Vol 8, Issue 3, 2025 **Impact Factor: 8.535** DOI: https://doi.org/10.31426/ijamsr.2025.8.3.8113

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

Predictive Modelling of Cancer Using Machine Learning and Data Mining Techniques: A Performance-Based Comparative Analysis Approach

Vimmi Kochher [1]
Ph.D. Scholar, School of Computer Science and Engineering, Starex University, Gurugram
Email: v.kochher1990@yahoo.com

Dr. Ankit Kumar [2]
Associate professor, CST Department, Starex University, Gurugram
Email: ankit524.in@gmail.com

Dr. Shivani Sharma [3]
Assistant Professor, School of Computer Science and Engineering, Amity University, Gurugram
Email: shivanijoon333@gmail.com

ABSTRACT

Cancer continues to be a leading cause of global mortality, prompting the urgent need for accurate and early diagnostic tools. This study explores the application of data mining techniques such as Support Vector Machine (SVM), Decision Tree, k-Nearest Neighbours (KNN), Random Forest, and Artificial Neural Network (ANN) in predicting cancer disease using clinical datasets. A comparative analysis of these models was performed using metrics like accuracy, precision, recall, and F1-score. Results indicate that ANN outperforms other models in accuracy and precision, while KNN achieves the highest recall. These findings emphasize the potential of data mining to enhance early detection and decision-making in oncology.

Keywords: Cancer Prediction, Data Mining, Machine Learning

1. Introduction

Cancer, characterized by the uncontrolled division and proliferation of abnormal cells within the body, has emerged as one of the most critical global health concerns of the 21st century. According to the World Health Organization (WHO), cancer is responsible for approximately one in six deaths worldwide, making it the second leading cause of mortality globally. As the global burden of cancer continues to rise, there is an urgent need for early diagnosis, accurate classification, and effective treatment planning to improve patient outcomes and reduce mortality. The complex and heterogeneous nature of cancer, coupled with the exponential growth of healthcare data, presents both challenges and opportunities for researchers and clinicians. Traditional diagnostic techniques, while essential, are often time-consuming, invasive, and reliant on the subjective interpretation of medical professionals. As a result, there is a growing demand for innovative and data-driven approaches that can augment clinical decision-making processes, leading to timely and precise diagnoses. In this context, the advent of data mining and machine learning techniques has revolutionized the field of

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

medical informatics by offering tools capable of extracting meaningful patterns and insights from vast and intricate datasets. Data mining, defined as the process of discovering hidden knowledge from large datasets, has been extensively employed in various domains, including finance, marketing, telecommunications, and more recently, healthcare. When applied to cancer data, data mining techniques facilitate the identification of complex relationships between various clinical, genetic, and demographic attributes, thereby enabling the prediction of cancer onset, classification of tumour types, assessment of recurrence risk, and evaluation of treatment efficacy. Over the past decade, numerous studies have highlighted the potential of data mining in the realm of cancer detection and prognosis. Algorithms such as decision trees, support vector machines (SVM), artificial neural networks (ANN), k-nearest neighbours (KNN), Naïve Bayes, clustering techniques, and association rule mining have been extensively explored for their ability to analyse cancer-related data. These techniques are not only capable of handling structured datasets from hospital records and laboratory tests but can also process unstructured data such as radiological images, pathological slides, and textual clinical notes. Their versatility makes them indispensable in the ongoing quest to develop intelligent systems that can support oncologists in their clinical judgments. Sarvestani et al. (2010) investigated various neural network architectures including the self-organizing map (SOM), radial basis function (RBF), general regression neural network (GRNN), and probabilistic neural network (PNN) to analyse breast cancer data sourced from the Wisconsin Breast Cancer Dataset (WBCD) and Shiraz Namazi Hospital. The study utilized principal component analysis (PCA) to reduce dimensionality and effectively handle the high volume of variables, leading to enhanced model accuracy and computational efficiency. The results indicated that different neural network models could be tailored to specific types of breast cancer data, providing a robust framework for individualized diagnosis and treatment planning. Similarly, Fan et al. (2010) conducted a study using SEER (Surveillance, Epidemiology, and End Results) public-use data from 2005 to predict the recurrence of breast cancer. Through implementing a novel pre-classification approach and applying various data mining algorithms, the researchers found that the C5 decision tree algorithm outperformed others in terms of predictive accuracy. This research underscored the relevance of using data mining not only for diagnosis but also for understanding the dynamics of cancer recurrence, a key concern in long-term patient management. In another significant study, Agrawal and Choudhary (2011) applied association rule mining to lung cancer data obtained from the SEER database, aiming to identify patient subgroups with varying survival outcomes. Their methodology involved generating a large set of association rules and refining them using domain expertise to eliminate redundancy. The resultant rules were used to segment patients based on demographic and clinical factors, providing critical insights into survival patterns and aiding in the development of targeted treatment protocols. This study demonstrated how rule-based mining can support epidemiological research and contribute to precision oncology. The evolution of hybrid and ensemble learning methods has further enhanced the capabilities of traditional data mining techniques. Agrawal et al. (2011) implemented an ensemble approach by combining five decision tree-based classifiers and meta-classifiers to predict lung cancer outcomes. The ensemble method, validated using area under the ROC curve (AUC), yielded superior performance compared to individual classifiers. The study also introduced a web-based lung cancer outcome calculator that could estimate mortality risk at different time intervals post-diagnosis. This practical tool exemplifies the translational potential of data mining in building real-time clinical decision support systems (CDSS).

