Interactive Dashboard, Machine Learning, Visualization Tool, Detection/Risk Management

I. INTRODUCTION

Breast cancer is one of the leading cancers globally among women and the second most common cause of cancer death in U.S women [17]. Early detection is a key factor for improving overall prognosis of the disease. Breast tumors are classified into two types – benign and malignant tumors [1]. As can be gleaned from Fig. 1, benign breast tumours are generally smooth, round, and not spreading, while malignant breast tumours are irregularly shaped with spiky borders and invasive [1, 16]. Although benign and malignant breast tumors are morphologically different, the process of diagnosing them is the same: physical examination, imaging tests (mammogram, ultrasound), and biopsy. These diagnostic steps help detect the presence, size, shape, and roughness of the lump.

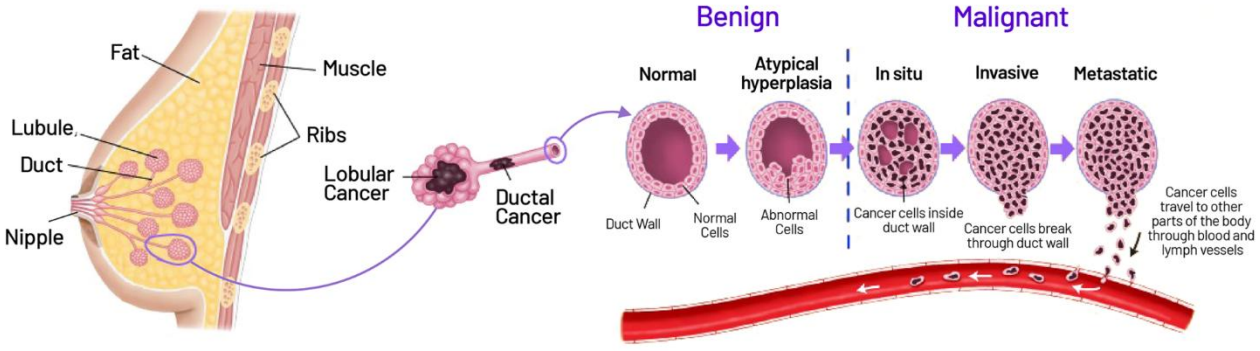

Fig. 1 Progression of Breast Cancer [1].

Despite the technological advancements and tools to diagnose breast cancer, challenges remain in breast cancer detection and characterization with the current diagnostic imaging modalities and tests [14]. Distinguishing between benign and malignant breast tumors often require additional interventions and often invasive tests that still may not yield definitive diagnosis. Even with more advanced medical technologies, such as imaging, misdiagnosis is still a common problem. This may result in subjecting benign tumour patients to unnecessary treatments or delays in early referrals for the management of malignant tumours that may require special attention [14].

Traditional diagnostic methods face significant challenges, including misdiagnosis, feature complexity, resource allocation, and issues with decision support [16]. For instance, misdiagnosis risk could result in false positives leading to

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

unnecessary treatments and patient anxiety, while false negatives result in delayed care and reduced survival rates. Healthcare professionals need better tools to interpret complex tumour characteristics, requiring a better and reliable decision support system. According to [16], the current state may lead to subjective interpretation, limited integration of quantitative tumor features in decision-making, and a lack of real-time analytical insights, as well as insufficient visualization tools for pattern recognition.

In this project, a set of variables was considered, including 32 clinical and morphological characteristics associated with tumour size, texture, and shape, and their diagnosis. A systematic examination of this data was undertaken to establish models that may improve the accuracy of diagnostics and decisions made by various healthcare workers. The cancer data set, which includes diverse measurements of tumors, is a valuable tool for assessing the major drivers that impact diagnostic outcomes. The critical characteristics of tumour size in terms of radius, roughness, perimeter, and area are described to learn patterns to improve breast cancer diagnosis. To this end, using statistical techniques and machine learning methods, as well as visualization tools, will help to establish relationships between the identified tumour attributes and their respective diagnoses. This paper provides a detailed discussion of the analysis done on the cancer dataset alongside graphical illustrations. The research will show the theoretical value of the proposed approach in enhancing diagnostic accuracy for patients with breast cancer.

II. METHOD

The cancer dataset used for this project is obtained from the University of California, Irvine (UCI) Machine Learning Repository, tagged the Breast Cancer Wisconsin (Diagnostic) dataset [15]. This dataset is publicly accessible online and contains additional attributes about breast tumors, including both benign and malignant classifications (Table 1). The raw data was collected through a series of measurements taken from the patients and reviewed by pathologists. Before the data was used for the research, the datasets were cleaned of any missing or inconsistent entries through techniques like imputation depending on the severity of inconsistency. Basic statistical analyses, such as mean, median, and standard deviation, were conducted to assess the distribution and identify anomalies.

