ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

Application of Data Mining Techniques in Biopsy Interpretation and Staging of Carcinoma Cancer Disease: A Case Study of Northeastern Nigeria

Ibrahim Hassan[1]; Ahmed Haruna Dokoro[2]; E. J. Garba[3]; A. S. Ahmadu[4]
[1,2]Department of Computer Science Technology, Gombe State Polytechnic Bajoga, Nigeria
[3,4]Department of Computer Science Modibbo Adama Univerisity, Yola, Nigeria

Abstract:-This paper introduces an innovative framework tailored for carcinoma cancer staging in northeastern Nigeria, employing data mining techniques and machine learning algorithms, with analysis conducted using the WEKA toolkit. Given the challenges in healthcare systems, especially in cancer diagnosis and treatment, this study aims to enhance diagnostic precision thorough analysis of biopsy reports. By assessing Support Vector Machine (SVM) and decision tree algorithms using datasets from six tertiary health institutions, the framework demonstrates promising outcomes in accurately determining cancer malignancy stage. SVM exhibits high accuracies in staging breast and prostate cancer, while decision tree algorithms show notable accuracy rates in staging skin cancer. These results highlight the potential of machine learning in advancing cancer diagnosis and treatment planning, particularly in resource-constrained settings, and pave the way for the development of computer-aided diagnostic tools tailored for carcinoma cancer staging in northeastern Nigeria.

Keywords:- *Biopsy, Carcinoma Cancer, Data mining, Framework, Staging*.

I. INTRODUCTION

In recent years, the field of data science has garnered significant attention worldwide, owing to its versatile applications across various domains. Datasets, spanning from marketing and transportation to social media and healthcare, play a pivotal role in shaping insights and decision-making processes within these diverse sectors [1]. Despite the potential of data-driven methodologies, challenges persist in effectively managing complex healthcare systems, particularly evident in Nigeria's ongoing struggle to improve health outcomes for its extensive population of over 200 million individuals [2]. Amidst these healthcare challenges, cancer has emerged as a formidable global health concern, with staggering statistics reflecting its pervasive impact. Annually, the world witnesses approximately 12.7 million new cancer cases, accompanied by 7.6 million cancer-related deaths, highlighting the urgent need for effective strategies in diagnosis and treatment [3]. Cancer, characterized by the uncontrolled growth and spread of malignant cells, manifests in various forms, posing significant threats to individuals' well-being and survival [4]. Among the myriad types of cancer, carcinoma cancer stands out as one of the most prevalent, originating in the epithelial cells that line the body's organs and cavities [5]. Lung cancer, breast cancer, and skin cancer are particularly prominent among the diverse spectrum of malignancies. In the pursuit of accurate cancer diagnosis, biopsy remains the gold standard, offering crucial insights into disease pathology [6].

In addressing the complexities of cancer diagnosis and classification, machine learning techniques, particularly Support Vector Machine (SVM) learning, have emerged as powerful tools. SVMs demonstrate remarkable efficacy in discerning intricate patterns within complex datasets, with applications ranging from handwriting recognition to cancer subtype identification through genomic feature analysis [7]. This paper delves into the utilization of SVM as a classifier in cancer research, aiming to explore its potential in enhancing diagnostic accuracy and refining treatment strategies. Cancer is difficult to diagnose at early stages or can easily relapse after treatment. Some cancers are difficult to detect in their early stages due to their vague symptoms and the indistinctive tell-tale signs on mammograms, biopsies, and scans [5]. Thus, biopsy is the main approach in which Doctors used in diagnosing cancer [3]. A biopsy is the removal of a small amount of tissue for examination under a microscope. Other tests can suggest that cancer is present but only biopsy can make a definite diagnosis [6]. The biopsy report obtained from a Pathologist is been interpreted to enable the Oncologist determine whether the tumor is benign (free of cancer) or malignant (presence of cancer). If the tumor happens to be malignant, it is further investigated to determine the TNM stage of the disease. This diagnosis is a prerequisite in administering any treatment. Therefore, the concept of data mining technique was employed in this study to ascertain a comprehensive interpretation of biopsy report obtained from six tertiary health institutions in Northeast Geopolitical zone of Nigeria to determine the stage of the disease and recommend treatment for (carcinoma) cancer patients. The health institutions are:

- Federal Teaching Hospital Gombe (FTHG).
- University of Maiduguri Teaching Hospital (UMTH).
- Abubakar Tafawabalewa Univerisity Teaching Hospital Bauchi (ATBUTHB).
- Federal Medical Centre Nguru (Yobe). (FMCN).
- Modibbo Adama University Teaching Hospital Yola (MAUTHY).

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

- Federal Medical Centre Jalingo. (FMCJ).