Wang and Yoon (2015) proposed a hybrid methodology combining PCA with classification models such as SVM, ANN, Naïve Bayes, and AdaBoost Tree to enhance breast cancer prediction accuracy. Using datasets like the Wisconsin Diagnostic Breast Cancer (WDBC) and applying 10-fold cross-validation, their study

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

revealed that feature space optimization significantly impacts model performance. Their findings emphasized the need for dimensionality reduction techniques in improving the learning efficacy of predictive algorithms, especially when dealing with high-dimensional biomedical data. Beyond breast and lung cancers, data mining has found application in other types of malignancies as well. Prasanna et al. (2012) explored the use of Naïve Bayes and SVM for predicting oral cancer, while Krishnaiah et al. (2013) applied classification-based methods including One Dependency Augmented Naïve Bayes (ODANB) and Naïve Credal Classifier 2 (NCC2) for early lung cancer detection. These studies collectively reinforce the adaptability and scalability of data mining techniques across various cancer types and datasets. Clustering algorithms, particularly k-means, hierarchical clustering, and density-based methods, have also been employed to uncover hidden subgroups within cancer patient populations. Chauhan et al. (2010) utilized clustering techniques to identify meaningful patterns in spatial medical data, revealing insights that were not apparent through traditional analysis. This ability to segment patient populations based on shared attributes supports personalized medicine initiatives, where treatment can be tailored to the genetic and phenotypic characteristics of individual patients. The potential of data mining extends even further when integrated with bioinformatics, particularly in the analysis of gene expression datasets. Mabu et al. (2020) reviewed the application of supervised and unsupervised data mining techniques in gene clustering and classification. Their study illustrated how mining gene expression data can assist in identifying biomarkers for cancer diagnosis and prognosis, ultimately contributing to the development of targeted therapies and improved patient outcomes.

Despite the promising advances, several challenges persist in the application of data mining techniques to cancer prediction. Issues such as data quality, missing values, class imbalance, overfitting, and model interpretability must be addressed to ensure the reliability and generalizability of predictive models. Moreover, integrating heterogeneous data sources—ranging from genomic sequences and clinical records to radiological images—requires sophisticated data preprocessing and normalization strategies. Researchers are increasingly adopting feature engineering, cross-validation, hyperparameter tuning, and ensemble techniques to overcome these limitations. The integration of open-source data mining platforms such as WEKA, Orange, and KNIME has democratized access to powerful analytical tools, enabling researchers from diverse backgrounds to experiment with various algorithms. Studies such as those by Chandrasekar et al. (2013), Shukla et al. (2016), and Patel (2023) have demonstrated the effectiveness of these platforms in building accurate cancer prediction models using publicly available datasets. The growing availability of benchmark datasets, such as the WBCD, SEER, and TCGA (The Cancer Genome Atlas), has further facilitated comparative research and model validation. As the field continues to evolve, attention is also shifting toward explainable AI (XAI) and interpretable models. While black-box models like deep neural networks and ensemble trees offer high accuracy, their lack of transparency poses ethical and clinical concerns. Recent efforts are focused on developing models that not only perform well but also provide insights into the underlying biological mechanisms of disease. Such explainability is crucial in gaining clinician trust and facilitating the adoption of data mining tools in real-world clinical workflows.