TABLE 1 A SNIPPET OF THE BREAST CANCER DATASET

	A	B	C	D	E	F	G	H	I	J
1	id	diagnosis	radius_me	texture_m	perimeter_	area_mea	smoothne	compactn	concavity_	concave p
2	842302	M	17.99	10.38	122.8	1001	0.1184	0.2776	0.3001	0.1471
3	842517	M	20.57	17.77	132.9	1326	0.08474	0.07864	0.0869	0.07017
4	84300903	M	19.69	21.25	130	1203	0.1096	0.1599	0.1974	0.1279
5	84348301	M	11.42	20.38	77.58	386.1	0.1425	0.2839	0.2414	0.1052
6	84358402	M	20.29	14.34	135.1	1297	0.1003	0.1328	0.198	0.1043
7	843786	M	12.45	15.7	82.57	477.1	0.1278	0.17	0.1578	0.08089
8	844359	M	18.25	19.98	119.6	1040	0.09463	0.109	0.1127	0.074
9	84458202	M	13.71	20.83	90.2	577.9	0.1189	0.1645	0.09366	0.05985
10	844981	M	13	21.82	87.5	519.8	0.1273	0.1932	0.1859	0.09353
11	84501001	M	12.46	24.04	83.97	475.9	0.1186	0.2396	0.2273	0.08543
12	845636	M	16.02	23.24	102.7	797.8	0.08206	0.06669	0.03299	0.03323
13	84610002	M	15.78	17.89	103.6	781	0.0971	0.1292	0.09954	0.06606
14	846226	M	19.17	24.8	132.4	1123	0.0974	0.2458	0.2065	0.1118
15	846381	M	15.85	23.95	103.7	782.7	0.08401	0.1002	0.09938	0.05364
16	84667401	M	13.73	22.61	93.6	578.3	0.1131	0.2293	0.2128	0.08025
17	84799002	M	14.54	27.54	96.73	658.8	0.1139	0.1595	0.1639	0.07364
18	848406	M	14.68	20.13	94.74	684.5	0.09867	0.072	0.07395	0.05259
19	84862001	M	16.13	20.68	108.1	798.8	0.117	0.2022	0.1722	0.1028
20	849014	M	19.81	22.15	130	1260	0.09831	0.1027	0.1479	0.09498
21	8510426	B	13.54	14.36	87.46	566.3	0.09779	0.08129	0.06664	0.04781
22	8510653	B	13.08	15.71	85.63	520	0.1075	0.127	0.04568	0.0311
23	8510824	B	9.504	12.44	60.34	273.9	0.1024	0.06492	0.02956	0.02076
24	8511133	M	15.34	14.26	102.5	704.4	0.1073	0.2135	0.2077	0.09756
25	851509	M	21.16	23.04	137.2	1404	0.09428	0.1022	0.1097	0.08632
26	852552	M	16.65	21.38	110	904.6	0.1121	0.1457	0.1525	0.0917
27	852631	M	17.14	16.4	116	912.7	0.1186	0.2276	0.2229	0.1401
28	852763	M	14.58	21.53	97.41	644.8	0.1054	0.1868	0.1425	0.08783

Cancer_Data

To analyze the cancer dataset, a two-pronged methodology was adopted: first, using R programming tools for static graphics generation, and second, using Microsoft Power BI for dynamic and interactive graphics generation. While static graphics, such as histograms, box plots, and heat maps, aided internal data exploration and model interpretation, the dynamic graphics were more user-focused. Through the use of interactive dashboards via Microsoft Power BI, data analysis became not only more accessible but also more helpful in clinical decision-making. The generated images will allow clinicians to directly interpret tumour characteristics, recognize risk patterns, and explore correlation features. Prompt updating, filtering, and graphical risk stratification will enable users to act upon insights in real-time. This work, therefore, combines static and dynamic visualizations for research purposes as well as real-time clinical use, supporting the need to integrate R programming tools in Microsoft Power BI.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

The data import process was different for the methodologies. The data was imported into R for analysis using the readxl package, which facilitates the reading of Excel files. The read-excel command was used to load the dataset into R, allowing for further manipulation and analysis of the tumor features. The import into the Power BI tool was facilitated by the Get-Data function, which allowed for the selection of file types, including Excel. Meanwhile, the Power Query editor assisted with the selection, transformation, and loading of data for analysis.

