Print ISSN 2682-3993
Online ISSN 2682-4000

Precancer Detection Based on Mutations in Codons 248 and 249 Using Decision Tree (DT) and XGBoost Deep Learning Model.

Hala S.Abuelmakarem [*1], Ahmed Majdy[1], George Maher[1], Hossam Khaled[1], Malak Emad[1], Esraa Asem Shaker[1]

[1] Systems and Biomedical Engineering Department, Higher Institute of Engineering, El-Shorouk Academy, Cairo, Egypt.
***Correspondence author:** h.saad@sha.edu.eg

Received 05 Nov. 2024 – Revised 09 Dec. 2024 – Accepted 11 Dec. 2024

Abstract:

The mutation serves as a biomarker for cancer diagnosis and prognosis, indicating early-stage cancer. Efforts are underway to develop advanced pre-cancer detection methods and therapeutic strategies to restore TP53 function or counteract its loss. This study evaluates the performance of different deep-learning techniques in mutation detection as a precancer classifier. The evaluation was conducted in two distinct phases. Following data processing and feature selection, the performance of several models—Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN)—was systematically assessed across accuracy, sensitivity, and specificity metrics for codons 248, 249, and a combined dataset of these codons. Each model's effectiveness was evaluated under two feature conditions, encompassing eight and thirty-three features. The Decision Tree, identified as the optimal performer, was further enhanced by integrating it with the XGBoost deep learning algorithm to maximize performance. Integrating a DT with XGBoost improves accuracy from 93.15% to 96.55%, sensitivity from 94% to 98%, and specificity from 92% to 96%, making it more effective in detecting precancers based on codon mutations. This combined model enhances both true positive and true negative identification. The detection of codon mutations shows promise for early cancer detection.

Keywords: Precancer Detection, Codon Mutations, Decision Tree, Deep Learning.

I. INTRODUCTION

Cancer remains a significant global health burden, with approximately 19.3 million new cases and 10 million cancer-related deaths reported in 2020, according to GLOBOCAN. The most common cancers worldwide include breast, lung, colorectal, prostate, and stomach cancers [1]. High-income countries tend to report higher incidence rates due to better screening and reporting systems, while low- and middle-income countries experience higher mortality rates due to limited healthcare access and late-stage diagnoses. Over half of all cancer diagnoses occur in individuals aged 65 and older, reflecting the impact of aging populations. Economically, cancer imposes a staggering cost of $1.16 trillion annually, encompassing treatment expenses, loss of productivity, and associated care. The global cancer burden is projected to rise to 28.4 million cases annually by 2040, a 47% increase driven by population growth, aging, and lifestyle factors such as tobacco use, poor diet, and physical inactivity. By 2040, low- and middle-income countries are expected to account for 70% of cancer-related deaths. However, with effective prevention strategies, such as reducing smoking and improving diets, coupled with early detection programs and advancements like AI-driven diagnostics and personalized medicine, the rising cancer burden can be mitigated. These measures have the potential to significantly reduce both incidence and mortality rates, highlighting the need for global collaboration in addressing this growing challenge[2, 3] .

Early detection can markedly improve prognosis and survival rates [4, 5]. Among the various biomarkers for cancer detection, the TP53 gene is called Guardian of the Genome. It safeguards the DNA integrity within the cell. TP53 is a guide for the cell development, aging, and differentiation[6]. TP53 encodes a tumor suppressor protein that regulates cell cycle progression, DNA repair, and apoptosis. The tumor suppressor protein *TP53* is a transcription factor frequently mutated in cancer [7].

Print ISSN 2682-3993
Online ISSN 2682-4000

Mutations in this gene are observed in approximately 50% of all human cancers, making it a focal point in cancer research[8]. Specifically, mutations at codons 248 and 249 of *TP53* have been implicated in the early stages of carcinogenesis, particularly in lung, liver, and colorectal cancers. **Figure 1** shows examples of codons 248 and 249 mutations. Codon 248 is one of the most commonly mutated sites in the TP53 gene, particularly in various cancers. This mutation often results in a substitution of the amino acid arginine with another amino acid, which disrupts the DNA-binding domain of the TP53 protein[9, 10]. Codon 249 is another mutation site within the TP53 gene and is frequently associated with various cancers, particularly hepatocellular carcinoma (HCC). The mutation at codon 249 often results in a substitution of the amino acid arginine with serine (R249S), which disrupts the normal function of TP53 as a tumor suppressor. This alteration impairs TP53′s DNA-binding capacity, leading to a loss of function in regulating cell growth, apoptosis, and genomic integrity[11–14]. Consequently, researchers have identified this mutation as a biomarker for cancer diagnosis and prognosis, and it is an indicator of the early stages of cancer[15–17]. Scientists aim to update pre-cancer detection methods and develop early therapeutic interventions to restore TP53 function or compensate for its loss[18].