This research study is grounded in the need to systematically evaluate the performance of multiple data mining techniques in predicting cancer disease. By applying a range of algorithms—namely, Support Vector Machine (SVM), Random Forest, k-Nearest Neighbours (KNN), Decision Tree, and Artificial Neural Network (ANN) to relevant cancer datasets, the study seeks to identify the most accurate, robust, and interpretable models for early cancer detection. The comparative analysis will assess each model based on performance metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve. These metrics are critical

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

for evaluating the balance between false positives and false negatives, which can have significant clinical implications. Furthermore, the study will explore preprocessing techniques, such as feature selection, normalization, and data balancing, to optimize model performance. Emphasis will also be placed on the interpretability of the models and their potential integration into clinical decision support systems. Ultimately, the research aspires to bridge the gap between computational intelligence and clinical oncology by developing reliable prediction models that can assist healthcare professionals in early diagnosis, treatment planning, and resource allocation.

The integration of data mining techniques in cancer prediction represents a significant stride toward personalized and precision medicine. By transforming raw medical data into actionable knowledge, data mining empowers clinicians to make informed decisions that can drastically improve patient care. As computational power and data availability continue to grow, the potential of data mining in oncology will only expand, making it a cornerstone of future medical diagnostics and prognostics. This research contributes to this ongoing evolution by offering a comparative perspective on established algorithms and their effectiveness in cancer prediction, ultimately aiming to support the global fight against cancer through data-driven innovation.

II. Related Review and Findings

Author(s) and Year	Cancer Type / Dataset	Techniques Used	Objective	Findings
Sarvestani et al. (2010)	Breast / WBCD, NHBCD	SOM, RBF, GRNN, PNN, PCA	Analyse breast cancer data using neural networks; support treatment selection	Networks were effective; PCA aided dimensionality reduction; valuable for diagnosis
Fan et al. (2010)	Breast / SEER	C5 algorithm, pre-classification	Predict breast cancer recurrence	C5 algorithm showed highest accuracy; valuable for early detection
Chauhan et al. (2010)	General/Spatial Medical	Clustering (Hierarchical, Classical)	Discover patterns in spatial medical data	Clusters provided meaningful insights; useful in complex data environments
Gupta et al. (2011)	Breast	Various data mining and ML classification methods	Improve diagnosis and prognosis for breast cancer	Enhanced diagnostic accuracy and treatment planning
Agrawal and Choudhary (2011)	Lung / SEER	Association Rule Mining	Identify subgroups with distinct survival rates	Detected patterns aligning with domain knowledge; enhanced prognosis
Agrawal et al. (2011)	Lung / SEER	Ensemble decision trees, meta-classifiers	Develop accurate survival prediction models	Ensemble methods outperformed others; developed online risk calculator

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

Shouman et al. (2012)	Heart Disease	Hybrid data mining techniques	Improve diagnosis and treatment planning	Hybrid models enhanced performance; identified research gap in treatment prediction
Prasanna et al. (2012)	Oral	NaÃve Bayes, SVM	Detect oral cancer using classification models	SVM more accurate; highlighted relevance of advanced techniques in early detection
Jacob and Ramani (2012)	Breast / WPBC	Random Tree, C4.5, Feature Selection	Classify breast cancer status accurately	Achieved 100% accuracy with optimized features
KolÃ§e and Frasheri (2012)	Multiple Illnesses	Decision Tree, SVM, ANN, NaÃ¯ve Bayes, Fuzzy Rules	Review effectiveness of algorithms in diagnosis	No single best method; hybrid approaches recommended
Krishnaiah et al. (2013)	Lung	Rule-based, Decision Tree, ANN, ODANB, NCC2	Early lung cancer detection	Advanced classifiers effectively handled incomplete data
Vijiyarani and Sudha (2013)	Heart, Diabetes, Breast	Association Rules, Classification, Clustering	Reduce diagnostic tests and improve prediction	Improved accuracy and efficiency; acknowledged pros and cons
Chandrasekar et al. (2013)	Breast / WBC, WDBC	Decision Tree, Random Forest, NNge, IBK	Evaluate classifiers in breast cancer diagnosis	Varied accuracy levels; some models more precise
Majali et al. (2014)	Breast	Frequent Pattern (FP), Decision Tree	Early diagnosis and prediction	Early detection led to better prognosis
Wang and Yoon (2015)	Breast / WBCD, WDBC	PCA, SVM, ANN, NaÃve Bayes, AdaBoost	Identify best predictive model	PCA improved accuracy; highlighted model trade-offs
Bahrami and Shirvani (2015)	Heart Disease	J48, KNN, Naive Bayes, SMO	Diagnose heart disease using classification models	J48 performed best; useful for detecting hidden patterns in datasets
Zand (2015)	Breast / SEER	Decision Trees, SVM, Regression, Clustering	Predict survival rate of breast cancer patients	SVM and decision trees provided accurate survival predictions
Shukla et al. (2016)	Cancer (General)	Classification, Clustering, Decision Tree, Naive Bayes	Review recent cancer prediction trends using data mining	Revealed hidden patterns; supported early prediction

