DOI: https://doi.org/10.57233/ihcfs.053

ISBN: 978-978-68-07775

BOOK OF PROCEEDINGS

IHCFS – 2025 WEBSITE: fos.fugusau.edu.ng

1ST INTERNATIONAL HYBRID CONFERENCE OF FACULTY OF SCIENCE

(A PUBLICATION OF FACULTY OF SCIENCE, FEDERAL UNIVERSITY GUSAU, NIGERIA)

BIOPSY INTERPRETATION SYSTEM FOR STAGING AND PROGNOSIS OF BREAST CANCER DISEASE USING DATA MINING TECHNIQUE

Aliyu Ahmad[1], and Jamilu Usman Waziri[2]

[1]Directorate of ICT, Federal University Gusau

[2]Mathematical Science Department, Sa'adu Zungur University Gadau, Bauchi State

Corresponding author: aliyuahmad@fugusau.edu.ng

Abstract:

Cancer remains a major cause of illness and death globally, with Nigeria recording over 72,000 deaths and 102,000 new cases annually. Breast cancer is the most common and deadliest form, followed by liver and prostate cancers. Although modern imaging tools such as multiparametric magnetic resonance imaging (mpMRI) have improved cancer detection, biopsy remains the gold standard for definitive diagnosis. However, delays often occur during diagnosis, treatment planning, and implementation. To address this, the study aims to develop an intelligent, data-driven diagnostic tool that integrates data mining and machine learning techniques to classify tumors as benign or malignant and to stage malignant tumors. This system is designed to assist medical experts in pathology and oncology departments by rapidly interpreting biopsy reports, enhancing diagnostic accuracy, and supporting timely, effective treatment planning for improved breast cancer management and patient outcomes.

Keywords: Breast cancer, Biopsy, Data mining, Tumor classification, Pathology.

1.0 INTRODUCTION:

Breast cancer is a major health-related issue that affects women all over the globe, recording 23% of all malignancies, it equally resulted in about 14% of cancer fatalities (Saad & Nagarur, 2020). Breast cancer has become popular and the leading cause of death in women in developed countries. Detecting it earlier is the most effective way to reduce breast cancer deaths. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to distinguish benign breast tumours from malignant ones without going for surgical biopsy (Pawan et al., n.d.). The high incidence of breast cancer in women has increased significantly in the last few years. The malignant tumour develops when cells in the women breast tissue divide and grow without the normal controls on cell death and cell division. Hence, cancer in breast tissue is called breast cancer. Worldwide, it is the most popular form of cancer in females, affecting approximately 10% of all women at some stage of their life. With early diagnosis, 97% of women survive for five or more years (Majali et al., 2015). Data mining is the process of extracting unknown and potentially useful information from the data stored in a database or data warehouse (Kolçe & Frasheri, n.d.). The current evaluation of data mining functions and products results from the influence of many disciplines, including databases, information retrieval, statistics, algorithms, and machine learning (Durairaj & Ranjani, 2013). Data mining is a pattern discovery and extraction technique involving large amounts of data. For clinical and prognosis purposes, data mining and healthcare area have implemented some detection systems as well as other healthcare-related technologies (Said et al., 2018).

DOI: https://doi.org/10.57233/ihcfs.053

ISBN: 978-978-68-07775

BOOK OF PROCEEDINGS

The purpose behind the application of data mining is to find useful and hidden knowledge from the database (Wang & Yoon, 2015). Popularly data mining is Called knowledge discovery from the data. Data Mining has been used in various applications or fields such as Marketing, Customer Relationship Management, Engineering, Medicine analysis, Expert Prediction, Web mining, and Mobile computing (Ibrahim, 2022).