After the data import, the main workflow in data preprocessing included ascertaining data accuracy and validation to maintain data quality. The diagnosis variable was converted into a binary format, where malignant tumors (M) were assigned a value of 1 and benign tumors (B) were assigned a value of 0. Then, statistical methods and machine learning algorithms were deployed to identify patterns within the dataset. The key steps included exploratory data analysis (EDA), which analyzed distributions, correlations, and relationships between features and the diagnosis outcome. Additionally, predictive models were further deployed to distinguish between malignant and benign tumors based on tumor characteristics, utilizing algorithms such as logistic regression and decision trees. The model performance was assessed with metrics such as accuracy, precision, recall, and the F1-score, ensuring that the selected model provides reliable predictions. The F1 score is a well-known metric for classification models, and in machine learning, it could be used to treat imbalanced datasets. It generates a single score (0 to 1) representing the harmonic mean of precision and recall. Here, recall refers to the number of relevant items selected, while precision describes the proportion of selected items that are relevant [3,5]. This metric is a robust measure for a medical analysis undertaken in this research, whereby a score above 0.85 is considered excellent, 0.75-0.85 is good, 0.6-0.75 is fair, and below 0.6 is poor [6, 8, 13].

III. RESULTS

This project addresses the challenge of improving breast cancer diagnosis using data analytics. R programming tools and Power BI dashboards are utilized to analyze cancer datasets, supporting clinical decision-making. The goal is to reduce misdiagnosis and improve patient care through visual insights. Exploratory data analysis was conducted to understand the nature and characteristics of the collected data. Descriptive statistics provide a quick profile of different attributes. The result indicated that the mean radius of the tumours was approximately 14.6, which aims to measure the typical size of the tumours in the sample dataset. Besides, the area and perimeter mean values reproduced basic measures related to tumour size, which is essential for determining tumour features. The results are presented in two parts: firstly, from the R programming tools, and secondly, from the Power BI dashboards. R programming tools generated static graphics, such as histograms, box plots, and heat maps, along with the models that helped explore internal data and interpret the results. In contrast, Power BI generated dynamic graphics that aided real-time use in clinical settings.

Examining the R-tools-generated static graphics first, Fig. 2 shows the histogram of radius-mean, depicting the distribution of tumour sizes based on diagnosis. Concerning the malignant tumours (1), the value of mean radii exceeded the value of mean radii of benign tumours (0), which can be a sign of a relationship between the size of a tumour and its malignancy [4, 7]. Additionally, Fig. 3 shows the box plot of area-mean, illustrating the differences in area measurement between malignant and benign tumors, which demonstrates that malignant tumors generally have larger areas [4, 7]. This tallies with the observations in the histogram diagram.

Fig. 3 shows a correlation matrix with a heatmap that suggests associations between the various features. For instance, the circle with the label 'radiusmean' was seen as highly correlated with the circle 'perimeter-mean,' and similarly, 'radius-mean' was highly correlated with the circle 'area-worst' as denoted by the darker patches in the heatmap. On the other hand, some features, such as id, did not show any strong association with other features, with various connections highlighted by the outer, less intense regions surrounding them. It is through these relationships that we can understand which of these features could make a significant and considerable contribution to the possible determination of tumour malignancy [9, 10].

A logistic regression model was constructed to predict tumour diagnosis based on various features (Fig. 4). The model's additional coefficients indicated the contribution of each feature to the likelihood of tumour malignancy. In particular, the coefficients of Smoothness Mean and Concave Points Mean were the largest, suggesting that these two variables were the most important predictors. On the other hand, some features, for example, id and area-mean, had coefficients close to zero, which may be useless in classification; hence, this proves that feature selection is crucial when developing a good predicted bottom line [9, 11].

Furthermore, the decision tree analysis variables pointed out that concave points-mean has significant values and is a significant predictor with threshold values of 0.048. In observations with points-mean $\geq$ 0.048 at concave point, the predicted result of Node 3 was high (1) for the target class, which most probably signified a malignant tumour, a norm expected in cancer datasets [12].