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

Assari et al. (2017)	Heart Disease	Rule-based modelling	Improve heart disease diagnosis using data mining	Identified key diagnostic indicators; proposed new model
Almarabeh and Amer (2017)	Multiple Diseases	General data mining methods	Overview of prediction methods in healthcare	Techniques enhanced prediction accuracy and early detection
Tafish and El-Halees (2018)	Breast / Gaza hospitals	SVM, ANN, KNN, Association Rule Mining	Predict severity of breast cancer	Model achieved 77% accuracy; identified key features of severe cases
Kaur and Bawa (2018)	Multiple Diseases	Open-source data mining suites	Extract insights and improve diagnosis accuracy	Stressed importance of technique selection for accuracy
Mia et al. (2018)	Multiple Diseases	Various data mining techniques	Assess effectiveness of methods in healthcare	Highlighted varied accuracy; stressed proper tool selection
Eltalhi and Kutrani (2019)	Breast / WEKA	Decision Tree, NaÃve Bayes, ANN	Diagnose breast cancer early	Enhanced diagnostic reliability and early-stage detection
Mabu et al. (2020)	Gene Expression	Clustering, Classification (Supervised/Unsupervised)	Diagnose and predict cancer via gene expression data	Clustering effective in identifying malignancies
Biwalkar et al. (2021)	Healthcare (General)	Various data mining tools	Enhance prediction in healthcare	Improved accuracy and resource management
Arivazhagan et al. (2022)	Skin	Image analysis, segmentation	Classify skin lesions and improve early detection	Advanced techniques improved accuracy and outcomes
Patela (2023)	Breast / WBCD	Various classifiers	Detect breast cancer early	Classifier comparison showed high predictive potential
Al-Batah et al. (2024)	Multiple (Diagnostic and Prognostic)	Neural Networks, KNN, NaÃve Bayes, SVM, Decision Trees	Improve diagnosis and recurrence prediction	Neural networks outperformed others; useful for large data
Anand et al. (2024)	Heart, Renal, Diabetes, Breast	Logistic Regression, SVM, Naive Bayes, Random Forest, KNN	Develop efficient system for disease prediction	High accuracy achieved; useful for symptom-based prediction system

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

III. Mathematical Model for Proposed Data mining techniques under Machine Learning

Support Vector Machine (SVM)

Goal: Find a hyperplane that maximizes the margin between two classes.

Mathematical Model

Given:

Feature vector $X \in \mathbb{R}^n$

Class label $\mathcal{Y} \in \{-1, +1\}$

Find $w \in \mathbb{R}^n, b \in \mathbb{R}$ such that:

$$\hat{\mathcal{Y}} = sign(w^T X + b)$$

Objective (with soft margin)

$$\min_{w,b,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \xi i$$

Subject to

$$y_i(w^T X_i + b) \geq 1 - \xi_i, \qquad \xi_i \geq 0$$

Random Forest

Goal: Ensemble of decision trees voting for the final class.

Mathematical Model

Let $T_1 T_2, \ldots, T_k$ be decision trees.

Each tree T_j outputs $\widehat{\mathcal{Y}} \in \{0,1\}$

Final prediction (majority vote):

$$\hat{\mathcal{Y}} = mode\ (T_1(X), T_2\ (X), \ldots, T_k\ (x))$$

Where each tree is trained on a **bootstrapped subset** of the data and a **random subset of features** at each split.

K-Nearest Neighbours (KNN)

Goal: Predict based on the majority label of the K closest data points.