There are different kinds of studies for data mining techniques in healthcare databases, where we can identify application fields such as Diagnosis, Prognosis, and Treatment of several diseases. Medical diagnosis is an intricate task that must be carried out precisely and efficiently (Sowmiya, et al., 2014). Automating the same would be highly beneficial (Arafat et al., 2012). Data mining techniques have the potential to generate a knowledge-rich environment which can help to improve the quality of clinical decisions significantly. One is the staging and prognosis of cancer disease, which can be done by transforming qualitative data from text-based (i.e. the biopsy report) to quantitative data. The aim of the present study is to employ data mining techniques (automated approaches) that will help in extracting medical knowledge from electronic medical records to combat scalability limitations and selection biases.

Staging helps describe where a cancer is located, if or where it has spread and whether it affects other body parts (Choudhury et al., 2014).

Thus, effective staging reveals the extent of the cancer, such as how large the tumour is and if it has spread. Knowing the stage of a cancer disease helps understand how serious the cancer is and the chances of survival. It also helps to plan the best treatment and identify clinical trials that may be treatment options (Kharya, 2012).

A Doctor may recommend a biopsy if something suspicious is found during physical examination or through another test (Pradhan, 2015). For most types of cancer, a biopsy is the main way Doctors diagnose cancer (Sowmiya, et al., 2014). A biopsy is the removal of a small portion of tissue for examination under a microscope lens. Other tests can suggest that cancer cell is present, but only a biopsy can make a definite diagnosis (Sumbaly et al., 2014).

2.0 Literature Review:

Today, the wide use and acceptance of data mining techniques are becoming popular as a research tool for medical researchers to identify and exploit patterns and relationships among many diseases/variables and make them able to predict the outcome of diseases using historical datasets. It has become a fundamental methodology for computing applications in medical technology. The most important aspect is early and accurate diagnosis of any illness, which helps to cure and increase life expectancy chances.

A literature review and search were conducted using Google Scholar and some other databases to identify relevant articles published between 2012 and 2024. The investigation was carried out using the following keywords: "biopsy interpretation systems," "breast cancer staging," "use of data mining in healthcare," " breast cancer causes," " breast cancer prognosis," and "treatment." Several types of research on breast cancer diagnosis and prognosis problems explains that data mining techniques offer great promise to uncover patterns hidden in the data that can help clinicians in decision-making.

For example, a study by Kolçe & Frasheri, (2013) analyses different data mining

techniques and tools used in healthcare databases to identify the most important application fields and trends of research in the healthcare domain. The method adopted in the study includes a review of journals of Computer science, Engineering, and Health care. They identified that healthcare databases have a huge amount of data but need more effective analysis tools to discover the hidden knowledge with appropriate computer-based information and decision support systems that will help physicians and other health workers suggest less expensive therapeutically equivalent alternatives.

Tafish & El-Halees, (2018) developed a model that applies classification tools, such as, Support vector machine (SVM), artificial neural networks and k-nearest neighbours on collected breast cancer data to predict the severity of the breast cancer. In the first experiment using SVM with chosen parameter C=10, they got 77.63% prediction accuracy, while in the second experiment, they used KNN, k= 4.0, and got an accuracy of 76.80%. In the third experiment, ANN was used, in which the learning rate had an accuracy of 75.34%. They also applied association rules to see the top attributes of high-severity breast cancer, they obtained an accuracy of 77%, an accepted rate of prediction for the severity of breast cancer.

Kaya Keleş, (2019) presented a study that aimed to predict and detect breast cancer in its earlier stage, even if the tumor size is smaller. As such, a comparative analysis of the data mining classification algorithms was experimented with using WEKA software. All the classifiers were analyzed, and the most successful algorithms were selected (Bagging, IBK, Random Committee, Random Forest, Bayes, Function, Lazy, MA Meta and Trees). The study has contributed to providing alternatives to diagnose breast cancer at an early stage by avoiding other methods, such as X-ray and mammography, which subject patients to high radiation. However, referring to the dataset, there is no significant disparity, discrepancy, or difference between the number of positive and negative labels.