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

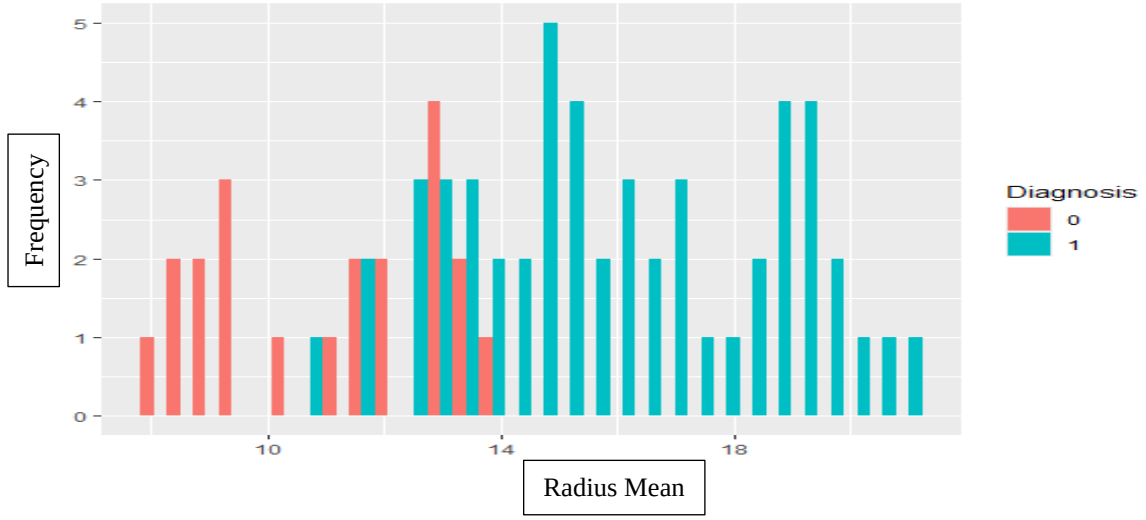

Fig. 2 Histogram of Radius Mean

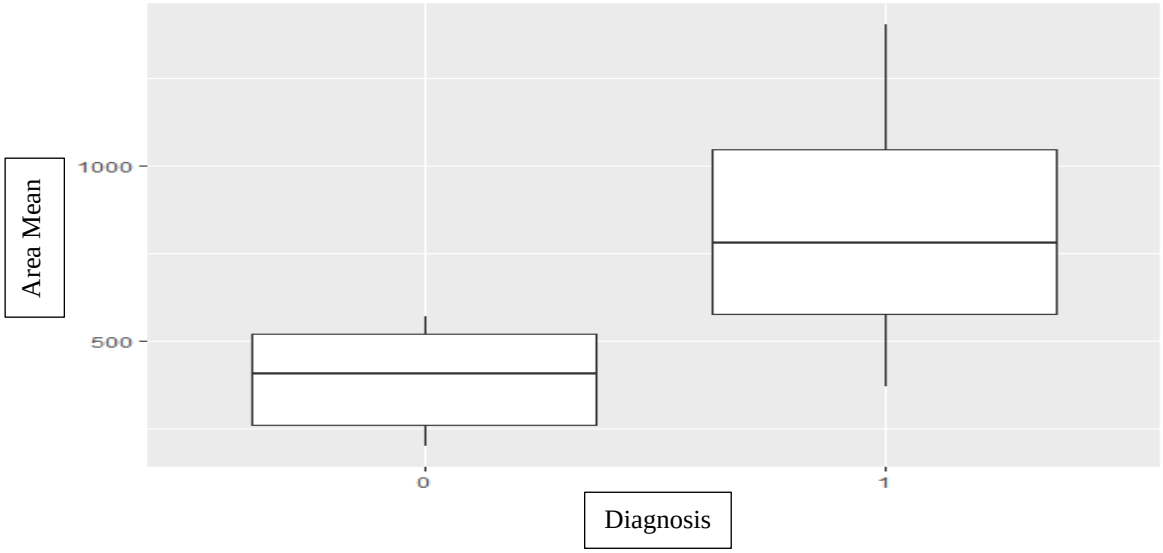

Fig 3. Boxplot of Area Mean by Diagnosis

The models were evaluated as well. In this case, several metrics were used to evaluate the performance of the estimated logistic regression model. The model yielded an accuracy of 80.95%, which means that the model could correctly classify a tumour 81 times out of 100. However, 60% of precision determined the attempts that were made for predicting malignant tumours as true positive ratios to the total extent of positive predictions, and it highlighted the fact that although the model can accurately diagnose some malignant tumours, the benign tumours could be misdiagnosed as malignant.

On the other hand, a recall of 100 percent showed how the model can capture all the malignant tumours in the dataset without missing any, thus showing how it performs among the true positives.The F1 score of 75% gave a fair balance between precision and recall, meaning that the performance in classifying malignant tumours was reasonable. Finally, the AUC (area under the curve) of 91.11% may be accepted as demonstrating the high ability of the model in the separation rate of malignant and benign tumours, where the value for this indicator is very close to 1, which will demonstrate high discriminatory power [12, 13]. Although the proposed model achieves superior performance measured by recall score and AUC, the low precision suggests there is still potential for model improvement towards reducing false positive cases. Some of these include feature engineering, hyperparameter tuning, and model validation on a different test dataset to further improve the final model performance and guarantee its usage in real clinical applications [2].