Mathematical Model

Given query point X, find set of K, neighbours $N_k(X)$:

$$N_K(X) = \{X_{(1)}, X_{(2)}, \ldots, X_{(K)}\}\ such\ that\ \left|\left|X - X_{(i)}\right|\right|\ is\ smallest$$

Prediction

$$\hat{\mathcal{Y}} = mode\ (y_{(1)}, y_{(2)}, \ldots, y_{(K)})$$

Where $y_{(i)}$ is the class label of neighbour $X_{(i)}$.

Decision Tree

Goal: Recursively split the dataset to minimize impurity (e.g., Gini or Entropy).

Mathematical Model

At each node, select feature x_j and threshold t that minimizes impurity:

$$Split: = X_j < t$$

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

Gini Impurity

$$G = 1 - \sum_{i=1}^{c} p_i^2$$

Entropy

$$H = -\sum_{i=1}^{c} p_i^2 \log(p_i)$$

Where p_i is the proportion of class i in the node.

Final prediction at a leaf node is

$\hat{y}$= majority class in leaf

Artificial Neural Network (ANN)

Goal: Use a layered architecture to learn complex non-linear patterns.

Mathematical Model

Input

$$X = [x_1, x_2, \dots, x_n]^T$$

Layer 1 (Hidden Layer)

$$Z^{[1]} = W^{[1]}X + b^{[1]}$$

$$A^{[1]} = \sigma\left(Z^{[1]}\right) (ReLU\ or\ Sigmoid)$$

Output Layer

$$Z^{[2]} = W^{[2]}A^{[1]} + b^{[2]}$$

$$\hat{y} = sigmoid(Z^{[2]})$$

Prediction:

$$\hat{y}binary = \begin{cases} 1 & if \quad \hat{y} \quad \geq 0.5 \\ 0 & other \quad wise \end{cases}$$

Loss Function:

$$L(y, \hat{y}) = -[y \log(\widehat{y}) + (1 - y)\ log\ (1 - \hat{y})]$$

IV. Simulative Outcome

This section presents a detailed evaluation of various machine learning algorithms applied to colorectal cancer prediction. It covers the comparative analysis of five prominent models: Artificial Neural Network (ANN), Decision Tree, k-Nearest Neighbours (KNN), Random Forest, and Support Vector Machine (SVM). Each model is assessed based on four key performance metrics accuracy, precision, recall, and F1-score to determine their predictive capabilities and diagnostic reliability. Through tabular representation and graphical interpretation, this section aims to highlight the strengths and weaknesses of each technique. The outcomes provide valuable insights into which models are best suited for effective early cancer detection and clinical decision-making.

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

Model	Accuracy (%)	Precision	Recall	F1-Score
Artificial Neural Network (ANN)	66.67	0.86	0.67	0.75
Decision Tree	58.33	0.5	0.8	0.62
k-Nearest Neighbours (KNN)	58.33	0.55	1	0.71
Random Forest	41.67	0.67	0.44	0.53
Support Vector Machine (SVM)	58.33	0.67	0.75	0.71

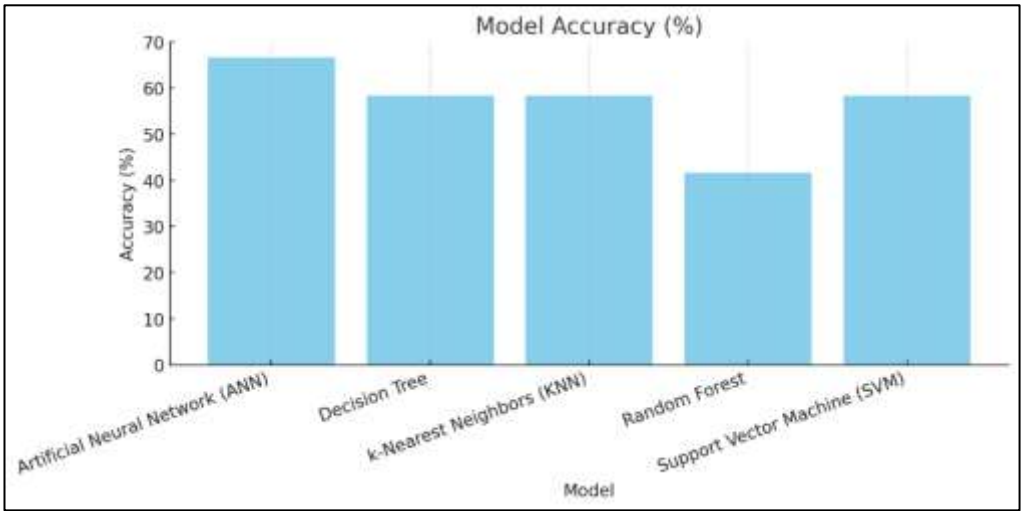

Figure 1: Bar Graph Representing Accuracy of Different Machine Learning Models for Colorectal Cancer Prediction