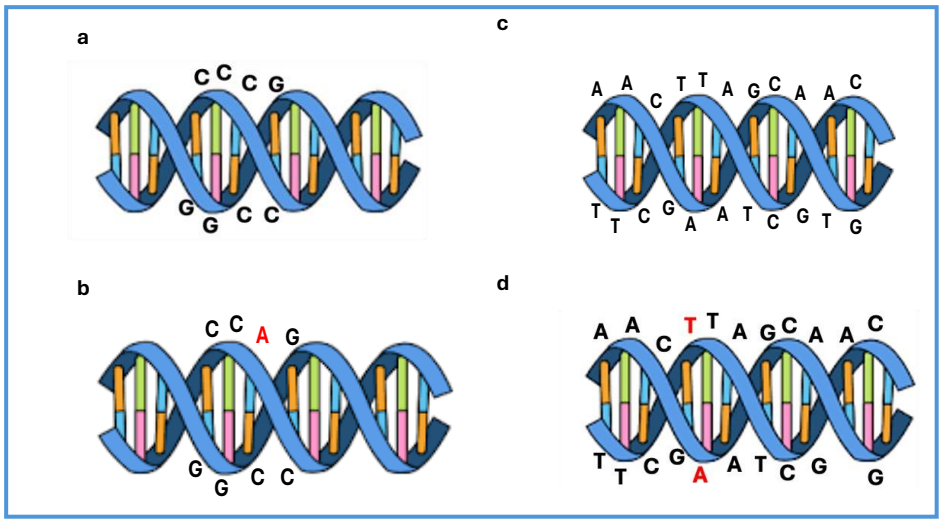

Figure 1: a. normal codon 248, b. mutated codon 248, c. normal codon 249 d. mutated codon 249.

This study aims to detect the mutations in the codons in the protein TP53 for precancer detection. This research evaluates the performance of several algorithms—Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN)—across accuracy, sensitivity, and specificity metrics for detection of the mutations in codons 248, 249, and a combined dataset of these codons. Each model's effectiveness was assessed under two feature conditions, involving both eight and thirty-three features. The Decision Tree (DT), identified as the best performer, was further optimized by integrating the XGBoost algorithm to enhance its performance.

The major contributions of this work can be summarized as follows:
1. Evaluate the performance of different algorithms to detect mutations in:
   - Codon 248
   - Codon 249
   - Codons 248 and 249 combined.
2. Study the effectiveness of using eight features versus thirty-three features.
3. Assess the performance of the DT model by integrating it with XGBoost to evaluate the best-performing model (Codons 248 + 249) using thirty-three features.

II. RELATED WORK

Artificial intelligence (AI) has improved healthcare in many domains [19–21]. It facilitates pre-cancer detection by indicating the codon mutations by checking the codon sequence. Genetic variants pose the challenge of modelling categorical attributes (the four nucleotides) within the context of surrounding nucleotide sequences[22]. Mahmood, D. Yousif, et al., demonstrated that neural network-based prediction effectively identifies cancerous and pre-cancerous conditions associated with mutations in codon 248, achieving high

Print ISSN 2682-3993
Online ISSN 2682-4000

predictive accuracy. Data mining preprocessing steps and pattern extraction techniques were employed to construct the prediction model, selecting 8 out of 132 TP53 gene fields to classify cases into cancer and pre-cancer pathology classes[23]. Zahraa N. et al, introduced an in silico molecular classification approach for breast and prostate cancers using a Back Propagation Network. This model leverages data from seven datasets, including five from the UMD TP53 database and two from the IARC TP53 database[24]. Chaurasia et al, conducted a performance analysis of various data mining techniques for cancer classification, particularly for breast cancer data. This classification approach can reveal disease patterns or uncover the shared characteristics of cancer. Several data mining algorithms were tested, including decision tree classifiers, K-Nearest Neighbor, Sequential Minimal Optimization (SMO), and Best First Tree. Their results indicated that SMO outperformed the other classifiers in terms of accuracy [25]. Prashant Gupta et al, represented a new technique called Continuous Representation of Codon Switches (CRCS) was developed. CRCS is a deep learning-based approach that generates numerical vector representations of mutations, supporting various machine learning applications. This method has three main applications: (1) detecting cancer-related somatic mutations without the need for matched normal samples, useful in assessing tumor mutation burden via cell-free DNA; (2) identifying and analyzing potential cancer driver genes, such as DMD, RSK4, OFD1, WDR44, and AFF2; and (3) scoring individual mutations within a tumor sample, which helps predict patient survival in cancers like bladder urothelial carcinoma, hepatocellular carcinoma, and lung adenocarcinoma[25]. Another study employed BioBERT; a transformer-based language model specialized for biomedical text. Adding a single layer to BioBERT achieved high performance across tasks like text classification, language inference, and question answering. The study also utilized models such as BERT, LSTM, and BiLSTM to classify genetic mutations based on textual data[26]. Xiaoxiao Wang trained a deep convolutional neural network (CNN) based on the ResNet architecture to predict gBRCA mutations in breast cancer using whole-slide images (WSIs) [27].

III. THE PROPOSED SYSTEM

Mutations alter the DNA-binding domain of the *TP53* protein, compromising its ability to control cell proliferation and respond to genomic instability. Recent advancements in molecular biology and genetic sequencing have enabled the detection of these mutations at earlier stages, opening avenues for pre-cancer detection and potentially improving outcomes. This research highlights novel methodology targeting the 248-249 codon mutation in *p53* using different deep learning-based approaches.

A. Dataset

The Universal Mutation Database (UMD_TP53) at (https://p53.fr/TP53-database) provides a comprehensive representation of the TP53 gene structure, isoform diversity, mutation nomenclature, and sequence data from over 5,000 tumor samples across 12 cancer types. The dataset highlights mutations specifically affecting the TP53 beta and gamma isoforms. The UMD TP53 mutation database, spanning more than 80,400 columns, includes details on germline mutations, cell lines, and more. Figure 2 shows the details of the Universal Mutation Database (UMD_TP53). missense mutations are the most common in the database, with 58,522 entries, indicating a significant impact on protein function. Indels and nonsense mutations are also prevalent, likely disrupting gene expression, while splice site and synonymous mutations are less common. Rare mutation types are minimal, with only 250 entries. This study focuses on mutations in codons 248 and 249. The target data comprises 5,120 records specifically within this domain, representing a dataset reduction at the row level. Additionally, the TP53 dataset includes information on mutations in codons 248 and 249, along with indications of whether the cells were cancerous or precancerous.