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

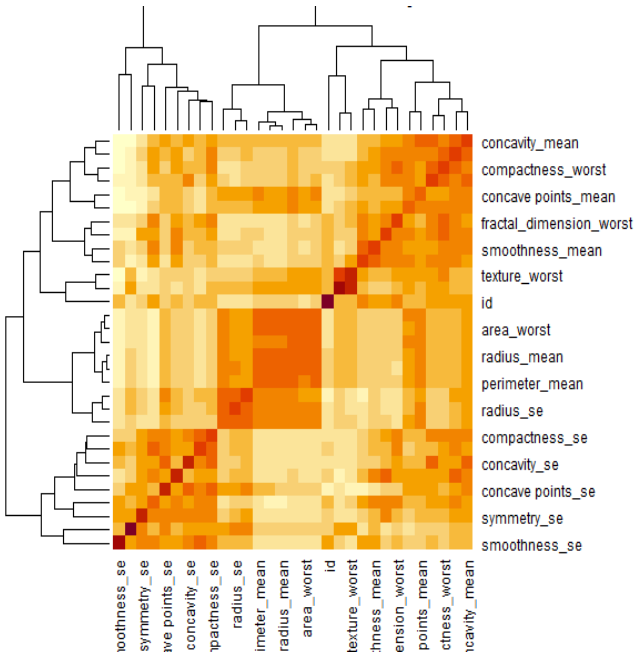

Fig. 4 Correlation Heatmap

The second part of the results focuses on the Power BI usage. As can be gleaned from Figure 6, the Power BI shows a clear visual representation of the tumour type distribution of 37.26% malignant versus 62.74 benign cases. The key insight is a clinical value of an immediate understanding of case severity distribution which could inform resource allocations and staffing requirements for the oncology departments. The risk category distribution shows low risk category of predominantly benign cases (~ 600 total count) while the high-risk category has a mixed distribution with notable malignant presence (~ 50 cases). The clinical workflow will enable prioritized patient triage and care protocols. The total cases by diagnosis (bar chart) reveals with visual clarity a direct comparison between benign (~ 350 cases) and malignant (~ 200 cases). As operational insight and quality assurance, this workload distribution validates dataset balance for reliable model training [10, 12]. Furthermore, the sum of radius-worst by diagnosis (Pie Chart) reveals a distribution of 48.4% versus 51.6% split showing balanced representation. As an analytical value, this provides insights into worst-case tumour characteristics that help identify cases requiring urgent intervention.

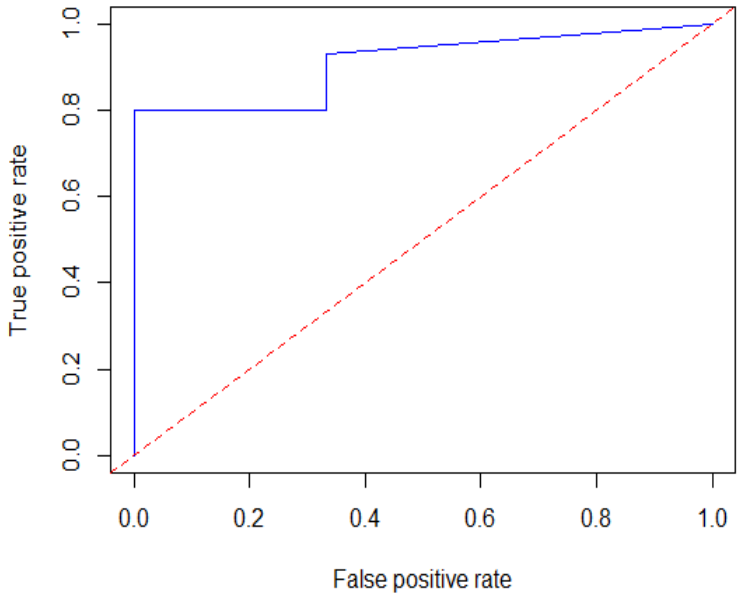

Fig. 5 Performance of the Logistic Regression Model
(Receiver Operating Characteristic (ROC) Curve)