Even though all the necessary steps must be taken in order to maintain the quality in healthcare data, data sharing is the major challenge [2]. Neither patients nor healthcare organizations are interested in sharing their private data. Due to this, planning to provide better treatments for a large population may not be possible, and difficulty in the detection of fraud and abuse in healthcare insurance companies etc [4]. Another challenge is to build the data warehouse where all the healthcare organizations within a country share their data is very costly and time-consuming process [6].

➢ *Objectives of the Study*

The primary aim of the research is to obtain an effective interpretation of carcinoma cancer disease (staging), through biopsy by evaluating two (2) Data Mining algorithms to ascertain the most appropriate. Apart from this fundamental objective, the research will look at the following supplementary objectives:

- To collect biopsy report (data of carcinoma cancer) from six tertiary health institutions from North-eastern Nigeria for a period of ten years (2012 to 2022).
- To design a framework that will determine whether the tumor is a benign (free of cancer cells which do not require staging) or malignant (which has cancer cells and require further biopsy to obtain the TNM stage of the disease).
- To interpret categorical data (from a biopsy report) into a quantitative data in order to obtain the TNM code of the disease.
- To compare the accuracy, sensitivity, and specificity of two (2) classical algorithms (decision tree and Support Vector Machine SVM) in staging of carcinoma cancer disease.

II. LITERATURE REVIEW

Reference [8] conducted a study aimed at identifying and categorizing benign and malignant skin lesions, with a specific focus on melanoma and non-melanoma cancers. The study utilized a range of classification algorithms and followed a multi-stage process involving preprocessing, segmentation, feature extraction, and classification. The findings emphasize the significance of early detection in enhancing cancer prognosis. Results revealed that the Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms achieved the highest accuracy rates, reaching 93.70% and 92.70%, respectively, highlighting their effectiveness in detecting skin cancer.

[9], conducted research addressing the urgent need for reliable automatic melanoma screening systems, given the significant public health burden of skin cancer, particularly melanoma, which contributes to a considerable number of annual deaths. Utilizing advancements in deep learning techniques, the study explored the development of an ensemble of deep convolutional neural networks to enhance the accuracy of classifying dermoscopy images into categories of melanoma, nevus, and seborrheic keratosis, despite having limited labeled data. Through fusion-based methods, the proposed ensemble framework exhibited enhanced classification performance, achieving an average area under the receiver operating characteristic curve of 0.891 on a well-known skin cancer recognition challenge dataset.

Whereas [10], introduced a robust method for skin cancer detection utilizing Support Vector Machine (SVM) in conjunction with Histogram of Oriented Gradients (HOG) features. This approach was applied to gray-scale images that underwent preprocessing with a median filter and rebalancing through image resampling. Employing Radial Basis Function (RBF) kernel-based SVM classification, the proposed method demonstrated notable performance, achieving 76% accuracy, 85% specificity, 84% precision, 76% recall, and 75% F1-score on the ISIC 2018 dataset.

While [11], addressed the urgent requirement for automated grading systems in diagnosing prostate cancer, considering its widespread occurrence and intricate nature. By employing 2D Discrete Wavelet Packet Decomposition and Support Vector Machines (SVM), the developed Computer Aided Diagnosis (CAD) system showcased encouraging outcomes, achieving a classification accuracy of 92.24% when tested on histological images of prostate cancer tissues. These automated systems have the potential to improve diagnostic accuracy and aid in treatment decision-making, representing a significant advancement in prostate cancer care.

[12] conducted a study on deep learning-based models aimed at detecting prostate cancer tissue in whole-slide images. Their approach involved leveraging a large annotated training dataset and utilizing a state-of-the-art convolutional network architecture (NASNetLarge). The proposed algorithm achieved an overall accuracy of 97.3% intumor detection, and over 98% when incorporating suggested augmentation strategies, demonstrating high performance comparable to that of pathologists. Furthermore, the research introduced a biologically meaningful deep learning algorithm for Gleason grading, showcasing performance on par with human experts in prognostic stratification of patients across various validation cohorts.

Also [13], created a swift and economical tool by leveraging an automated machine learning platform (AutoML) to improve the detection of clinically significant prostate cancer (csPCa) using routine clinical examinations. Through the utilization of multicenter cohorts and rigorous algorithmic selection, the Prostate Cancer Artificial Intelligence Diagnostic System (PCAIDS) demonstrated robust diagnostic performance, with an area under the curve (AUC) ranging from 0.807 to 0.853 across cohorts. This performance surpasses that of conventional markers such as PSA or fPSA/tPSA in identifying csPCa.

Furthermre, [14] underscore cancer as a significant global health challenge, projecting an increase in mortality rates by 2030. They delve into the realm of data mining, specifically focusing on classification algorithms, to predict

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

diseases in healthcare. By comparing Bayesian Network and J48 algorithms using WEKA software, their study evaluates classification accuracy, execution time, and other parameters, providing insights into their effectiveness for healthcare applications.