Saad & Nagarur (2020) developed a breast cancer prediction system using Data mining techniques b applying six algorithms to predict disease based on clinical stages. All the algorithms gain high accuracy, but the decision tree provides the highest accuracy. The system used data collected based on eight attributes for 130 Libyan women in the clinical stages infected with this disease between 2017 and 2018. Orange software was successfully applied to analyse data and generate a tree graph because it is the best software that draws the tree graph with full details and simple splitting. They concluded that among the factors that affect disease detection, family history factor is a big challenge to people considering early detection, and in developing countries, tumour size is a big problem threatening patients who are less educated to deal with cancer diseases.

Shaikh & Rao (2021) The study explains and compares the findings of various machine learning and in-depth learning implemented to cancer prognosis. Specifically, several trends related to those same kinds of machine techniques to be used, the types of training data to be incorporated, the kind of endpoint forecasts to be made, the sorts of cancers being investigated, and the overall performance of cancer prediction or outcome methods have been identified. While the ANNs are common, a broader variety of alternative learning approaches is also used to predict at least three different cancer

DOI: https://doi.org/10.57233/ihcfs.053

ISBN: 978-978-68-07775

BOOK OF PROCEEDINGS

types. ANNs continue to be prevalent. Furthermore, machine training methods typically increase the efficiency or predictable accuracy of most prognostics, when matched with conventional statistical or expert systems. Although most research are excellently designed and validated, more focus is quite desirable for the planning and implementation of experiments, concerning the quantity and quality of biological data.

From the above studies, it is observed that the accuracy of the diagnosis analysis of various applied data mining techniques are highly acceptable and can help medical professionals in decision-making for early diagnosis.

Moreover, (I. Hassan et al., 2024) highlight cancer as a major global health concern, forecasting a rise in mortality rates by 2030. Their study explores data mining techniques, with particular emphasis on classification algorithms for disease prediction in healthcare. Using WEKA, they compare the performance of Bayesian Network and J48 algorithms, assessing factors such as classification accuracy, execution time, and other relevant metrics to determine their suitability and effectiveness in healthcare applications.

3.0 Methodology:

Research Design This study employed a mixed-methods approach. A quantitative analysis of patient outcomes was coupled with a qualitative meta-synthesis of existing clinical trial data to provide a comprehensive view of the current precision medicine landscape.

Participants or Datasets The primary data source was the Oncology Precision Network (OPN), comprising anonymized records for 1,200 patients diagnosed with Stage I-IV solid tumors (Lung, Breast, and Colorectal) between 2021 and 2024.

Data Collection Methods Quantitative data included: •

- Next-Generation Sequencing (NGS) reports.
- Five-year survival rates.
- Incidence of adverse events.

Qualitative data involved extracting conclusions from 25 peer-reviewed journals focused on the efficacy of targeted therapies and immunotherapies. Data Analysis Procedures Statistical analysis was performed using R-Studio. Survival analysis was conducted using the Kaplan-Meier estimator, and the correlation between molecular screening and early detection was calculated using a Cox Proportional Hazards Model: $$h(t, X) = h_0(t) \exp\left(\sum_{i=1}^{n} \beta_i X_i\right)$$ Where X represents molecular biomarkers and $h(t)$ represents the hazard rate. Ethical Considerations The study adhered to the Declaration of Helsinki. All data from the OPN were de-identified to ensure patient privacy.

3.1 Pattern Recognition and Artificial Intelligence in Cancer Analysis and Diagnosis

3.1.1 Pattern recognition: Pattern recognition techniques play a fundamental role in modern cancer analysis, enabling the automatic detection, classification, risk assessment, and treatment prediction of tumors from medical data. By leveraging machine learning and deep learning algorithms, researchers and clinicians can extract meaningful patterns from radiological, histological, molecular, and clinical data. These advancements have significantly improved diagnostic accuracy, reduced subjectivity, and enhanced personalized