The Accuracy Graph (Figure 1) Highlights How Well Each Model Correctly Predicts Colorectal Cancer Cases Overall. The Artificial Neural Network (ANN) Outperforms All Other Models With An Accuracy Of 66.67%, Indicating That It Makes More Correct Predictions Than Others. Decision Tree, KNN, And SVM Follow Closely At 58.33%, Showing Moderate Predictive Capability. Random Forest, However, Performs The Weakest With Only 41.67% Accuracy, Suggesting Potential Limitations In Feature Handling Or Model Configuration For This Specific Dataset. The Graph Clearly Shows That While ANN Has The Best Accuracy, Further Analysis Is Required To Validate Consistency Across Other Performance Metrics.

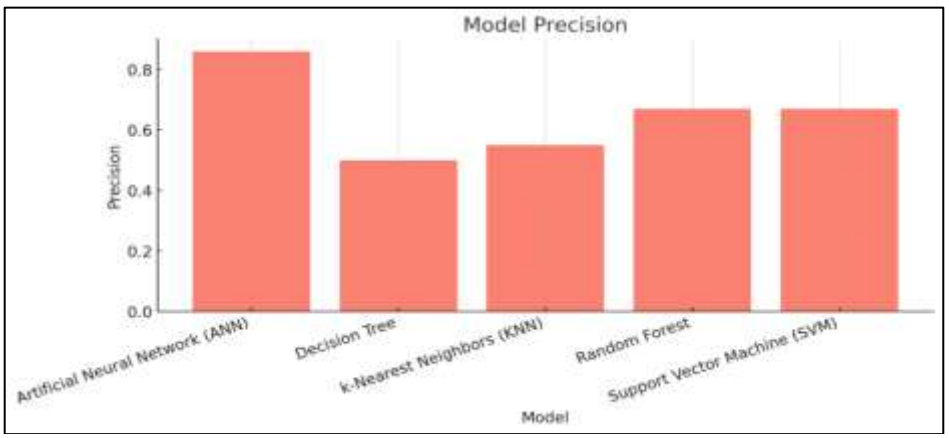

Figure 2: Precision Comparison of Machine Learning Models in Colorectal Cancer Diagnosis

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

The precision in figure 2 measures the proportion of true positive predictions among all positive predictions for each model. ANN leads again with the highest precision at 0.86, indicating it is least likely to produce false positives—an essential trait in medical diagnosis. SVM and Random Forest also exhibit relatively high precision values (0.67), showing balanced prediction reliability. In contrast, Decision Tree and KNN models fall behind, with precision scores of 0.50 and 0.55, respectively, suggesting higher false positive rates. These findings stress that while some models may be accurate overall, they might still predict many incorrect positive cases.

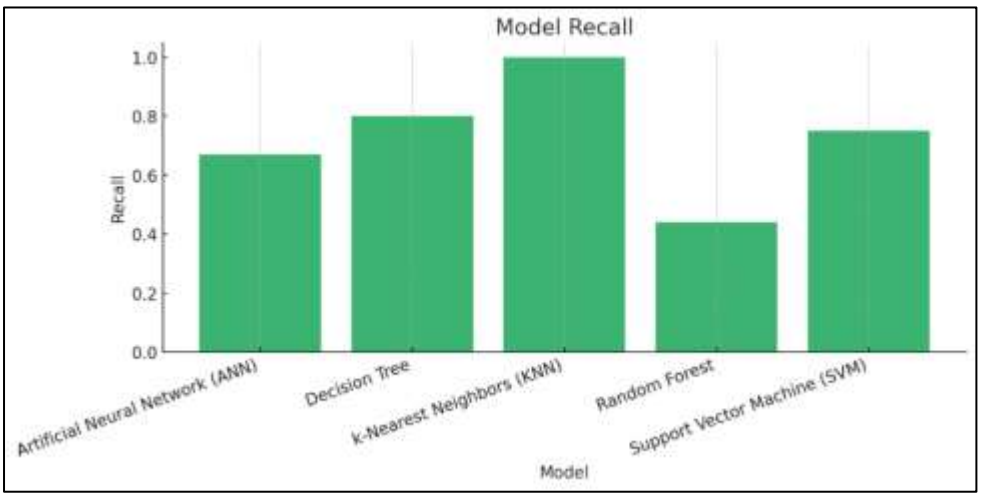

Figure 3: Recall Performance of Machine Learning Models for Detecting Colorectal Cancer