[15], tackle the notable issue of breast cancer, highlighting its prevalence and the difficulties patients encounter due to treatments and mortality rates. They introduce a novel approach called BCD-WERT, which employs the Extremely Randomized Tree and Whale Optimization Algorithm (WOA) for efficient feature selection and classification. The methodology involves experimenting with various machine learning algorithms on an extensive dataset, demonstrating BCD-WERT's superior performance with a 99.30% accuracy rate, closely followed by SVM at 98.60%. These findings underscore the efficacy of feature selection techniques in improving prediction accuracy, offering a promising solution for breast cancer prediction.

[17] introduce a hybrid framework designed for both classifying and segmenting breast scans. This framework utilizes different convolutional neural network (CNN) architectures through transfer learning, incorporating multiple datasets from various modalities. Despite encountering challenges related to data availability for training and segmentation purposes, the reported outcomes highlight the effectiveness of the framework. Notably, it achieves a segmentation accuracy of 95.58% with the Attention U-Net architecture.

III. METHODOLOGY

This study is a retrospective diagnostic investigation conducted across multiple centers, involving consecutive clinical patients from six tertiary medical institutions located in the northeastern region of Nigeria.

A. Area of Study and Population

The research seeks to examine biopsies from three carcinoma ailments: Breast Cancer Disease (BBCD), Prostate Cancer Disease (BPCD), and Skin Cancer Disease (BSCD), with the objective of ascertaining their individual stages. Tailored proforma documents were devised for each disease in collaboration with healthcare experts from the six institutions in northeastern Nigeria over a span of ten years. Stringent measures were implemented to safeguard patient confidentiality by excluding any personal details from the dataset, relying solely on pertinent clinical information. Table 1 presents the documented cases spanning from 2012 to 2022.

Table 1: Dataset and Approval for Ethical Consideration

Health Institution	Approval for Ethical Consideration	BBCD	BPCD	BSCD	TOTAL
ATBUTHB	4th Sept 2023	602	523	476	1,601
FTHG	3rd Feb 2022	725	418	509	1,652
FMCN	20th Sept 2023	390	285	187	862
FMCJ	28th Sept 2023	376	212	190	778
MAUTHY	23 Oct 2023	412	322	387	1,121
UMTH	1st Nov 2023	816	511	624	1,951
Total Cases Recorded		3,321	2,271	2,373	7,965

B. Attributes Definition of the Dataset

The detailed descriptions of the various attributes of the datasets pertaining to BBCD, BPCD, and BSCD are provided in Table 2, Table 3, and Table 4, respectively.

Table 2: Attributes of the BBCD Datset

Feature	Description	Type
ID	A unique identifier for every individual	Numeric
AGE	Age of the individual diagnosed with breast cancer (in years)	Numeric
SEX	Gender of the individual (male or female Value 1: Male Value 0: Female	Nominal
Tumor Size	Size of the tumor in millimeters	Numeric
HG:	Degree of abnormality in cancer cells Low:0 Intermediate:1 High:2	Ordinal
LNI	Presence or absence of cancer in nearby lymph nodes Absence:0 Presence:1	Nominal
MS	Indicates if cancer has spread to distant organs Abence:0 Presence:1	Nominal

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

HRS	Presence or absence of hormone receptors (e.g., estrogen, progesterone) Negative:0 Positive:1	Nominal
HER2/Status	Molecular marker influencing treatment decisions Negative:0 Positive:1	Nominal
Ki-67 PI	Measurement indicating cancer growth rate	Numeric
D/LC	Cancer subtype classification Ductal:1 Labular:1 In-situ:3	Nominal
MoE	Information about the completeness of tissue removal Clear:1 Involved:2	Nominal
IC	Features extracted from imaging studies (e.g., mammograms, MRIs)	Binary
GM	Presence or absence of specific genetic markers (e.g., BRCA1, BRCA2) Absence:0 Presence:1	Nominal
TH	Previous treatments and their outcomes	Binary

Table 3: Attributes of the BPCD Datset

Feature	Description	Type
ID	patients' unique hospital number	Numeric
AGE	Diagnosed patients' age (in years)	Numeric
GLEASON SCORE	A grading system that assesses the aggressiveness of prostate cancer cells.	Numeric
PSA LEVEL	Blood marker indicating the presence and progression of prostate cancer. Typically measured in nanograms per milliliter (ng/mL), with normal levels below 4 ng/mL	Numeric
CLINICAL STAGE	The extent of the cancer based on physical examination and imaging studies. T1 =1, T2 =2, T3 =3, T4=4	Nominal
BR	Outcome of the prostate biopsy indicating the presence or absence of cancer . Negative =0 Positive =1	Nominal
PI	Presence or absence of cancer invading nerves in the prostate. No = 0 Yes= 1	Nominal
CI	invasion of cancer beyond the prostate capsule No= 0 Yes =1	Nominal
SVI	Presence or absence of cancer involvement in the seminal vesicles. No= 0 Yes =1	Nominal
ETHINICITY	Ethnic background of the patient Hausa =1 Igbo = 2 Yoruba = 3 Others =4	Nominal
FHoPC	Whether the patient has a family history of prostate cancer. No =0 Yes = 1	Nominal
PTH	Record of any prior treatments for prostate cancer. Negative =0 Positive = 1 Neutral =2	Nominal