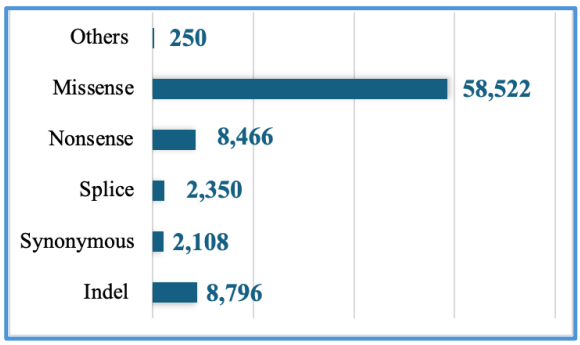

Figure 2: Universal Mutation Database (UMD_TP53)

Print ISSN 2682-3993
Online ISSN 2682-4000

B. Data Preprocessing

The dataset was cleaned by addressing missing values and removing duplicate entries. Data normalization was employed to ensure comparability across features for subsequent analyses.

C. Feature Selection:

A sequence of genetic and molecular changes across several features identifies the mutation. This research is run in tracks, first track employs eight features, while the second one employs thirty-three features. The eight input set features are Start_cdna, EndcDNA, Genome base coding, cDNA variant, Mutant AA-1, Mutant AA-3, Mutant codon, and sample original. Thirty-three features are selected as the following: cDNA position (Start_cDNA, End_cDNA), the Genome Base Coding, cDNA Variant, and Codon position using *TP53* alpha. Also, Mutant AA_1 (Mutant amino acid: 1-letter nomenclature), Mutant AA_3 (Mutant amino acid: 3-letter nomenclature), Base_Change_Size, Wild Type Codon, Mutant Codon, Mutational_Event, Mutation Type, CpG, Py-Py Doublets, Variant Type. Other features as, Disease, Sample pathology, Complexity, and records number. Features refer to the tumor status as Leukaemia_State, Solid Status, Tumor Status Cell line State, Somatic State, Germline_Stat. Tumor Repetition, and Publication_Repetition, Sift_Score, Mutassessor_score, Polyphen, MutPred_Splice_General_Score, PCA_Outliers, PCA_score. Those are selected according to the heatmap that works as an indicator for the important features. The data is split to train, test, and validate as the following 60%, 20%, and 20%.

D. Model implementation

The system is evaluated in two steps: after data processing and feature selection, splitting, the performance of different algorithms—Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN)—is assessed in terms of accuracy, sensitivity, and specificity on codons 248, 249, and a combined dataset of codons 248 and 249. Each model's performance is evaluated twice, based on eight features and thirty-three features. The best-performing model is DT. So, it is reevaluated by combining it with the XGBoost deep learning algorithm to enhance performance. Figure 3 shows the block diagram of the Model.

The decision tree is used machine learning model that classifies or predicts outcomes by splitting data into branches based on feature values[28]. It structures decision-making in a tree-like format, where each internal node represents a decision rule, each branch represents an outcome of the decision, and each leaf node represents a final classification or prediction. Decision trees are intuitive and interpretable, as the sequence of splits mirrors human decision-making processes[29]. They are effective for handling numerical and categorical data and can capture complex feature interactions. However, decision trees can be prone to overfitting, especially with deep trees, which may capture noise rather than underlying patterns[30]. XGBoost, or Extreme Gradient Boosting, is a highly efficient and scalable machine learning algorithm, particularly well-suited for structured or tabular data. It operates as an ensemble learning method, where multiple decision trees are sequentially built to minimize prediction errors. Unlike traditional boosting methods, XGBoost employs a more optimized implementation with features like regularization (to prevent overfitting), parallel processing, and handling of missing values. These enhancements enable it to deliver high accuracy, often outperforming other algorithms in classification and regression tasks. XGBoost has become widely used in competitive machine learning, due to its speed and predictive power, making it a top choice for complex data challenges across various domains, including finance, biology, and marketing[31]. The measurements are calculated in terms of accuracy, Sensitivity, precision, and it is calculated as

$$\text{Accuracy} = (\text{TN} + \text{TP}) / (\text{TN} + \text{FP} + \text{FN} + \text{TP})$$
$$\text{Sensitivity} = TP / (TP + FN).$$
$$Precision = TP / (TP + FP).$$

Where True Positives (TP) = The number of correctly identified positive cases, and False Negatives (FN) = The number of positive cases incorrectly identified as negative.

Print ISSN 2682-3993
Online ISSN 2682-4000

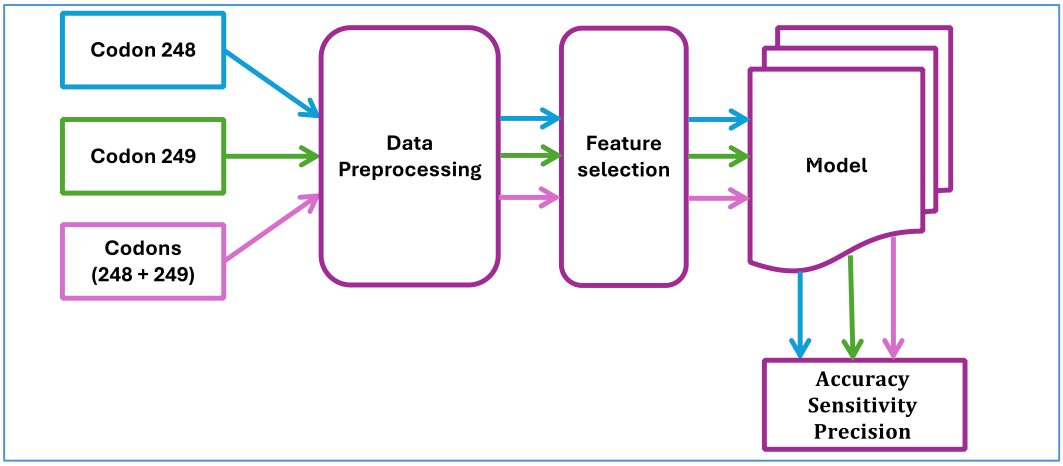

Figure 3: Flow diagram for the proposed models.