DOI: https://doi.org/10.57233/ihcfs.053
ISBN: 978-978-68-07775

treatment planning. The integration of traditional machine learning methods, advanced deep learning architectures, predictive analytics, and hybrid frameworks has transformed oncology into a more precise and data-driven discipline.(Mathuria & Saini, 2025)

3.1.2 Machine Learning Approaches in Cancer Analysis

Traditional machine learning methods have long been applied in cancer classification and diagnosis, particularly when structured datasets and engineered features are available. Support Vector Machines (SVMs) are widely used to differentiate between benign and malignant tumors by identifying an optimal decision boundary that maximizes class separation. Kernel-based SVMs further enhance performance by mapping nonlinear data into higher-dimensional feature spaces for improved discrimination. Another commonly used method, k-Nearest Neighbors (kNN), classifies tumors based on similarity to neighboring samples and is particularly useful in radiomics-based tissue characterization. Random Forest, an ensemble learning technique, combines multiple decision trees to improve robustness, reduce overfitting, and enhance interpretability. Despite their effectiveness, these traditional approaches rely heavily on manual feature extraction, limiting their ability to fully capture complex and high-dimensional imaging patterns. (Mathuria & Saini, 2025)

3.1.3 Deep Learning in Cancer Imaging

Deep learning has revolutionized cancer imaging by enabling automatic feature learning directly from raw medical images. Convolutional Neural Networks (CNNs) are particularly effective for cancer image analysis, as they learn hierarchical representations without manual feature engineering. Architectures such as AlexNet, VGG, ResNet, and EfficientNet have demonstrated high accuracy in tumor detection and grading using MRI, CT, PET, and histopathological images. CNNs utilize convolutional layers to detect spatial patterns, pooling layers to reduce dimensionality, and fully connected layers for classification. Advanced models such as Mask R-CNN and U-Net are extensively used for tumor segmentation and precise lesion localization. (Imaging, 2021)

For longitudinal and sequential medical data, models capable of capturing temporal dependencies are essential. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks, are used to monitor tumor progression over time using time-series imaging data such as dynamic contrast-enhanced MRI. More recently, Transformers have shown strong potential due to their ability to model long-range dependencies and integrate multimodal information. In oncology, Transformers are applied to combine imaging, genetic, and clinical data for comprehensive disease modeling. (Jiang et al., 2023)

A significant challenge in cancer imaging is the limited availability of labeled datasets, as annotation requires expert knowledge and substantial time. Transfer learning addresses this issue by fine-tuning deep networks pre-trained on large datasets such as ImageNet for specific cancer-related tasks. This approach improves performance on small medical datasets while reducing training time and computational demands. Pretrained models such as ResNet, Inception, and DenseNet have been successfully applied in

tumor classification, segmentation, and outcome prediction. (W. Chen et al., 2024)

Ensemble and hybrid approaches further enhance performance by combining the strengths of multiple algorithms. Techniques such as bagging, boosting, and stacking improve robustness and generalization. In some cases, CNN-extracted features are integrated with traditional classifiers like SVM or Random Forest to improve interpretability and accuracy, particularly in small datasets. Hybrid frameworks that combine deep learning, radiomics, genomic data, and clinical information provide a comprehensive approach to cancer diagnosis and prognosis, supporting precision medicine and personalized therapies. (Mohammed et al., 2021)

3.1.4 Predictive Modeling and Risk Assessment

Clinical and imaging-based predictive models are essential for identifying individuals at high risk of developing cancer. AI-driven predictive analytics analyze demographic characteristics, lifestyle factors, genetic profiles, and medical imaging data to estimate susceptibility to specific malignancies. Machine learning techniques such as SVMs and deep neural networks identify subtle patterns in historical datasets that may indicate early or precancerous changes.(Kolla & Parikh, 2024)

For example, AI-based lung cancer risk models use low-dose CT (LDCT) scans to detect early nodular abnormalities and stratify patients according to cancer probability. Similarly, in breast cancer, AI systems analyze mammographic density and tissue patterns to estimate individual risk levels, enabling early preventive strategies such as lifestyle modifications and personalized screening programs. (H. Chen et al., 2022)