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

Table 4: Attributes of the BSCD Datset

Feature	Description	Type
ID	Patient's unique ID	Numeric
AGE	Diagnosed patients' age (in years)	Numeric
SEX	Gender of diagnosed patients with two values: Value 1: indicates male Value 0: indicates female	Nominal
BT	Depth of invasion of the melanoma tumor into the skin Measured in millimeters, typically ranging from 0 to a few millimeters.	Numeric
CL	Classification based on the depth of invasion into the layers of the skin. Levels I to V, with higher levels indicating deeper invasion.	Nominal
T	tumor size (T).	Nominal
N	lymph node involvement (N)	Nominal
M	metastasis (M)	Nominal
US	Presence or absence of ulceration on the skin surface. No=0 Yes=1	Nominal
MR	Rate of cell division within the melanoma tumor. Measured as the number of mitoses per square millimeter.	Numeric
RS	Presence or absence of regression, where part of the melanoma may have regressed No=0 Yes=1	Nominal
TILs	Presence and density of immune cells within the tumor Non Visible:0 Partially visible =1 Visible = 2	Nominal
MSI STATUS	Presence or absence of microsatellite instability, which can indicate a deficiency in DNA repair mechanisms. Unstable=0 Stable=1	Nominal
MB STATUS	Presence or absence of a specific genetic mutation.	Nominal
PTS	Specific location of the primary melanoma tumor on the skin.	Nominal

C. Method of Data Analysis

The widely-used open-source data mining tool WEKA (version 3.7.1) was employed for conducting the data analysis. WEKA, short for Waikato Environment for Knowledge Analysis, is software utilized for data mining analysis. Originating from the University of Waikato in New Zealand, it was developed using the Java programming language and is compatible with various operating systems [1].

D. Proposed System Framework Architecture

The system framework's design follows the sequential phases of data mining. Its aim is to categorize patients with tumors and discern whether the tumor is malignant or benign. Moreover, the framework seeks to identify the degree of malignancy and then suggest an appropriate treatment strategy based on the severity of the condition. Two traditional algorithms, namely decision tree and support vector machine, were employed to construct the models. The main objective was to create classifiers utilizing three datasets obtained from the case study, each meticulously collected, pre-processed, and refined for analysis.

E. Experimental Setup

During this phase, experimentation involved constructing models for each of the three datasets using the selected algorithms (J48 and SVM). A validation technique employing stratified k-fold cross-validation was employed to assess the classifiers. The dataset was partitioned into k equal-sized segments, with k−1 groups utilized for training and the remaining segment for evaluation in each iteration. A value of 10 was chosen for k, referred to as "10-fold cross-validation," to enhance performance. Throughout this process, 90% of the data was allocated for training, while the remaining 10% was reserved for testing, repeating 10 times for each fold. Instances were distributed randomly across the dataset for both training and testing sets. Subsequent to the 10-fold procedure, the averages of all performance metrics were computed.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