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

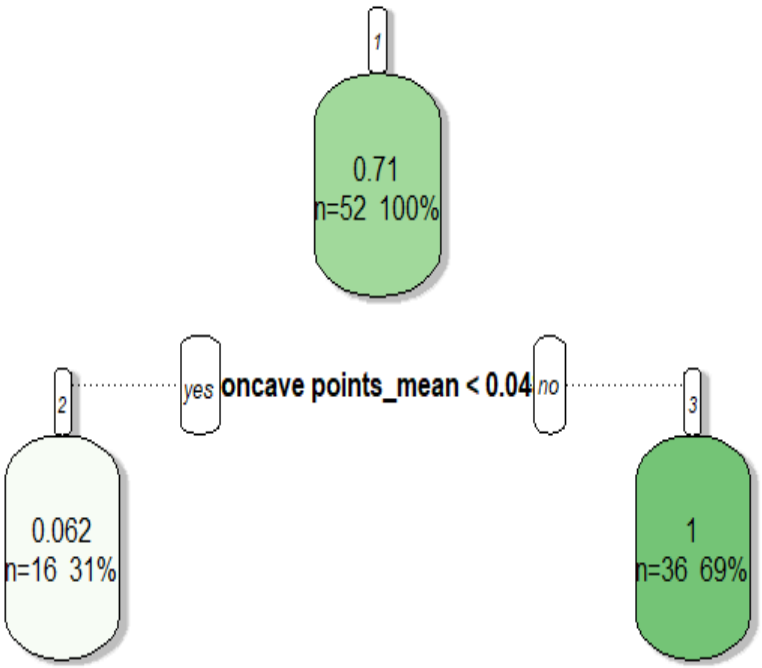

Fig. 6 Decision Tree Analysis

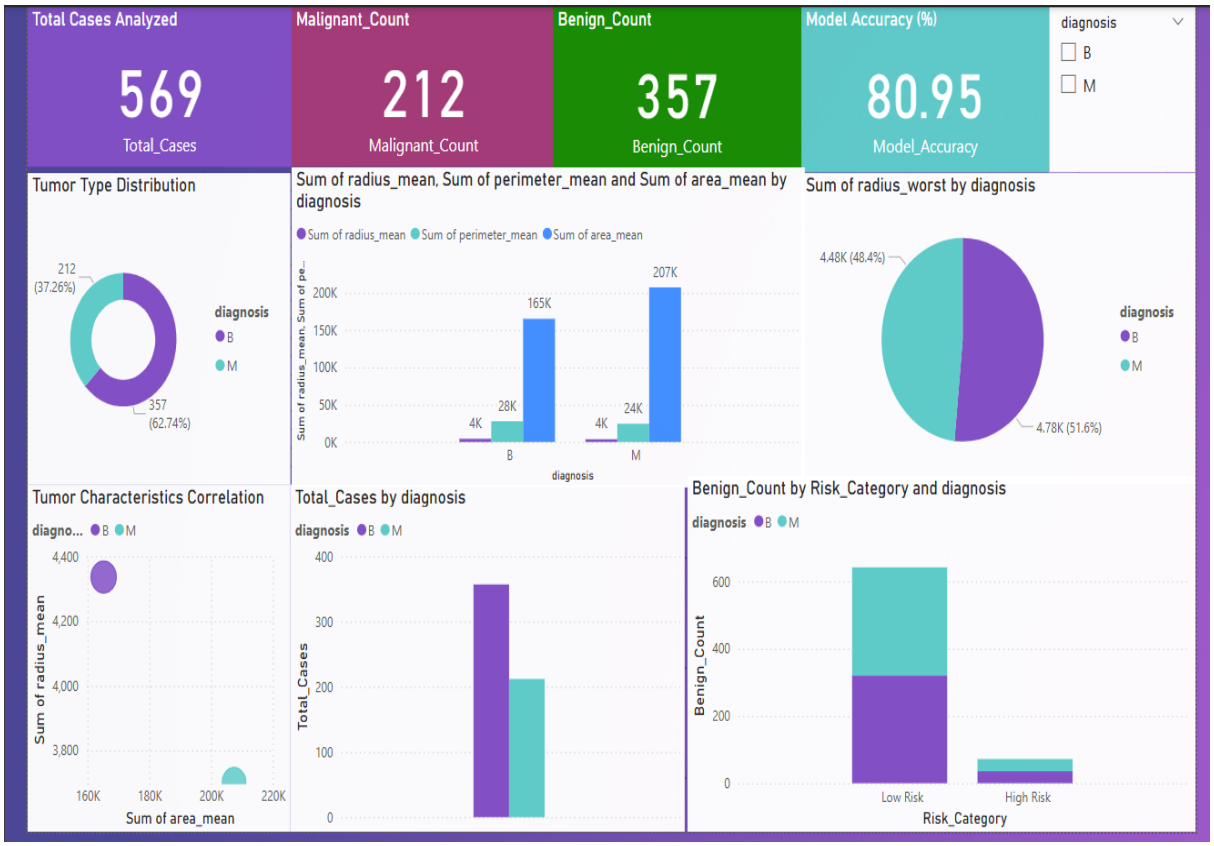

Fig. 7 Diagnosis Multi-Panel View - Power BI Visualizations Analysis

There is a strong positive correlation (r = 0.3) between radius and area measurements, providing a clear visualization of interdependent tumor characteristics that support feature selection and engineering for improved predictions. Strong correlations (r = 0.3M) between dimensional characteristics confirm the validity of the integrated approach [4]. Furthermore, the tumour characteristics analysis (multi-panel view) for the sum of radius, perimeter, and area by diagnosis shows that for malignant tumours significantly higher values across all dimensional metrics, viz.: radius sum is ~207K for malignant versus ~4K for benign; perimeter sum and area sum show a similar dramatic difference pattern.

e-ISSN 3108-026X
International Multidisciplinary Research Journal Reviews (IMRJR)
A Peer-reviewed journal
Volume 3, Issue 1, January 2026 | DOI 10.17148/IMRJR.2026.030107

Clinical significance confirms a correlation between size and malignancy, supporting informed clinical decision-making [6]. The organizational impact is noticeable as visual analytics reduce misdiagnosis rates through clear pattern recognition. Early accurate detection and risk stratification enable optimal staff allocation and equipment usage, while the interactive dashboards encourage continuous monitoring and data-driven improvement and excellence. The strategic value includes immediate access to critical diagnostic metrics for 212 cases, clear visualization of malignant (37.26%) vs. benign (62.74%) distribution, integrated risk assessment supporting clinical workflow, and a foundation for predictive modelling and advanced analytics. Moving forward, continuous model refinement and expanded feature analysis will ensure this solution remains at the forefront of cancer diagnostic technology, ultimately contributing to better patient outcomes and more efficient healthcare delivery across the organization.