Table 1: The initial parameters for the Convolutional Neural Network and Artificial Neural Network

Parameter	CNN	ANN
Input Neurons	128*3	10
Hidden Layers	64*3 - 32*3 – 16*3	6
Output Neurons	1	1
Activation Function	ReLU	ReLU
Learning Rate	0.001	0.001
Epochs	200	200
Optimizer	Adam	Adam

Table 1 shows the initial parameters (Input Neurons, Hidden Layers, Output, Neurons, Activation Function, Learning Rate, Epochs, and Optimizer) of the CNN and ANN that are adjusted for this study. For the KNN classifier, the number of neighbors is adjusted to 7. The Decision Tree model used as the base learner in the XGBoost ensemble was configured with a carefully selected set of hyperparameters to optimize performance. The maximum tree depth, set to 1, restricts the model's complexity to prevent overfitting while maintaining interpretability. The learning rate was set to 0.3, allowing the model to make incremental updates that contribute to stable convergence. A gamma value of 0 was selected, indicating no minimum loss reduction requirement for split formation, thus allowing splits based on impurity alone. The minimum child weight parameter was assigned a value of 1, ensuring that a minimum level of observation weight is required to form a new leaf node, which controls model complexity.

IV. RESULTS AND DISCUSSION

Figure 3 represents the performance measurements —accuracy, sensitivity, and specificity—for various models (Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbors(KNN), Convolutional Neural Network (CNN)) on codons 248, 249, and a combined dataset of codons 248 and 249. Each model performance is evaluated based on eight features.

For codon 248, the models show varying levels of performance: (ANN) indicates moderate performance with slightly higher specificity. (DT) demonstrates the highest performance among the models, with 82% accuracy, 79% sensitivity, and 86% specificity. (KNN) has similar accuracy (77%) but slightly lower sensitivity (71%) and specificity (81%) compared to ANN. (CNN) performs the best in this group with 84% accuracy as DT, 80% sensitivity, and 89% specificity, indicating balanced and robust performance across all metrics.

For codon 249, ANN shows a slight reduction in performance compared to codon 248, with 75% accuracy, 71% sensitivity, and 80% specificity. DT maintains higher measurements compared to ANN and KNN. CNN

Print ISSN 2682-3993
Online ISSN 2682-4000

continues to perform well, with 82% accuracy, 79% sensitivity, and 83% specificity, although it falls slightly short of its performance on codon 248.

When combining data from both codons (248 + 249), ANN improves overall accuracy (80%) and sensitivity (77%) compared to individual codons. KNN shows an improvement in sensitivity (77%) compared to codon 249 alone, maintaining a specificity of 81%. CNN once again delivers the highest performance with 84% accuracy, 80% sensitivity, and 89% specificity, solidifying its robustness when handling combined data.

Figure 4 analysis shows the performance of the four machine learning algorithms—ANN, DT, KNN, and CNN—applied to datasets representing mutations at Codons 248 and 249, as well as a combined dataset of both codons, with a feature space of thirty-three attributes.

For Codon 248, the Decision Tree (DT) algorithm exhibited the highest accuracy at 86%, with a sensitivity of 82% and a specificity of 82%. This indicates robust performance in identifying true positives while maintaining a moderate level of specificity. For Codon 249, ANN displayed an accuracy of 77%, sensitivity of 72%, and specificity of 82%, showing a slightly improved specificity for Codon 248 compared to Codon 249. CNN performed on par with DT in terms of accuracy (83%) but achieved a slightly lower sensitivity (81%) and specificity (89%), making it a solid alternative depending on the specific requirements of sensitivity versus specificity. KNN achieved an accuracy of 81% but slightly reduced sensitivity (71%) compared to DT and specificity at 88%, showing a balanced but slightly less sensitive performance.