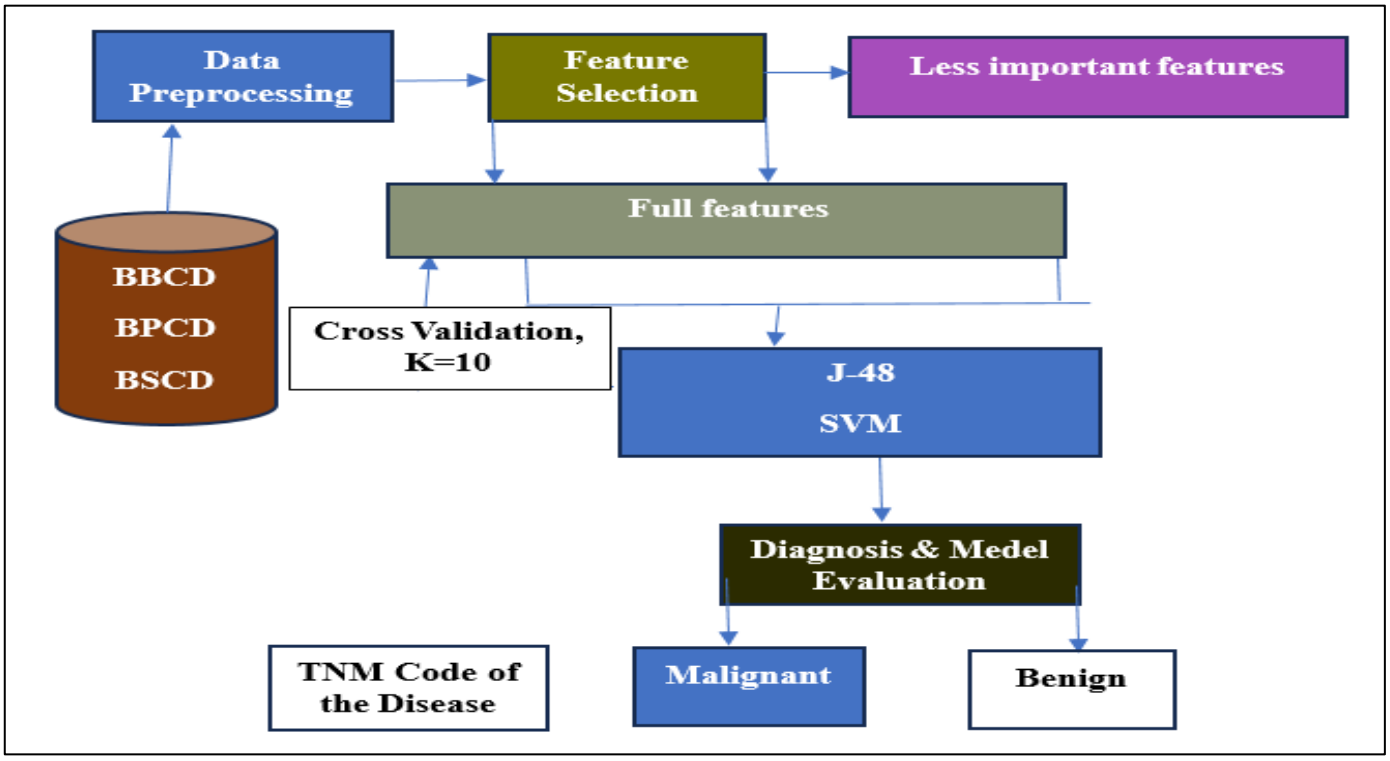

Fig 1: Model of the Biopsy Interpreter Framework

IV. RESULTS AND DISCUSSION

The experiment entailed training and evaluating algorithms using the complete set of features within the three datasets: BBCD, BPCD, and BSCD, to develop the models.

Table 5: Performances of Classifiers on BBCD Dataset

Classifier	Accuracy (%)	Sensitivity (%)	Precision (%)	MCC (%)	ROC (%)	Execution Time (s)	Frequency 1/T
J-48	98	99	99	99	99	0.09	11.11
SVM	98	99	99	99	99	0.03	33.33