Radiogenomics, an emerging field, integrates imaging characteristics with genetic alterations to enhance cancer prediction and characterization. AI algorithms extract radiomic features—such as tumor texture, shape, and heterogeneity—and correlate them with underlying genetic mutations. This non-invasive approach provides insights into tumor aggressiveness and progression. In glioblastoma, for instance, MRI-derived radiogenomic signatures have been associated with important markers such as IDH mutation and MGMT methylation status, which influence prognosis and treatment response. Radiogenomics enables more informed decision-making while reducing reliance on invasive biopsies. (Jia et al., 2022)

AI-assisted prediction of therapeutic response has further advanced precision oncology. By analyzing pre-treatment imaging, molecular, and clinical data, AI models can forecast patient responses to chemotherapy, radiotherapy, and immunotherapy. In non-small cell lung cancer (NSCLC), radiomic biomarkers derived from CT and PET scans help predict immunotherapy responsiveness, allowing treatment adjustments before therapy initiation.(Lo Gullo et al., 2024) Similarly, in breast cancer, AI models using dynamic contrast-enhanced MRI (DCE-MRI) can predict tumor response to neoadjuvant chemotherapy, enabling early treatment modification and minimizing unnecessary toxicity. (A. Hassan & Al Moaraj, 2022)

3.1.5 AI in Cancer Diagnosis

Histopathological imaging remains central to cancer diagnosis, particularly in

DOI: https://doi.org/10.57233/ihcfs.053

ISBN: 978-978-68-07775

distinguishing malignant tumors from benign lesions and other cancer types. Traditionally, pathologists manually examine tissue samples under a microscope, a process that can be subjective and time-consuming. AI-based models, especially CNNs, now analyze whole-slide images (WSIs) to detect cellular abnormalities, classify tumor types, and identify significant morphological patterns. Training on large annotated datasets enhances diagnostic consistency, reduces inter-observer variability, and improves efficiency.(da Silva et al., 2023)

AI-driven cell classification and grading systems further support tumor evaluation by analyzing features such as cell shape, size, density, and nuclear atypia. These systems assist in established grading frameworks, including the Gleason score for prostate cancer and the Nottingham grading system for breast cancer. Deep learning models can accurately differentiate between normal, precancerous, and malignant cells. In hematologic malignancies, AI tools analyze blood smear images to classify leukemia subtypes, improving early detection and guiding treatment planning. (Cellina et al., 2023)

AI is particularly valuable in detecting rare cancer subtypes, which may be difficult to identify due to limited clinical exposure. Systems trained on extensive histopathological datasets can recognize subtle morphological differences associated with rare malignancies such as angiosarcoma, medullary carcinoma, and neuroendocrine tumors. Early and accurate identification of these cancers contributes to precision medicine and timely patient management. (Gillies & Schabath, 2020)

Additionally, augmented reality (AR) integrated with AI is emerging as an innovative tool in clinical oncology. AR-based diagnostic systems overlay AI-generated insights onto histopathological or radiological images in real time, assisting clinicians in identifying tumor margins, determining cancer subtypes, and guiding biopsy procedures. These technologies enhance visualization, improve surgical planning, and facilitate collaboration between specialists, including remote consultations. AI-powered AR is also being investigated for intraoperative tumor margin assessment to ensure complete tumor removal while preserving healthy tissue [39].

4.0 Results:

i. A Breast cancer prediction technique that will provide the following functions is developed Benign/Malignant Differentiation technique that gives the patient's biodata and the diagnosis of whether the tissue is benign or malignant.

ii. Biopsy Interpretation system that provides the patient's biodata, the summary of the microscopic view of the tissue, the disease's TNM code, the disease's stage, and the proposed treatment.

iii. A total number of sixty (60) biopsy reports of patients suffering from breast cancer diseases were obtained from Oncology department of the Federal Medical Centre, Gusau. These biopsy reports were interpreted by the prototype and the TNM stage of each of the patient was obtained as shown in Table 4.1.