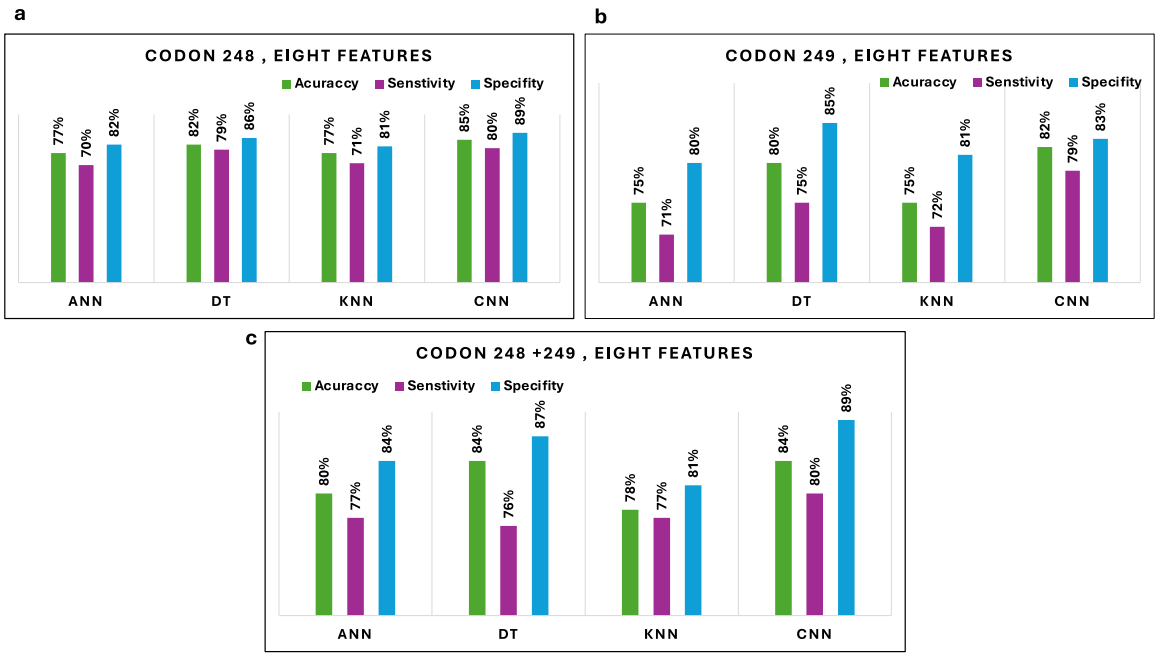

Figure 3: The performance (accuracy, sensitivity, specificity) for various models (Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN)) for eight features for Codon a) 248, b) Codon 249, c) codons (248 and 249).

Print ISSN 2682-3993
Online ISSN 2682-4000

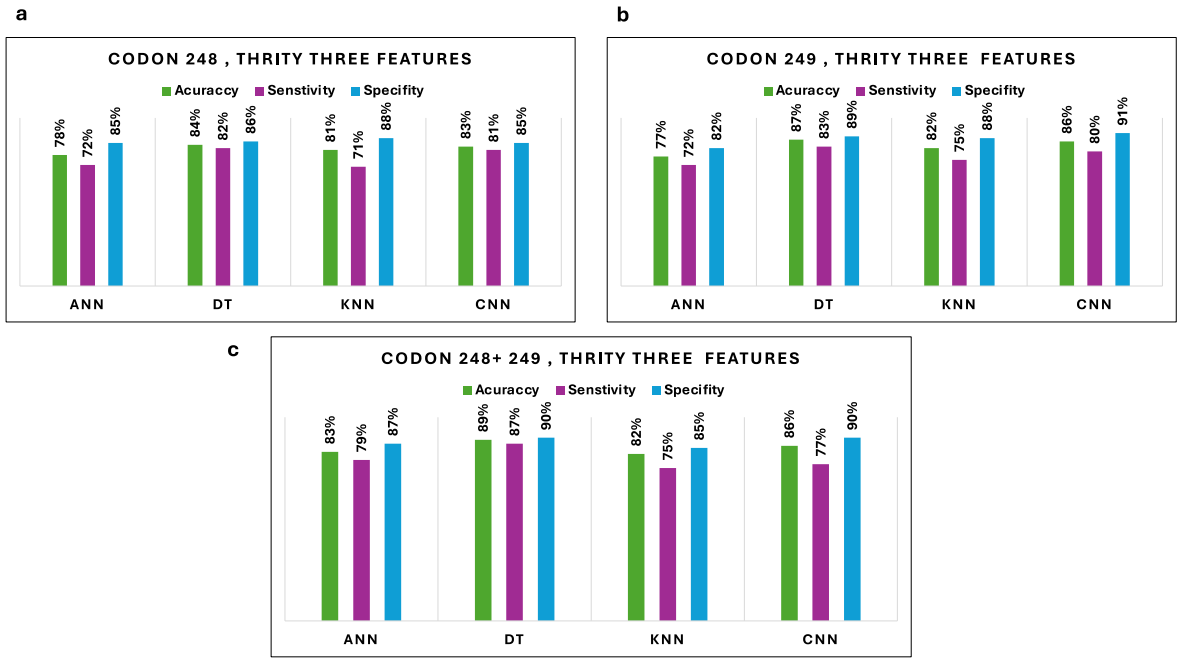

Figure 4: The performance (accuracy, sensitivity, specificity) for various models (Artificial Neural Network (ANN), Decision Tree (DT), K-Nearest Neighbors (KNN), Convolutional Neural Network (CNN)) for Thirty-Three features for Codon a) 248, b) Codon 249, c) codons (248 and 249).

Table 2: Performance of the XGBoost and the performance of the XGBoost+ Decision Tree

	XGBoost	XGBoost + DT
Accuracy	93.15%	96.55%
sensitivity	94%	98%
specificity	92%	96%

Combining data from Codons 248 and 249 creates a synergistic effect, amplifying performance across all models: CNN has an outstanding 90% accuracy, sensitivity of 86%, and specificity of 90%. K- KNN remains competitive with 86% accuracy, a slightly lower sensitivity of 77%, but a high specificity of 90%. The high specificity is indicative of KNN's ability to avoid false positives, which may be advantageous in critical applications. ANN shows substantial improvement in the combined dataset. Decision Tree (DT) follows closely, maintaining an impressive 89% accuracy, sensitivity of 87%, and specificity of 90%. This analysis highlights that DT consistently outperforms other models, particularly in the combined dataset, suggesting that it best captures complex feature relationships. So, it was our candidate to improve its performance by the XGBoost algorithm.