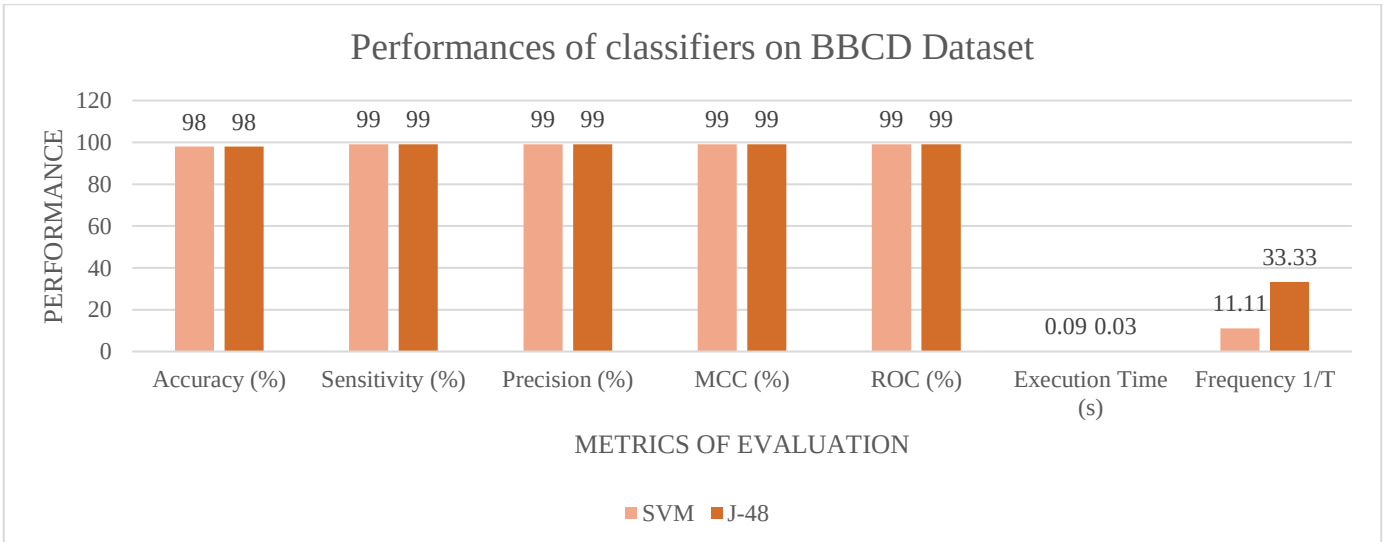

Fig 2: Performances of Classifiers on BBCD Dataset

Table 5 and Figure 2 illustrated the performance of the two classifiers in interpreting biopsy reports for breast cancer staging, revealing consistently high accuracy rates of 98%. However, SVM exhibited superior performance across all metrics and the shortest execution time of 0.03 seconds compared to [1].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

Table 6: Performances of Classifiers on BPCD Dataset

Classifier	Accuracy (%)	Sensitivity (%)	Precision (%)	MCC (%)	ROC (%)	Execution Time (s)	Frequency (1/T)
SVM	100	90	90	63	69	0.05	20
J-48	98	83	90	67	80	0.02	50

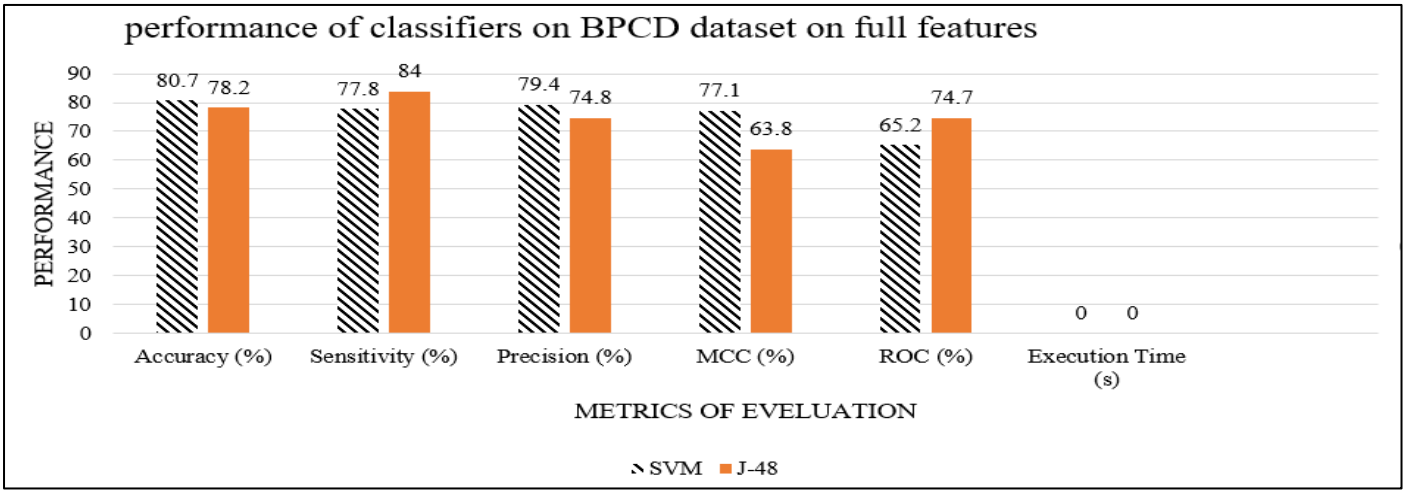

Fig 3: Performances of Classifiers on BPCD Dataset

The two classifiers were trained and assessed using the complete features of the BPCD dataset. The outcomes demonstrate strong performance across various metrics, particularly with SVM achieving perfect accuracy of 100%, 90% sensitivity and precision, 63% MCC, and 69% ROC, with a processing time of 0.05 seconds. Conversely, J-48 attained 98% accuracy, 83% sensitivity, 90% precision, 67% MCC, and 80% ROC, with an execution time of 0.02 seconds, as detailed in Table 6 and visually represented in Figure 3. These results surpass those reported by [20]

Table 7: Performances of Classifiers on BSCD Dataset

Classifier	Accuracy (%)	Sensitivity (%)	Precision (%)	MCC (%)	ROC (%)	Execution Time (s)	Frequency (1/T)
SVM	94	90	90	81	84	0.05	20
J-48	98	93	90	86	80	0.20	5