IV. DISCUSSION

The findings from the logistic regression model and decision tree analysis provide a promising starting point for predicting tumour malignancy levels based on the features. Subsequently, the accuracy (80.95%) and specificity score (100%) of the model proves its efficiency in identifying all malignant tumors. Since missing a malignant appearance can be catastrophic, especially in clinical practice, this feature plays a central role in the detection process. However, given that the precision achieved is only 60%, concerns may arise about the model's effectiveness in reducing the number of false positives. This trade-off between recalling all the data and not missing any, while minimizing the chances of wrongly including data, is a common issue in medical diagnosis, where sensitivity is tested against specificity [10].

The decision tree analysis builds on this by adding another critical facet to the tumour classification – concave points-mean. The threshold established (0.048) may be helpful in clinical management, as it can assist healthcare professionals in determining the parameters of tumors that may require additional evaluation. However, the result may reduce the usefulness of this model as tumour classification usually employs multiple features. Thus, it is crucial to identify the relationships between different features to improve the model's stability [9].

Further, as the findings suggest improvement, specific issues require attention. First, the overfitting problem, concerning the model training process, must be solved. Such a risk can be mitigated by employing cross-validation techniques, as well as regularization, to prevent overfitting of the model to the data [11]. In addition, another analysis revealed the potential for further enhancing the models through feature selection and engineering, which can increase the importance of the most relevant features while neglecting the least relevant ones.

The final methodological aspect is the possibility of using the model for forecasting real-life conditions. It is worth noting that, though diverse for this analysis, the dataset may not capture all the tumours that may be found in cancer practice. Hence, the model must be validated on an independent dataset containing more diverse tumor varieties and properties. This step will enhance confidence in the model's overall accuracy and ensure that it remains versatile across various clinical settings. Collectively, the findings of this study support the notion that machine learning models can aid in cancer diagnosis. Studies on logistic regression and decision-tree methods for classifying tumor types are helpful and should be continued with further improvement and examination for clinical purposes [10]. Future research should address these limitations, analyze other modeling methods, and incorporate clinical knowledge to enhance the accuracy and validity of the models.

This data analytics and visualization project represents a significant advancement in breast cancer diagnostic support. The Power BI dashboard analysis reveals critical insights that directly support clinical decision-making [7]. Integration of R-programming tools with Microsoft Power BI will facilitate the transition from research visualizations to clinical applications. For instance, R-based predictive modeling can be integrated with Power BI using data flows and Power Query, while Power BI Streaming Datasets or DirectQuery can facilitate live integration with hospital databases. Microsoft Power BI offers dashboard automation, real-time integration, forecasting, diagnostics, alerts, and predictive modeling capabilities

V. CONCLUSION

This project applied data analytics to the field of oncology. The analysis of breast cancer prognosis using machine learning has unveiled some fundamental observations as follows. Features such as radius-mean, perimeter-mean, and area-worst exhibit a very high correlation coefficient, indicating their overall contribution to accurate modeling. Due to the multiple features, special attention must be paid to feature selection to prevent the problem of overfitting. When using the decision tree model, the conclusion was that the concave points-mean was the highly relevant predictor, with a threshold of 0.048 to distinguish between malignant and benign tumors. This model's predictive accuracy was high; therefore, it will be necessary to use it in clinical practice. The logistic regression model achieved an accuracy of 80.95%

and an F1 score of 0.75, indicating good overall classification performance, but a precision score of 0.60, which demonstrates a moderate capability to minimize false predictions. Thus, the work proves the possibility of using machine learning algorithms to improve the diagnosis of breast cancer. Tackling issues such as precision enhancement and feature engineering for deployed models will be critical for future enhancements in real-life usage. Thus, the possible improvements for future research include the usage of various datasets and complex methodologies to enhance the consistency of predictive results. The Microsoft Power BI-driven research aspect provides a powerful, interactive, and real-time decision-support tool for breast cancer diagnosis. By visualizing complex tumor metrics and model predictions, clinicians are empowered to make faster, data-driven decisions, reduce the risk of misdiagnosis, achieve operational efficiency gains, and benefit from a scalable platform for future predictive analytics.

In conclusion, the presented analysis demonstrates that machine learning models, including logistic regression and decision trees, can be effectively applied for predicting tumour malignancy with the aid of key features, such as concave points and mean. The models yielded high accuracy and recall; thus, they can be used as reliable clinical decision support systems for reducing the number of false negatives in cases of malignant tumors. Although they show good performance on the chosen dataset, issues such as sacrificing precision values and the possibility of overtraining require cross-checking with different datasets. With these limitations addressed and clinical knowledge integrated into these models, the potential for cancer diagnosis and patient benefits could be improved.