Table 2 shows the performance of the XGBoost algorithm both as a standalone model and in combination with the DT ensemble. These metrics provide a comprehensive view of the model's capability in terms of overall predictive accuracy, its ability to correctly identify positive cases (sensitivity), and its effectiveness in avoiding false positives (specificity). The accuracy metric measures the model's overall success rate in making correct predictions and balancing in positive and negative cases. Sensitivity, or recall, reflects the model's ability to identify positive cases correctly. This metric is critical in applications where correctly classifying positives is more important than negatives, such as in medical diagnoses. For the accuracy measurements, XGBoost alone Achieves an accuracy of 93.15%, indicating a high level of reliability in general classification tasks. This accuracy suggests that XGBoost is already adept at identifying patterns in the dataset, making it a strong standalone classifier. When combining XGBoost with a Decision Tree, the accuracy increases significantly to 96.55%.

This improvement suggests a synergistic effect where the ensemble captures additional nuances in the data that XGBoost alone may not fully address. The 3.4% increase in accuracy underscores the effectiveness of this

Print ISSN 2682-3993
Online ISSN 2682-4000

ensemble approach, particularly in applications requiring a higher level of precision. XGBoost Alone shows a sensitivity of 94%, demonstrating a strong ability to detect positive instances. This high sensitivity rate indicates that XGBoost is reliable in identifying true positives with minimal oversight. The sensitivity of the combination (XGBoost and Decision Tree Ensemble) rises to 98% with the ensemble, marking a substantial 4% improvement. This near-perfect sensitivity suggests that the addition of the DT enables the model to capture a broader range of positive patterns, thereby minimizing false negatives. This characteristic is particularly advantageous in high-stakes scenarios where missing positive cases could have serious consequences. Specificity measures the model's effectiveness in correctly identifying negative cases or its ability to minimize false positives. High specificity is essential in contexts where false positives must be reduced.

To the best of our knowledge, no previous studies have combined codons 248 and 249 for precancer detection. Table 3 presents previous studies that employed deep learning techniques for mutation detection, and our proposed model outperformed them. the ANN model in the study [25] outperformed other approaches with 98% accuracy, while CNN and SVM also demonstrated strong but slightly lower accuracy rates. Each study emphasizes different machine-learning techniques and datasets, highlighting diverse approaches to mutation detection in cancer research. Abdullah Abdul [32] utilizes RNA-Seq datasets (highlights AK2 and CD68 as potential biomarkers and therapeutic targets for breast cancer), machine learning algorithms, and feature selection techniques to predict breast cancer prognosis and identify relevant biomarkers. Quadratic Discriminant Analysis (QDA) achieved the highest accuracy of 96%, outperforming Neural Networks, Random Forest, Linear SVM, and Logistic Regression, each with 91% accuracy. Davies et al[33], investigates the inefficiency of nucleotide excision repair (NER) in recognizing BPDE-DNA adducts by combining molecular dynamics and machine learning. A random forest model, trained on helical distortion data, achieved over 91% accuracy, identifying base pair rotation and regional GC content as key factors distinguishing repair hotspots and non-hotspots.

Table 3: Comparison the DL models results with the relevant studies.

Reference	Dataset	Model	Accuracy
Pre-cancer Diagnosis via TP53 Gene Mutations by Using Bioinformatics & Neural Network [34]	http://p53.fr/TP53-database/mutation-database "UMD TP53 Mutation Database" Oct, 2017	Classification Using ANN Feed Forward Back Propagation Neural Network	98%
Deep convolutional neural networks for accurate somatic mutation detection[35]	the European Nucleotide Archive : https://www.ebi.ac.uk/ena	Classification Using CNN	93.5%
Evaluating machine learning methodologies for identification of cancer driver genes[36]	https://www.uniprot.org/	Classification (SVM)	91.0%
		ANN	87.2%
Using machine learning algorithms to find novel biomarkers for breast cancer using RNA-seq dataset [32]	RNA-Seq datasets	Neural Networks, Random Forest, Linear SVM, and Logistic Regression.	91%
		Quadratic Discriminant Analysis (QDA)	96 %
Utilizing biological experimental data and molecular dynamics for the classification of mutational hotspots through machine learning[33]	helical data extracted from duplexes representing both BPDE hotspot and nonhotspot sites within the TP53 gene	random forest model	91%

The clinical applications and scalability of precancer detection based on TP53 mutations using AI hold immense potential for transforming healthcare. AI can enable early diagnosis by detecting TP53 mutations, allowing timely interventions that improve patient outcomes. It also facilitates personalized medicine by identifying mutation patterns that guide tailored treatment strategies, while serving as a prognostic tool to predict disease progression.

Print ISSN 2682-3993
Online ISSN 2682-4000

Furthermore, insights gained from TP53-related pathways can aid in developing targeted therapies for cancers with these mutations. The scalability of AI-driven detection makes it suitable for mass screening programs by analyzing large-scale datasets efficiently, reducing costs through automation, and democratizing access to advanced diagnostics, particularly in under-resourced areas [37]. Integrating AI tools into clinical workflows and electronic health records ensures seamless adoption by healthcare professionals. The real-world impact of these advancements includes reduced cancer mortality rates, better allocation of healthcare resources, and accelerated research into TP53 mutations, driving breakthroughs in cancer prevention and treatment. This AI-powered approach has the potential to revolutionize cancer care globally.

CONCLUSION

This study demonstrates the potential of deep learning techniques, particularly the integration of Decision Tree and XGBoost, for effective mutation detection in precancer classification. By combining codons 248 and 249, the proposed model achieves notable improvements in accuracy, sensitivity, and specificity, offering a reliable tool for early-stage cancer diagnosis. The optimized model not only enhances true positive and negative identification but also represents a significant advancement in leveraging mutation biomarkers for precancer detection, potentially guiding future diagnostic and therapeutic strategies targeting TP53-related cancers.