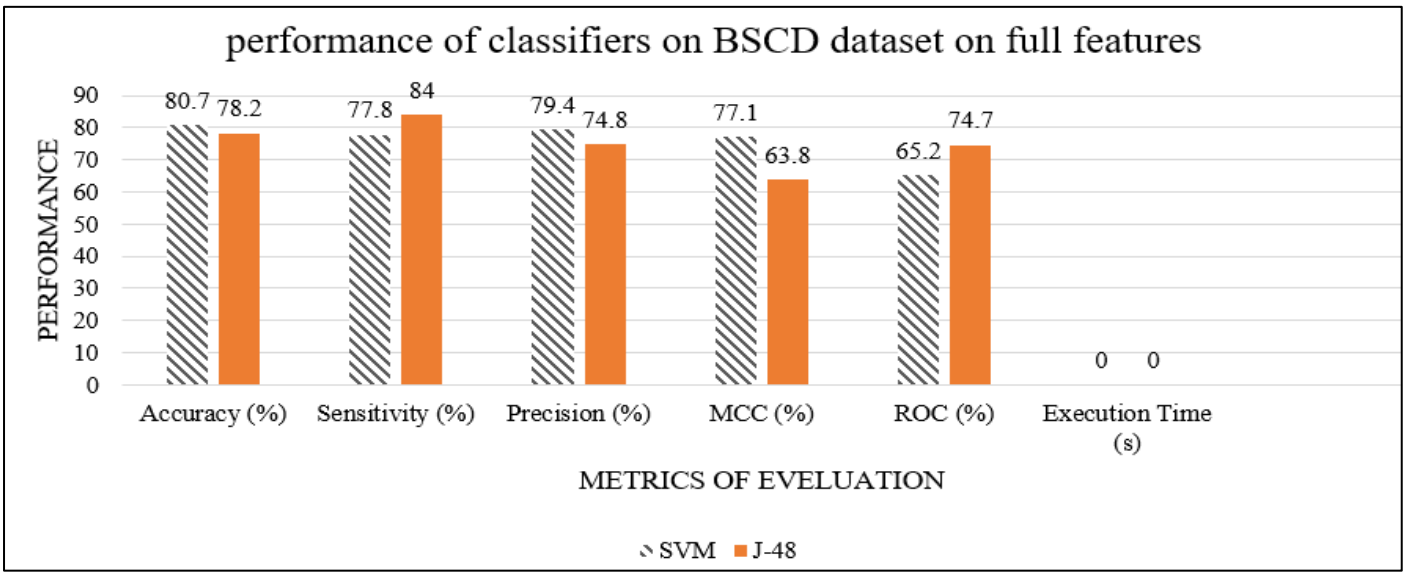

Fig 4: Performances of Classifiers on BSCD Dataset

The evaluation revealed that SVM achieved an accuracy of 94%, with 90% sensitivity and precision, an 81% MCC, and an 84% ROC, completing processing within 0.05 seconds. In comparison, J-48 attained a higher accuracy of 98%, alongside 93% sensitivity, 90% precision, an 86% MCC, and 80% ROC, with an execution time of 0.19 seconds. These findings outperformed those reported by Victor and [21].

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

V. CONCLUSION

This research presents an innovative approach to carcinoma cancer staging in northeastern Nigeria, employing data mining techniques and machine learning algorithms. Through the analysis of biopsy reports from six tertiary health institutions, our framework accurately determined the stage of cancer malignancy. Utilizing Support Vector Machine (SVM), we achieved high accuracies in staging breast and prostate cancer, reaching 98% and 100% respectively, with minimal execution times of 0.03 seconds and 0.05 seconds respectively. Additionally, decision tree algorithms demonstrated high accuracy rates in staging skin cancer at 98% accuracy, with an execution time of 0.02 seconds. These results highlight the potential of these techniques in enhancing cancer diagnosis and treatment planning, particularly in settings with limited resources. The study lays the foundation for the development of computer-aided diagnostic tools for carcinoma cancer staging and serves as a valuable reference for medical training in interpreting biopsy reports.

ACKNOLEDGEMENTS

We extend our sincere gratitude to Tetfund for generously sponsoring this research project, particularly in facilitating data collection from the six tertiary health institutions. Their support has played a crucial role in enabling our exploration of biopsy interpretation of carcinoma cancer staging using data mining methodologies.

REFERENCES

[1]. M. Fatih, "A Comparative Analysis of Breast Cancer Detection and Diagnosis Using Data Visualization and Machine Learning Applications," *HealthCare,* pp. 1-23, 2020.

[2]. O. A. Makinde and D. Boone, "Distribution of health facilities in Nigeria: Implications and options for Universal Health Coverage," *International Journal of Health Planning and Management,* pp. 1-25, August 2018.

[3]. E. Jedy-Agba, M. P. Curado, O. Ogunbiyi, E. Oga, T. Fabowalec and F. Igbinoba, "Cancer Incidence in Nigeria: A Report from Population-based Cancer Registries," *NIH-PA Author manuscript,* pp. 1-17, 2013.