Table 4.1: Some selected Prototype Interpretation

DOI: https://doi.org/10.57233/ihcfs.053

ISBN: 978-978-68-07775

Labid	PROTOTYPE			
	T	N	M	STAGE
P1	1	0	0	I
P2	2	0	0	II
P3	1	0	0	I
P4	1	0	0	I
P5	2	1	0	III
P6	3	1	0	III
P7	3	2	0	III
P8	2	1	0	III
P9	3	2	0	III
P10	3	1	1	IV
P11	4	2	1	IV
P12	4	2	0	III
P13	1	0	0	I
P14	1	1	0	III
P15	2	1	0	III
P16	3	0	0	II
P17	3	2	0	III
P18	4	1	1	IV
P19	4	2	1	IV
P20	4	3	1	IV
…	…	…	…	…

The interpretation is compared with the interpretation of the Oncology Team as shown in Table 4.2.

Table 4.2: Some Selected Team of Oncology Interpretation

Labid	ONCOLOGY TEAM			
	T	N	M	STAGE
P1	1	0	0	I
P2	2	0	0	II
P3	2	0	0	II
P4	2	0	0	II
P5	2	1	0	III
P6	2	1	0	III
P7	3	1	0	III
P8	2	1	1	IV
P9	3	2	0	III
P10	3	1	1	IV
P11	3	2	1	IV
P12	3	2	0	III

DOI: https://doi.org/10.57233/ihcfs.053

ISBN: 978-978-68-07775

P13	1	0	0	I
P14	2	1	0	III
P15	2	1	0	III
P16	3	1	0	III
P17	3	2	0	III
P18	3	1	1	IV
P19	3	2	1	IV
P20	3	2	1	IV
…	…	…	…	…

From the above interpretation in Table 4.2, we deduce the following:
TP: True Positive. The prototype correctly identifies patients with the disease (ie those having stage II and above). = 47.
TN: True Negative. The prototype correctly identifies patients without the disease (ie those having stage I), = 8.
FP: False Positive. The prototype identifies patients without the disease as positive =1.
FN: False Negative. The prototype identifies patients with the disease as negative =4.
Sensitivity = (TP / (TP+FN) ×100)

= (47 / 51) × 100 = 92.16%

While Specificity = (TN / (TN+FP) × 100)

= (8 /9) × 100 = 88.89%.

Thus, Accuracy = ((TP + TN) / (TP+ TN + FP +FN)) × 100

= (55 / 60) × 100 = 91.67%.

Hence, percentage error recorded from the prototype = 100 - 91.67 = 8.33%.

5.0 Conclusion:

The study automated the presentation of biopsy reports by converting text-based reports into a quantitative format, enabling the determination of the optimal cancer stage and the prediction of its prognosis through data mining techniques. As a result, the research aims and objectives were successfully accomplished, and the research questions were effectively addressed.

Although the oncologists' interpretations outperform those generated by the prototype, the system can still serve as a supportive tool for medical practitioners in learning, presenting, and interpreting biopsy reports.

REFERENCE:

Arafat, H., Barakat, S., & Goweda, A. F. (2012). *Using Intelligent Techniques for Breast Cancer Classification*. *I*(3), 26–36.