Reference

1. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 71, 209–249 (2021). https://doi.org/10.3322/caac.21660
2. Abuelmakarem, H.S., Hamdy, O., Sliem, M.A., El-Azab, J., Om-Hashem, M.A., Ahmed, W.A.: Colonic Carcinoma Diagnosis using Chitosan Nanoparticles Based on the Optical Properties. J Phys Conf Ser. 1472, 012001 (2020). https://doi.org/10.1088/1742-6596/1472/1/012001
3. Abuelmakarem, H.S., Hamdy, O., Sliem, M.A., El-Azab, J., Ahmed, W.A.: Early cancer detection using the fluorescent Ashwagandha chitosan nanoparticles combined with near-infrared light diffusion characterization: in vitro study. Lasers Med Sci. 38, 37 (2023). https://doi.org/10.1007/s10103-022-03678-x
4. Abuelmakarem, H., Sliem, M., El-Azab, J., Farghaly, M., Ahmed, W.: Toward Highly Efficient Cancer Imaging and Therapy Using the Environment-Friendly Chitosan Nanoparticles and NIR Laser. Biosensors (Basel). 9, 28 (2019). https://doi.org/10.3390/bios9010028
5. Abuelmakarem, H.S., Sliem, M.A., El-Azab, J., Farghaly, M., Om-Hashem, M.A., Ahmed, W.A.: The Fluorescent Effect of Withania Somnifera Chitosan Nanocomposite as an Effective Contrast Agent for Cancer Theragnostic: Experimental Study in Vitro. J Fluoresc. 32, 949–960 (2022). https://doi.org/10.1007/s10895-022-02895-5
6. Jain, A.K., Barton, M.C.: p53: emerging roles in stem cells, development and beyond. Development. 145, (2018). https://doi.org/10.1242/dev.158360
7. Paz Hernandez, M., Oses, C., Peña, D., Criollo, A., Morselli, E.: Mutant p53 Located in the Cytoplasm Inhibits Autophagy. In: Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging. pp. 189–203. Elsevier (2016)
8. Vaddavalli, P.L., Schumacher, B.: The p53 network: cellular and systemic DNA damage responses in cancer and aging. Trends in Genetics. 38, 598–612 (2022). https://doi.org/10.1016/j.tig.2022.02.010
9. Prior, I.A., Lewis, P.D., Mattos, C.: A Comprehensive Survey of Ras Mutations in Cancer. Cancer Res. 72, 2457–2467 (2012). https://doi.org/10.1158/0008-5472.CAN-11-2612
10. Chen, X., Zhang, T., Su, W., Dou, Z., Zhao, D., Jin, X., Lei, H., Wang, J., Xie, X., Cheng, B., Li, Q., Zhang, H., Di, C.: Mutant p53 in cancer: from molecular mechanism to therapeutic modulation, (2022)
11. Scott, A.J., Lieu, C.H., Messersmith, W.A.: Therapeutic Approaches to RAS Mutation. The Cancer Journal. 22, 165–174 (2016). https://doi.org/10.1097/PPO.0000000000000187