[4]. S. G. Morounke, J. B. Ayorinde, A. O. Benedict, F. F. Adedayo, F. O. Adewale, I. Oluwadamilare, S. S. Sokunle and A. Benjamin, "Epidemiology and Incidence of Common Cancers in Nigeria," *Journal of Cancer Biology & Research ,* pp. 1-17, 2017. Population-based Cancer Registries," *NIH-PA Author manuscript,* pp. 1-17, 2013.

[5]. D. Talaat, F. Zada and R. Kadry, "Staging of Clear Cell Renal Cell Carcinoma Using Random Forest and Support vector machine," *Journal of physics: comference Series,* pp. 1-9, 2019.

[6]. O. Oyewo and O. Boyinbode, "Prediction of Prostate Cancer using Ensemble of Machine Learning Techniques," *International Journal of Advanced Computer Science and Applications,* vol. 11, no. 3, pp. 149-154, 2020.

[7]. S. Huang, N. Cai, P. P. Pacheco, S. Narandes, Y. Wang and W. Xu, "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics," *CANCER GENOMICS & PROTEOMICS,* pp. 41-51, 2018.

[8]. A. Victor and M. R. Ghalib, "Automatic Detection and Classification of Skin Cancer," *Internation Journal of Engineering and Systems,* vol. 10, no. 3, pp. 444-451, 1 March 2017.

[9]. B. Harangi, "Skin Lesion Classification With Ensembles of Deep Convolutional Neural Network," *Journal of Biomedical Informatics,* pp. 25-32, 2018.

[10]. G. N. K. Babu and V. J. Peter, "SKIN CANCER DETECTION USING SUPPORT VECTOR MACHINE WITH HISTOGRAM OF ORIENTED GRADIENTS FEATURES," *ICTACT JOURNAL ON SOFT COMPUTING,* vol. 11, no. 2, pp. 2301-2305, 2021.

[11]. U. Ali, A. Shaukat, M. Hussain, J. Ali1, K. Khan, M. B. Khan and M. A. Shah, "Automatic Cancerous Tissue Classification using Discrete Wavelet Transformation and Support Vector Machine," *Journal of Basic and Applied Scientific Research,* pp. 1-10, 2016.

[12]. Y. Tolkach, T. Dohmgörgen, M. Toma and G. Kristiansen, "High-Accuracy Prostate Cancer Pathology Using Deep Learning," *Nature Machine Intelligence,* pp. 1-13, 2021.

[13]. D. C. Nalini and D.Meera, "Breast cancer prediction system using Data mining methods," *International Journal of Pure and Applied Mathematics,* vol. 119, no. 12, pp. 10901-10911, 2018.

[14]. M. Botlagunta, M. Botlagunta, M. B. Myneni, D. Lakshmi, A. Nayyar, J. SaiGullapalli and M. Shah, "Classifcation and Diagnostic Prediction of Breast Cancer Metastasis on Clinical Data Using Machine Learning Algorithms," *Scientific Reports,* pp. 1-17, 2023.

[15]. Zhang, Huiyong, J. Ji, Z. Liu, H. Lu, C. Qian, C. Qian, S. Chen, W. Lu, C. Wang and H. Xu, "Artifcial Intelligence for the Diagnosis of Clinically Signifcant Prostate Cancer Based on Multimodal Data: a Multicenter Study," *BMC Medicine,* pp. 1-11, 2023.

[16]. H. M. Balaha, E. R. Affan, M. M. Saafan and E. M. El-Gandy, "A comprehensive Framework Towards Segmenting and Classifying Breast Cancer Patients Using Deep Learning and Aquila Optimizer," *Journal of Ambient Intelligence and Humanized Computing,* pp. 1-21, 2023.

[17]. M. Salvi, F. Molinari, U. R. Acharya, L. Molinaro and K. M. Meiburger, "Impact of stain normalization and patch selection on the performance of convolutional neural networks in histological breast and prostate cancer classification," *Computer Methods and Programs in Biomedicine Update ,* pp. 1-6, 2021.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24APR2607

[18]. M. A. Jabbar, "Breast cancer data classification using ensemble machine learning," *Engineering and Applied Science Research,* vol. 48, no. 1, pp. 61-72, 2021.
[19]. A. Mosayebi, B. Mojaradi, A. B. Naeini and S. H. K. Hosseini, "Modeling and comparing data mining algorithms for prediction of recurrence of breast cancer," *Plus One,* vol. 15, no. 10, pp. 1-23, 2020.
[20]. D. C. Nalini and D.Meera, "Breast cancer prediction system using Data mining methods," *International Journal of Pure and Applied Mathematics,* vol. 119, no. 12, pp. 10901-10911, 2018.
[21]. A. Victor and M. R. Ghalib, "Automatic Detection and Classification of Skin Cancer," *International Journal of Intelligent Engineering & Systems,* vol. 10, no. 3, pp. 444-451, 2017.