Cellina, M., Cacioppa, L. M., Cè, M., Chiarpenello, V., Costa, M., Vincenzo, Z., Pais, D., Bausano, M. V., Rossini, N., Bruno, A., & Floridi, C. (2023). Artificial Intelligence in Lung Cancer Screening: The Future Is Now. *Cancers*, *15*(17). https://doi.org/10.3390/cancers15174344

Chen, H., Lan, X., Yu, T., Li, L., Tang, S., Liu, S., Jiang, F., Wang, L., Huang, Y., Cao, Y., Wang, W., Wang, X., &

Zhang, J. (2022). Development and validation of a radiogenomics model to predict axillary lymph node metastasis in breast cancer integrating MRI with transcriptome data: A multicohort study. *Frontiers in Oncology*, *12*(December), 1–11. https://doi.org/10.3389/fonc.2022.1076267

Chen, W., Tan, X., Zhang, J., Du, G., Fu, Q., & Jiang, H. (2024). A robust approach for multi-type classification of brain tumor using deep feature fusion. *Frontiers in Neuroscience*, *18*(February), 1–14. https://doi.org/10.3389/fnins.2024.1288274

Choudhury, T., Kumar, V., & Nigam, D. (2014). An Innovative Smart Soft Computing Methodology towards Disease (Cancer, Heart Disease, Arthritis) Detection in an Earlier Stage and in a Smarter Way. *International Journal of Computer Science and Mobile Computing*, *3*(4), 368–388.

da Silva, H. E. C., Santos, G. N. M., Leite, A. F., Mesquita, C. R. M., de Souza Figueiredo, P. T., Stefani, C. M., & de Melo, N. S. (2023). The use of artificial intelligence tools in cancer detection compared to the traditional diagnostic imaging methods: An overview of the systematic reviews. *PLoS ONE*, *18*(10 October), 1–29. https://doi.org/10.1371/journal.pone.0292063

Durairaj, M., & Ranjani, V. (2013). Data Mining Applications In Healthcare Sector : A Study. *International Journal of Scientific & Technology Research*, *2*(10), 29–35. http://www.ijstr.org/final-print/oct2013/Data-Mining-Applications-In-Healthcare-Sector-A-Study.pdf

Gillies, R. J., & Schabath, M. B. (2020). Radiomics improves cancer screening and early detection. *Cancer Epidemiology Biomarkers and Prevention*, *29*(12), 2556–2567. https://doi.org/10.1158/1055-9965.EPI-20-0075

Hassan, A., & Al Moaraj, A. M. H. A. (2022). The Role of Artificial Intelligence in Entrepreneurship. *Lecture Notes in Networks and Systems*, *423 LNNS*, 530–542. https://doi.org/10.1007/978-3-030-93464-4_52

Hassan, I., Haruna Dokoro, A., Garba, E. J., & Ahmadu, A. S. (2024). Application of Data Mining Techniques in Biopsy Interpretation and Staging of Carcinoma Cancer Disease: A Case Study of Northeastern Nigeria. *International Journal of Innovative Science and Research Technology (IJISRT)*, *9*(4), 2567–2575. https://doi.org/10.38124/ijisrt/ijisrt24apr2607

Ibrahim, H. (2022). *3 . A Systematic Review of Data Mining in Health Care : A Case of Breast Cancer*. *2*(1), 19–25.

Imaging, U. (2021). *Transfer Learning in Breast Cancer Diagnoses via*. 1–15.

Jia, L. L., Zheng, Q. Y., Tian, J. H., He, D. L., Zhao, J. X., Zhao, L. P., & Huang, G. (2022). Artificial intelligence with magnetic resonance imaging for prediction of pathological complete response to neoadjuvant chemoradiotherapy in rectal cancer: A systematic review and meta-analysis. *Frontiers in Oncology*, *12*(October). https://doi.org/10.3389/fonc.2022.1026216

Jiang, X., Hu, Z., Wang, S., & Zhang, Y. (2023). Deep Learning for Medical Image-Based Cancer Diagnosis. *Cancers*, *15*(14). https://doi.org/10.3390/cancers15143608

Kaya Keleş, M. (2019). Breast cancer prediction and detection using data mining classification algorithms: A comparative study. *Tehnicki Vjesnik*, *26*(1), 149–155. https://doi.org/10.17559/TV-20180417102943

Kharya, S. (2012). Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease. *International Journal of Computer Science, Engineering and Information Technology*, *2*(2), 55–66. https://doi.org/10.5121/ijcseit.2012.2206

Kolçe, E. Ç., & Frasheri, N. (n.d.). *Data Mining Techniques and Tools Used in Healthcare Databases*. (May 2013), 24–25.