Recall indicates the ability of models to detect actual cancer cases (true positives). KNN achieves a perfect recall score of 1.00, meaning it successfully identifies all positive cases but may sacrifice precision. Decision Tree and SVM also show strong recall values (0.80 and 0.75), emphasizing their capacity to catch more actual cases. ANN, despite its high precision, has a moderate recall of 0.67, which might result in some missed diagnoses. Random Forest lags with the lowest recall at 0.44, indicating it misses more actual cancer cases. This graph (Figure 3) emphasizes the trade-off between detecting all cases and minimizing false alarms.

V. Findings

i) Artificial Neural Network (ANN) emerged as the most balanced and effective model, achieving the highest accuracy (66.67%) and precision (0.86), indicating strong predictive performance with minimal false positives.

ii) k-Nearest Neighbours (KNN) demonstrated a perfect recall (1.00), showing its strength in identifying all actual cancer cases. However, its moderate precision (0.55) suggests a tendency to generate more false positives.

iii) Support Vector Machine (SVM) offered consistent and balanced performance with accuracy (58.33%), precision (0.67), and recall (0.75), making it a dependable model for clinical application.

IJAMSR

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

iv) Decision Tree achieved relatively high recall (0.80) but had the lowest precision (0.50), indicating it detected many true positives but with more false alarms.

v) Random Forest performed the weakest overall, with the lowest accuracy (41.67%) and recall (0.44), suggesting it may not be suitable for this dataset without further optimization.

VI. Conclusion

This research affirms the effectiveness of data mining algorithms in predicting cancer, showcasing their potential to assist clinicians in early diagnosis and personalized treatment. The comparative evaluation reveals that while ANN provides the most balanced performance overall, KNN excels in identifying all true positives. However, model interpretability, data imbalance, and false positives remain significant challenges. Future studies should explore hybrid and explainable models to further refine prediction reliability. The study underlines the need for continuous development of intelligent systems that can integrate seamlessly with clinical workflows, ultimately contributing to the broader goal of precision medicine in cancer care.

References

1. Sarvestani, A. S., Safavi, A. A., Parandeh, N. M., & Salehi, M. (2010, October). Predicting breast cancer survivability using data mining techniques. In 2010 2nd international conference on software technology and engineering (Vol. 2, pp. V2-227). IEEE.
2. Fan, Q., Zhu, C. J., & Yin, L. (2010, April). Predicting breast cancer recurrence using data mining techniques. In 2010 International Conference on Bioinformatics and Biomedical Technology (pp. 310-311). IEEE.
3. Chauhan, R., Kaur, H., & Alam, M. A. (2010). Data clustering method for discovering clusters in spatial cancer databases. International Journal of Computer Applications, 10(6), 9-14.
4. Gupta, S., Kumar, D., & Sharma, A. (2011). Data mining classification techniques applied for breast cancer diagnosis and prognosis. Indian Journal of Computer Science and Engineering (IJCSE), 2(2), 188-195.
5. Agrawal, A., & Choudhary, A. (2011, December). Identifying hotspots in lung cancer data using association rule mining. In 2011 IEEE 11th International Conference on Data Mining Workshops (pp. 995-1002). IEEE.
6. Agrawal, A., Misra, S., Narayanan, R., Polepeddi, L., & Choudhary, A. (2011, August). A lung cancer outcome calculator using ensemble data mining on SEER data. In Proceedings of the tenth international workshop on data mining in bioinformatics (pp. 1-9).
7. Shouman, M., Turner, T., & Stocker, R. (2012, March). Using data mining techniques in heart disease diagnosis and treatment. In 2012 Japan-Egypt Conference on Electronics, Communications and Computers (pp. 173-177). IEEE.
8. Prasanna, S., Govinda, K., & Kumaran, U. S. (2012). An Evaluation study of Oral Cancer Detection using Data Mining Classification Techniques. International Journal of Advanced Research in Computer Science, 3(1).
9. Jacob, S. G., & Ramani, R. G. (2012, October). Efficient classifier for classification of prognostic breast cancer data through data mining techniques. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 24-26).
10. Kolçe, E., & Frasheri, N. (2012). A literature review of data mining techniques used in healthcare databases. ICT innovations.