ACKNOWLEDGMENT

The authors express special thanks to the departmental colleagues (Texas A&M University, Johns Hopkins University, and the University of Kentucky) for solid support and encouragement during the research process. The departmental resources facilitated the project's completion.

REFERENCES

[1]. Saint John's Cancer Institute. Types of Cancer. Retrieved from https://www.saintjohnscancer.org/breast/breast-cancer/types-of-breast-cancer/, 2025.
[2]. S. Douville, et al., "Data Visualization Expert Lessons Learned: Implications for Program Evaluators", *American Journal of Evaluation*, vol. 46, no. 3, https://doi.org/10.1177/10982140241290744, 2025.
[3]. H. A. Ganatra, "Machine Learning in Pediatric Healthcare: Current Trends, Challenges, and Future Directions", *Journal of Clinical Medicine*, vol. 14, no. 3, 807, https://doi.org/10.3390/jcm14030807, 2025.
[4]. J. B. Oehler, W. Rajapaksha and H. Albrecht, "Emerging Applications of Nanoparticles in the Diagnosis and Treatment of Breast Cancer", *Journal of personalized medicine*, vol. 14, no. 7, 723, https://doi.org/10.3390/jpm14070723, 2024.
[5]. J. G. C. Ramírez, M. M. Islam and A. I. H. Even, "Machine learning applications in healthcare: Current trends and future prospects", *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, vol. 1, no. 1, 2024.
[6]. H. Tanveer et al., "AI-Powered Diagnosis: A Machine Learning Approach to Early Detection of Breast Cancer" *International Journal of Engineering Development and Research*, vol. 13, no. 2, pp. 153-166. 2025.
[7]. H. Wang and & C. Zhang, "Visualization Analysis of Literature in the Design Issues", Journal from the Perspective of Sustainable Development. In *International Conference on Human-Computer Interaction*, pp. 103-114, 2025.
[8]. M. A. Naji et al., "Machine Learning Algorithms for Breast Cancer Prediction and Diagnosis", *Procedia Computer Science*, vol. *191*, 487-492, 2021.
[9]. G. Alfian et al., "Predicting Breast Cancer from Risk Factors using SVM and Extra-trees-based Feature Selection Method", *Computers*, *11*(9), 136. 2022.
[10]. R. J. Ellis, M. Sander and A. Limon, "Twelve Key Challenges in Medical Machine Learning and Solutions" Solutions", Intelligence-Based Medicine, Vol.6, 100068, 2022.
[11]. I. Tougui, A. Jilbab and J. El Mhamdi, "Impact of the Choice of Cross-Validation Techniques on the Results of Machine Learning-based Diagnostic Applications", Healthcare informatics research, vol. 27, no. 3, pp. 189-199, 2021.
[12]. R. Hazra, M. Banerjee and L. Badia, "Machine Learning for Breast Cancer Classification with ANN and Decision Tree. In 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 0522-0527, 2020.
[13]. R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and f1 Score for more thorough Evaluation of Classification Models. In Proceedings of the first workshop on evaluation and comparison of NLP systems, pp. 79-91, 2020.
[14]. A. Q. Khan, M. Touseeq, S. Rehman, M. Tahir, M. Ashfaq, E. Jaffar and S. F. Abbasi, Advances in breast cancer diagnosis: a comprehensive review of imaging, biosensors, and emerging wearable technologies. Frontiers in oncology, 15, 1587517. https://doi.org/10.3389/fonc.2025.1587517, 2025.
[15]. M. H. Yasser and W. H. Wolberg, Breast Cancer Dataset. Retrieved from

https://www.kaggle.com/datasets/nancyalaswad90/breast-cancer-dataset). 2025.
[16]. S. C. Clasen and M. E. Miller, "How We Monitor Cardiac Health in Breast Cancer Survivors", *Journal of Clinical Oncology*, vol. 43, no. 1, pp. 4-9, 2025.
[17]. National Cancer Institute. Cancer Stat Facts: Female Breast Cancer Subtypes. https://seer.cancer.gov/statfacts/html/breast-subtypes.html, 2025.

BIOGRAPHY

Dr. Emmanuel Udoh is a professor of Information technology at the Texas A&M University Central, Texas, USA. He holds two doctoral degrees, one in Information Technology from Capella University, USA and one in Geology from the University of Erlangen-Nuremberg, Germany.

Dr. Tachi Udoh is a resident medical doctor at the Johns Hopkins Osler Internal Medicine Hospital, Johns Hopkins University, Baltimore, USA. Her area of research is related to cancers and environmental exposures to toxins/heavy metals.

Liz Udoh is a double major student in the Neuroscience/Psychology department at the University of Kentucky, Lexington, USA.