Print ISSN 2682-3993
Online ISSN 2682-4000

12. Huang, X.-H., Sun, L.-H., Lu, D.-D., Sun, Y., Ma, L.-J., Zhang, X.-R., Huang, J., Yu, L.: Codon 249 mutation in exon 7 of p53 gene in plasma DNA: maybe a new early diagnostic marker of hepatocellular carcinoma in Qidong risk area, China. World J Gastroenterol. 9, (2003)
13. Hernández Borrero, L.J., El-Deiry, W.S.: Tumor suppressor p53: Biology, signaling pathways, and therapeutic targeting, (2021)
14. Wang, H., Liao, P., Zeng, S.X., Lu, H.: It takes a team: a gain-of-function story of p53-R249S. J Mol Cell Biol. 11, 277–283 (2019). https://doi.org/10.1093/jmcb/mjy086
15. Sah, S., Surendiran, B., Dhanalakshmi, R., Mohanty, S.N.: Mutation prediction and phylogenetic analysis of SARS-CoV2 protein sequences using LSTM based encoder-decoder model. Arab J Basic Appl Sci. 30, 103–121 (2023). https://doi.org/10.1080/25765299.2023.2188677
16. Cheng, P., Mao, C., Tang, J., Yang, S., Cheng, Y., Wang, W., Gu, Q., Han, W., Chen, H., Li, S., Chen, Y., Zhou, J., Li, W., Pan, A., Zhao, S., Huang, X., Zhu, S., Zhang, J., Shu, W., Wang, S.: Zero-shot prediction of mutation effects with multimodal deep representation learning guides protein engineering. Cell Res. 34, 630–647 (2024). https://doi.org/10.1038/s41422-024-00989-2
17. Constant, D.A., Gutierrez, J.M., Sastry, A. V., Viazzo, R., Smith, N.R., Hossain, J., Spencer, D.A., Carter, H., Ventura, A.B., Louie, M.T.M., Kohnert, C., Consbruck, R., Bennett, J., Crawford, K.A., Sutton, J.M., Morrison, A., Steiger, A.K., Jackson, K.A., Stanton, J.T., Abdulhaqq, S., Hannum, G., Meier, J., Weinstock, M., Gander, M.: Deep learning-based codon optimization with large-scale synonymous variant datasets enables generalized tunable protein expression, (2023)
18. Paz Hernandez, M., Oses, C., Peña, D., Criollo, A., Morselli, E.: Mutant p53 Located in the Cytoplasm Inhibits Autophagy. In: Autophagy: Cancer, Other Pathologies, Inflammation, Immunity, Infection, and Aging. pp. 189–203. Elsevier (2016)
19. Abuelmakarem, H.S., Fawzi, S.A., Quriba, A., Elbialy, A., Kandil, A.H.: COMPUTER-AIDED THERAPY USING AUTOMATIC SPEECH RECOGNITION TECHNIQUE FOR DELAYED LANGUAGE DEVELOPMENT CHILDREN. Biomed Eng (Singapore). 34, (2022). https://doi.org/10.4015/S1016237222500235
20. Mahdy, S., Abuelmakarem, H.S.: Alzheimer's disease progression detection based on optical fluence rate measurements using alternative laser wavelengths. Int J Numer Method Biomed Eng. (2024). https://doi.org/10.1002/cnm.3816
21. Ashraf, F. Bin, Akter, S., Mumu, S.H., Islam, M.U., Uddin, J.: Bio-activity prediction of drug candidate compounds targeting SARS-Cov-2 using machine learning approaches. PLoS One. 18, e0288053 (2023). https://doi.org/10.1371/journal.pone.0288053
22. Chen, X., Zhang, T., Su, W., Dou, Z., Zhao, D., Jin, X., Lei, H., Wang, J., Xie, X., Cheng, B., Li, Q., Zhang, H., Di, C.: Mutant p53 in cancer: from molecular mechanism to therapeutic modulation. Cell Death Dis. 13, 974 (2022). https://doi.org/10.1038/s41419-022-05408-1
23. Mahmood, D.Y., Ismaeel, A.G., Taqi, A.H.: Mining method for cancer and pre-cancer detection caused by mutant codon 248 in TP53. Periodicals of Engineering and Natural Sciences (PEN). 7, 522 (2019). https://doi.org/10.21533/pen.v7i2.546
24. Naser, Z., Weli, S., Ramadhan, S.: In Silico Molecular Classification of Breast and Prostate Cancers using Back Propagation Neural Network. (2017). https://doi.org/10.7537/marscbj070317.01
25. Chaurasia, V., Pal, S.: A Novel Approach for Breast Cancer Detection using Data Mining Techniques. Article in International Journal of Innovative Research in Computer and Communication Engineering. 3297, (2014)
26. Elsamahy, E.A., Ahmed, A.E., Shoala, T., Maghraby, F.A.: Deep-GenMut: Automated genetic mutation classification in oncology: A deep learning comparative study. Heliyon. 10, (2024). https://doi.org/10.1016/j.heliyon.2024.e32279
27. Wang, X., Zou, C., Zhang, Y., Li, X., Wang, C., Ke, F., Chen, J., Wang, W., Wang, D., Xu, X., Xie, L., Zhang, Y.: Prediction of BRCA Gene Mutation in Breast Cancer Based on Deep Learning and Histopathology Images. Front Genet. 12, (2021). https://doi.org/10.3389/fgene.2021.661109
28. Rameh, V., Löbel, U., D'Arco, F., Bhatia, A., Mankad, K., Poussaint, T.Y., Alves, C.A.: Cortically-based Brain Tumors in Children: A Decision-tree Approach in the Radiology Reading Room. American Journal of Neuroradiology. ajnr.A8477 (2024). https://doi.org/10.3174/ajnr.A8477

Print ISSN 2682-3993
Online ISSN 2682-4000

29. Putra, P.P., Anam, M.K., Defit, S., Yunianta, A.: Enhancing the Decision Tree Algorithm to Improve Performance Across Various Datasets. INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi. 8, 200–212 (2024). https://doi.org/10.29407/intensif.v8i2.22280
30. Rokhsati, H., Rezaee, K., Abbasi, A.A., Belhaouari, S.B., Shafi, J., Liu, Y., Gheisari, M., Movassagh, A.A., Kosari, S.: An efficient computer-aided diagnosis model for classifying melanoma cancer using fuzzy-ID3-pvalue decision tree algorithm. Multimed Tools Appl. 83, 76731–76751 (2024). https://doi.org/10.1007/s11042-024-18314-9
31. Patel, S.K., Wekalao, J., Mandela, N., Al-Zahrani, F.A.: Design of encoded graphene-gold metasurface-based circular ring and square sensors for brain tumor detection and optimization using XGBoost algorithm. Diam Relat Mater. 148, 111439 (2024). https://doi.org/10.1016/j.diamond.2024.111439
32. Abdullah Abdul: USING MACHINE LEARNING ALGORITHMS TO FIND NOVEL BIOMARKERS FOR BREAST CANCER USING RNA-SEQ DATASET, (2023)
33. Davies, J.G., Menzies, G.E.: Utilizing biological experimental data and molecular dynamics for the classification of mutational hotspots through machine learning. Bioinformatics Advances. 4, (2024). https://doi.org/10.1093/bioadv/vbae125
34. Mikhail, D.Y.: Pre-cancer Diagnosis via TP53 Gene Mutations by Using Bioinformatics & Neural Network. In: 2019 International Engineering Conference (IEC). pp. 136–141. IEEE (2019)
35. Sahraeian, S.M.E., Liu, R., Lau, B., Podesta, K., Mohiyuddin, M., Lam, H.Y.K.: Deep convolutional neural networks for accurate somatic mutation detection. Nat Commun. 10, 1041 (2019). https://doi.org/10.1038/s41467-019-09027-x
36. Malebary, S.J., Khan, Y.D.: Evaluating machine learning methodologies for identification of cancer driver genes. Sci Rep. 11, 12281 (2021). https://doi.org/10.1038/s41598-021-91656-8
37. Chillotti, S., Maloberti, T., Degiovanni, A., Malvi, D., D'Errico, A., de Biase, D., Vasuri, F.: Hepatocellular Carcinomas with Concomitant Mutations of TERT, TP53, and CTNNB1: Is There a Role for Artificial Intelligence? Crit Rev Oncog. 28, 31–35 (2023). https://doi.org/10.1615/CritRevOncog.2023049650