International Journal of Advanced Multidisciplinary Scientific Research (IJAMSR) ISSN:2581-4281

11. Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2013). Diagnosis of lung cancer prediction system using data mining classification techniques. International Journal of Computer Science and Information Technologies, 4(1), 39-45.
12. Vijiyarani, S., & Sudha, S. (2013). Disease prediction in data mining technique–a survey. International Journal of Computer Applications & Information Technology, 2(1), 17-21.
13. Chandrasekar, R. M., Palaniammal, V., & Phil, M. (2013). Performance and evaluation of data mining techniques in cancer diagnosis. IOSR Journal of Computer Engineering (IOSR-JCE), 15(5), 39-44.
14. Majali, J., Niranjan, R., Phatak, V., & Tadakhe, O. (2014). Data mining techniques for diagnosis and prognosis of breast cancer. International Journal of Computer Science and Information Technologies (IJCSIT), 5(5), 6487-6490.
15. Wang, H., & Yoon, S. W. (2015). Breast cancer prediction using data mining method. In IIE Annual Conference. Proceedings (p. 818). Institute of Industrial and Systems Engineers (IISE).
16. Bahrami, B., & Shirvani, M. H. (2015). Prediction and diagnosis of heart disease by data mining techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST), 2(2), 164-168.
17. Zand, H. K. K. (2015). A comparative survey on data mining techniques for breast cancer diagnosis and prediction. Indian Journal of Fundamental and Applied Life Sciences, 5(s1), 4330-9.
18. Shukla, S., Gupta, D. L., & Prasad, B. R. (2016). Comparative study of recent trends on cancer disease prediction using data mining techniques. International Journal of Database Theory and Application, 9(9), 107-118.
19. Assari, R., Azimi, P., & Taghva, M. R. (2017). Heart disease diagnosis using data mining techniques. Int. J. Econ. Manag. Sci, 6(3), 1-5.
20. Almarabeh, H., & Amer, E. (2017). A study of data mining techniques accuracy for healthcare. International Journal of Computer Applications, 168(3), 12-17.
21. Tafish, M. H., & El-Halees, A. M. (2018, October). Breast cancer severity degree predication using data mining techniques in the gaza strip. In 2018 International Conference on Promising Electronic Technologies (ICPET) (pp. 124-128). IEEE.
22. Kaur, S., & Bawa, R. K. (2018, August). Review on data mining techniques in healthcare sector. In 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on (pp. 224-228). IEEE.
23. Mia, M. R., Hossain, S. A., Chhoton, A. C., & Chakraborty, N. R. (2018, February). A comprehensive study of data mining techniques in health-care, medical, and bioinformatics. In 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2) (pp. 1-4). IEEE.
24. Eltalhi, S., & Kutrani, H. (2019). Breast cancer diagnosis and prediction using machine learning and data mining techniques: A review. IOSR Journal of Dental and Medical Sciences, 18(4), 85-94.
25. Mabu, A. M., Prasad, R., & Yadav, R. (2020). Mining gene expression data using data mining techniques: A critical review. Journal of Information and Optimization Sciences, 41(3), 723-742.
26. Biwalkar, A., Gupta, R., & Dharadhar, S. (2021, May). An Empirical Study of Data Mining Techniques in the Healthcare Sector. In 2021 2nd International Conference for Emerging Technology (INCET) (pp. 1-8). IEEE.
27. Arivazhagan, N., Mukunthan, M. A., Sundaranarayana, D., Shankar, A., Vinoth Kumar, S., Kesavan, R., ... & Abebe, T. G. (2022). Analysis of skin cancer and patient healthcare using data mining techniques. Computational Intelligence and Neuroscience, 2022(1), 2250275.
28. Patela, R. (2023). Diagnosis of Breast Cancer using Data Mining Techniques.
29. Al-Batah, M., Alzboon, M. S., & Muhyeeddin Alqaraleh, F. A. (2024). Comparative Analysis of Advanced Data Mining Methods for Enhancing Medical Diagnosis and Prognosis. Data Metadata, 3, 465.
30. Anand, H., Prakash, S., Sandhia, G. K., & Prabha, J. R. (2024, July). Chronic disease prediction using data mining algorithms. In AIP Conference Proceedings (Vol. 3075, No. 1). AIP Publishing.