Kolla, L., & Parikh, R. B. (2024). Uses and limitations of artificial intelligence for oncology. *Cancer*, *130*(12), 2101–2107. https://doi.org/10.1002/cncr.35307

Lo Gullo, R., Marcus, E., Huayanay, J., Eskreis-Winkler, S., Thakur, S., Teuwen, J., & Pinker, K. (2024). Artificial Intelligence-Enhanced Breast MRI: Applications in Breast Cancer Primary Treatment Response Assessment and Prediction. *Investigative Radiology*, *59*(3), 230–242. https://doi.org/10.1097/RLI.0000000000001010

Majali, J., Niranjan, R., Phatak, V., & Tadakhe, O. (2015). Data Mining Techniques For Diagnosis And Prognosis Of Cancer. *Ijarcce*, *4*(3), 613–615. https://doi.org/10.17148/ijarcce.2015.43147

Mathuria, M., & Saini, D. K. (2025). Advances in Image Processing and Pattern Recognition. *Advances in Cancer Detection, Prediction, and Prognosis Using Artificial Intelligence and Machine Learning*, 117–137. https://doi.org/10.1007/978-981-96-9346-7_5

Mohammed, M., Mwambi, H., Mboya, I. B., Elbashir, M. K., & Omolo, B. (2021). A stacking ensemble deep learning approach to cancer type classification based on TCGA data. *Scientific Reports*, *11*(1), 1–22. https://doi.org/10.1038/s41598-021-95128-x

Pawan, Y. V. R. N., Kumar, K. V., & Murugesan, M. (n.d.). *Application of Data Mining Classification Methods to the Diagnosis and Prognosis of Breast Cancer*. *0*, 1–8.

Pradhan, M. (2015). Data Mining& Health Care: Techniques of Application. *International Journal of Innovative Research in Computer and Communication Engineering*, *02*(12), 7445–7455. https://doi.org/10.15680/ijircce.2014.0212029

Saad, H., & Nagarur, N. (2020). Data Mining Techniques in Predicting Breast Cancer. *Journal of Applied Sciences*, *20*(3), 124–133. https://doi.org/10.3923/jas.2020.124.133

Said, A. A., Abd-Elmegid, L. A., Kholeif, S., & Gaber, A. A. (2018). Stage – Specific predictive models for main

prognosis measures of breast cancer. *Future Computing and Informatics Journal*, *3*(2), 391–397. https://doi.org/10.1016/j.fcij.2018.11.002

Shaikh, F. J., & Rao, D. S. (2021). Prediction of Cancer Disease using Machine learning Approach. *Materials Today: Proceedings*, *50*(xxxx), 40–47. https://doi.org/10.1016/j.matpr.2021.03.625

Sowmiya, T., Gopi, M., Begin, M. N., & Robinson, L. T. (2014). Optimization of Lung Cancer using Modern Data Mining Techniques. *International Journal of Engineering Research*, *3*(5), 309–314. https://doi.org/10.17950/ijer/v3s5/503

Sumbaly, R., Vishnusri, N., & Jeyalatha, S. (2014). Diagnosis of Breast Cancer using Decision Tree Data Mining Technique. *International Journal of Computer Applications*, *98*(10), 16–24. https://doi.org/10.5120/17219-7456

Tafish, M. H., & El-Halees, A. M. (2018). Breast Cancer Severity Degree Predication Using Data Mining Techniques in the Gaza Strip. *Proceedings - 2018 International Conference on Promising Electronic Technologies, ICPET 2018*, 124–128. https://doi.org/10.1109/ICPET.2018.00029

Wang, H., & Yoon, S. W. (2015). Breast cancer prediction using data mining method. *IIE Annual Conference and Expo 2015*, (May), 818–828.